Technische Universität München
TUM School of Life Sciences

Transcriptomic analysis of Human Endogenous Retroviruses (HERVs) activities in the pathological context

Hengyuan Liu

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Fabian J. Theis

Prüfer*innen der Dissertation:

1. Prof. Dr. Dmitrij Frischmann
2. Prof. Dr. Ulrike Protzer

Die Dissertation wurde am 21.02.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 29.06.2023 angenommen.

Abstract

Human endogenous retroviruses (HERVs) are the remnants of ancient exogenous retroviruses infecting the human germline cells and remaining in the genome. These sequences comprise as much as 8% of the human genome. Most HERV elements are silenced because of the methylation, while many environmental factors like neurological diseases and infection of exogenous viruses can induce their activities. This dissertation mainly discusses HERV activities and their interactions with the host in these pathological contexts.

In chapter 2, we first investigate the transcriptional activity of HERVs in the human cell upon infections of three Influenza A viruses, including two wild-type strains (PR8M and SC35M) and one mutated strain (SC35MΔNS1). We detected large sums of differentially expressed HERVs (DEHERVS) and genes (DEGs) in the infected cells, most of which were upregulated. Through the DEHERV-G coregulation analysis of them, we found that the genes that co-upregulated genes with HERVs are significantly enriched in the immunological processes and pathways, which indicates the ubiquitous interactions of HERV activities with the cell immune system. Furthermore, we identified the LTR12C element, which is the LTR of the HERVW9 group, to be a possible trigger of the immune response by providing promoter sequences to two essential immunity-relative genes.

To gain a deeper knowledge of the widely reported HERV-K overexpression in the many kinds of neurodegeneration diseases, in chapter 3, we intentionally reactivated the expression of HERV-K of H9 cells and differentiated them into neuron cortical cells. We found the reactivation of HERV-K resulted in considerable disruption of the cortical neuron after 60 days of cell differentiation. Through monitoring cell transcriptional activities overtimes and further wet-lab validation, we found that the overexpression of the NTRK3 gene induced by the HERV-K reactivation to be the crucial factor disrupting the cortical neuron growth, as knocking down the NTRK3 expression will prevent HERV-K reactivation from disrupting the cortical neuron differentiation. Based on this investigation, we further discussed the potential pathogen role of HERV-K in Alzheimer's disease.

Since this dissertation is mainly based on bioinformatics analysis, we elucidated the

main challenge of in silico transcriptional analysis of HERVs caused by their repetitive property and different strategies. We also discussed the outlook from the newly developed bioinformatics algorithm and sequencing techniques to a more solid HERV expression analysis.

Zusammenfassung

Humane endogene Retroviren (HERVs) sind die Überreste alter exogener Retroviren, die die menschlichen Keimzellen infizieren und im Genom verbleiben. Diese Sequenzen umfassen bis zu 8% des menschlichen Genoms. Die meisten HERV-Elemente werden aufgrund der Methylierung zum Schweigen gebracht, während viele Umweltfaktoren wie neurologische Erkrankungen und Infektionen mit exogenen Viren ihre Aktivitäten induzieren können. Diese Dissertation beschäftigt sich hauptsächlich mit HERV-Aktivitäten und deren Interaktionen mit dem Wirt in diesen pathologischen Kontexten.

In Kapitel 2 untersuchen wir zunächst die transkriptionelle Aktivität von HERVs in der menschlichen Zelle bei Infektionen von drei Influenza-A-Viren, darunter zwei Wildtyp-Stämme (PR8M und SC35M) und einem mutierten Stamm (SC35MΔNS1). In den infizierten Zellen konnten große Summen differentiell exprimierter HERVs (DEHERVS) und Gene (DEGs) nachgewiesen werden, von denen die meisten hochreguliert waren. Durch die DEHERV-G-Koregulierungsanalyse dieser Gene konnten wir feststellen, dass Gene, die Gene mit HERVs koporeguliert haben, signifikant in den immunologischen Prozessen und Signalwegen angereichert sind, was auf die allgegenwärtigen Wechselwirkungen von HERV-Aktivitäten mit dem Zellimmunsystem hindeutet. Darüber hinaus identifizierten wir das LTR12C-Element, das LTR der HERVW9-Gruppe ist, als möglicher Auslöser der Immunantwort, indem es Promotorsequenzen für zwei essentielle immunrelative Gene bereitstellt.

Um ein tieferes Verständnis der weit verbreiteten HERV-K-Überexpression bei den vielen Arten von Neurodegenerationskrankheiten zu erlangen, haben wir in Kapitel 3 absichtlich die Expression von HERV-K von H9-Zellen reaktiviert und in neuronale kortikale Zellen differenziert. Wir fanden, dass die Reaktivierung von HERV-K nach 60-tägiger Zelldifferenzierung zu einer erheblichen Störung des kortikalen Neurons führte. Durch die Überwachung der Zelltranskriptionsaktivität und weitere Validierung im Nasslabor fanden wir heraus, dass die Überexpression des NTRK3-Gens, induziert durch die HERV-K-Reaktivierung, der entscheidende Faktor für die Störung des kortikalen Neuronenwachstums ist, da das Niederschlagen der NTRK3-Expression verhindert, dass die HERV-K-Reaktivierung die Differenzierung der kortikalen Neuronen stört. Basierend auf dieser Untersuchung diskutierten wir die mögliche Rolle von HERV-K bei Alzheimer.

Da diese Dissertation hauptsächlich auf bioinformatischer Analyse basiert, haben wir die Hauptherausforderung der in silico transkriptionellen Analyse von HERVs aufgeklärt, die durch ihre repetitiven Eigenschaften und unterschiedliche Strategien verursacht werden. Wir diskutierten auch den Ausblick vom neu entwickelten bioinformatischen Algorithmus und Sequenzierungstechniken zu einer solideren HERV-Expressionsanalyse.

Publications

1. Liu, H.; Bergant, V.; Frishman, G.; Ruepp, A.; Pichlmair, A.; Vincendeau, M.; Frishman, D. Influenza A Virus Infection Reactivates Human Endogenous Retroviruses Associated with Modulation of Antiviral Immunity. Viruses 2022, 14, doi:10.3390/v14071591.

2. Padmanabhan Nair, V.; Liu, H.; Ciceri, G.; Jungverdorben, J.; Frishman, G.; Tchieu, J.; Cederquist, G.Y.; Rothenaigner, I.; Schorpp, K.; Klepper, L.; et al. Activation of HERV-K(HML-2) Disrupts Cortical Patterning and Neuronal Differentiation by Increasing NTRK3. Cell Stem Cell 2021, 28, 1566-1581.e8, doi:10.1016/j.stem.2021.04.009.

Acknowledgments

It has been years since I started my Ph.D. program. During the long journey, I encountered many problems and hard times either in the scientific work or my life. But there are always lovely guys guiding, supporting, and encouraging me. Sincerely, I would like to express my gratitude to all of them.

Firstly, I would like to thank my supervisor, Prof. Dmitrij Frishman. I'm grateful for the opportunity he offered to start my Ph.D. program. I really appreciate for his patient scientific training, and teaching. And he also helped for making my decision in life. Thank you, Dmitrij, for all of these.

I would also like to thank all the collaborators in different scientific projects. Special thanks to Dr. Michelle Vincendeau for her excellent help in the HERV-K reactivation and Influenza A projects. Thanks to Vidya Padmanabhan Nair for her outstanding wet-lab experiment in the HERV-K reactivation project. Thanks to Prof. Jens Mayer for offering his insightful scientific opinion, which benefits me very much. Thanks to Prof. Paul Lingor and Dr. Lucas Caldi Gomes for their valuable and beneficial discussion about HERV-H research.

I am very grateful to Prof. Dr. Fabian Theis and Prof. Dr. Ulrike Protzer, who kindly agreed to be the chair and examiner of my examination committee.

Thanks very much to all the secretaries in the department for their administrative help. Thanks, Mrs. Léonie Corry, Mrs. Roswitha Weinbrunn, and Mrs. Martina Rüttger, they all offered me much help in life. Thanks, Mr. Drazen Jalsovec for managing the computing sever, which is essential for the bioinformatics research. Thanks also to the China Scholarship Council for their financial support.

I'm also thankful for all my colleagues in the group. Thanks to Fei Qi, Hongen Xu, Yanping Zhang, Weiyun Tsai, Jianfeng Sun, Yingying Dong, Wen Liang, Michael Kiening, Peter Hönigschmid, Marina Parr, Jan Zaucha, Evans Kataka, Pinar Karabulut, Martina Weigl, Usman Saeed, Andrey Chursov, Anja Mösch for all their help. It was a really good time to have valuable scientific discussions with them.

I wish to thank all my friends who come together with me. Thanks to Renjie Guo, Li Gao, Yu Chen, Danyang Liu, Ming Li, Jin Peng, Yawei Dai, Yancong Zhang,

Xuan Li, Chiye Yuan, Jinlong Ru, Haoran Du, Zixuan Fan. I enjoyed all the time we spent together.

Finally, I wish to thank my parents, Huaiguang Liu and Sumei Huang. Thank you very much for your supporting, understanding, and helping. I could not make it without your standing with me.

Contents

List of Tables

List of Figures

Abbreviations

AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
BH	Benjamini-Hochberg
BP	biological process
CCS	circular consensus sequencing
CPM	counts per million
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
CTD	Comparative Toxicogenomics Database
DE	differential expression
DEG	differentially expressed gene
DEHERV	differential expressed HERV
FPKM	fragments per kilobase of exon per million mapped fragments
GO	gene ontology
gRNA	guide RNA
H-protein	hemagglutinin protein
HERV	human endogenous retrovirus
HIV	human immunodeficiency virus
HTLV	human T-lymphotropic virus
IAV	Influenza A virus
kbp	kilobase pair
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	gene knockout
LFC	log2 fold change
lincRNA	long intervening non-coding RNA
lncRNA	long non-coding RNA
LTR	Long terminal repeat
MaLR	mammalian apparent LTR retrotransposon

miRNA	microRNA
MOI	multiplicity of infection
MPO	Mammalian Phenotype Ontology
MS	Multiple sclerosis
N-protein	Neuraminidase protein
NGS	next generation sequencing
NT-3	neurotrophin-3
PA	pathologic aging
PBS	primer-binding site
PCA	principal component analysis
PD	Parkinson's disease
PPT	polypurine tract
PR8M	influenza A/PR/8/1934/H1N1
PSP	progressive supranuclear palsy
qPCR	quantitative PCR
RIG-like receptor	retinoic acid-inducible gene-I-like receptors
RIN	RNA integrity number
RNA-seq	RNA sequencing
SC35M	influenza A/seal/Mass/1-SC35M/1980/H7N7
shRNA	short hairpin RNAs
TCX	temporal cortex
TGS	third-generation sequencing
TLR	toll-like receptor
TPM	transcripts per million
TSS	transcription start site

Chapter 1 Background

1.1 Human endogenous retroviruses

Human endogenous retroviruses (HERVs) are the remnants of ancient exogenous retroviruses infecting human germline cells and remaining in the genome. Normative HERV element generally consists of the internal portion encoding the viral genes and two flanking LTRs (Long terminal repeat) (Figure 1.1). The insertion of many HERV elements can be traced back 30 million years ago (Bannert and Kurth 2006). Thus, due to the accumulation of point or structure mutations, many HERV elements in the genomes are not intact. They present as non-LTR proteins, solo-LTRs, or only small segments of them, all of which form as much as 8% of the human genome (Belshaw et al. 2004), and solo LTRs are most common in the human genome.

Similar to many exogenous infective retroviruses, most types of HERVs contain gag, pol, and env three viral genes in the internal portions. The gag gene encodes the capsid, which forms the main structure of the virus (Yamashita and Emerman 2004). The pol gene encodes the necessary enzymes for reverse transcription (reverse transcriptase), integration (integrase), and some protease for viral protein maturation (Krausslich et al. 1989). Env gene encodes the proteins forming the envelope of the virus, which was involved in the signal identification, and membrane fusion when infecting the host and induces the host's immune response (Kamps, Lin, and Wong 1991). In addition to the viral genes, the internal region also contains a primer-binding site (PBS) and a polypurine tract (PPT) (Figure 1.1). The PBS is about 8-21 bp in length and located between the 5' LTR and the gag gene, which acts as the binding site for the cellular tRNA and initiates the minus-strand DNA synthesis. While the PPT is an 8-22 bp long purine-rich region at the 3' end of the internal portion, serving as the primer of the plus strand synthesis (Grandi and Tramontano 2018; Wilhelm et al. 1999).

Each LTR flanking the internal portion consists of the U3, R, and U5 regions. The U3 region harbors enhancer and promoter, so the transcript of HERV provirus was initialized in this region of 5' LTR. R region contains the signal for polyadenylation.

As a result, the transcript ends in the R region of the 3’ LTR. The function of the U5 region was not clearly identified, while some researchers inferred that the U5 region of 3’ LTR assisted the polyadenylation process (Böhnlein, Hauber, and Cullen 1989).

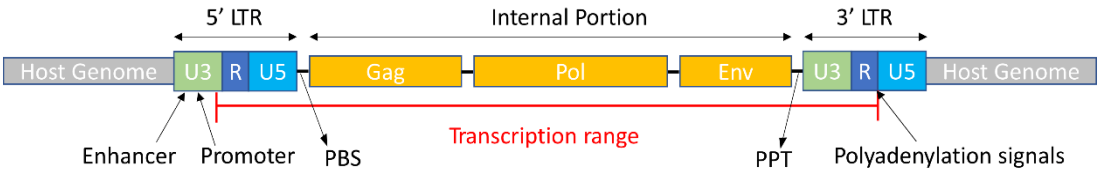


Figure 1.1 Schematic diagram of Human endogenous retroviruses

1.2 Classification of Human endogenous retroviruses

In the past, human endogenous retroviruses were broadly classified into three classes according to their similarity with different kinds of retroviruses (Gifford and Tristem 2003). Class I are HERVs that are phylogenetically clustered with Gammaretrovirus and Epsilonretrovirus. Class II and Class III are HERVs, respectively, clustered with Betaretrovirus and Spumaretrovirus. As more kinds of HERVs were identified, two more classes were introduced: Gypsy Class contains HERVs closely related to Errantivirus, and all the other HERVs that were not clearly classified fall into the last “Unclassifiable class” (Vargiu et al. 2016).

While the five-class classification system is too simplistic for the study of HERVs further detailed classification has long been confusing. Many HERV elements in human genomes accumulated many mutations since integration and were presented as partial segments. Different individual HERV elements also went through recombination during history, which made it even harder for classification. On the other hand, for a long time, there has been a lack of a unique naming system for HERVs, leading to various HERV names originating from different research groups, research purposes, or research methods. Commonly, the wet-lab experiments identified intact HERV elements in the genome with both internal portion and flanking LTRs existing. HERVs identified in this methodology were usually grouped with the consideration of the whole sequences and were usually named according to the identification method.

In comparison, most bioinformatic methodologies for HERV identification were

designed to recognize the repetitive segments in the genomes like RepeatMasker or RetroTector (Smit 1999; Sperber et al. 2007). As a result, HERVs identified in this way, as those recorded in the Repbase (Jurka 2000; Jurka et al. 2005), the LTRs were grouped and named independently from the internal portions. Although this strategy is convenient for the clustering of LTR and internal portion, it loses the association between them and causes confusion when combining these data with HERVs originating from wet-lab identification.

To address the mussy classification and ambiguous naming system of HERVs, Vargiu et al. induced "Simage" (similarity image) methodology. They classified 3173 "canonical" and "noncanonical" (HERVs that went through recombination events) HERV elements detected by RetroTector software (Sperber et al. 2007) into 39 HERV groups based on the five classes (Vargiu et al. 2016). They organized the naming of each group and aligned the internal portions with the LTR elements detected in silico. Afterward, these groups and alignments were further optimized the according to the Repbase database by Kojima et al. (Kojima 2018). Nowadays, public databases like DFam kept rolling updates to gradually include the newly discovered HERV elements and improve the precise alignment between internal portion and LTRs (Hubley et al. 2016).

Taking advantage of Kojima's work and the Dfam database, we can explore the activities and biological characteristics of each group of HERVs using the HERV database (Paces et al. 2004; Paces, Pavlícek, and Paces 2002). The HERVd research group optimized the fragmented HERV segments in the Repbase and RepeatMasker and provided a precise annotation of coordinates and other information of each HERV locus. However, the offline annotation obtained from HERVd does not contain information on the association between the internal portions and LTRs, which makes it challenging to analyze the partial HERV elements like solo LTRs. Thus, we introduced the association information in the HERVd annotation by mapping the HERV types that appeared in HERVd to the HERV groups defined in the study of Kojima et al. HERV types that were not successfully mapped to Kojima's HERV group were manually corrected according to the Dfam database. The detailed organized results is listed in Table S2.1.

1.3 Activities of Human endogenous retroviruses

It has been reported that endogenous retroviruses remain infectious in other mammals. For example, researchers have observed the mobilization of murine endogenous retroviruses upon exogenous retrovirus infections (Bock et al. 2018; Boi et al. 2016; Evans et al. 2009). And the infectious capability of porcine endogenous retroviruses can also impact organ xenotransplantation (Güell et al. 2017; Miyagawa et al. 2005). In contrast, although many human endogenous retrovirus loci are equipped with the complete basis of retrotransposon-like LTRs and viral genes, their transposon capability was strictly restrained. No cases of HERV moving or replicating have been reported ever since it was first discovered. Nevertheless HERV influence the host in many other ways. Some HERV activities offer positive effects on the physiological operation. For example, some of the cis-regulatory signals harbored in HERV LTRs have been integrated into the host expression regulation network (Feschotte and Gilbert 2012; Jung et al. 2017; Kurth and Bannert 2010). It has also been reported that HERV protein expression can assist cell antiviral defense: HERV protein could occupy the membrane receptors to defend the cell from exogenous retrovirus infection (Blanco-Melo, Gifford, and Bieniasz 2017; Hilditch et al. 2011). However, HERV activities bringing advantages to the host only happen in some limited circumstances. More common activities are the abnormal expression of HERVs observed in the pathological context.

1.3.1 Abnormal expression of HERVs in pathological disorders

Although there are few solid proofs of human endogenous retroviruses directly inducing diseases (Young, Stoye, and Kassiotis 2013), abnormal expression of different types of HERVs have been widely reported in the pathological context like tumors, viral infections, or neurological disorders, etc. (Ferrari et al. 2019; Küry et al. 2018; Nellåker et al. 2006). In comparison, in most healthy human tissues the expression of most HERV loci are strictly suppressed at a low level due to the methylation (Gimenez et al. 2010; Lavie et al. 2005). In tumors, Most HERV expression cases in tumors are tumor-associated rather than tumor-specific; in other words, most observed HERV expressions in tumors are also detected expressing at

a low level in health tissues (Gimenez et al. 2010; Gloor et al. 2017). For example, the rec and np9 transcripts originated from the abnormal splicing of env mRNA, which was suggested to contribute to tumor development (Armbruester et al. 2002; Lower et al. 1995; Magin, Löwer, and Löwer 1999; Yang et al. 1999), were also observed in normal human cells (Schmitt et al. 2015). As mentioned, little evidence supported the direct causal role of HERV over-expression to oncogenesis. And the ubiquitous hypomethylation in the tumor tissue was suggested to be the major lead to the activation of HERVs (Attermann et al. 2018; Kassiotis 2014). Nevertheless, due to the overexpression of viral particles, HERVs are still suggested to be potential biomarkers for cancer screening or targets for immunotherapy (Curty et al. 2020; Golkaram et al. 2021; Smith et al. 2018).

Except for tumors, HERV activities are closely related to neurological disorders. The elevated expressions of HERV protein or nucleotide acid were widely observed in neurological diseases like Amyotrophic lateral sclerosis (ALS), Multiple sclerosis (MS), Alzheimer's disease (AD), etc. (Golkaram et al. 2021; Gröger and Cynis 2018; Li et al. 2015; Mayer et al. 2018). And the homozygous of HERV-K18 was reported to lead to three times the risk of Multiple sclerosis (Antony et al. 2011). HERV sequences usually harbored transcription binding sites which can recruit transcription binding factors and regulate the cellular immune response (Imhof et al. 1999; Wang, Wang, and Adamo 2000). Additionally, several studies reported the different HERV expression patterns between old-age and young-age people (Autio et al. 2020; Nevalainen et al. 2018), and the disordered expression of HERV-derived proteins, mostly env protein, can also stimulate the host immune response (Römer 2021). The chronic imbalanced activation of the immune system is inferred to induce severe neuroinflammatory, which is considered a chief pathological component of the majority of neurodegenerative diseases (Bjelobaba, Savic, and Lavrnja 2017; Gröger and Cynis 2018; Morris et al. 2019). And this is a consensus that most neurological diseases happen in old age. In this dissertation, we also reported the detrimental impact of HERV-K activation on cortical neuron differentiation. We identified the over-expression of NTRK3 caused by HERV-K activation are the critical factor leading to the disruption of the cortical neuron. This project gives a cell-level knowledge of HERV activities and neuron development, which hopefully provides a deeper understanding of the potential pathogenic role

of HERV in neurological diseases.

HERVs also exhibit intense activities in the context of multiple kinds of viral infections, such as human immunodeficiency virus (HIV), influenza virus, human T-lymphotropic virus (HTLV), etc. (Chen, Foroozesh, and Qin 2019; Contreras-Galindo et al. 2007; Nellåker et al. 2006; Toufaily et al. 2011). Generally, the mechanisms of exogenous infection elevating HERV expression include 1) viral elements affect directly or indirectly the transcription binding sites harbored in the LTRs; 2) viral infection alters the epigenetic modification; 3) the responding cell anti-viral immune system affects the HERV expression (Chen et al. 2019). The interactions of HERVs with the cell immune system were widely reported. In some circumstances, the expression of the env gene could competitively combine with the receptor on the celluar membrane to restrain the exogenous infection as mentioned (Ponferrada, Mauck, and Wooley 2003). On the other hand, the expression of the env gene of certain types of HERVs, such as HERV-K and HERV-H, also induces immunosuppression (Lemaître et al. 2017; Mangeney et al. 2001). Through identifying the DEHERV-G pairs, Wang et al. reported a large amount of co-regulation of HERV loci and immune genes upon multiple kinds of exogenous viral infections (M. Wang et al. 2020, 2021). Following this method, in this dissertation, we reported similar co-activation of immune genes and HERVs upon three different strains of influenza infections. This work took a further step to separately analyze the effects on the immune relative genes caused by different HERV groups. We also discussed the possible mechanism by which HERV activities affect the immune system. We detected that LTR12C likely induces the expression of immune genes upon viral infection by providing the promoters.

1.3.2 Transcriptomic analysis of Human endogenous retroviruses and its challenge

The improvement of biotechnology offers researchers diverse technical choices to explore the expression of HERVs. Although the high throughput sequencing technology gradually replaces traditional quantitative PCR (qPCR) analysis, the latter is still cost-effective for getting the expression level of certain HERV groups when the study target is clear. In comparison, the next-generation sequencing (NGS) technology provides the benefit of a large-scale comprehensive analysis of HERV

expressions and their relations with gene expression. And multiple well-established public databases for HERVs, such as HERVd, Repbase, Dfam etc., provide the detailed annotation of HERV loci in the human reference genomes along with the consensus sequences of different HERV types for the downstream bioinformatic analysis (Hubley et al. 2016; Jurka et al. 2005; Paces et al. 2004, 2002). For comparative expression analysis, the traditional read count normalization methods based on feature length and sequencing depth, such as calculating the FPKM or TPM values, were still applied in some studies (Bhardwaj et al. 2015; Deniz et al. 2020; Montesion et al. 2018). In contrast, methods based on the negative binomial distribution, such as DESeq2, edgeR provide more statistically accurate differential expression (DE) analysis of HERVs and gradually become the primary analysis trend (Chen, Lun, and Smyth 2016; Love, Huber, and Anders 2014).

The main challenge of high-throughput HERV transcriptomic analysis happens during the processing of reads alignment. Due to the repetitive property of HERV elements, there are always HERV expression reads mapped to multiple HERV loci of the genome during the alignment process. And this effect is especially critical in the "young" HERV groups, which accumulated fewer mutations and different loci are highly similar. In practice, this multi-mapping issue was addressed using different approaches. A simple way is to discard the multi-mapped reads and use only the unique-mapped reads for quantitative analysis (Engel et al. 2021; M. Wang et al. 2020). This method guarantees alignment accuracy while scarifies many high-quality reads and will lead to global underestimation of HERV expression. Thus, more researchers tend to use statistical approaches to estimate the approximate read counts mapped to HERV elements, and there are several bioinformatic tools available for that purpose. For example, the general-usage software RSEM is designed to deal with multi-mapped reads on different genes or isoforms, which estimates the maximum likelihood value of the mapped reads counts for each feature for the downstream analysis (Li and Dewey 2011). In the project described in chapter 2, we have applied RSEM for the locus-specific differential expression analysis of HERVs. Beyond this, there are also bioinformatic tools designed explicitly for transposon elements like TEtranscripts, SalmonTE, Telescope, etc. (Bendall et al. 2019; Jeong et al. 2018; Jin et al. 2015), which offer more options for HERV expression analysis and Schwarz et al. gave a detailed comparison of

these packages (Schwarz et al. 2022).

The other strategy to avoid the multiple mapping issue is to analyze the expression of different HERV types, which is applied in both projects of this dissertation. This strategy makes sense since the multi-mapped reads are mostly mapped to the sample HERV group. Thus, the sum number of multi-mapped reads along with the unique mapped read will generally reflect the overall expression of a specific group of HERV. Alternatively, expression of different HERV groups can be obtained by mapping the NGS reads to the artificial "HERV genomes" as described in the work of Bhardwaj et al. (Bhardwaj et al. 2015), in which they use extracted sequences of HERV loci to create faux "HERV-K genomes". In the project described in Chapter 2, we created artificial "HERV genomes" of different groups using the consensus sequencing of each type of HERV element obtained from Dfam database (Hubley et al. 2016). This method yields the read count data by mapping the reads to a single individual sequence, which is more properly fitted with negative binomial distribution and suitable for the downstream differential expression analysis with the aid of packages like DESeq2. In summary, this strategy provides reliable group analysis of type-specific HERV expression. It suits the studies focusing on different HERV types and can also be a complement or verification of locus-specific HERV expression analysis.

1.4 Outline of this dissertation

As mentioned in 1.3.1, HERV exhibits abnormal expression in the context of viral infections or neurological disorders, while the knowledge of detailed mechanisms and pathways involved remains to be improved. This dissertation lists two projects discussing the interaction of HERV activities with host cells and their potential role in pathogeny. Chapter 2 addressed the expression pattern of HERVs and their coregulation with host genes in cells infected by three influenza A virus strains. We discussed the HERVs' active effect on the cell anti-viral immune system. In the project of chapter 3, we activated the expression of HERV-K purposely in the cortical neuron cell. We monitored the corresponding cell reaction and transcriptome shift during cell differentiation. We observed the disruption of cortical neuron development upon HERV-K reactivation. We identified that the overexpression caused by the HERV-K reactivation as the critical factor leading to

the interruption.

This dissertation mainly discusses the bioinformatic analysis of HERV expression based on the RNA sequencing (RNA-seq) data and the involved methods. In chapter 2, except for the cell incubation, viral infection, and library preparation for RNA sequencing, most work of this project is in silico analysis, and the results and methods are all detail reported. While chapter 3, on the other hand, described a project that thoroughly combined the wet lab and in silico analysis. The bioinformatics analysis of this project was detailed in chapter 3, while the wet-lab analysis parts were only summarized as a conclusion. In the last chapter, we summarized the work of the two projects and pointed out the limitations of bioinformatic analysis on HERV activities. We discussed the possible improvement of the bioinformatic pipelines to strengthen the solidity of the results and the wet-lab strategies to verify the hypothesis based on the in-silico methods.

Chapter 2 Study of activities of HERV expression in Influenza infected cells

2.1. Introduction

2.1.1 Influenza A viruses

Influenza A viruses (IAVs) are single-stranded RNA viruses that cause respiratory diseases in many avians and mammals (Dou et al. 2018). The subtype of Influenza A viruses was determined by the types of two classes of antigens on the viral surface, the Hemagglutinin protein (H-protein) and the Neuraminidase protein (N-protein). For example, the H7N9 subtype stands for the Influenza A subtype with type 7 H-protein and type 9 N-protein on its surface. So far, scientists have identified 18 types of H-proteins and 11 N-proteins, so theoretically, there are 198 Influenza A subtypes in total. In contrast, not all of them have been discovered in nature (Krammer et al. 2018).

Several subtypes of IAVs are capable of infecting human cells and are used to cause severe pandemics with a large number of deaths and hospitalizations worldwide.

For example, the H1N1 subtype was the pathogen of “The Spanish flu” in 1918 and “Swine flu” in 2009, while the H2N2 caused "The Asian flu" in the 1950s (Javanian et al. 2021). The infections of IAVs often induce the host defense response and lead to severe inflammatory symptoms in the human body, which assists in the restraining of viral replication but also results in lung injury, morbidity, and death (Tavares, Teixeira, and Garcia 2017). To optimize the therapeutic strategy of IAVs caused diseases, a deep understanding of cell-level biological processes or pathways involved in the IAV infection and corresponding cellular immune response is needed.

2.1.2 Human endogenous retroviruses and immune functions

Up to 8% of the human genome are formed by human endogenous retroviruses (HERVs) (Jakobsson and Vincendeau 2022), which are remnants of ancient exogenous retroviral integrations. Normative HERV element consists of the internal portion of encoding the necessary proteins of retroviruses flanked by two long terminal repeats (LTRs), assisting the viral integration processes to the host genome. HERV LTRs often harbor cis-elements and can regulate host gene activities via multiple mechanisms (Jakobsson and Vincendeau 2022). Although most HERV elements are silenced by mutations or epigenetically controlled, they can be reactivated by environmental conditions, including infections of exogenous viruses such as HIV-1 (Vincendeau et al. 2015), hepatitis C (C. Liu et al. 2017; Weber et al. 2021), or IAVs (Li et al. 2014; Nellåker et al. 2006). Previous studies reported that human endogenous retroviruses (HERVs) might be involved in regulating the immune functions upon reactivation during viral infection through different mechanisms (Chuong, Elde, and Feschotte 2016; Hale 2022; Kassiotis and Stoye 2016; Srinivasachar Badarinarayan et al. 2020). HERV-derived nucleic acids can induce the pattern recognition receptors like retinoic acid-inducible gene-I-like receptors (RIG-like receptors) and toll-like receptors (TLR) (Srinivasachar Badarinarayan and Sauter 2021). The former is known as the fundamental pathway of antiviral response against IAVs (Rehwinkel and Gack 2020). On the other hand, proteins encoded by HERV internal portion, such as the HERV-W envelope, have also been reported to induce cytokine production via TLR signaling (Rolland et al. 2006; X. Wang et al. 2021). And these HERV-derived proteins can also be

recognized by the adaptive immune system and induce T or B-cell response (Bonaventura et al. 2022; Saini et al. 2020). Additionally, promoters or enhancers harbored in the LTR regions can impact the expression of inflammatory genes, which can also influence antiviral immunity (Chuong, Elde, and Feschotte 2017; Jakobsson and Vincendeau 2022; Srinivasachar Badarinarayan et al. 2020).

In this work, we studied the effect of IAV infections on HERV expression and the relative regulation of the host gene network, especially the antiviral immune responses. We sequenced the transcriptome of A549 cells infected by three different IAVs: influenza A/PR/8/1934/H1N1 (PR8M), influenza A/seal/Mass/1-SC35M/1980/H7N7 (SC35M), as well as an NS1 deleted version of the latter strain (SC35MΔNS1). We found multiple HERV elements (internal region or LTRs) upregulated in the IAV-infected cell lines compared with mock control, including LTR5 and LTR12C, which were previously reported to infect the immune activities. Additionally, by identifying the DEHERV-G pairs defined by Wang et al. (M. Wang et al. 2020), we explored the co-regulations of host genes and HERVs. The further functional enrichment analysis of the genes that appeared in the DEHERV-G pairs revealed the potential ability of some types of HERV elements to regulate the immune system's response to the viral infection.

2.2. Materials and Methods

2.2.1. Transcriptomic Sequencing of IAV-Infected A549 Cells and Control

Three separate wells of A549 cells were respectively infected for 24h with IAV strains PR8M, SC35M (Scheiblauer, Kendal, and Rott 1995), or SC35M with genetic ablation of NS1 (SC35MΔNS1) (Kochs et al. 2007). All three cells were infected at the same MOI (MOI 3), which allows the viruses to propagate in their natural way. Then the IAV-infected A549 cells and control were sent to single-end RNA sequencing on Illumina NextSeq 2000 platform with 75bp read length.

2.2.2. HERV Annotation

The annotation of each HERV locus was obtained from the Human Endogenous Retroviruses Database (HERVd) (Paces et al. 2004, 2002), including the HERV ID, coordinates and a rough classification of each locus to five top-level superfamilies.

HERVd does not give the association of the internal portion and their corresponding LTRs. To build such association, we organized the information from Kojima's publication (Kojima 2018) and divided the five superfamilies into 22 groups. The association of the internal portion and corresponding LTRs was provided within each group. For the information on HERV elements that were not provided in Kojima's work, we manually made the annotation using the DFAM database (Hubley et al. 2016). Two types of internal proviruses were not associated with any LTRs, and 122 LTR types were not identified as linked to any internal region, which were excluded from this study. Finally, we annotated 103 types of HERV internal portions with 346 associated LTR types. Detailed information is provided in Table S2.1.

2.2.3. RNA-Seq Data Processing

We did quality control and low-quality filtering of the single-end RNA-seq data using the Fastp software (Chen et al. 2018). Reads were mapped to the human genes and HERV loci using the RSEM package (Li and Dewey 2011), with Ensembl version 99 as the human gene annotation (Yates et al. 2020) and HERVd for HERV annotation. RSEM does the read mapping work using the STAR aligner (Dobin et al. 2013), during which the proper parameters were settled so that multiple mapped reads will not be discarded. RSEM estimated the maximum likelihood value of read count for each gene or HERV locus to ease the bias caused by the multiple mapped reads. And it also avoids underestimating the overall HERV expression for downstream analysis, like in the studies that use only uniquely mapping reads on HERVs.

Type-specific expression of HERVs was measured by creating artificial "HERV genomes" for different types of HERVs from the consensus sequences obtained from the Dfam database (Hubley et al. 2016). Then, the RNA-Seq reads were mapped to the faux genome with RSEM to calculate the expected read count value for each type of HERVs. In the downstream analysis, these type-specific read counts were handled together with the human gene read counts, while the HERV loci were analyzed separately. All the human genes or HERVs (including HERV types or HERV loci) that have no read mapped in all samples were not included in the analysis.

2.2.4. Differential Expression Analysis

The differential expression (DE) analysis of genes, HERV loci, or different types of HERV elements was identified using the DESeq2 R package (Love et al. 2014). Two read count matrices were built for the separated DE analysis as mentioned above: one contains read counts mapped to each HERV locus among all samples, while the other has read counts mapped to human genes and different types of HERV consensus sequences. Accordingly, a design matrix was created identifying the information of each sample, including cell type and strains of infected viruses. The calculated $\log_2$ fold change (LFC) values were further shrunk using the "apeglm" algorithm provided in the DESeq2 package (Zhu, Ibrahim, and Love 2019). Entries with shrunk LFC values > ±1 and s-values of $< 1 \times 10^{-3}$ were considered as differentially expressed. For the expression analysis of each type of HERV element, we kept only the types with average value mapped read counts of all samples larger than 10 (basemean > 10).

2.2.5. Association of differentially expressed LTRs types with genes

We selected 18 types of HERV LTRs (Table S2.2), each of which was either detected up-regulated itself or its associated internal portion was up-regulated. We performed a series of Fisher's exact tests to check the potential association of each up-regulated LTR type with nearby DEGs. The p-values of Fisher's test were further adjusted for multiple tests using Benjamini–Hochberg (BH) method. A detailed scheme of Fisher's exact test is listed in Table 1. LTR type with significant fisher odds value larger than 1 mean that this type of LTR have a higher probability occuring within 100 kbp of DEG's transcription start sites (TSSs) than randomly, indicating a potential regulating role of this LTR type.

Model of the Fisher's exact test	Numbers of genes that have specific DE HERV-LTR (e.g., LTR13) appearing within 100 kbp of TSS	Numbers of genes that do not have specific DE HERV-LTR (e.g., LTR13) not appearing within 100 kbp of DEG TSS
Significantly differentially expressed genes	Number of DEGs having LTR13 within 100 kbp of their TSS	Number of DEGs having no LTR13 within 100 kbp of their TSS

Genes that were NOT significantly differentially expressed	Number of non-regulated genes having LTR13 within 100 kbp of their TSS	Number of non-regulated genes having no LTR13 within 100 kbp of their TSS

Table 2.1 Model of the Fisher's exact test for a given DE HERV-LTR type (using LTR13 as an example).

2.2.6. Motif Analysis of Up-Regulated HERV-LTRs

We performed motif enrichment analysis using HOMER package (version 4.11) (Heinz et al. 2010). The coordinates of the selected HERV loci were provided to "findMotifsGenome.pl" HOMER script as the "target sequence set", and HOMER automatically selected the "background sequence set" from the human reference genome. Through searching all the possible oligos in the target set and counting the occurrence, HOMER determines the enrichment of each motif through fisher's exact test. Since HERV elements are repetitive sequences, the parameter "-mask" was not called in this analysis.

2.2.7. Identification of Differentially Expressed HERV and Gene Pairs (DEHERV-G Pairs)

The DEHERV-G pair was first defined in the study of Wang et al. (M. Wang et al. 2020). Following their method, we found the nearest differentially expressed gene of each DE HERV locus within 100 kbp (distance of the closest edge of the gene and HERV locus) on the same strand. If a DE HERV locus occurred within the region of a DEG and they are on the same strand, they will also be counted as a DEHERV-G pair.

2.2.8. Functional Enrichment Analysis

All the GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses in this work were handled using the modEnrichr website Suite (Kuleshov et al. 2019).

2.3. Results

2.3.1. Identification of Differentially Expressed Genes (DEGs) and HERVs (DEHERVs) of IAV infected cells

We studied the transcriptome profile of IAVs infected cells using the A549 cell line. We sequenced the transcript ones of three cell lines infected by three different strains of IAVs as well as the mock control using single-strand RNA-seq. The three IAVs contain two wild types of IAV strains, PR8M and SC35M, along with SC35MΔNS1, a mutant of SC35M. Compared with SC35M, SC35MΔNS1 lacks the NS1 protein, which can interfere with the cell expression of antiviral factor PAF1 and suppress cell immune response (Hale et al. 2008; Marazzi et al. 2012). So, theoretically, it is predictable that the infection of SC35MΔNS1 would lead to a more significant immune response than SC35M.

We built two feature-sample read count matrices and subjected them to the DESeq2 package to explore the transcriptome profiles of each sample. The first matrix contains read counts for the host genes and HERV consensus sequences obtained from Dfam database, and the other lists the read counts for each individual HERV locus (see methods for the details). We first did principal component analysis (PCA) on the two read matrices to generally explore the expression differences among samples. PCA analysis on both matrices clustered samples in different situations (Figure 2.1 A,C), except for one mock control sample exhibited as an outlier. After removing the outlier sample, the PCA analysis yielded a better cluster (Figure 2.1 B,D). Thus, we excluded that sample for the downstream analysis.

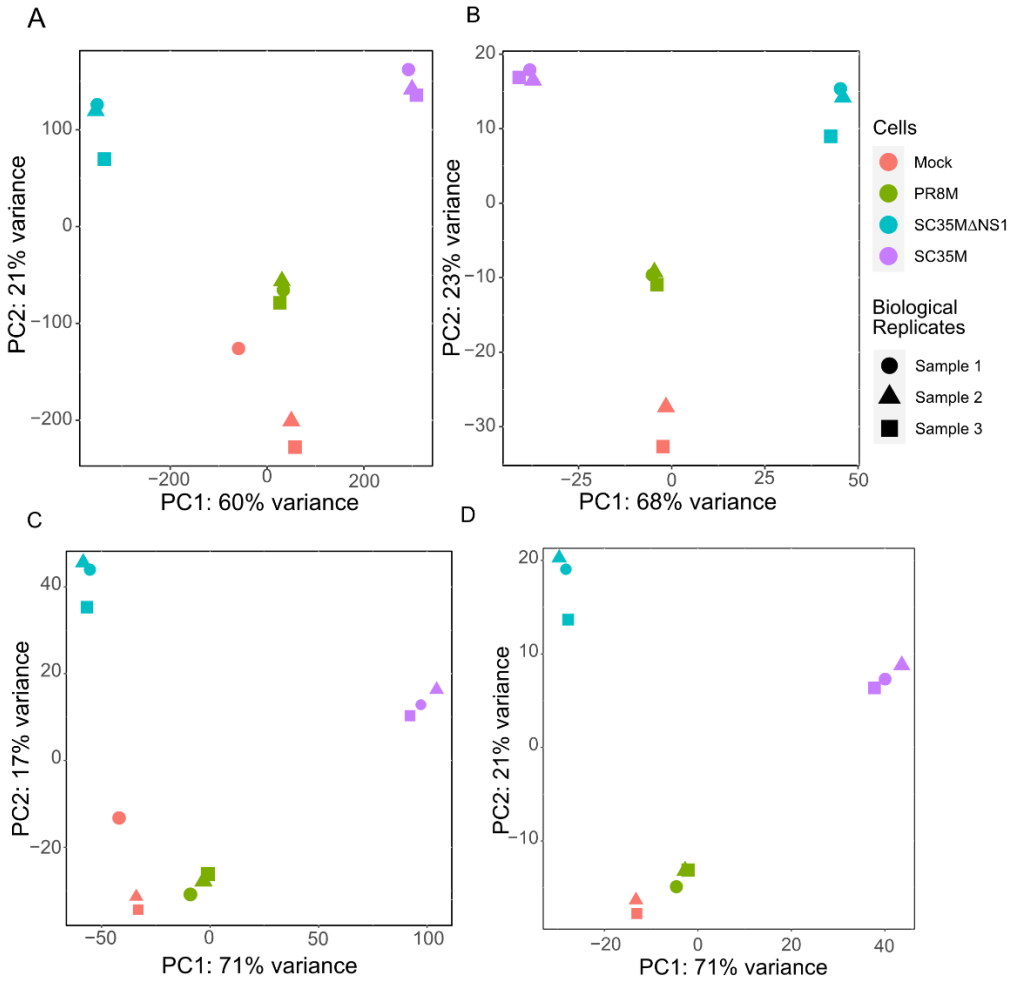


Figure 2.1 Principal component analysis (PCA) plots of the gene matrix (A, B) and HERV loci matrix (C, D). (A, C) PCA plots with all samples included. (B, D) PCA plots with the control sample 1 removed. Only one technical replicate is shown in all plots. See Figures S1 and S2 for the gene matrices of the other two technical replicates.

Next, we performed the differential expression analysis of human genes and HERVs between each virus-infected cell and the control cell with the DESeq2 package (Love et al. 2014). Genes or HERV loci with shrunk Log_2 fold change > ±1 and s-value < 0.001 were identified as DEGs or DEHERVs, and their amounts of each sample comparison were shown as Volcano plots (Figure 2.2). Most of the DEGs and DEHERVs in all three infected cells are upregulated compared with the mock control cells. The protein-coding genes take the lead portion of all the DEGs (86% in PR8M, 75% in SC35M, and 90% in SC35MΔNS1), followed by the non-coding RNAs as the second top (Table S2.3). In comparison, there are no common

leading DEHERV groups among the three infected cells. Most of the DEHERVs in PR8M and SC35M are mammalian apparent LTR retrotransposons (MaLRs) from the ERV3 superfamily, while in SC35MΔNS1, HERVW9 group from ERV1 takes the dominant (Table S2.4, Figure 2.3). There are 495 and 19 common DEGs and DEHERVs among the three infected cells (Figure 2.4). Compared with the overlap, the numbers genes or HERV loci that differentially expressed exclusively in one cell are much higher, indicateing the significant difference in responding to regulating networks of cell upon different viral strain infections. For the functional analysis, we did GO (gene ontology) biological process (BP) enrichment analysis of the joint 495 DEGs and each cell-exclusive DEG set. The overlapping DEG set was enriched in viral immunity relative processes and pathways according to the enrichment analysis of GO and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Figure 2.5). As for cell exclusive DEG set, all the top 10 enriched BP entries are closely related to cell immune response, while the other two cell-exclusive DEG sets did not have any immune relative BP terms among their top 10 enrichment results (Figure 2.6), which is as predicted that the SC35MΔNS1 infection caused the largest cell immune response due to the lack of NS1 protein.

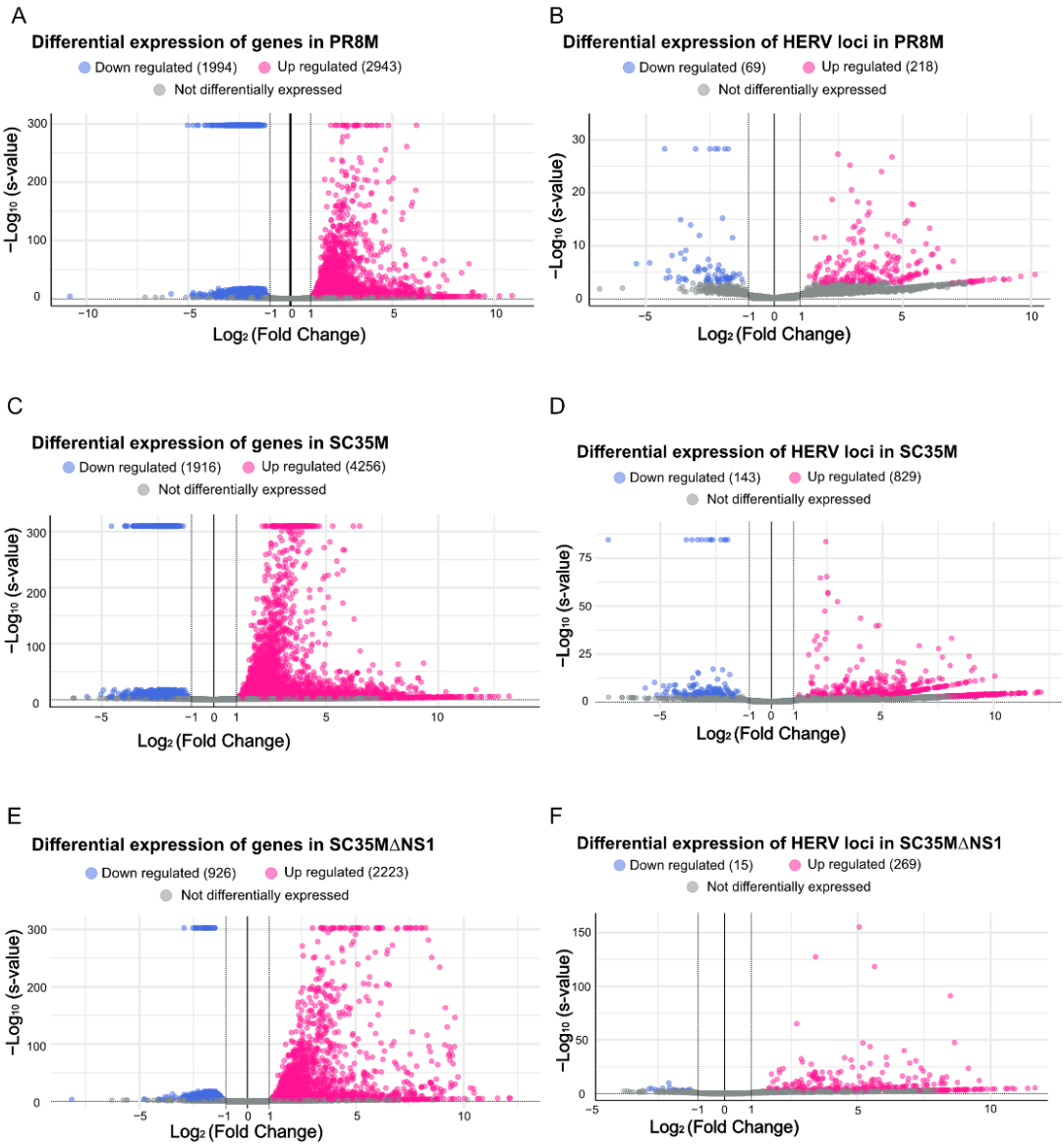


Figure 2.2 Volcano plots illustrating the differential expression of genes and HERV loci in all three cell lines compared to the mock control. (A,B) Differential expression genes and HERV loci in PR8M. (C,D) Differential expression of genes and HERV loci in SC35M. (E,F) Differential expression genes and HERV loci in SC35MΔNS1. DESeq2 converts very small s-values to 0, which leads to 580 genes in PR8M, 343 in genes in SC35M, and 75 genes in SC35MΔNS1 all having the same -Log10 (s-value) and appearing at the top.

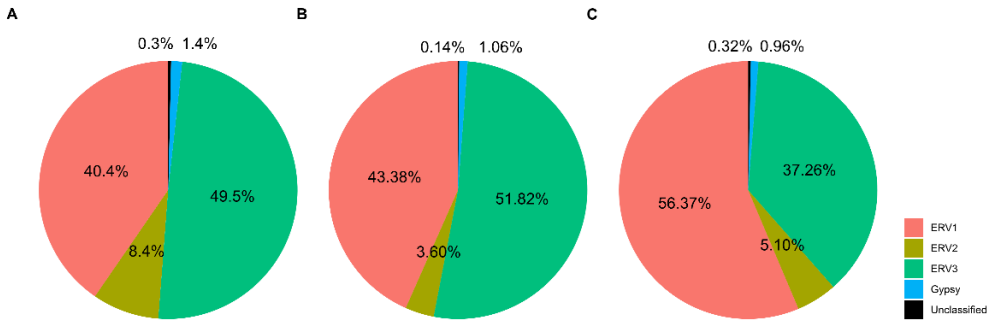


Figure 2.3 Percentage of DEHERVs in each HERV superfamily in the cells infected by PR8M (A), SC35M (B), and SC35MΔNS1 (C).

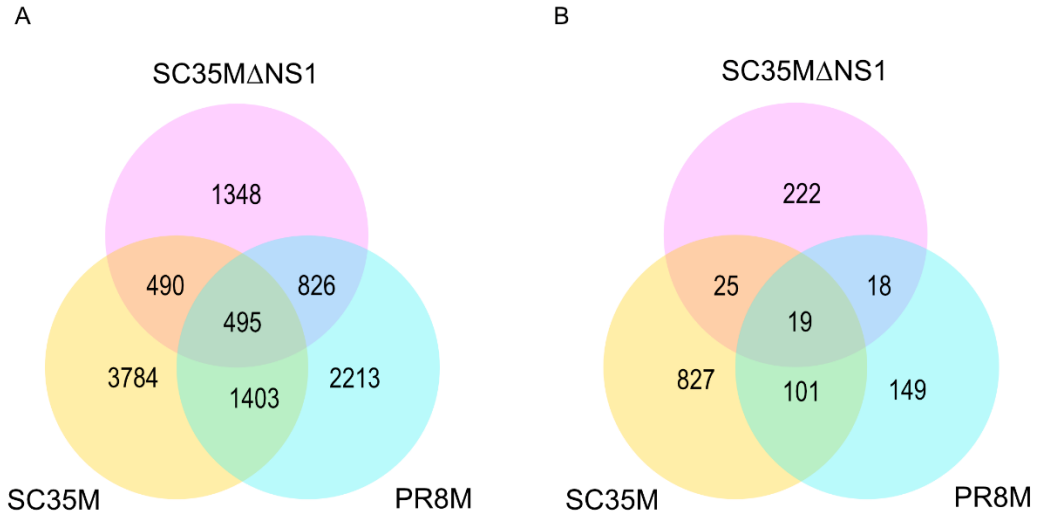


Figure 2.4 Venn diagrams showing the overlap between the differentially expressed genes and HERV loci in the three infected cells (SC35M, SC35MΔNS1, and PR8M) relative to control. A) Differentially expressed genes. B) Differentially expressed HERV loci.

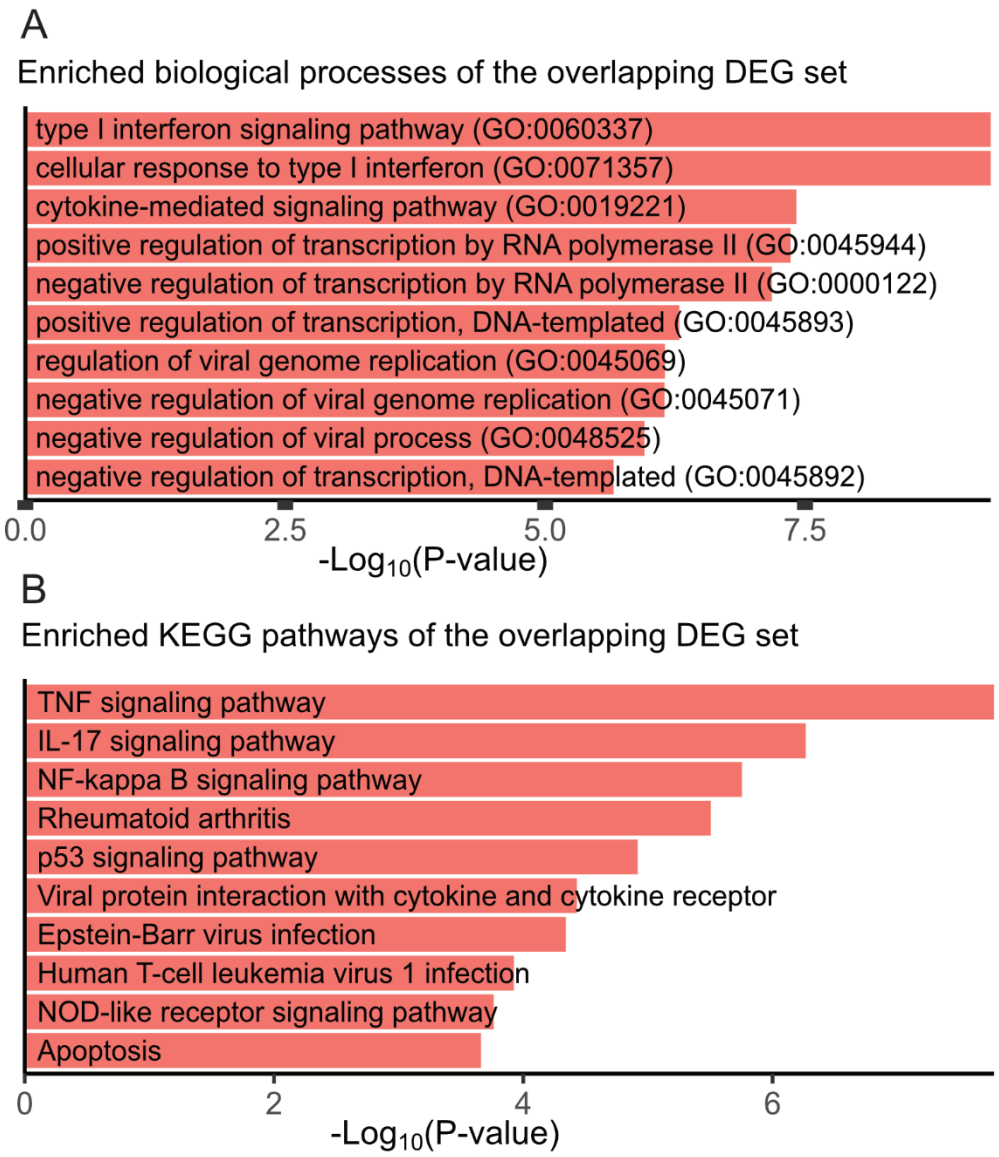


Figure 2.5 Top 10 most significantly enriched biological process GO terms (A) and KEGG pathways (B) associated with the overlapping exclusively differentially expressed gene set of all three infected cells. Terms are ranked according to the p-value.

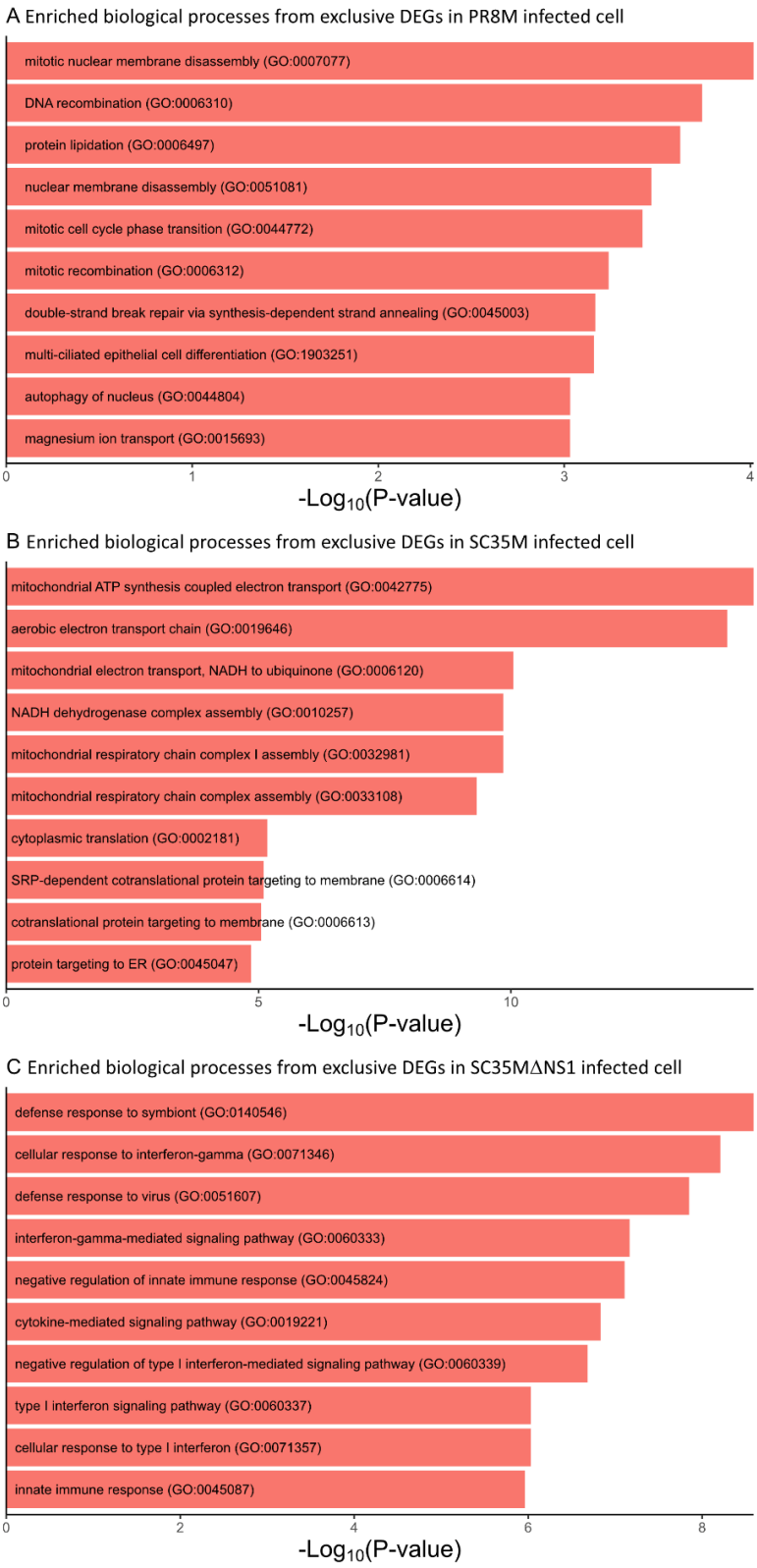


Figure 2.6 Top 10 most significantly enriched biological process GO terms associated with the genes exclusively differentially expressed in A) PR8M infected cells, B) SC35M infected cells, C) SC35MΔNS1 infected cells. Terms are ranked according to the p-value.

2.3.2. Type-specific differential expression analysis of HERVs and the regulatory potential of DEHERVs

In the last section, we described our DE analysis of each HERV locus annotated in HERVd (Paces et al. 2004, 2002). Beyond that, we explored the type-specific expression regulation HERVs by mapping the RNA-seq reads to the consensus sequences of different HERVs. This strategy provides us with a different perspective to understand the roles among different kinds of HERV elements in the gene regulation network. All the significant DE HERV types identified are upregulated cases (Table 2.2-2.4). Notably, the two wild strains infected cells yielded similar results, while the mutant strain (SC35MΔNS1) infection caused less diversity of DEHERV types.

HERV element	Base mean	Log_2(Fold change)	s-value
LTR5_Hs	222.37	1.6	4.92E-19
LTR12C	127.68	1.61	1.87E-06
LTR5	85.17	1.97	1.24E-16
HERV9NC-int	36.45	2.21	6.10E-06
LTR13	33.07	1.52	2.19E-05
LTR5A	30.53	1.66	3.94E-06
LTR7B	23.45	1.68	8.20E-04
LTR10C	20.28	2.8	2.58E-06

Table 2.2 Differentially expressed HERV elements in PR8M vs MOCK. Entries with low read counts (base mean < 20) were discarded.

HERV element	Base mean	Log_2(Fold change)	s-value
LTR5_Hs	222.37	3.46	1.66E-233
HERVH-int	164.25	3.02	3.92E-109
LTR12C	127.68	3.58	6.74E-60
LTR5	85.17	3.59	4.55E-86
HERV9NC-int	36.45	3.21	4.48E-11
LTR13	33.07	2.93	1.53E-31
LTR5A	30.53	3.57	3.68E-46
LTR7B	23.45	5.03	2.02E-48
LTR10C	20.28	6.37	3.57E-34

Table 2.3 Differentially expressed HERV elements in SC35M (wild type) vs MOCK. Entries with low read counts (base mean < 20) were discarded.

HERV element	Base mean	Log_2(Fold change)	s-value
LTR5_Hs	222.37	1.5	4.04E-13
LTR12C	127.68	4.46	1.00E-124
LTR5	85.17	2.32	3.80E-27
HERV9NC-int	36.45	5.7	9.47E-53

Table 2.4 Differentially expressed HERV elements in SC35MΔNS1 (SC35M lacking viral protein NS1) vs MOCK. Entries with low read counts (base mean < 20) were discarded.

Except for a few internal portions like HERV9NC-int or HERVH-int, the majority of significantly upregulated HERV types are LTRs. Since HERV-LTRs usually contain cis-regulatory elements like enhancers and promoters (Deniz et al. 2020; Kim 2012; Schön et al. 2009), we decided to investigate the potential co-regulation of the DE LTRs and DEGs. We did a series of fisher's exact tests to examine whether the loci of certain up-regulated types are prone to exist within the close region (100 kbp upstream or downstream) of DEGs' transcript binding sites (TSS). If one type of LTR yielded a significant result of fisher's test (BH adjusted p-value < 0.05) with the odds value larger than 1, that LTR type was considered positively associated with DEGs and was more likely to appear near the DEGs' TSSs (Table 2.5 as an example). Eighteen types of HERV LTR were subjected to these fisher's tests (Table S2.2). Five and one were respectively identified as positively associated with DEGs in SC35M and SC35MΔNS1 infected cells (Table S2.5). No positive results were detected in PR8M infected cells. Interestingly, the only type of LTR detected in SC35MΔNS1, LTR12C, was also among the five positive results in SC35M. We further performed HOMER motif analysis on the up-regulated HERV loci for the 18 HERV-LTRs (Heinz et al. 2010). The most enriched motif in up-regulated HERV loci in all three infected cells was the NFY(CCAAT), known as a promoter.

	PR8M	SC35M	SC35MΔNS1
Number of up-regulated DEGs having up-regulated LTR12C loci within 100 kbp of their TSS	10	26	22
Number of genes having up-regulated LTR12C loci within 100 kbp of their TSS	59	84	122
Number of up-regulated DEGs	2943	4256	2233
Fisher's test odds value	1.63	2.34	3.25
Fisher's test *p*-value	0.15	8.34×10^{-4}	8.74×10^{-7}
Adjusted *p*-value	1	3.54×10^{-3}	1.49×10^{-5}

Table 2.5 LTR12C occurrence information used in the Fisher's test. The total number of genes considered in this test is 26423. See Methods for details on Fisher's test model.

2.3.3. Pairwise co-regulation of Differentially Expressed HERVs and Genes (DEGERV-G)

We next studied the coregulation of DEGs and DEHERVs on the locus level using DEHERV-G pairs mentioned in the research of Wang et al. (M. Wang et al. 2020). For each DEHERV locus, we searched for the nearest DEG within 100 kbp on the same DNA strand to form a DEHERV-G pair. In SC35M, we identified 593 DEHERV-G pairs, which is much more than the other two cells (168 in PR8M and 115 in SC35MΔNS1), which was supposed to be due to the most occurring of DEGs and DEHERVs in SC35M infected cell (Figure 2.2). The identified DEHERV-G pairs or the genes involved in the DEHERV-G pairs differed greatly among the three infected cells (Figure 2.8). In most DEHERV-G pairs, the contained gene and HERV locus are synchronously up-regulated (Table 2.6), which is also in line with the previous results that the majority of differentially expressed genes or HERV loci are up-regulated ones. The distance distribution of the DEG and DEHERV in the DEHERV-G exhibits a downward trend (Figure 2.7).

Cells	DEHERV-G Pairs				Involved Genes	
	G+H+	G+H-	G-H+	G-H-	Up Regulated	Down Regulated
PR8M	103	8	26	31	97	54
SC35M	451	32	84	26	391	100

SC35MΔNS1	104	1	5	5	88	10

Table 2.6 Differential expression status of DEHERV-G pairs in the three cell types. Symbols +/- denote up/down-regulation of genes or HERVs, respectively. e.g., G+H+ indicates pairs in which both a HERV locus and a gene are up-regulated.

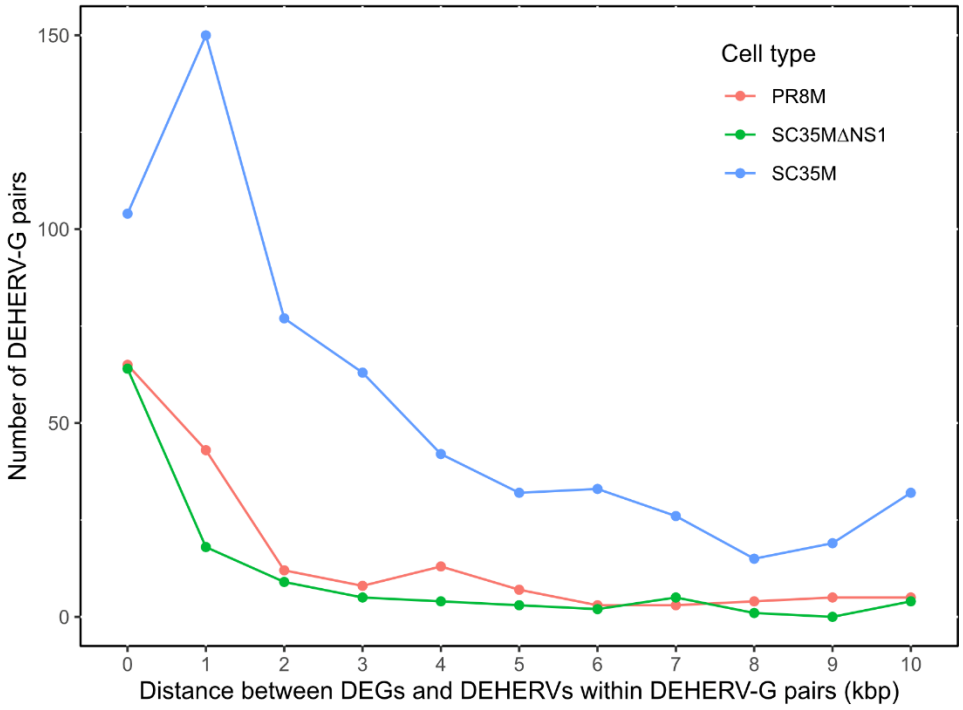


Figure 2.7 Distribution of distances between DEGs and DEHERVs within DEHERV-G pairs. Zero distances correspond to overlapping DEGs and DEHERVs.

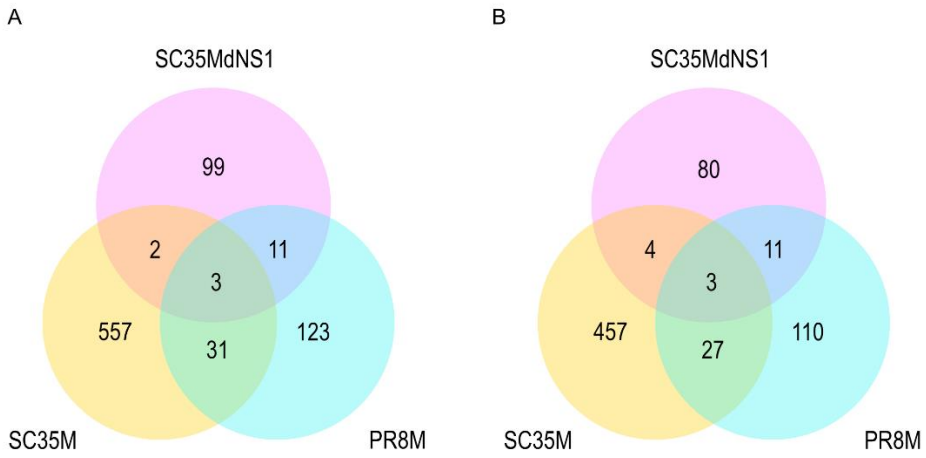


Figure 2.8 Overlaps between DEHERV-G pairs (A) and genes involved in DEHERV-G pairs (B) among the three IAV infected conditions vs mock.

We further did functional analyses of the identified DEHERV-G pairs in the three infected conditions. The protein-coding genes that appeared in the DEHERV-G pairs with both genes and HERV loci up-regulated (G+H+ pairs) were subjected to the GO and KEGG enrichment analysis. The top ten most significantly enriched biological processes and KEGG pathways were listed in (Figure 2.9, 2.10). We found that all the ten most significant biological processes in SC35MΔNS1 and more than half in PR8M infected cells were closely related to immune or inflammatory processes. In comparison none was directly associated with those in SC35M. The KEGG enrichment analysis of this cell yields three enriched immune or inflammation-related pathways from the DEHERV-G proteins in SC35MΔNS1: “NF-kappa B signaling pathway”, “RIG-I-like receptor signaling pathway”, and “NOD-like receptor signaling pathway” inflammation (Brisse and Ly 2019; T. Liu et al. 2017; Saxena and Yeretssian 2014). By contrast, no immune pathway was

detected in the DEHERV-G proteins from the other two viral-infected cells. Similarly, the proteins from the SC35MΔNS1's DEHERV-G pairs were enriched in multiple virus-associated pathways (including "Influenza A pathway"), while in the other two cell types, only one virus-related pathway was significantly enriched ("Coronavirus disease" in SC35M).

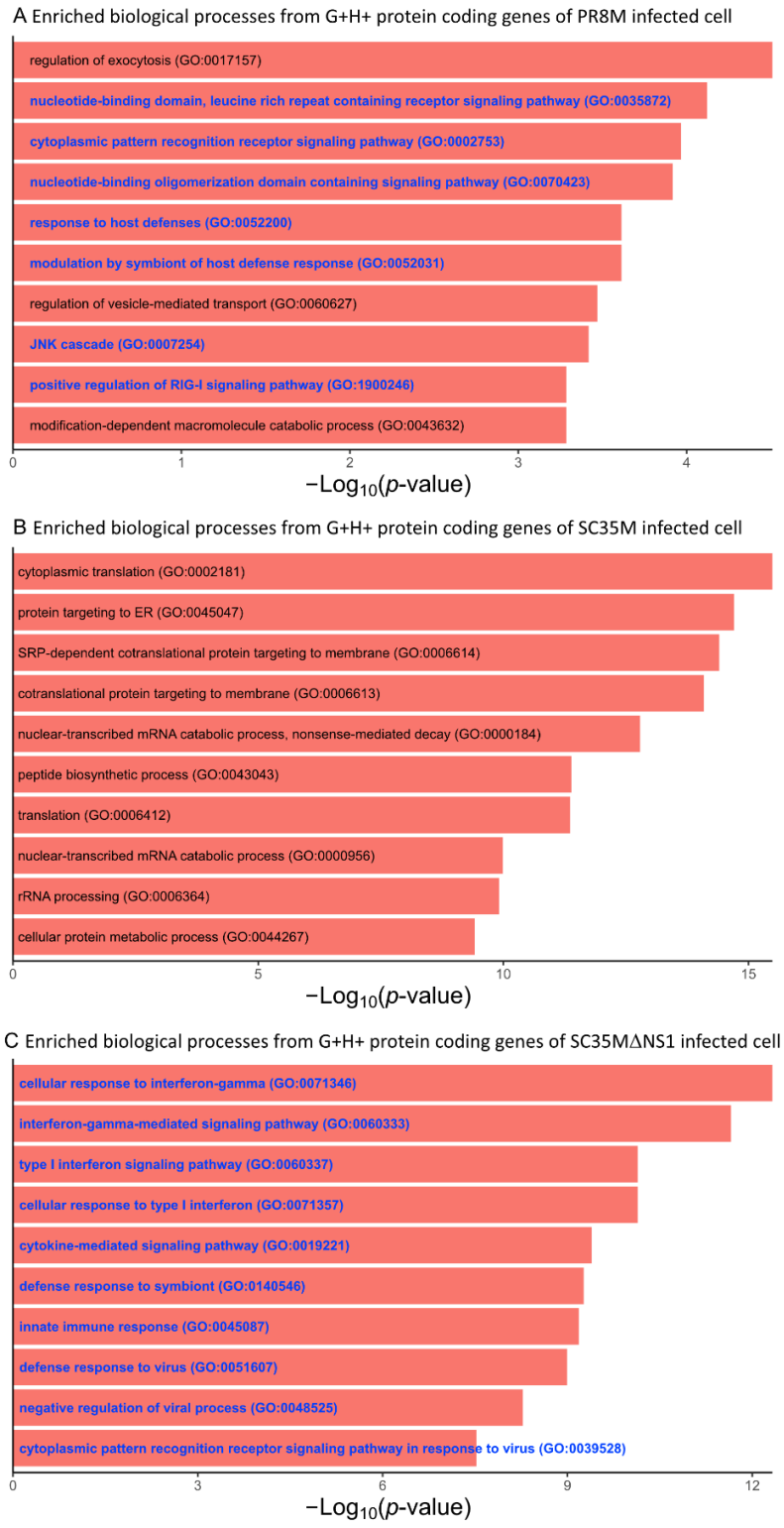


Figure 2.9 Top 10 most significantly enriched biological process GO terms associated with the protein-coding genes of the G+H+ DEHERV-G pairs. Terms are ranked according to the p-value. (A) PR8M infected cells, (B) SC35M infected cells, and (C) SC35MΔNS1 infected cells. Go terms related to immune or inflammation processes are shown in blue color.

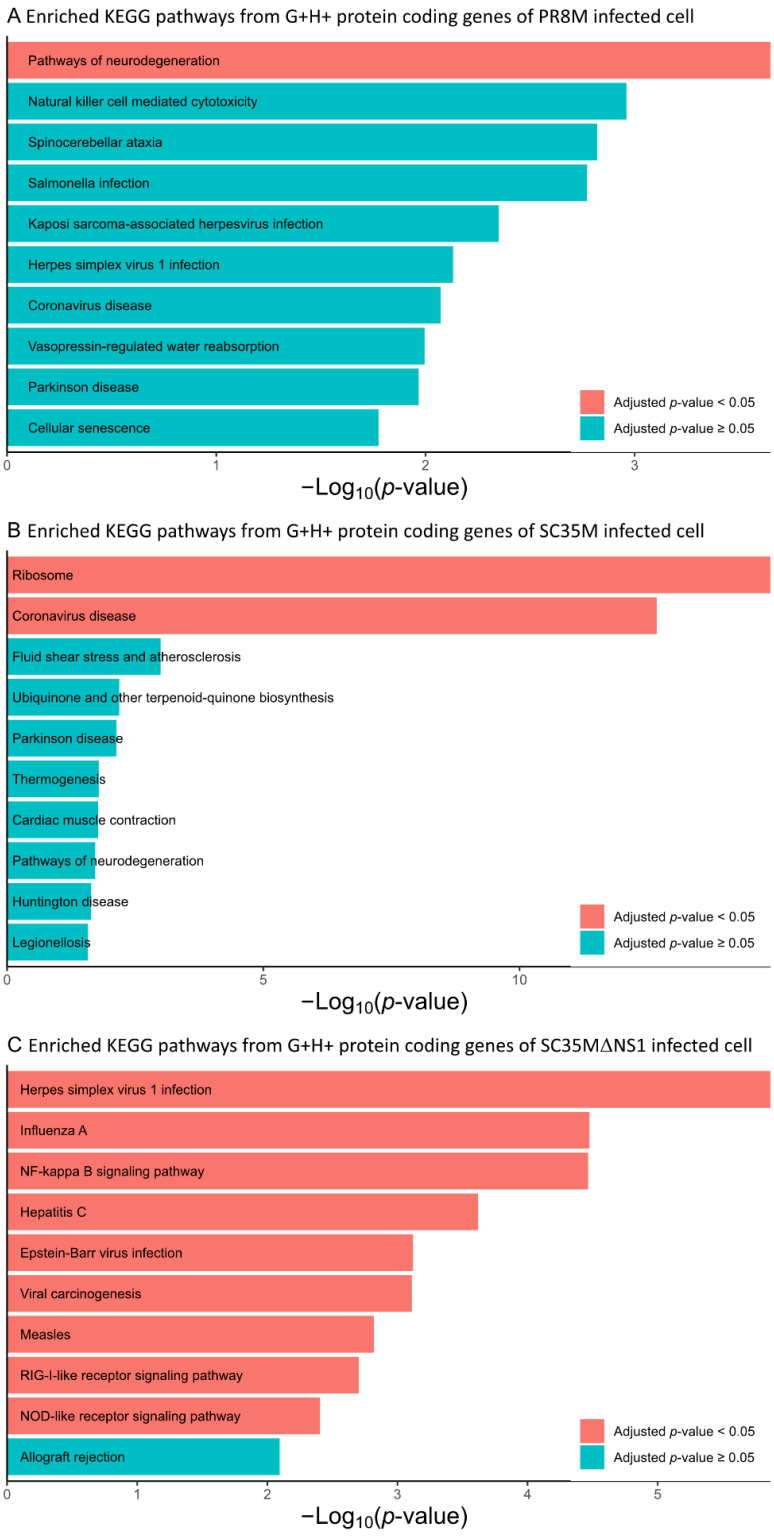


Figure 2.10 Top 10 most significantly enriched KEGG pathways associated with the protein-coding genes of the G+H+ DEHERV-G pairs. Terms are ranked according to the p-value and bars filled in blue means their p-value lower than 0.05 while BH corrected p-values larger than 0.05. (A) PR8M infected cells, (B) SC35M infected cells, (C) SC35MΔNS1 infected cells.

2.4. Discussion

IAVs infection induces inflammation in the host (Tavares et al. 2017). The inflammatory process is part of the immune response that can help to stop the virus replication and spreading. But on the other hand, an overactive inflammatory response can also damage the host or even be lethal (Tavares et al. 2017). It is essential to understand how IAVs infections trigger the inflammatory response. Previous studies widely reported that exogenous viral infections induce abnormal HERV activities, many of which were detected to affect the inflammation processes (Chiappinelli et al. 2015; de Cubas et al. 2020; Lee et al. 2020; Roulois et al. 2015; Schmidt et al. 2019). And the cis-regulatory elements harbored in HERV LTRs were also reported to influence the expression of inflammatory genes (Chuong et al. 2017; Jakobsson and Vincendeau 2022; Srinivasachar Badarinarayan et al. 2020). (Li et al. 2014; C. Liu et al. 2017; Vincendeau et al. 2015; Weber et al. 2021). Thus, in this study, we seek to check the correlation ship HERV activities and cell immunity upon IAVs infection and verify the potential roles of HERV elements regulating cell inflammation. We infected the human A549 cells with three different IAVs strains and detected massive up-regulation of genes and HERV loci. According to the functional enrichment analysis, all three infected cells exhibited immune responses upon the infections since the common DEGs shared among three different infective conditions are enriched in anti-viral immunity. And the SC35MΔNS1 infection induces more severe immune activities than the two wild strains due to the lack of NS1 protein, which was verified by the functional enrichment analysis on the exclusive DEG set of each cell: only the SC35MΔNS1 exclusive DEGs were enriched cell immune response. In addition to the up-regulation of host genes, we also detected abundant up-regulation of HERV loci in all three infected cells. Through identifying the DEHERV-G pairs, we filtered the genes co-upregulating with the activated DEHERVs. We found them enriched in cell immune processes, illustrating the HERV activities with immunity.

The type-specific differential expression analysis of HERV offers a comprehensive understanding of the different effects of different HERV groups. By mapping the RNA-seq reads to the consensus sequences, this method can ease the potential bias caused by the low read count noise in the locus-specific DE analysis and offer a

liable differential expression monitor of each HERV. Consistent with the locus-specific DEHERV analysis results, all significant DEHERV types are upregulated, among which two HERV-K LTRs (LTR5, LTR5_Hs) and LTR12C with its internal portion HERV9NC were detected in all three infective conditions (Table 2.2-2.4). HERV-K (HML-2) is one of the youngest and most active HERV elements in the human genome (Mayer et al. 1999; Subramanian et al. 2011). Several proteins encoded by the HERV-K internal portion were supposed to affect the cell immunity in different mechanisms. For example, the HERV-K-derived dUTPase was reported to activate NF-κB and induce the TH1 and TH17 cytokine response (Ariza and Williams 2011). And the HERV-K gag protein was described to co-assemble with HIV-1 gag proteins and thus reduced its infectivity. Except for the difference between HIV-1 and IAVs, it is reasonable to infer a similar cell effect of HERV-K gags upon IAVs infections.

We also detected a prominent association of LTR12C activation with cell immune activities. Although LTR12C was detected upregulated in all three infected cells, it displayed highest regulation level in the SC35MΔNS1 infected cell (highest LFC value in the type-specific DEHERV analysis and most up-regulated loci in locus-specific analysis), which exhibited the most severe immune activities due to the lack of the NS1 protein as mentioned. Except for the positive correlation of its expression level and the cell immunity level, we also detected that the DEGs in SC35MΔNS1 infected cells are prone to have LTR12C loci within 100 kbp of their TSSs, suggesting a potential regulating ability of LTR12C. This assumption is further convinced by the motif analysis, where we identified that the LTR12C loci were enriched in NFY(CCAAT) motif, which is known to serve as a promoter. Correspondingly, in the SC35MΔNS1 infected cell, we detected two activated LTR12C locus occur closely upstream to two activated immunity relative genes: GBP5 and CXCL11. GBP5 is widely known as a restriction factor against viral infection. The CXCL11 gene encodes a cytokine belonging to the CXC chemokine family, which plays a vital role in the development and function of the immune system. And there are also previous studies pointing to the regulatory ability of LTR12C (Beyer et al. 2016; Jung et al. 2011, 2017; Krönung et al. 2016). All things considered, it is reasonable to believe that the expressions of these two genes are regulated by LTR12C rather than that the LTR12C loci were upregulated due to the

Hitchhiking effect.

In summary, this study explores the regulation of HERV expression upon IAVs infection and their effects on the gene regulation network, especially those contributing to cell defense activities. Adopting locus-specific and type-specific HERV expression analysis provides us with two perspectives on this issue. For the locus-specific HERV expression analysis, we conducted the DEHERV-G pair analysis defined in Wang et al. s studies. We verified the close relationship between HERV expression and cellular immune response. On the other hand, the type-specific analysis relieved much low expression noises in the locus-specific analysis, and we did identify several HERV types possibly affecting cell immune responses. Although pure bioinformatic analysis cannot forge a solid conclusion, this study provides a clear guideline for the further wet-lab experiment uncovering the exact mechanism of HERV activities affecting the cellular immune response.

2.5 Supplementary materials

Table S2.1 Association of different types of HERV internal portion and LTRs

Superfamily	Group	Internal portion	Associated LTRs
ERV1	HEPSI	ERV24B_Prim-int	LTR24B, LTR24C
ERV1	HEPSI	ERV24_Prim-int	LTR24, LTR24C
ERV1	HERVHF	HERV-Fc1-int	HERV-Fc1_LTR1, HERV-Fc1_LTR2, HERV-Fc1_LTR3
ERV1	HERVHF	HERV-Fc2-int	HERV-Fc2_LTR
ERV1	HERVERI	HERV15-int	LTR15
ERV1	HERVW9	HERV17-int	LTR17
ERV1	HERVERI	HERV1_I-int	HERV1_LTRa, HERV1_LTRb, HERV1_LTRc, HERV1_LTRd, HERV1_LTRe
ERV1	HERVERI	HERV3-int	LTR4, LTR76, LTR61
ERV1	HERVW9	HERV30-int	LTR30
ERV1	HERVW9	HERV35I-int	LTR35, LTR35A, LTR35B
ERV1	HERVFRDlike	HERV4_I-int	MER51A, MER51B, MER51C, MER51D, MER51E, MER61D
ERV1	HERVW9	HERV9-int	LTR12, LTR12B, LTR12C, LTR12D, LTR12E, LTR12F, LTR12_
ERV1	HERVW9	HERV9N-int	LTR12, LTR12B, LTR12C, LTR12D, LTR12E, LTR12F, LTR12_
ERV1	HERVW9	HERV9NC-int	LTR12, LTR12B, LTR12C,

			LTR12D, LTR12E, LTR12F, LTR12_
ERV1	HERVERI	HERVE-int	LTR2
ERV1	HERVERI	HERVE_a-int	LTR2B, LTR2C
ERV1	HERVHF	HERVFH19-int	LTR19
ERV1	HERVHF	HERVFH21-int	LTR21B, LTR21C
ERV1	HERVHF	HERVH-int	LTR7A, LTR7B, LTR7C, LTR7Y, LTR7
ERV1	HERVHF	HERVH48-int	MER48, LTR21A, MER72, MER72B
ERV1	HERVERI	HERVI-int	LTR10B, LTR10B1, LTR10A, LTR10B2, LTR10C, LTR10D, LTR10E, LTR10G
ERV1	HERVIPADP	HERVIP10B3-int	LTR10F
ERV1	HERVIPADP	HERVIP10F-int	LTR10F
ERV1	HERVIPADP	HERVIP10FH-int	LTR10F
ERV1	HERVIPADP	HERVP71A-int	LTR71A, LTR71B
ERV1	MLLV	HERVS71-int	LTR6A, LTR6B
ERV1	HUERSP	HUERS-P1-int	LTR8, LTR8A, LTR8B, LTR73
ERV1	HUERSP	HUERS-P2-int	LTR1, LTR1A1, LTR1A2, LTR1B, LTR1B0, LTR1B1, LTR1C, LTR1C1, LTR1C2, LTR1C3, LTR1D, LTR1D1, LTR1E, LTR1F, LTR1F1, LTR1F2, LTR28, LTR28B, LTR28C
ERV1	HUERSP	HUERS-P3-int	LTR9A1, LTR9B, LTR9C, LTR9D, MER61A, MER61B, MER61E, MER61F
ERV1	HUERSP	HUERS-P3b-int	LTR9, LTR25
ERV1	HERVERI	Harlequin-int	LTR2, LTR2B, LTR2C
ERV1	HEPSI	LOR1-int	LOR1, LOR1a, LOR1b
ERV1	HERVHF	LTR19-int	LTR19A, LTR19B, LTR19C
ERV1	Unclassified	LTR23-int	LTR23
ERV1	HUERSP	LTR25-int	LTR25
ERV1	HEPSI	LTR37-int	LTR37A, LTR37B
ERV1	HEPSI	LTR38-int	LTR38, LTR38A1, LTR38B, LTR38C
ERV1	HERVFRDlike	LTR39-int	LTR39
ERV1	HEPSI	LTR43-int	LTR43, LTR43B
ERV1	HERVHF	LTR46-int	LTR46
ERV1	Unclassified	LTR49-int	LTR49
ERV1	HEPSI	MER101-int	MER101, MER101B
ERV1	Unclassified	MER110-int	MER110, MER110A
ERV1	HEPSI	MER21-int	MER21, MER21A, MER21B, MER21C, MER21C_BT
ERV1	HEPSI	MER31-int	MER31, MER67A, MER67B, MER67C, MER67D
ERV1	HEPSI	MER34-int	MER34, MER34A,

			MER34A1, MER34C, MER34C2, MER34C_, MER34D
ERV1	HEPSI	MER34B-int	MER34B, MER34A, MER34A1, MER34C, MER34C2, MER34C_, MER34D
ERV1	HEPSI	MER4-int	MER4A, MER4A1, MER4A1_, MER4C, MER4CL34, MER4D0, MER4D1, MER4D, MER4E, MER4E1, MER4B
ERV1	HERVW9	MER41-int	MER41A, MER41B, MER41D, MER41E, MER41F, MER41G
ERV1	HEPSI	MER4B-int	MER4A, MER4A1, MER4A1_, MER4C, MER4CL34, MER4D0, MER4D1, MER4D, MER4E, MER4E1, MER4B
ERV1	MER50like	MER50-int	MER50, MER50B, MER50C
ERV1	HERVFRDlike	MER51-int	MER51A, MER51B, MER51C, MER51D, MER51E, MER61D
ERV1	HUERSP	MER52-int	MER52A, MER52B, MER52C, MER52D, LTR27D, LTR27E, LTR27, LTR27B, LTR27C
ERV1	MER50like	MER57-int	MER57A1, MER57B1, MER57B2, MER57C1, MER57C2, MER57D, MER57E1, MER57E2, MER57E3, MER57F
ERV1	MER50like	MER57A-int	MER57A1, MER57B1, MER57B2, MER57C1, MER57C2, MER57D, MER57E1, MER57E2, MER57E3, MER57F
ERV1	HEPSI	MER61-int	MER61C
ERV1	HEPSI	MER65-int	MER65C, MER65A, MER65B, MER65D
ERV1	HERVFRDlike	MER66-int	MER66C, MER66A, MER66B, MER66D
ERV1	HUERSP	MER83A-int	MER83
ERV1	HUERSP	MER83B-int	MER83B, MER83C
ERV1	MER50like	MER84-int	MER84
ERV1	HEPSI	MER89-int	MER89
ERV1	Unclassified	MER92-int	MER92A, MER92B, MER92C, MER92D
ERV1	HERVFRDlike	PABL_A-int	PABL_A
ERV1	HERVFRDlike	PABL_B-int	PABL_B
ERV1	HEPSI	PRIMA4-int	PRIMA4_LTR
ERV1	HERVFRDlike	PRIMA41-int	MER41C, MamRep1151
ERV2	HML2	HERVK-int	LTR5, LTR5A, LTR5B, LTR5_Hs
ERV2	HML8	HERVK11-int	MER11A, MER11B, MER11C

ERV2	HML7	HERVK11D-int	MER11D
ERV2	HML4	HERVK13-int	LTR13, LTR13A, LTR13_
ERV2	HML1	HERVK14-int	LTR14A, LTR14B
ERV2	HML9	HERVK14C-int	LTR14C
ERV2	HML5	HERVK22-int	LTR22A, LTR22B, LTR22B1, LTR22B2, LTR22C, LTR22C0, LTR22C2, LTR22, LTR22E
ERV2	HML6	HERVK3-int	LTR3, LTR3A, LTR3B, LTR3B_
ERV2	HML3	HERVK9-int	MER9a1, MER9a2, MER9a3, MER9B
ERV2	HML10	HERVKC4-int	LTR14
ERV3	HERVL	ERV3-16A3_I-int	ERV3-16A3_LTR, LTR16A, LTR16A1, LTR16A2, LTR16B, LTR16B1, LTR16B2, LTR16C, LTR16D, LTR16D1, LTR16D2, LTR16E, LTR16E1, LTR16E2
ERV3	HERVL	ERVL-B4-int	MLT2B4, MLT2B1, MLT2B5, MLT2C1
ERV3	HERVL	ERVL-E-int	MLT2E, LTR33, LTR33A_, LTR33A, LTR33B, LTR33C, MLT2B1, MLT2B5, MLT2C1
ERV3	HERVL	ERVL-int	MLT2B2, MLT2B1, MLT2B5, MLT2C1
ERV3	HERVL	ERVL47-int	LTR47B, LTR47B2, LTR47B3, LTR47B4, LTR47A, LTR47A2
ERV3	HERVL	HERV16-int	LTR16
ERV3	HERVL	HERVL-int	MLT2A1, MLT2A2, MLT2B3, MLT2C2, MLT2D, MLT2F, MLT2B1, MLT2B5, MLT2C1
ERV3	HERVS	HERVL18-int	LTR18A, LTR18B, LTR18C
ERV3	Unclassified	HERVL32-int	LTR32
ERV3	Unclassified	HERVL40-int	LTR40A, LTR40A1, LTR40B, LTR40C, LTR40a, LTR40b, LTR40c
ERV3	HERVS	HERVL66-int	LTR66
ERV3	Unclassified	HERVL74-int	MER74C, MER74A, MER74B
ERV3	Unclassified	LTR52-int	LTR52
ERV3	Unclassified	LTR53-int	LTR53, LTR53B
ERV3	Unclassified	LTR57-int	LTR57
ERV3	Unclassified	MER68-int	MER68, MER68B, MER68C
ERV3	Unclassified	MER70-int	MER70A, MER70B, MER70C
ERV3	Unclassified	MER76-int	MER76
ERV3	MaLR	MLT1-int	MLT1C, MLT1C1, MLT1C2, MLT1E, MLT1E1, MLT1E1A, MLT1E2, MLT1B, MLT1D, MLT1G, MLT1G1, MLT1G2,

			MLT1G3, MLT1K, MLT1L, MLT1M, MLT1N2, MLT1O, MLT1A, MLT1A0, MLT1A1, MLT1E3, MLT1I
ERV3	MaLR	MLT1F-int	MLT1F, MLT1F1, MLT1F2, MLT1A, MLT1A0, MLT1A1, MLT1E3, MLT1I
ERV3	MaLR	MLT1H-int	MLT1H, MLT1H1, MLT1H2, MLT1A, MLT1A0, MLT1A1, MLT1E3, MLT1I
ERV3	MaLR	MLT1J-int	MLT1J, MLT1J1, MLT1J2, MLT1A, MLT1A0, MLT1A1, MLT1E3, MLT1I
ERV3	MaLR	MST-int	MSTA, MSTA1, MSTB2, MSTB, MSTB1, MSTC, MSTD
ERV3	MaLR	THE1-int	THE1A, THE1B, THE1C, THE1D, MamRep605
Gypsy	Unclassified	MamGyp-int	MamGypLTR1a, MamGypLTR1b, MamGypLTR1c, MamGypLTR1d, MamGypLTR2, MamGypLTR2b, MamGypLTR2c
Gypsy	Unclassified	MamGypsy2-int	MamGypsy2-LTR, MamGypsy2-I

Table S2.2 HERV LTR types involved in the analysis of the association between up-regulated HERV-LTR types and DEGs

HERV LTR	group	superfamily	associated internal portion
LTR13	HML4	ERV2	HERVK13-int
LTR13A	HML4	ERV2	HERVK13-int
LTR7A	HERVHF	ERV1	HERVH-int
LTR7B	HERVHF	ERV1	HERVH-int
LTR7C	HERVHF	ERV1	HERVH-int
LTR7Y	HERVHF	ERV1	HERVH-int
LTR7	HERVHF	ERV1	HERVH-int
LTR10C	HERVERI	ERV1	HERVI-int
LTR12C	HERVW9	ERV1	HERV9NC-int
LTR12B	HERVW9	ERV1	HERV9NC-int
LTR12C	HERVW9	ERV1	HERV9NC-int
LTR12D	HERVW9	ERV1	HERV9NC-int
LTR12E	HERVW9	ERV1	HERV9NC-int
LTR12F	HERVW9	ERV1	HERV9NC-int
LTR12_	HERVW9	ERV1	HERV9NC-int
LTR5_Hs	HML2	ERV2	HERVK-int
LTR5A	HML2	ERV2	HERVK-int
LTR5	HML2	ERV2	HERVK-int

Table S2.3 Counts of differential expression of each type of gene in each cell

Gene type	SC35M	PR8	SC35MΔNS1
protein_coding	4614	4266	2834
lncRNA	956	481	189
processed_pseudogene	267	44	24
transcribed_unprocessed_pseudogene	93	54	45
TEC	64	37	13
unprocessed_pseudogene	39	7	10
snRNA	34	1	1
snoRNA	20	6	5
transcribed_unitary_pseudogene	20	6	3
transcribed_processed_pseudogene	19	12	13
misc_RNA	17	4	4
rRNA	8	8	8
polymorphic_pseudogene	6	3	5
Mt_tRNA	3	1	0
miRNA	3	3	3
Mt_rRNA	2	2	1
ribozyme	2	1	1
scaRNA	2	0	0
unitary_pseudogene	2	0	0
scRNA	1	1	0

Table S2.4 Counts of DEHERV loci of each HERV class in each cell

HERV Group	PR8M	SC35M	SC35MΔNS1
MaLR	131	425	81
HEPSI	47	227	36
HERVL	37	149	32
Unclassified	13	85	14
HERVW9	21	52	112
MER50like	8	41	19
HERVFRDlike	14	39	10
HUERSP	21	37	5
HERVHF	18	33	14
HERVERI	13	23	11
HERVIPADP	4	11	2
HML2	2	10	2
HML4	11	10	6
HML3	7	8	4
HML1	3	5	1
HML8	4	5	2
HML5	0	4	5
HERVS	0	3	0
MLLV	4	3	4
HML10	0	2	0
HML6	6	0	2
HML7	0	0	1

Table S2.5 List of up-regulated HERV-LTR types that tend to appear within 100kbp vicinity of up-regulated DEGs' TSSs

LTR element	HERV group	superfamily	odds	p-value	adjusted p-value
SC35M: 6 types of HERV-LTRs are more likely to appear near the DEGs' TSSs					
LTR12C	HERVW9	ERV1	3.25	8.74E-07	1.48614E-05
LTR13	HML4	ERV2	8.73	5.29E-13	8.98E-12
LTR5_Hs	HML2	ERV2	7.31	3.02E-06	2.57E-05
LTR10C	HERVERI	ERV1	4.52	0.0001633	9.25E-04
LTR12C	HERVW9	ERV1	2.34	0.000834	3.54E-03
LTR5A	HML2	ERV2	3.79	0.006369	2.17E-02
SC35MΔNS1: 1 type of HERV-LTRs are more likely to appear near the DEGs' TSSs					
LTR12C	HERVW9	ERV1	3.25	8.74E-07	1.48614E-05
PR8M: 0 type of HERV-LTRs are more likely to appear near the DEGs' TSSs					

Figure S2.1 Principal component analysis (PCA) plots of the gene matrix (A,B) and HERV loci matrix (C,D) of second technical replicates. (A,C) PCA plots with all samples included. (B,D) PCA plots with the control sample 1 removed.

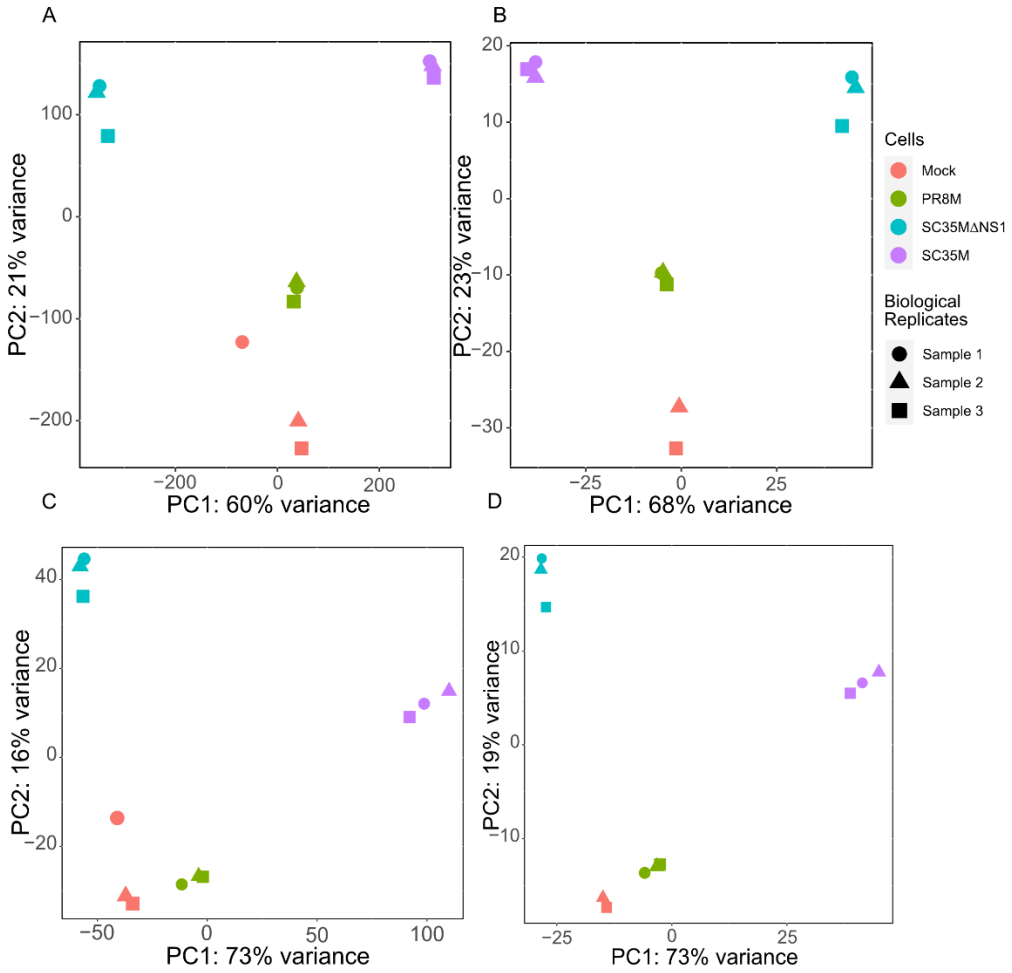
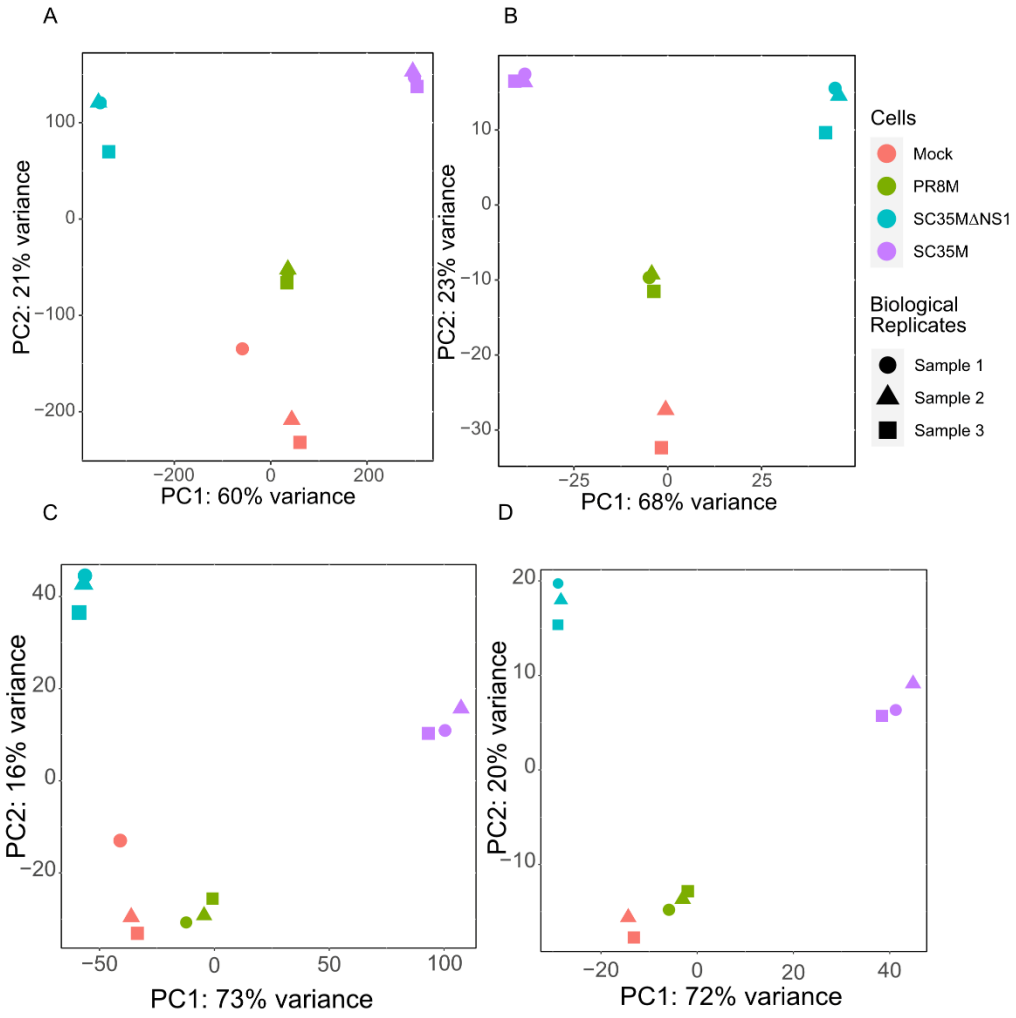


Figure S2.2 Principal component analysis (PCA) plots of the gene matrix (A,B) and HERV loci matrix (C,D) of third technical replicates. (A,C) PCA plots with all samples included. (B,D) PCA plots with the control sample 1 removed.



Chapter 3 Impact of HERV-K activation in neural development

The abnormal expressions of Human endogenous retroviruses (HERVs) in the context of neurological disorders have been widely reported. But the association and the causality of HERV and neurological disease remain to be fully understood. HERV-K (HML-2) is one the youngest and most active HERV groups in the human genome, and its overexpression has been detected in several different neurological diseases such as Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS) or Parkinson's disease (PD). In order to get a deeper understanding of HERV-K's role in these disorders, we purposely elevated the expression of HERV-K elements using CRISPR activation technology, observing its impact on the cell transcriptome during neuron development. We detected an apparent detrimental impact on cortical neuron development and massive differential expressed genes (DEGs) upon HERV-K reactivation. After the filtering and further wet-lab experiments verification, we finally identified the up-regulating of NTRK3 induced by the HERV-K activation to be the critical factor interrupting the neuron function and development.

3.1 Introduction

Human endogenous retroviruses (HERVs) comprise as much as 8% of the human genome and affect the cell and organism in multiple ways. HERV-derived nucleic acid and proteins were reported to influence different biological processes. On the other hand, the cis-elements like promoters or enhancers harbored in the LTR region can also regulate the expression of host genes and thus affect biological functions. HERVs are widely mentioned to be implicated in neurological processes. A few cases of HERVs' functional integration during neurogenesis were reported (Mortelmans, Wang-Johanning, and Johanning 2016; T. Wang et al. 2020). However, more reports were about the links of abnormal expression of HERV elements in neurological disorders such as multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), or Parkinson's disease (PD). Most neurological disorders happen in old-age people; it is hard to determine the causality of these HERV activities and these diseases from the evolution perspective because even the overexpressed HERV elements do play a causal role in these

diseases, they can still survive from the purifying nature selection. Thus, detailed mechanisms and pathways that HERV activities involved in the genesis of the neurological disease remain to be revealed.

HERV-K (HML-2) is the youngest and most active HERV group in the human genomes, which keeps relatively intact virus genes and LTRs (Hohn, Hanke, and Bannert 2013; Tönjes et al. 1996). Although the majority of HERV-K loci stay silent in most healthy conditions, the reactivation of this group of HERV is supposed to be implicated in different types of biological processes, including embryogenesis (Grow et al. 2015), immune response (Arru et al. 2021; Greenig 2019), and diseases like different types of tumors or neurodegenerative diseases as mentioned (Douville and Nath 2014; Küry et al. 2018). For example, through analyzing the transcriptome data of temporal cortex (TCX) samples from AD patients provided by Mayo Clinic Brain Bank, Dembny et al. detected a slight over-expression of the internal portion of the HERV-K group (Dembny, Andrew G Newman, et al. 2020). To reveal the causality of the HERV-K over-expression and these neurological disorders, we intentionally reactivated the expression of HERV-K in the H9 cell and then induced them into cortical neurons to observe the corresponding cell level effects.

In the pilot study, we activate HERV-K expression in H9 cells through CRISPR activation technology (Figure 2.1 A) (Maeder et al. 2013; Perez-Pinera et al. 2013) and then differentiate HERV-K activated cells (H9-dCas9-VP64-gRNA (HERVK-K)) and the non-specific control cells (H9-dCas9-VP64-gRNA(control)) into cortical neurons. After 60 days of cell growth, we observed noticeable morphology changes under the electron microscope and a pronounced reduction of MAP2 staining (Figure 3.1 C, D), indicating large-scale of cell apoptosis (Soltani et al. 2005). By contrast, the H9-derived neuron cells with control CRISPR treatment were not disrupted at the same growing time point (Figure 3.1 C, D). This observation clearly illustrated the impact of HERV-K reactivation on the differentiation of cortical neurons and indicated a potential pathogenic role of HERV-K in neurological disorders.

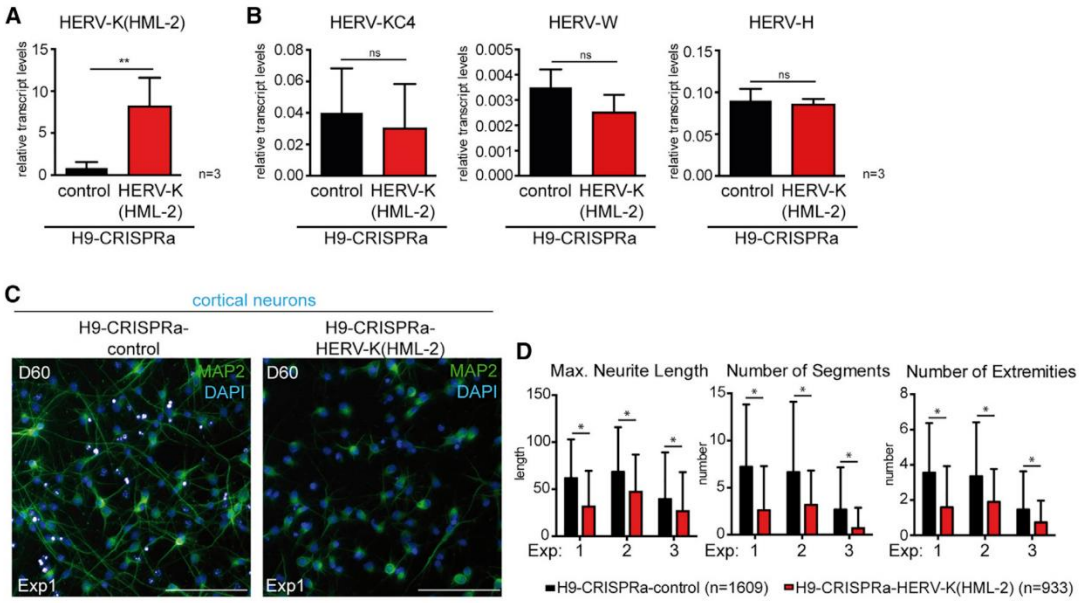


Figure 3.1 (A) qPCR measured expression level of HERV-K after HERV-K reactivation by CRISPR activation. (B) qPCR measured expression level of HERV-KC4, HERV-W and HERV-H after HERV-K reactivation using CRISPR activation. (C) Immunofluorescence analysis of MAP2 expression of cortical neurons derived from H9-CRISPRa-HERV-K cells and control cells. Electron micrograph was taken on day 60 after cell differentiation. (D) Morphological summary of cortical neurons upon HERV-K(HML-2) activation.

To deeply understand the effects of the reactivation of HERV-K expression during the neuron cell development, we sequenced the transcriptome of the H9-derived HERV-K activated neuron cell along with the control neuron cell at different time points. We detected a large amount of differentially expressed genes corresponding to the HERV-K reactivation. We selected 28 top-rank up-regulated DEGs in the last two time points and found that these genes are closely related to neurological functions. Through further wet-lab experiments, we successfully identified the up-regulation of the NTRK3 gene as the crucial factor disrupting neuron cell growth. In addition, we also explore the transcriptome datasets of temporal cortex (TCX) samples from neurological patients provided by the Mayo Clinic Brain Bank. We verified the observation of Dembny et al. that the expression of HERV-K was elevated in the AD patients' brain samples compared with control samples. Moreover, we also detected the overexpression of NTRK3 gene in the old (age of death over 80 years old) female AD samples, which indicates the potential implication of the NTRK3 in the genesis of AD disease.

3.2 Methods

3.2.1 HERV-K annotation

We extracted 1239 HERV-K entries from the HERVd database (Paces et al. 2002), including 61 complete sequences containing the protein coding region flanked by two LTRs, 47 HERV-K elements that lost one LTR on either side, 21 elements with only the internal region preserved, and 1110 elements constituted by solo LTRs.

3.2.2 RNA-Seq data processing

The RNA integrity numbers (RINs) of total RNA samples were measured with the Agilent 2100 Bioanalyzer system. RNA samples with a RIN value higher than eight were sent for cDNA libraries preparation with the Illumina TruSeq RNA Sample Preparation kit. And the paired-end sequencing process was finished on Illumina HiSeq4000 platform. For preprocessing the RNA-seq data, we used fastp software with default parameters for quality control and reads filtering (Chen et al. 2018). The reads were mapped to the human GRCh38 reference genome using the STAR package with all default parameters (Dobin et al. 2013). The comprehensive gene annotation for the primary assembly (chromosomes and scaffolds) was obtained from the GENCODE database version 27 (Harrow et al. 2012). The numbers of reads mapped to genes were counted using featureCounts software (Liao, Smyth, and Shi 2014), during which only read pairs that mapped to the same chromosome were counted (command line options -B, -C, -p set), and multi-mapped read pairs were discarded (option -M not specified). Genes with very low read counts (<10) in all samples were excluded from consideration.

3.2.3 Differential expression analysis between the H9-dCas9-VP64-gRNA (HERVK-K) neuron and H9-dCas9-VP64-gRNA(control) cell

The differential expression (DE) analysis of HERV-K activated cortical cell (H9-dCas9-VP64-gRNA (HERVK-K)) and the control cortical cell (H9-dCas9-VP64-gRNA(control)) on different time points was handled using DESeq2 package (Love et al. 2014). A count matrix containing the read count of each gene across all samples, and a design matrix identifying the sample information, were built and subjected to the DESeq2. For the overall exploration of the dataset, we first applied a regularized-logarithm transformation to the raw read count data to stabilize the

variance across the means of the genes. Then the transformed data were clustered using the Principal Component Analysis (PCA) to visualize the overall differences among each group of samples. For accessing the differential expressed genes, DESeq2 modeled the read counts using a negative binomial distribution with the consideration of sample information in the design matrix. The significance of DEGs was determined using the Wald test, and the p-values were adjusted using Benjamini-Hochberg (BH) methods. The $\log_2$ fold change values were further shrunk using the "apeglm" algorithm to reduce the bias of LFC value in the gene with a low count of mapped reads. For the overall comparison of expression profiles among each timepoint, we applied a rough threshold to determine the DEGs (Log_2fold change > 1 and adjusted p-value < 0.1).

The most up-regulated genes during day 41-60 after cell differentiation were identified according to both the mapped read count and the LFC level. The maximal mapped read count number determined the mapped read count level for each gene in the control cell over the five time points. For example, if for a certain gene the read counts for the control cell on day 0, 10, 27, 41, and 60 were 1512, 1693, 1996, 2063, and 1905, respectively, that gene was represented by the maximum read count number 2063. As for the LFC level, we ranked the LFC values of up-regulated genes in the HERV-K-activated cells compared with control cells separately for each time point. Subsequently, the most up-regulated genes were selected as genes that: a) appear among the top 25% LFC gene list on Day 41; b) appear among the top 25% LFC gene list on Day 60; c) appear among the top 25% read count list defined above. According to this standard, the top 28 up-regulated genes were finally determined (Figure 3.3). Based on the gene and HERV coordinates provided in the human genome annotation and the HERVd database, respectively, we identified the HERV-K LTRs located within a 500 kb vicinity of the transcript start sites (TSS) of these genes (Table S3.2).

3.2.4 Gene function and network analysis

Gene Ontology (GO) and Mammalian Phenotype Ontology (MPO) enrichment analysis of the target-interested differentially expressed genes was performed using the website interface of the modEnrichr suite (Kuleshov et al. 2019; Smith and Eppig 2009; The Gene Ontology Consortium 2018). Disease enrichment analysis

was effected with the help of the Set Analyzer tool provided by the Comparative Toxicogenomics Database (CTD) (Davis et al. 2010). Gene network analysis was performed and visualized by GeneMANIA. We manually curated an interaction network between the identified genes and neurodegenerative diseases based on the information from the CIDeR database (Lechner et al. 2012) and three publicly available text-mining tools: Chilibot (Chen and Sharp 2004), Polysearch2 (Liu, Liang, and Wishart 2015), and Google Scholar. The pathway enrichment analysis was performed based on the Reactome database (Fabregat et al. 2017).

3.2.5 Differential expression analysis of samples from neurological disease patients

We obtained the transcriptome data of neurological disease patients provided by the Mayo Clinic Brain Bank from the Synapse database (Synapse ID: syn7344223). Two organizations contributed to this project, and only the samples from "MayoBrainBank Dickson" was taken for the analysis. It contains 234 temporal cortex (TCX) samples Alzheimer's Disease (AD), progressive supranuclear palsy (PSP), pathologic aging (PA), or elderly controls (CON). Samples with low RNA Integrity Number (RIN <= 7) were discarded due to the low quality. Thus, 82 AD samples, 21 Control samples, 30 PA samples, and 84 PSP samples were left for the downstream analysis. In addition, following (Dembny, Andrew G. Newman, et al. 2020), the PSP samples were regarded as unrelated neurodegenerative disease control, so this group was merged with the control group in the differential expression analysis.

The file type of RNA-seq data provided by the Mayo Clinic Brain Bank was in SNAPR-mapped bam files mapped to GRCh37 reference genomes. Thus, we first extracted the raw reads sequences using SAMtools (Li et al. 2009). Then the reads were mapped to the GRCh38 reference genome using STAR (Dobin et al. 2013). The numbers of reads mapped to genes were counted using featureCounts software (Liao et al. 2014). Only read pairs mapped to the same chromosome were calculated (command line options -B, -C, -p set), and multi-mapped read pairs were discarded (option -M not specified). Then, a count matrix and design matrix were built for the downstream DESeq2 analysis. The read count of the HERV-K internal portion was represented by the sum of reads mapped to all the HERV-K internal loci in the

genome. We obtained the coordinates of all the intact HERV-K internal portions (consensus missing rate lower than 50%) on the human reference genome from HERVd. The read count mapped to each locus was accessed using featureCounts with the same parameters as in the read counting of genes. Then, the read counts of all HERV-K loci were summed to stand for the HERV-K expression of each sample. The read counts of HERV-K in each sample were added to the gene matrix, and DESeq2 was used to determine the significance of HERV-K differential expression.

3.3 Results

3.3.1 The overall expression profiles of the H9-dCas9-VP64-gRNA (HERVK-K) and H9-dCas9-VP64-gRNA(control) diverged largely after 10 days since cell differentiation

We generated HERV-K activated cell line by treating H9 cells using dCas9-VP64 transcription activation factor with HERV-K gRNAs. Similarly, a control cell line was created through dCas9-VP64 treatment with control gRNAs. Thus, the expression of HERV-K in the control cell will not be activated. We first checked the mRNA level using a qPCR experiment and verified the elevation of HERV-K expression. In comparison, the expressions of other HERV elements like HERV-KC4, HERV-W and HERV-H that also originated from gammaretroviruses were not affected by the CRISPR activation treatment (Figure 2.1 A, B) (Padmanabhan Nair et al. 2021).

We then differentiated both cell lines into cortical neurons. The total transcriptomes of both cell lines were sequenced using pair-end RNA-seq technology at day 0, 10, 27, 41, and 60 after differentiation. After preprocessing the data, we first explored the expression pattern of these time-series RNA-seq data using Principal components analysis (PCA). Several types of genes, including protein-coding genes, bidirectional promoter-derived long non-coding RNA (lncRNA), microRNA (miRNA), antisense RNA, and long intervening non-coding RNA (lincRNA), exhibited overall similarity in the PCA analysis: that the main variation of the expression data was contributed by the time after cell differentiation, which indicates the large divergence of gene expression pattern in the first 27 days of neuron cell growth (Figure 3.2). Interestingly, only the protein-coding gene exhibited a clear difference between the HERV-K activated neuron cell and the

control cell after Day 10 (Figure 3.2A). Since Day 27, the distances between cell lines have become more extensive than those caused by differentiation time. Similarly, with a roughly set threshold (Log_2fold change > 1 and adjusted p-value < 0.1), the numbers of significantly differentially expressed protein-coding genes were least at day 0 and day 10 (Table 2.1). These results indicate that the differential expression of protein coding genes between the HERV-K activated cell and control possibly explains the observed HERV-K reactivated cell apoptosis and morphology changes under electron microscope on Day 60.

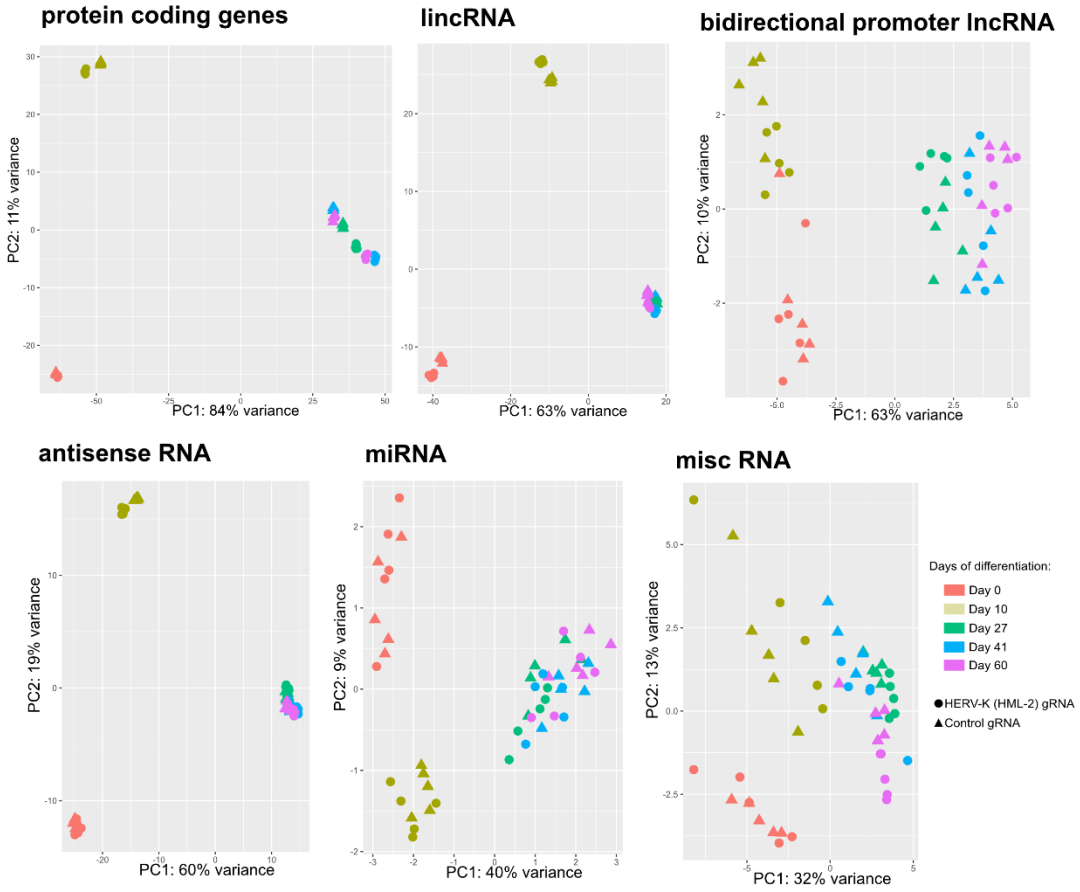


Figure 3.2 Principal components analysis of the expression of different types of genes.

3.3.2. Identification of most significantly up-regulated genes on Day 41 and Day 60

As mentioned, under the electron microscope, we observed a sizeable MDP2 reduction and morphological change in the HERV-K activated neuron cells on day 60 after differentiation. Thus we focused on the differentially expressed protein-

coding genes on day 41 and day 60. To achieve the background mechanism, we filtered of differentially expressed protein coding genes and focused on those strongly upregulated on both day 41 and day 60. Arbitrarily setting an LFC threshold will bias the results towards genes with fewer read counts mapped. Thus, in addition to the LFC value, read counts mapped to the control sample were also considered in the filtering strategy (See method in 3.2.3). After filtering, we finally obtained 28 genes that exhibit the strongest upregulating on both Day 41 and Day 60 (Figure 2.3, Table S3.1).

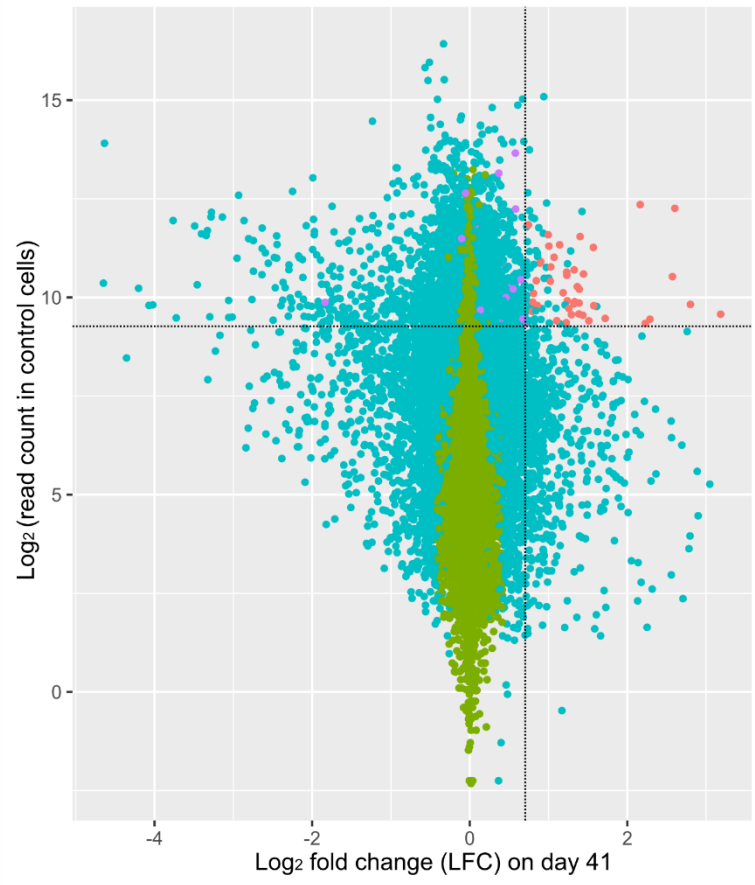


Figure 3.3 Filtered 28 most significantly upregulated genes in HERV-K activated cells compared to control cells on Day 41 and Day 60. Each dot stands for a protein-coding gene. Points above the horizontal dotted line represent the top 25% genes with the highest read count (see method in 3.2.3). The green color stands for the genes that are not significantly identified as DEGs on Day 41 while the blue colors are the significant ones. Genes on the right side of the vertical dotted line are the top 25% upregulated genes on day 41. Purple color corresponds to genes with read counts among the top 25% and that are among the top 25% upregulated genes on day 60. Red dots depict the top 28 genes with read counts among the top 25% that are also among the top 25% upregulated genes on both days 41 and 60.

The gene ontology (GO) enrichment analysis suggested that the 28 genes are closely related to neurobiology like neurogenesis (Figure 2.4). Besides, these genes were also significantly enrolled with several neurological diseases such as neurodegenerative, mental disorder, bipolar disorder, etc., according to the disease enrichment analysis (Figure 2.5). In addition, we further organized the interaction network of the 28 genes related to neurodegenerative diseases in the multifactorial database CIDeR (Lechner et al. 2012) (Figure S3.3). Interestingly, 25 out of the 28 most substantial up-regulated genes are linked directly or indirectly to neurological disorders, which implied that the critical factor leading to this cell disruption was among the 25 genes.

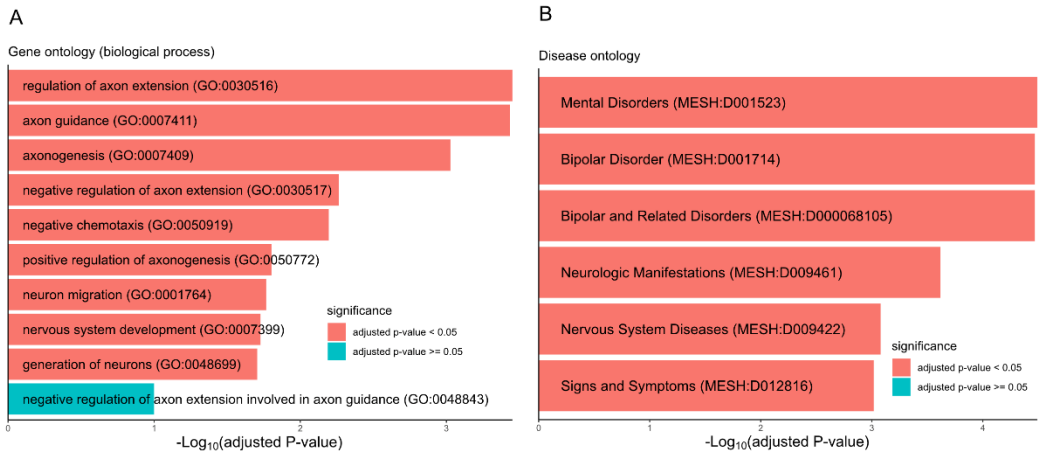


Figure 3.4 (A) Most enriched biological processes of the top 28 most upregulated genes on Day 41 and 60. Entries are ranked according to the adjusted p-value. Longest bar stands for the smallest adjusted p-value (B) Disease enrichment analysis on the top 28 most upregulated genes on Day 41 and 60.

3.3.3 Overexpression of NTRK3 caused by the HERV-K reactivation is the crucial factor disrupting the cortical differentiation

We selected four genes closely related to neuron function from the 28 up-regulated genes for further analysis: CLSTN2 (calsyntenin 2), CHRDL1 (chordin-like 1), EPHA4 (EPH receptor A4) and NTRK3 (neurotrophic tyrosine receptor kinase 3). QPCR experiments verified their differential expressions, and all four genes kept a constantly high differential level of expressing on day 41 and day 60 (Figure S3.1). We simulated the overexpression of the four genes in the h9 cells using the

CRISPRa (CRISPR activation) technology. Then we differentiated them into cortical neurons. During day 40-60 after differentiation, we observed neuron morphology change and the reduced expression of neuron marker gene MAP2 in the NTRK3-activated neuron cell. This observation is similar to what we observed in the HERV-K neuron cell on day 60 (Figure 3.5). In contrast, the other three CRISPRa-treated cells exhibited no or less effect on the expression of MAP2 and cell morphology (Figure 3.5). These results indicated that the elevated expression of the NTRK3 gene in the HERV-K activated is vital in disturbing neuron differentiation. This assumption is further confirmed by knocking out the NTRK3 gene in HERV-K activated neurons. After knocking down the expression of NTRK3, the effects of MDP2 reduction and cell morphology changes were largely reduced in the HERV-K reactivated cells (Figure 3.6). Thus, we can conclude that the disruption of HERV-K activated neuron cell is due to the elevation of NTRK3 expression. This finding has already passed the peer review, and the detailed wet-lab information and discussion can be accessed in (Padmanabhan Nair et al. 2021).

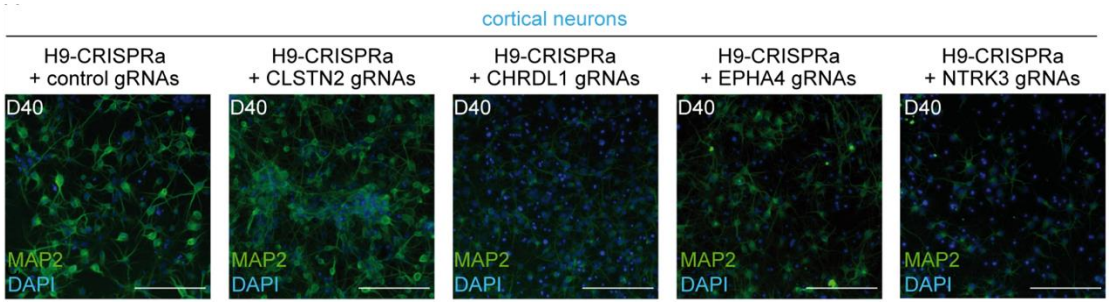


Figure 3.5 MAP2 expression was analyzed by immunofluorescence in neurons and control cells with CRISPR activated CLSTN2, CHRDL1, EPH4 or NTRK3.

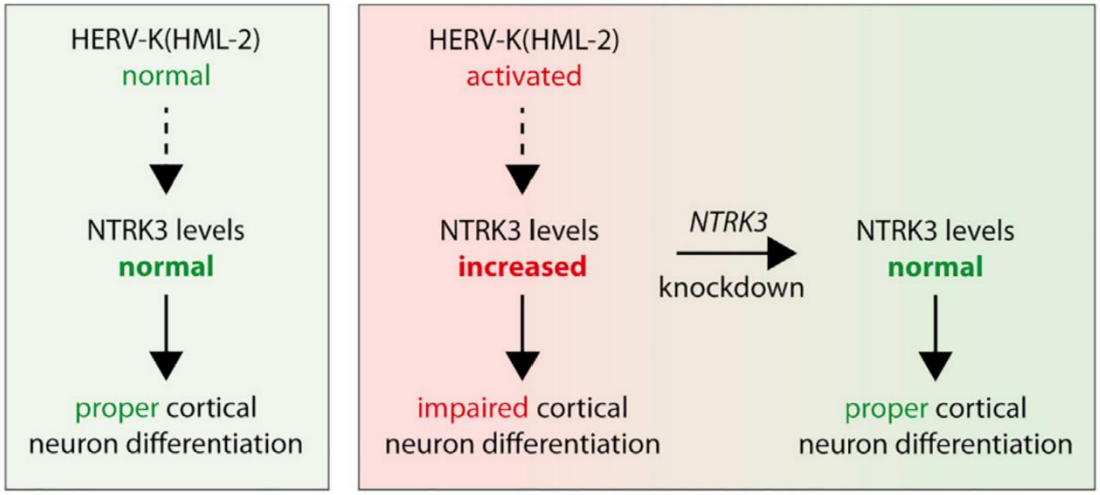


Figure 3.6 Schematic diagram of how HERV-K reactivation is supposed to affect cortical neuron development.

3.3.4. Reactivation of HERV-K induces the overexpression of NTRK3 indirectly

HERV activities affect the host gene expression in different ways. Many HERV-LTRs harbor cis-regulatory elements like promoters or enhancers, and they can directly regulate host gene expression. On the other hand, the overexpression of HERVs can also indirectly affect the expression of host genes. For example, the expression of HERV internal portion, either as HERV-derived proteins or nucleotide acids, can induce the cell's anti-viral immune response (Alcazer, Bonaventura, and Depil 2020; Grandi and Tramontano 2018) and globally affect the host gene expression. In this project, we designed the guide RNA (gRNA) for the CRISPR activation experiment using the internal portion of HERV-K elements rather than the flanking LTRs. Thus, we supposed that the elevated HERV-K-derived proteins or nucleotide acid triggered the up-regulation of NTRK3 in the HERV-K reactivated cell rather than caused by the cis-regulatory effects such as promoting or enhancing harbored in the HERV-K LTRs. To check this hypothesis, we first explored the flanking region of the 28 identified top-level up-regulation proteins. We detected only a limited number of HERV-K LTRs appeared within the 500 kbp vicinity of their transcript start sites (Table S3.2). Notably, an antisense HERV-K LTR5 (LTR5_hs) is located at about 283 kbp upstream of NTRK3 gene. At the same time, theoretically, an antisense promoter can't take effect at this distance. In comparison, enhancers can impact on genes at a far stretch at 500 kbp. To check whether the elevation of NTRK3 expression was caused by the enhancing effect caused by the upstream LTR5_hs, we explored the RNA-seq dataset published by Fuentes et al. (Fuentes, Swigut, and Wysocka 2018). In their study, they parallelly activated and deactivated the LTRs of HERV-K element (LTR5_hs) using CRISPRa and CRISPRi (CRISPR interference) techniques and performed the total RNA sequencing of the LTR5_hs activated/deactivated cells along with non-specific target control cell. Based on their dataset, we did differential expression analysis of the treated cells and control (Table 3.1). Gene NTRK3 was differentially expressed in either LTR5_hs activated or deactivated cells. In combination with the DEG analysis from our data, we found only 9 of the 28 top up-regulated genes

overlapped with the up-regulated genes in the LTR5_hs activated cell. More strictly, if we used the "LTR5_hs controlled gene set" defined in their study (genes that up-regulated in the LTR5_hs activated cell and down-regulated in the LTR5_hs deactivated cell), none of the 28 top up-regulated genes were overlapped. These results illustrated the different transcriptome changes upon reactivation of the HERV-K internal portion (as in this project) and HERV-K LTRs (as in the study of Fuentes et al.). And the overexpression of NTRK3 in the H9-dCas9-VP64-gRNA (HERVK-K) cell was not because of the activation of its upstream LTR5_hs locus.

Cell	up regulated genes	down regulated genes	LTR5_hs controlled genes
LTR5_hs activated	2291	1928	242
LTR5_hs deactivated	345	487	

Table 3.1 Numbers of differentially expressed genes identified based on the RNA-seq data from (Fuentes et al. 2018). The LTR5_hs controlled genes refers to genes that up-regulated in the LTR5_hs activated cell and down-regulated in the LTR5_hs deactivated cell.

3.3.5 Up-regulated HERV-K and NTRK3 in old female patients with Alzheimer's disease

We have identified the overexpression of NTRK3 induced by HERV-K reactivation can negatively impact neuronal differentiation. To take a more profound step, we explored the potential association of this effect with neurological diseases. As mentioned, Dembny et al. detected a slight over-expression of the HERV-K internal portion in the temporal cortex of patients with Alzheimer's disease (AD) using the transcriptome data provided by the Mayo Clinic Brain Bank (citation). We firstly verified the upregulation of HERV-K internal portion in AD patients reported by Dembny et al. (Dembny, Andrew G. Newman, et al. 2020) with an optimized bioinformatics pipeline engaging DESeq2. In contrast with simply comparing the CPM value of HERV-K applied in their study, DESeq2 normalizes the read counts based on the negative binomial distribution and can reduce the batch effect better. DESeq2 also detected significant upregulating HERV-K expression in the AD patients compared with the control group (FDR < 0.1). However, using the same patient-control groups, NTRK3 was not identified as significantly up-regulated in the patient group. If we limited the sample space to only female AD patients whose death age is above 80, we could observe significantly higher expression of NTRK3

and HERV-K internal portion.

3.4 Discussion

HERV-K has been widely reported to be associated with neurological disorders. In this project, we discussed the impact of HERV-K reactivation on neuron cell growth from the cell level, seeking to provide evidence of HERV-K activities inducing neurological disorders. Taking advantage of CRISPR activation technology, we reactivated the expression of the HERV-K internal portion. We monitored the corresponding cell growth conditions and transcriptome profiles at a series of time points. After 60 days after cell differentiation, we observed apparent cortical neuron cell disruption on the HERV-K activated cells while the non-specific control cell stayed normal. And we figured out the overexpression of NTRK3 induced by the HERV-K reactivation to be the critical factor leading to the disruption of neuron differentiation.

Gene NTRK3 encodes tropomyosin receptor kinase C, which binds the neurotrophin-3 (NT-3) and plays an essential role in cortical development (Bartkowska et al. 2007; Nikoletopoulou et al. 2010; Szobota et al. 2019). It is reported that the suppression of NTRK3 will prevent the neuron cell from migrating into the intermediate zone and cortical plate and disturb the cortical precursor cell proliferation (Bartkowska et al. 2007). On the other hand, in this study, we detected that reactivated HERV-K disrupts neuronal differentiation by upregulating the NTRK3 expression. Thus, it can be inferred that the expression of NTRK3 along with NT-3 during cortical development is strictly programmed: too much or insufficient NTRK3 and NT-3 at a particular stage of differentiation will lead to severe injury to the cortical. In this work, we observed the disruption of cortical neurons 60 days after differentiation, while the over-expression of NTRK3 can be traced early from day 27 (Figure S3.2). Thus, we hypothesize that the reactivation of HERV-K induced the too-early up-regulation of NTRK3 and disturbed the rhythm of the cortical differentiation process. Interestingly, by analyzing the transcriptome dataset Mayo Clinic Brain Bank provided, we found the elevation of HERV-K internal portion and NTRK3 in the old female patients with Alzheimer's disease. Although we cannot confirm the causality of HERV-K and NTRK3 over-expression, which is consensus with previous studies reporting that other types of

HERVs are more prone to induce neurological diseases in females (Garcia-Montojo et al. 2013, 2014). This result went against previous studies which reported the downregulation of NTRK3 in Alzheimer's disease (Mufson et al. 2019), which further supposed the expression of NTRK3 in the neuro system is strictly orchestrated. However, the complexity of the association between NTRK3 and neurological diseases remains to be decrypted.

HERV activities affect the expression of host genes in multiple ways. On the one hand, the cis-regulatory elements harbored in HERVs (usually in LTRs), like promoters and enhancers, can induce the expression host gene directly. On the other hand, HERV-derived proteins or nucleotide acids can indirectly influence the cell immune system and thus affect the whole transcriptome pattern. In this project, we got more evidence supporting that the HERV-K reactivation indirectly induced the overexpression of NTRK3. Firstly, in the experiment design we selected the internal portion of HERV-K to make guild RNA (gRNA) for CRISPR activation. As mentioned, most cis-regulatory factors of HERVs locate in the LTRs. Thus, cis-regulatory elements are not likely to be activated by design. Secondly, if the cis-regulatory factors like promoters or enhancers induced the overexpression of host genes, it should take immediate effect after their activation. However, on day 0 after cell differentiation, we detected only three up-regulated genes and NTRK3 was not among them(LFC > 1, BH adjusted p-value < 0.1). And according to the PCA analysis, the HERV-K activated cell and control cell cannot be separated into different clusters on day 0. Thirdly, although an antisense LTR5_hs locus locates in 283 kbp upstream of the NTRK3 gene, we found it cannot affect the expression of NTRK3 according to the analysis of the transcriptome dataset provided by Fuentes et al. (Fuentes et al. 2018). All things considered, we concluded that the cis-regulatory abilities of the HERV-K LTRs take only a minor effect in this experiment. Instead, HERV-K-derived proteins or nucleotide acids induced the upregulation of NTRK3, which started from day 10 until the end of the experiment. The detailed mechanism or pathway involved remains to be revealed.

We optimized the bioinformatics pipeline for the differential expression analysis in this project. Firstly, we adapted our filtering standards identifying DEGs with the consideration of read count numbers mapped to the genes. Arbitrarily setting a fixed

threshold of logarithm fold-change value will ignore the overall differential expression conditions at a particular time point and will cause the biased results to the genes with fewer reads mapped. Shrinking the LFC values will considerably reduce the LFC bias of the low read mapped genes, but simply with that, we cannot find the key factor leading to cortical disruption. Thus, we ranked the genes in consideration of both the shrunk LFC values and the mapped read numbers and successfully selected the top 28 significant up regulated genes closely associated with neurogenesis. We found the over-expression of the NTRK3 gene induced by HERV-K activation as the critical factor leading to cortical disruption in the context of HERV-K reactivation. For the analysis of RNA-seq data of patients with neurological disorders provided by the Mayo Clinic Brain Bank, Dembny et al. simply compared the CPM values of the HERV-K internal portion between patient and control group using the Wilcoxon's test. We verified their results with the aid of DESeq2 package. Taking the advantage of the accurate dispersion estimated from all protein coding genes, DESeq2 made a suitable normalization of the HERV-K read counts data and yielded more convincing differential expression identification.

3.5 Supplementary Materials

Figure S3.1 qPCR verification of the upregulation of the four selected genes on Day 41 and 60 after differentiation

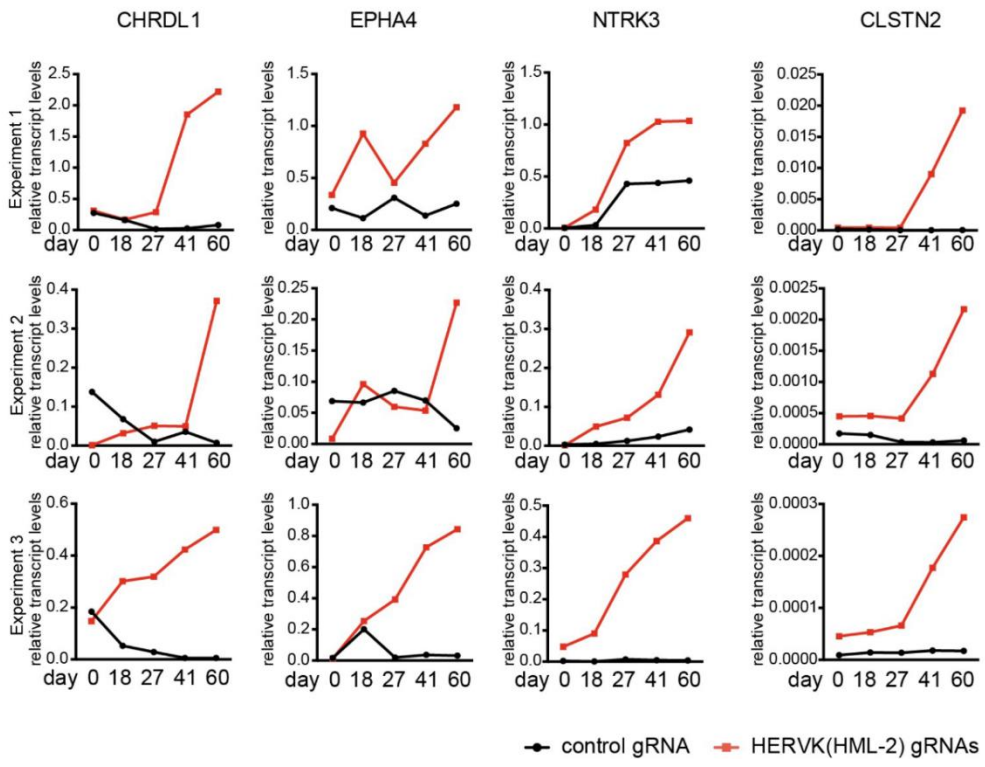


Figure S3.2 Expression tendency of NTRK3 after differentiation

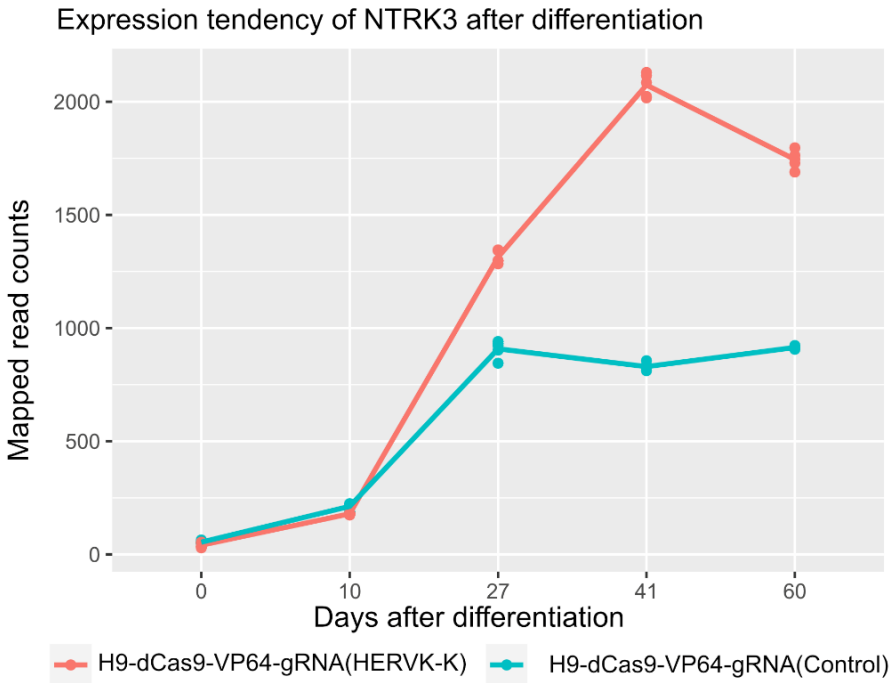


Figure S3.3 Manually curated interaction network of the relationship between the top up-regulated 28 genes during Day 41 to 60 with the neurodegenerative diseases

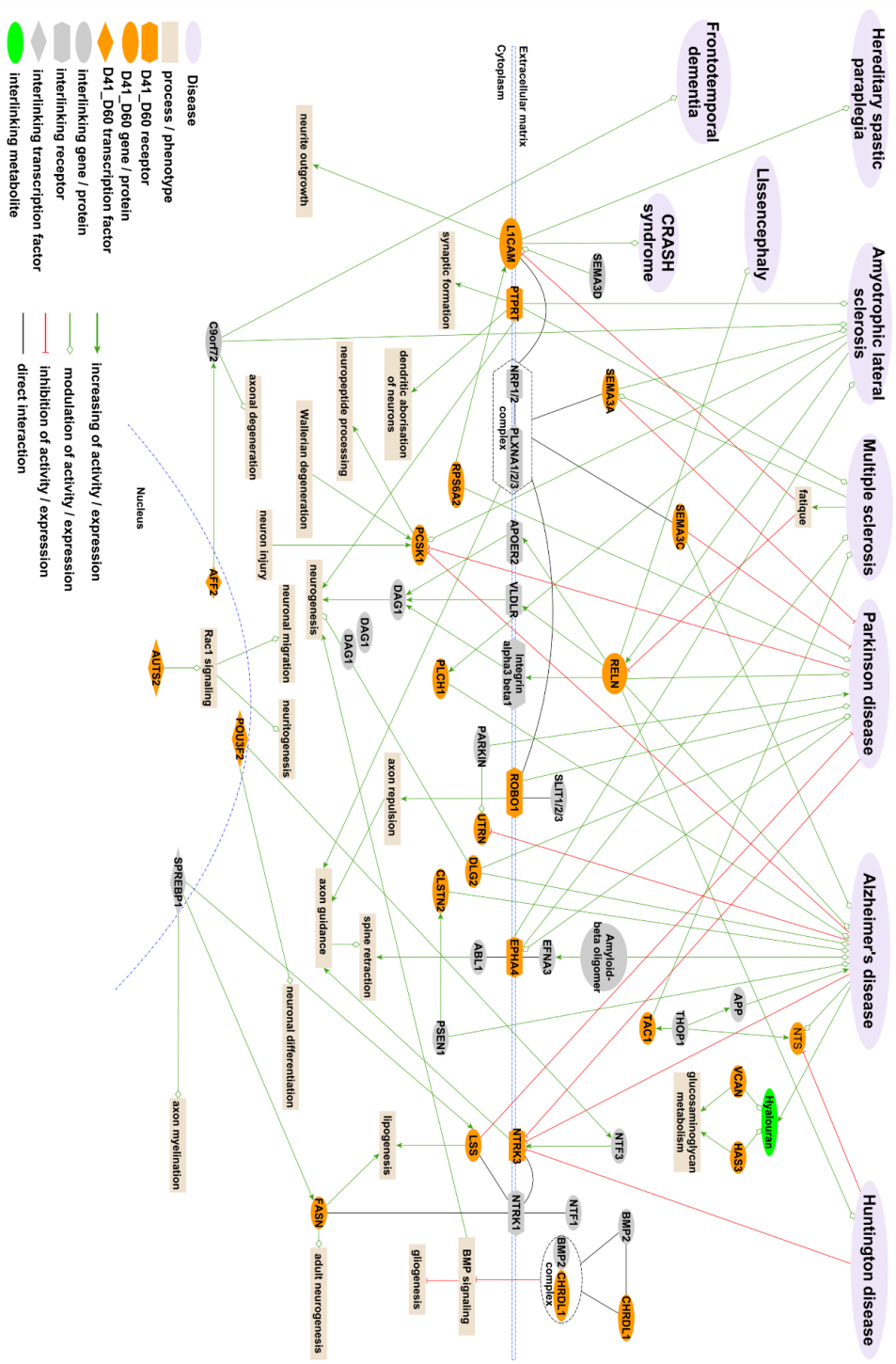


Table S3.1 List of the selected top up regulated genes during Day 41 to 60 after cell differentiation

Ensembl gene id	Gene name	Gene description
ENSG00000157570	TSPAN18	tetraspanin 18
ENSG00000075223	SEMA3C	semaphorin 3C
ENSG00000114805	PLCH1	phospholipase C eta 1
ENSG00000189056	RELN	reelin
ENSG00000160285	LSS	lanosterol synthase
ENSG00000158258	CLSTN2	calsyntenin 2
ENSG00000140538	NTRK3	neurotrophic receptor tyrosine kinase 3
ENSG00000184486	POU3F2	POU class 3 homeobox 2
ENSG00000167178	ISLR2	immunoglobulin superfamily containing leucine rich repeat 2
ENSG00000075213	SEMA3A	semaphorin 3A
ENSG00000198910	L1CAM	L1 cell adhesion molecule
ENSG00000071242	RPS6KA2	ribosomal protein S6 kinase A2
ENSG00000150672	DLG2	discs large MAGUK scaffold protein 2
ENSG00000093072	ADA2	adenosine deaminase 2
ENSG00000101938	CHRDL1	chordin like 1
ENSG00000006128	TAC1	tachykinin precursor 1
ENSG00000103044	HAS3	hyaluronan synthase 3
ENSG00000175426	PCSK1	proprotein convertase subtilisin/kexin type 1
ENSG00000133636	NTS	neurotensin
ENSG00000169855	ROBO1	roundabout guidance receptor 1
ENSG00000116106	EPHA4	EPH receptor A4
ENSG00000196090	PTPRT	protein tyrosine phosphatase%2C receptor type T
ENSG00000169710	FASN	fatty acid synthase
ENSG00000131016	AKAP12	A-kinase anchoring protein 12
ENSG00000152818	UTRN	utrophin
ENSG00000158321	AUTS2	AUTS2%2C activator of transcription and developmental regulator
ENSG00000155966	AFF2	AF4/FMR2 family member 2
ENSG00000038427	VCAN	versican

Table S3.2 HERV-K LTRs appeared within 500 kbp vicinity of the transcript start sites (TSS) of the top 28 up-regulated genes

Ensembl gene id	Gene name	Relative location	HERV id	LTR	Strand	Distance to TSS
ENSG00000140538	NTRK3	upstream	ERV_1326451	LTR5_Hs	different	283784
ENSG00000093072	ADA2	downstream	ERV_2425454	LTR5A	different	198068
ENSG00000006128	TAC1	downstream	ERV_3754277	LTR5A	different	208934
ENSG00000133636	NTS	upstream	ERV_0953958	LTR5_Hs	different	404392
ENSG00000169855	ROBO1	upstream	ERV_2564550	LTR5B	different	302915
ENSG00000131016	AKAP12	upstream	ERV_3539935	LTR5B	same	377529
ENSG00000131016	AKAP12	downstream	ERV_3400954	LTR5_Hs	different	215244

"Relative location" refers to the location of LTR relative to the gene (e.g. upstream stands for that the LTR located at the gene's upstream). "HERV id" refers to the id in database HERVd.

Chapter 4 Summary and outlook

This dissertation mainly explores the activities of Human endogenous retroviruses in different cell conditions and their potential roles in the pathological context. In chapter 2, we examined the differential expression of host genes and HERVs in A549 cells in three strains of IAV infections (PR8M, SC35M, SC35MΔNS1). We observed a considerable upregulation of genes and HERV loci upon IAVs infection. The induced DEG and the DEHERV sets differed significantly among the three infected cells. In all three infected cells, we observed the upregulation of immunity-relative genes. And as predicted, SC35MΔNS1 infection resulted in the most intense immune response as it lacks NS1 protein. According to the functional analysis of the involved genes of the DEHERV-G pairs, we discovered ubiquitous co-upregulation of HERV loci and immune genes in PR8M and SC35MΔNS1 infected cells, indicating the close relationship of HERVs and cell immune system. Further analysis suggested that the cis-regulatory elements harbored in the HERV LTRs, such as promoter-like NFY(CCAAT) motif enriched in the up-regulated HERV LTRs, potentially activated essential immune relative genes and thus induced the immune responding. The project discussed in chapter 3 aimed to gain a deeper understanding of the biological relationship between the activities of HERV-K, the youngest HERV group, and neuron differentiation since various studies had identified overexpression of HERV-K in different kinds of neurological disorders. We observed the reactivation of HERV-K expression disturbed the cortical neuron differentiation and identified the overexpression of NTRK3 induced by HERV-K reactivation to be the critical factor in this process. By studying the regulation of HERV-K and NTRK3 in patients suffering from different neurological diseases, we discussed the possibility and pathway of HERV-K's role in Alzheimer's disease.

Pure differential expression analysis with bioinformatics methods can hardly tell the causality of the co-regulation of HERV elements and host genes. For example, in chapter 2, we applied the pipeline of identifying DEHERV-G pairs defined in the study of Wang et al., which helped to discover the ubiquitous co-regulation of HERV-LTRs and immune genes. However, we cannot solidly conclude whether the activities of HERV-LTRs induce the expression of immune genes or are induced by

the immune response. Most identified DEHERV-G pairs, either in this work or in (M. Wang et al. 2020, 2021), are in the schema that the DEG is located upstream of the DEHERV, which indicates the hitchhiking effect of the HERV up-regulation resulted from the elevating of the host gene expression. Nevertheless, we still tried to reveal the possible pathways implicated in the interaction of HERVs and the immune system. We conducted a series of fisher's tests and found that LTR12C loci are more prone to occur in close proximity to the TSS of up-regulated genes. The motif analysis also indicated that the enriched NFY(CCAAT) motif harbored in the LTR12C could potentially serve as the promoter. Intriguingly, we identified two activated LTR12C loci at the close upstream of two essential up-regulated immune genes: GBP5 and CXCL11. Considering the widely reported regulation ability of LTR12C in the previous studies (Beyer et al. 2016; Jung et al. 2011, 2017; Krönung et al. 2016), it is reasonable to hypothesize that the LTR12C loci assisted the cell viral immunity upon IAVs infection by inducing the expression of GBP5 and CXCL11.

Chapter 3 tells an entire scientific story combining the bioinformatic analysis and wet-lab experiment. In addition to solely exploring the HERV-K expression in different neuron disorder contexts, we initiatively activated the expression of the HERV-K internal portion in cortical neurons. We monitored the cell growth condition and the changes in transcriptome pattern. This project's bioinformatics work mainly contributes to DEG identification and functional analysis. Notably, we successfully filtered the noise results and selected the most up-regulated genes possibly affecting neuron differentiation. In further wet-lab validation, through comparative observation of HERV-K reactivation on normal cortical cells and cells with each chosen candidate gene knocked out, we finally discovered the crucial role of NTRK3 over-expression in disrupting cortical differentiation upon HERV-K reactivation. The Gene knockdown technique provided a solid verification of the role of NTRK3 in cortical neuron disruption upon HERV-K reactivation. We knocked the selected candidate genes using Short hairpin RNAs (shRNAs) (Grimm and Kay 2006; Sheng, Flood, and Xie 2020), which silence the target gene by identifying and cleaving the target mRNA. We detected that 70% knocking down of NTRK3 would prevent the disruption of cortical neurons in the context of HERV-K reactivation. Thus, we confirmed the critical role of NTRK3 in the cortical

disruption process. Similar strategy can also be applied to verify the function of LTR12C on the viral immune system upon IAVs infection mentioned in chapter 2. In the future, this technology can also be applied to check the functionalities of HERV-derived nucleotide acid or proteins in the host cell upon IAVs infection as a complement validation of chapter 2.

However, the gene knockdown technology is not enough for the LTR functionality verification as LTR12C mentioned in chapter 2 since we supposed them mainly to serve as cis-regulatory elements during the IAVs infection, which induced the immune genes. In this case, the gene knockout (KO) technology, which can completely remove the target sequence from the genome (Charbogne, Kieffer, and Befort 2014; Tai et al. 2007), should be applied to eliminate the cis-regulatory function. By comparing the activities of relative immune genes upon IAVs infection of normal cells and the KO cells with specific LTR12C locus was removed, we can verify the cis-regulatory ability of the target locus. Multiple studies have already successfully knocked out some other retrotransposons in humans or other mammals (Le Bas-Bernardet et al. 2011; Ko et al. 2022; Zhang et al. 2019), so it is feasible to make this verification analysis in the future.

Due to the repetitive property, the accurate expression estimation of HERV elements has long been challenging. This dissertation addresses this issue by applying a type-specific analysis of HERV expression. The expression of each type of HERV per sample was measured by a single integer count number. And the HERV expression count matrix was further modeled and normalized along with the protein-coding gene matrix using DESeq2 to estimate the intermediate parameters like sample-specific parameters. In the influenza project, we measured the type-specific HERV expression count by counting the RNA-seq reads aligned to the consensus sequence of each type of HERV provided by the Dfam database (Hubley et al. 2016). In comparison, in the HERV-K activation project, the expression of HERV type was represented by summing the reads mapped to any intact loci of that type. The latter strategy measures the actual read counts mapped to the HERVs. However, bias can be introduced into estimating intermediate parameters if the summed number of reads is too large. This issue can be largely solved by discarding the multiple mapped reads. In comparison, the former type-specific strategy yields

an average estimation of each type of HERVs and prevents bias. This analysis is similar to the widely used qPCR experiment, which uses a consensus sequence to design primers and measure the overall expression of target sequences. So, the result obtained from this strategy is more reliable and was accepted by a peer-reviewed published paper (Liu et al. 2022).

The locus-specific expression analysis of HERVs is also important as it reflects the activity of each individual HERV locus. The optimization of the bioinformatic algorithm can partially address the multiple alignment issue caused by the repetitive property of HERVs, either by the generally purposed alignment tools like RSEM or tools designed explicitly for transposon elements like TEtranscripts, SalmonTE, Telescope, etc. (Bendall et al. 2019; Jeong et al. 2018; Jin et al. 2015). And the development of public HERV databases will also improve the quality of HERV expression analysis. For instance, HERVd recently released its third version, which offers more accurate results by introducing the Dfam database and the ssearch algorithm (Smith and Waterman 1981). Predictably, the more precise annotation information provided by HERVd will result in a better quality of HERV RNA-seq reads alignment. In addition, new sequencing technology can also contribute to better accuracy of locus-specific HERV expression measurement. For example, the third-generation sequencing (TGS) technology for RNA sequencing (aka. Iso-seq) developed by Pacific Biosciences (PacBio) has solved the long-existing TGS drawback, the low accuracy of the read (Au et al. 2012; May et al. 2021). Through engaging circular consensus sequencing (CCS) mode in the RNA-seq process, Iso-seq can generate high-quality Hi-fi reads (>99.9% read accuracy), and the single read length can reach 10 kb or longer (Gonzalez-Garay 2016). This long read can effectively distinguish the intact HERV loci and those that accumulated many point or structure mutations, providing a higher resolution of the HERV transcriptome pattern. By now, due to the high cost of TGS, Iso-seq was seldom used for quantitative transcriptome analysis. Nevertheless, attempts have been made to combine the NGS and TGS technologies (aka. hybrid sequencing) (Au et al. 2013), which takes advantage of the long read length of TGS and deep sequencing depth of NGS at low expense. We can expect in the future the implementation of this method in the future HERV transcriptome analysis in the future.

Bibliography

Alcazer, Vincent, Paola Bonaventura, and Stephane Depil. 2020. “Human Endogenous Retroviruses (HERVs): Shaping the Innate Immune Response in Cancers.” *Cancers* 12(3). doi: 10.3390/cancers12030610.

Antony, Joseph M., Andre M. Deslauriers, Rakesh K. Bhat, Kristofer K. Ellestad, and Christopher Power. 2011. “Human Endogenous Retroviruses and Multiple Sclerosis: Innocent Bystanders or Disease Determinants?” *Biochimica et Biophysica Acta* 1812(2):162–76. doi: 10.1016/j.bbadis.2010.07.016.

Ariza, Maria-Eugenia, and Marshall V Williams. 2011. “A Human Endogenous Retrovirus K DUTPase Triggers a TH1, TH17 Cytokine Response: Does It Have a Role in Psoriasis?” *The Journal of Investigative Dermatology* 131(12):2419–27. doi: 10.1038/jid.2011.217.

Armbruester, Vivienne, Marlies Sauter, Ellen Krautkraemer, Eckart Meese, Anya Kleiman, Barbara Best, Klaus Roemer, and Nikolaus Mueller-Lantzsch. 2002. “A Novel Gene from the Human Endogenous Retrovirus K Expressed in Transformed Cells.” *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 8(6):1800–1807.

Arru, Giannina, Grazia Galleri, Giovanni A. Deiana, Ignazio R. Zarbo, Elia Sechi, Marco Bo, Maria Piera L. Cadoni, Davide G. Corda, Claudia Frau, Elena R. Simula, Maria Antonietta Manca, Franca Galistu, Paolo Solla, Roberto Manetti, Gian Pietro Sechi, and Leonardo A. Sechi. 2021. “HERV-K Modulates the Immune Response in ALS Patients.” *Microorganisms* 9(8). doi: 10.3390/microorganisms9081784.

Attermann, A. S., A. M. Bjerregaard, S. K. Saini, K. Grønbæk, and S. R. Hadrup. 2018. “Human Endogenous Retroviruses and Their Implication for Immunotherapeutics of Cancer.” *Annals of Oncology* 29(11):2183–91. doi: 10.1093/annonc/mdy413.

Au, Kin Fai, Vittorio Sebastiano, Pegah Tootoonchi Afshar, Jens Durruthy Durruthy, Lawrence Lee, Brian A. Williams, Harm van Bakel, Eric E. Schadt, Renee A. Reijo-Pera, and Jason G. Underwood. 2013. “Characterization of the Human ESC Transcriptome by Hybrid Sequencing.” *Proceedings of the National Academy of Sciences* 110(50):E4821–30.

Au, Kin Fai, Jason G. Underwood, Lawrence Lee, and Wing Hung Wong. 2012. “Improving PacBio Long Read Accuracy by Short Read Alignment.” *PLOS ONE* 7(10):e46679.

Autio, Arttu, Tapio Nevalainen, Binisha H. Mishra, Marja Jylhä, Heini Flinck, and Mikko Hurme. 2020. “Effect of Aging on the Transcriptomic Changes Associated with the Expression of the HERV-K (HML-2) Provirus at 1q22.” *Immunity and Ageing* 17(1):1–8. doi: 10.1186/s12979-020-00182-0.

Bannert, Norbert, and Reinhard Kurth. 2006. “The Evolutionary Dynamics of Human Endogenous Retroviral Families.” *Annual Review of Genomics and Human Genetics* 7:149–73. doi: 10.1146/annurev.genom.7.080505.115700.

Bartkowska, Katarzyna, Annie Paquin, Andrée S. Gauthier, David R. Kaplan, and Freda D. Miller. 2007. “Trk Signaling Regulates Neural Precursor Cell Proliferation and Differentiation during Cortical Development.” *Development* 134(24):4369–80. doi: 10.1242/dev.008227.

Le Bas-Bernardet, S., X. Tillou, N. Poirier, N. Dilek, M. Chatelais, J. Devallière, B. Charreau, D. Minault, J. Hervouet, K. Renaudin, C. Crossan, L. Scobie, P. J. Cowan, A. J. F. d’Apice, C. Galli, E. Cozzi, J. P. Soulillou, B. Vanhove, and G. Blancho. 2011. “Xenotransplantation of Galactosyl-Transferase Knockout, CD55, CD59, CD39, and Fucosyl-Transferase Transgenic Pig Kidneys Into Baboons.” *Transplantation Proceedings* 43(9):3426–30. doi: https://doi.org/10.1016/j.transproceed.2011.09.024.

Belshaw, Robert, Vini Pereira, Aris Katzourakis, Gillian Talbot, Jan Paces, Austin Burt, and Michael Tristem. 2004. “Long-Term Reinfection of the Human Genome by Endogenous Retroviruses.” *Proceedings of the National Academy of Sciences of the United States of America* 101(14):4894–99. doi: 10.1073/pnas.0307800101.

Bendall, Matthew L., Miguel De Mulder, Luis Pedro Iñiguez, Aarón Lecanda-Sánchez, Marcos Pérez-Losada, Mario A. Ostrowski, R. Brad Jones, Lubbertus C. F. Mulder, Gustavo Reyes-Terán, Keith A. Crandall, Christopher E. Ormsby, and Douglas F. Nixon. 2019. “Telescope: Characterization of the Retrotranscriptome by Accurate Estimation of Transposable Element Expression.” *PLoS Computational Biology* 15(9):1–25. doi: 10.1371/journal.pcbi.1006453.

Beyer, U., S. K. Krönung, A. Leha, L. Walter, and M. Dobbelstein. 2016. “Comprehensive Identification of Genes Driven by ERV9-LTRs Reveals TNFRSF10B as a Re-Activatable Mediator of Testicular Cancer Cell Death.” *Cell Death and Differentiation* 23(1):64–75. doi: 10.1038/cdd.2015.68.

Bhardwaj, Neeru, Meagan Montesio, Farrah Roy, and John M. Coffin. 2015. “Differential Expression of HERV-K (HML-2) Proviruses in Cells and Virions of the Teratocarcinoma Cell Line Tera-1.” *Viruses* 7(3):939–68. doi: 10.3390/v7030939.

Bjelobaba, Ivana, Danijela Savic, and Irena Lavrnja. 2017. “Multiple Sclerosis and Neuroinflammation: The Overview of Current and Prospective Therapies.” *Current Pharmaceutical Design* 23(5):693–730. doi: 10.2174/1381612822666161214153108.

Blanco-Melo, Daniel, Robert J. Gifford, and Paul D. Bieniasz. 2017. “Co-Option of an Endogenous Retrovirus Envelope for Host Defense in Hominid Ancestors.” *ELife* 6:1–19. doi: 10.7554/eLife.22519.

Bock, Stephanie, Christina S. Mullins, Ernst Klar, Philippe Pérot, Claudia Maletzki, and Michael Linnebacher. 2018. “Murine Endogenous Retroviruses Are Detectable in Patient-Derived Xenografts but Not in Patient-Individual Cell Lines of Human Colorectal Cancer.” *Frontiers in Microbiology* 9(APR):1–9. doi: 10.3389/fmicb.2018.00789.

Böhnlein, S., J. Hauber, and B. R. Cullen. 1989. “Identification of a U5-Specific

Sequence Required for Efficient Polyadenylation within the Human Immunodeficiency Virus Long Terminal Repeat." *Journal of Virology* 63(1):421–24. doi: 10.1128/jvi.63.1.421-424.1989.

Boi, Stefano, Kyle Rosenke, Ethan Hansen, Duncan Hendrick, Frank Malik, and Leonard H. Evans. 2016. "Endogenous Retroviruses Mobilized during Friend Murine Leukemia Virus Infection." *Virology* 499:136–43. doi: 10.1016/j.virol.2016.07.009.

Bonaventura, Paola, Vincent Alcazer, Virginie Mutez, Laurie Tonon, Juliette Martin, Nicolas Chuvin, Emilie Michel, Rasha E. Boulos, Yann Estornes, Jenny Valladeau-Guilemond, Alain Viari, Qing Wang, Christophe Caux, and Stéphane Depil. 2022. "Identification of Shared Tumor Epitopes from Endogenous Retroviruses Inducing High-Avidity Cytotoxic T Cells for Cancer Immunotherapy." *Science Advances* 8(4):eabj3671. doi: 10.1126/sciadv.abj3671.

Brisse, Morgan, and Hinh Ly. 2019. "Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5." *Frontiers in Immunology* 10. doi: 10.3389/fimmu.2019.01586.

Charbogne, Pauline, Brigitte L. Kieffer, and Katia Befort. 2014. "15 Years of Genetic Approaches in Vivo for Addiction Research: Opioid Receptor and Peptide Gene Knockout in Mouse Models of Drug Abuse." *Neuropharmacology* 76:204–17.

Chen, Hao, and Burt M. Sharp. 2004. "Content-Rich Biological Network Constructed by Mining PubMed Abstracts." *BMC Bioinformatics* 5(1):147. doi: 10.1186/1471-2105-5-147.

Chen, Jungang, Maryam Foroozesh, and Zhiqiang Qin. 2019. "Transactivation of Human Endogenous Retroviruses by Tumor Viruses and Their Functions in Virus-Associated Malignancies." *Oncogenesis* 8(1). doi: 10.1038/s41389-018-0114-y.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34(17):i884–90. doi: 10.1093/bioinformatics/bty560.

Chen, Yunshun, Aaron T. L. Lun, and Gordon K. Smyth. 2016. "From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the EdgeR Quasi-Likelihood Pipeline." *F1000Research* 5:1438. doi: 10.12688/f1000research.8987.2.

Chiappinelli, Katherine B., Pamela L. Strissel, Alexis Desrichard, Huili Li, Christine Henke, Benjamin Akman, Alexander Hein, Neal S. Rote, Leslie M. Cope, Alexandra Snyder, Vladimir Makarov, Sadna Budhu, Dennis J. Slamon, Jedd D. Wolchok, Drew M. Pardoll, Matthias W. Beckmann, Cynthia A. Zahnow, Taha Merghoub, Timothy A. Chan, Stephen B. Baylin, and Reiner Strick. 2015. "Inhibiting DNA Methylation Causes an Interferon Response in Cancer via DsRNA Including Endogenous Retroviruses." *Cell* 162(5):974–86. doi: 10.1016/j.cell.2015.07.011.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2016. "Regulatory

Evolution of Innate Immunity through Co-Option of Endogenous Retroviruses." *Science* 351(6277):1083–87. doi: 10.1126/science.aad5497.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2017. "Regulatory Activities of Transposable Elements: From Conflicts to Benefits." *Nature Reviews. Genetics* 18(2):71–86. doi: 10.1038/nrg.2016.139.

Contreras-Galindo, Rafael, Pablo López, Rosa Vélez, and Yasuhiro Yamamura. 2007. "HIV-1 Infection Increases the Expression of Human Endogenous Retroviruses Type K (HERV-K) in Vitro." *AIDS Research and Human Retroviruses* 23(1):116–22. doi: 10.1089/aid.2006.0117.

de Cubas, Aguirre A., William Dunker, Andrew Zaninovich, Rachel A. Hongo, Anuj Bhatia, Anshuman Panda, Kathryn E. Beckermann, Gyan Bhanot, Shridar Ganesan, John Karijolich, and W. Kimryn Rathmell. 2020. "DNA Hypomethylation Promotes Transposable Element Expression and Activation of Immune Signaling in Renal Cell Cancer." *JCI Insight* 5(11). doi: 10.1172/jci.insight.137569.

Curty, Gislaine, Jez L. Marston, Miguel De Mulder Rougvie, Fabio E. Leal, Douglas F. Nixon, and Marcelo A. Soares. 2020. "Human Endogenous Retrovirus K in Cancer: A Potential Biomarker and Immunotherapeutic Target." *Viruses* 12(7):14–16. doi: 10.3390/v12070726.

Davis, Allan Peter, Benjamin L. King, Susan Mockus, Cynthia G. Murphy, Cynthia Saraceni-Richards, Michael Rosenstein, Thomas Wiegers, and Carolyn J. Mattingly. 2010. "The Comparative Toxicogenomics Database: Update 2011." *Nucleic Acids Research* 39(suppl_1):D1067–72. doi: 10.1093/nar/gkq813.

Dembny, Paul, Andrew G Newman, Manvendra Singh, Michael Hinz, Michal Szczepek, Christina Krüger, Robert Adalbert, Omar Dzaye, Thorsten Trimbuch, Thomas Wallach, Gunnar Kleinau, Katja Derkow, Bernhard C. Richard, Carola Schipke, Claus Scheidereit, Harald Stachelscheid, Douglas Golenbock, Oliver Peters, Michael Coleman, Frank L. Heppner, Patrick Scheerer, Victor Tarabykin, Klemens Ruprecht, Zsuzsanna Izsvák, Jens Mayer, and Seija Lehnardt. 2020. "Human Endogenous Retrovirus HERV-K(HML-2) RNA Causes Neurodegeneration through Toll-like Receptors." *JCI Insight* 5(7). doi: 10.1172/jci.insight.131093.

Dembny, Paul, Andrew G. Newman, Manvendra Singh, Michael Hinz, Michal Szczepek, Christina Krüger, Robert Adalbert, Omar Dzaye, Thorsten Trimbuch, Thomas Wallach, Gunnar Kleinau, Katja Derkow, Bernhard C. Richard, Carola Schipke, Claus Scheidereit, Harald Stachelscheid, Douglas Golenbock, Oliver Peters, Michael Coleman, Frank L. Heppner, Patrick Scheerer, Victor Tarabykin, Klemens Ruprecht, Zsuzsanna Izsvák, Jens Mayer, and Seija Lehnardt. 2020. "Human Endogenous Retrovirus HERVK(HML-2) RNA Causes Neurodegeneration through Toll-like Receptors." *JCI Insight* 5(7). doi: 10.1172/JCI.INSIGHT.131093.

Deniz, Özgen, Mamataz Ahmed, Christopher D. Todd, Ana Rio-Machin, Mark A. Dawson, and Miguel R. Branco. 2020. "Endogenous Retroviruses Are a Source of Enhancers with Oncogenic Potential in Acute Myeloid Leukaemia." *Nature Communications* 11(1):1–14. doi: 10.1038/s41467-020-17206-4.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29(1):15–21. doi: 10.1093/bioinformatics/bts635.

Dou, Dan, Rebecca Revol, Henrik Östbye, Hao Wang, and Robert Daniels. 2018. "Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement." *Frontiers in Immunology* 9:1581. doi: 10.3389/fimmu.2018.01581.

Douville, Renée N., and Avindra Nath. 2014. "Human Endogenous Retroviruses and the Nervous System." Pp. 465–85 in *Physiology & behavior*. Vol. 176.

Engel, Kristina, Lisa Wieland, Anna Krüger, Ines Volkmer, Holger Cynis, Alexander Emmer, and Martin S. Staege. 2021. "Identification of Differentially Expressed Human Endogenous Retrovirus Families in Human Leukemia and Lymphoma Cell Lines and Stem Cells." *Frontiers in Oncology* 11:637981. doi: 10.3389/fonc.2021.637981.

Evans, Leonard H., A. S. M. Alamgir, Nick Owens, Nick Weber, Kimmo Virtaneva, Kent Barbian, Amenah Babar, Frank Malik, and Kyle Rosenke. 2009. "Mobilization of Endogenous Retroviruses in Mice after Infection with an Exogenous Retrovirus." *Journal of Virology* 83(6):2429–35. doi: 10.1128/jvi.01926-08.

Fabregat, Antonio, Konstantinos Sidiropoulos, Guilherme Viteri, Oscar Forner, Pablo Marin-Garcia, Vicente Arnau, Peter D'Eustachio, Lincoln Stein, and Henning Hermjakob. 2017. "Reactome Pathway Analysis: A High-Performance in-Memory Approach." *BMC Bioinformatics* 18(1):142. doi: 10.1186/s12859-017-1559-2.

Ferrari, Luca, Marco Cafora, Federica Rota, Mirjam Hoxha, Simona Iodice, Letizia Tarantini, Maria Dolci, Serena Delbue, Anna Pistocchi, and Valentina Bollati. 2019. "Extracellular Vesicles Released by Colorectal Cancer Cell Lines Modulate Innate Immune Response in Zebrafish Model: The Possible Role of Human Endogenous Retroviruses." *International Journal of Molecular Sciences* 20(15). doi: 10.3390/ijms20153669.

Feschotte, Cédric, and Clément Gilbert. 2012. "Endogenous Viruses: Insights into Viral Evolution and Impact on Host Biology." *Nature Reviews Genetics* 13(4):283–96. doi: 10.1038/nrg3199.

Fuentes, Daniel R., Tomek Swigut, and Joanna Wysocka. 2018. "Systematic Perturbation of Retroviral LTRs Reveals Widespread Long-Range Effects on Human Gene Regulation." *ELife* 7:1–29. doi: 10.7554/elife.35989.

Garcia-Montojo, Marta, María Dominguez-Mozo, Ana Arias-Leal, Ángel Garcia-Martinez, Virginia De las Heras, Ignacio Casanova, Raphaël Faucard, Nadège Gehin, Alexandra Madeira, Rafael Arroyo, François Curtin, Roberto Alvarez-Lafuente, and Hervé Perron. 2013. "The DNA Copy Number of Human Endogenous Retrovirus-W (MSRV-Type) Is Increased in Multiple Sclerosis Patients and Is Influenced by Gender and Disease Severity." *PloS One* 8(1):e53623. doi: 10.1371/journal.pone.0053623.

Garcia-Montojo, Marta, Belén de la Hera, Jezabel Varadé, Ana de la Encarnación,

Iris Camacho, María Domínguez-Mozo, Ana Árias-Leal, Angel García-Martínez, Ignacio Casanova, Guillermo Izquierdo, Miguel Lucas, Maria Fedetz, Antonio Alcina, Rafael Arroyo, Fuencisla Matesanz, Elena Urcelay, and Roberto Alvarez-Lafuente. 2014. "HERV-W Polymorphism in Chromosome X Is Associated with Multiple Sclerosis Risk and with Differential Expression of MSRV." *Retrovirology* 11:2. doi: 10.1186/1742-4690-11-2.

Gifford, Robert, and Michael Tristem. 2003. "The Evolution, Distribution and Diversity of Endogenous Retroviruses." *Virus Genes* 26(3):291–315. doi: 10.1023/A:1024455415443.

Gimenez, Juliette, Cécile Montgiraud, Jean Philippe Pichon, Bertrand Bonnaud, Maud Arsac, Karine Ruel, Olivier Bouton, and François Mallet. 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(7):2229–46. doi: 10.1093/nar/gkp1214.

Gloor, Gregory B., Jean M. Macklaim, Vera Pawlowsky-Glahn, and Juan J. Egozcue. 2017. "Microbiome Datasets Are Compositional: And This Is Not Optional." *Frontiers in Microbiology* 8(NOV):1–6. doi: 10.3389/fmicb.2017.02224.

Golkaram, Mahdi, Michael L. Salmans, Shannon Kaplan, Raakhee Vijayaraghavan, Marta Martins, Nafeesa Khan, Cassandra Garbutt, Aaron Wise, Joyee Yao, Sandra Casimiro, Catarina Abreu, Daniela Macedo, Ana Lúcia Costa, Cecília Alvim, André Mansinho, Pedro Filipe, Pedro Marques da Costa, Afonso Fernandes, Paula Borralho, Cristina Ferreira, Fernando Aldeia, João Malaquias, Jim Godsey, Alex So, Traci Pawlowski, Luis Costa, Shile Zhang, and Li Liu. 2021. "HERVs Establish a Distinct Molecular Subtype in Stage II/III Colorectal Cancer with Poor Outcome." *NPJ Genomic Medicine* 6(1):13. doi: 10.1038/s41525-021-00177-w.

Gonzalez-Garay, Manuel L. 2016. "Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq) BT - Transcriptomics and Gene Regulation." Pp. 141–60 in, edited by J. Wu. Dordrecht: Springer Netherlands.

Grandi, Nicole, and Enzo Tramontano. 2018. "HERV Envelope Proteins: Physiological Role and Pathogenic Potential in Cancer and Autoimmunity." *Frontiers in Microbiology* 9:462. doi: 10.3389/fmicb.2018.00462.

Greenig, Matthew. 2019. "HERVs, Immunity, and Autoimmunity: Understanding the Connection." *PeerJ* 7:e6711. doi: 10.7717/peerj.6711.

Grimm, D., and M. A. Kay. 2006. "Therapeutic Short Hairpin RNA Expression in the Liver: Viral Targets and Vectors." *Gene Therapy* 13(6):563–75. doi: 10.1038/sj.gt.3302727.

Gröger, Victoria, and Holger Cynis. 2018. "Human Endogenous Retroviruses and Their Putative Role in the Development of Autoimmune Disorders Such as Multiple Sclerosis." *Frontiers in Microbiology* 9:265. doi: 10.3389/fmicb.2018.00265.

Grow, Edward J., Ryan A. Flynn, Shawn L. Chavez, Nicholas L. Bayless, Mark Wossidlo, Daniel J. Wesche, Lance Martin, Carol B. Ware, Catherine A. Blish, Howard Y. Chang, Renee A. Reijo Pera, and Joanna Wysocka. 2015. "Intrinsic Retroviral Reactivation in Human Preimplantation Embryos and Pluripotent Cells." *Nature* 522(7555):221–25. doi: 10.1038/nature14308.

Güell, Marc, Dong Niu, Yinan Kan, Haydy George, Tao Wang, I. Hsiu Lee, Gang Wang, George Church, and Luhan Yang. 2017. "PERV Inactivation Is Necessary to Guarantee Absence of Pig-to-Patient PERVs Transmission in Xenotransplantation." *Xenotransplantation* 24(6):e12366.

Hale, Benjamin G. 2022. "Antiviral Immunity Triggered by Infection-Induced Host Transposable Elements." *Current Opinion in Virology* 52:211–16. doi: 10.1016/j.coviro.2021.12.006.

Hale, Benjamin G., Richard E. Randall, Juan Ortín, and David Jackson. 2008. "The Multifunctional NS1 Protein of Influenza A Viruses." *The Journal of General Virology* 89(Pt 10):2359–76. doi: 10.1099/vir.0.2008/004606-0.

Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, and F. Kokocinski. 2012. "GENCODE: The Reference Human Genome Annotation for The ENCODE Project." *Genome Res* 22:1760–74. doi: 10.1101/gr.135350.111.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Molecular Cell* 38(4):576–89. doi: 10.1016/j.molcel.2010.05.004.

Hilditch, Laura, Rishi Matadeen, David C. Goldstone, Peter B. Rosenthal, Ian A. Taylor, and Jonathan P. Stoye. 2011. "Ordered Assembly of Murine Leukemia Virus Capsid Protein on Lipid Nanotubes Directs Specific Binding by the Restriction Factor, Fv1." *Proceedings of the National Academy of Sciences of the United States of America* 108(14):5771–76. doi: 10.1073/pnas.1100118108.

Hohn, Oliver, Kirsten Hanke, and Norbert Bannert. 2013. "HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease." *Frontiers in Oncology* 3(September):246. doi: 10.3389/fonc.2013.00246.

Hubley, Robert, Robert D. Finn, Jody Clements, Sean R. Eddy, Thomas A. Jones, Weidong Bao, Arian F. A. Smit, and Travis J. Wheeler. 2016. "The Dfam Database of Repetitive DNA Families." *Nucleic Acids Research* 44(D1):D81–89. doi: 10.1093/nar/gkv1272.

Imhof, A., M. Schuierer, O. Werner, M. Moser, C. Roth, R. Bauer, and R. Buettner. 1999. "Transcriptional Regulation of the AP-2alpha Promoter by BTEB-1 and AP-2rep, a Novel Wt-1/Egr-Related Zinc Finger Repressor." *Molecular and Cellular Biology* 19(1):194–204. doi: 10.1128/MCB.19.1.194.

Jakobsson, Johan, and Michelle Vincendeau. 2022. "SnapShot: Human Endogenous Retroviruses." *Cell* 185(2):400-400.e1. doi:

10.1016/j.cell.2021.12.028.

Javanian, Mostafa, Mohammad Barary, Sam Ghebrehewet, Veerendra Koppolu, VeneelaKrishnaRekha Vasigala, and Soheil Ebrahimpour. 2021. “A Brief Review of Influenza Virus Infection.” *Journal of Medical Virology* 93(8):4638–46. doi: 10.1002/jmv.26990.

Jeong, Hyun-Hwan, Hari Krishna Yalamanchili, Caiwei Guo, Joshua M. Shulman, and Zhandong Liu. 2018. “An Ultra-Fast and Scalable Quantification Pipeline for Transposable Elements from next Generation Sequencing Data.” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 23:168–79.

Jin, Ying, Oliver H. Tam, Eric Paniagua, and Molly Hammell. 2015. “TEtranscripts: A Package for Including Transposable Elements in Differential Expression Analysis of RNA-Seq Datasets.” *Bioinformatics (Oxford, England)* 31(22):3593–99. doi: 10.1093/bioinformatics/btv422.

Jung, Yi Deun, Jae Won Huh, Dae Soo Kim, Yun Ji Kim, Kung Ahn, Hong Seok Ha, Ja Rang Lee, Joo Mi Yi, Jae Woo Moon, Tae Oh Kim, Geun Am Song, Kyudong Han, and Heui Soo Kim. 2011. “Quantitative Analysis of Transcript Variants of CHM Gene Containing LTR12C Element in Humans.” *Gene* 489(1):1–5. doi: 10.1016/j.gene.2011.09.001.

Jung, Yi Deun, Hee Eun Lee, Ara Jo, Imai Hiroo, Hee Jae Cha, and Heui Soo Kim. 2017. “Activity Analysis of LTR12C as an Effective Regulatory Element of the RAE1 Gene.” *Gene* 634(August):22–28. doi: 10.1016/j.gene.2017.08.037.

Jurka, J. 2000. “Repbase Update: A Database and an Electronic Journal of Repetitive Elements.” *Trends in Genetics : TIG* 16(9):418–20. doi: 10.1016/s0168-9525(00)02093-x.

Jurka, J., V. V Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz. 2005. “Repbase Update, a Database of Eukaryotic Repetitive Elements.” *Cytogenetic and Genome Research* 110(1–4):462–67. doi: 10.1159/000084979.

Kamps, C. A., Y. C. Lin, and P. K. Wong. 1991. “Oligomerization and Transport of the Envelope Protein of Moloney Murine Leukemia Virus-TB and of Ts1, a Neurovirulent Temperature-Sensitive Mutant of MoMuLV-TB.” *Virology* 184(2):687–94. doi: 10.1016/0042-6822(91)90438-h.

Kassiotis, George. 2014. “Endogenous Retroviruses and the Development of Cancer.” *The Journal of Immunology* 192(4):1343–49. doi: 10.4049/jimmunol.1302972.

Kassiotis, George, and Jonathan P. Stoye. 2016. “Immune Responses to Endogenous Retroelements: Taking the Bad with the Good.” *Nature Reviews. Immunology* 16(4):207–19. doi: 10.1038/nri.2016.27.

Kim, Heui Soo. 2012. “Genomic Impact, Chromosomal Distribution and Transcriptional Regulation of HERV Elements.” *Molecules and Cells* 33(6):539–44. doi: 10.1007/s10059-012-0037-y.

Ko, Eun-Ji, Eun Taeg Kim, Heungyeol Kim, Chul Min Lee, Suk Bong Koh, Wan

Kyu Eo, Hongbae Kim, Young Lim Oh, Mee Sun Ock, Ki Hyung Kim, and Hee-Jae Cha. 2022. “Effect of Human Endogenous Retrovirus-K Env Gene Knockout on Proliferation of Ovarian Cancer Cells.” *Genes & Genomics* 44(9):1091–97. doi: 10.1007/s13258-022-01280-7.

Kochs, Georg, Iris Koerner, Lena Thiel, Sonja Kothlow, Bernd Kaspers, Nicolas Ruggli, Artur Summerfield, Jovan Pavlovic, Jürgen Stech, and Peter Staeheli. 2007. “Properties of H7N7 Influenza A Virus Strain SC35M Lacking Interferon Antagonist NS1 in Mice and Chickens.” *The Journal of General Virology* 88(Pt 5):1403–9. doi: 10.1099/vir.0.82764-0.

Kojima, Kenji K. 2018. “Human Transposable Elements in Repbase: Genomic Footprints from Fish to Humans.” *Mobile DNA* 9(1):1–14. doi: 10.1186/s13100-017-0107-y.

Krammer, Florian, Gavin J. D. Smith, Ron A. M. Fouchier, Malik Peiris, Katherine Kedzierska, Peter C. Doherty, Peter Palese, Megan L. Shaw, John Treanor, Robert G. Webster, and Adolfo García-Sastre. 2018. “Influenza.” *Nature Reviews Disease Primers* 4(1):1–21. doi: 10.1038/s41572-018-0002-y.

Krausslich, H. G., R. H. Ingraham, M. T. Skoog, E. Wimmer, P. V. Pallai, and C. A. Carter. 1989. “Activity of Purified Biosynthetic Proteinase of Human Immunodeficiency Virus on Natural Substrates and Synthetic Peptides.” *Proceedings of the National Academy of Sciences of the United States of America* 86(3):807–11. doi: 10.1073/pnas.86.3.807.

Krönung, Sonja K., Ulrike Beyer, Maria Luisa Chiaramonte, Diletta Dolfini, Roberto Mantovani, and Matthias Dobbelstein. 2016. “LTR12 Promoter Activation in a Broad Range of Human Tumor Cells by HDAC Inhibition.” *Oncotarget* 7(23):33484–97. doi: 10.18632/oncotarget.9255.

Kuleshov, Maxim V, Jennifer E. L. Diaz, Zachary N. Flamholz, Alexandra B. Keenan, Alexander Lachmann, Megan L. Wojciechowicz, Ross L. Cagan, and Avi Ma’ayan. 2019. “ModEnrichr: A Suite of Gene Set Enrichment Analysis Tools for Model Organisms.” *Nucleic Acids Research* 47(W1):W183–90. doi: 10.1093/nar/gkz347.

Kurth, Reinhard, and Norbert Bannert. 2010. “Beneficial and Detrimental Effects of Human Endogenous Retroviruses.” *International Journal of Cancer* 126(2):306–14. doi: 10.1002/ijc.24902.

Küry, Patrick, Avindra Nath, Alain Créange, Antonina Dolei, Patrice Marche, Julian Gold, Gavin Giovannoni, Hans Peter Hartung, and Hervé Perron. 2018. “Human Endogenous Retroviruses in Neurological Diseases.” *Trends in Molecular Medicine* 24(4):379–94. doi: 10.1016/j.molmed.2018.02.007.

Lavie, Laurence, Milena Kitova, Esther Maldener, Eckart Meese, and Jens Mayer. 2005. “CpG Methylation Directly Regulates Transcriptional Activity of the Human Endogenous Retrovirus Family HERV-K(HML-2).” *Journal of Virology* 79(2):876–83. doi: 10.1128/jvi.79.2.876-883.2005.

Lechner, Martin, Veit Höhn, Barbara Brauner, Irmtraud Dunger, Gisela Fobo, Goar Frishman, Corinna Montrone, Gabi Kastenmüller, Brigitte Waegele, and Andreas Ruepp. 2012. “CIDeR: Multifactorial Interaction Networks in Human

Diseases." *Genome Biology* 13(7):R62. doi: 10.1186/gb-2012-13-7-r62.

Lee, Andrew K., Dong Pan, Xuhui Bao, Mengjie Hu, Fang Li, and Chuan-Yuan Li. 2020. "Endogenous Retrovirus Activation as a Key Mechanism of Anti-Tumor Immune Response in Radiotherapy." *Radiation Research* 193(4):305–17. doi: 10.1667/RADE-20-00013.

Lemaître, Cécile, Jhen Tsang, Caroline Bireau, Thierry Heidmann, and Marie Dewannieux. 2017. "A Human Endogenous Retrovirus-Derived Gene That Can Contribute to Oncogenesis by Activating the ERK Pathway and Inducing Migration and Invasion." *PLoS Pathogens* 13(6):e1006451. doi: 10.1371/journal.ppat.1006451.

Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12(1):323. doi: 10.1186/1471-2105-12-323.

Li, Fang, Christoffer Nellåker, Sarven Sabunciyan, Robert H. Yolken, Lorraine Jones-Brando, Anne-Sofie Johansson, Björn Owe-Larsson, and Håkan Karlsson. 2014. "Transcriptional Derepression of the ERVWE1 Locus Following Influenza A Virus Infection." *Journal of Virology* 88(8):4328–37. doi: 10.1128/JVI.03628-13.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics (Oxford, England)* 25(16):2078–79. doi: 10.1093/bioinformatics/btp352.

Li, W., M. H. Lee, L. Henderson, R. Tyagi, M. Bachani, J. Steiner, E. Campanac, D. A. Hoffman, G. von Geldern, K. Johnson, D. Maric, H. D. Morris, M. Lentz, K. Pak, A. Mammen, L. Ostrow, J. Rothstein, and A. Nath. 2015. "Human Endogenous Retrovirus-K Contributes to Motor Neuron Disease." *Science Translational Medicine* 7(307):307ra153-307ra153. doi: 10.1126/scitranslmed.aac8201.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. "FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30(7):923–30. doi: 10.1093/bioinformatics/btt656.

Liu, Cong, Lijuan Liu, Xiuling Wang, Youyi Liu, Miao Wang, and Fan Zhu. 2017. "HBV X Protein Induces Overexpression of HERV-W Env through NF-KB in HepG2 Cells." *Virus Genes* 53(6):797–806. doi: 10.1007/s11262-017-1479-2.

Liu, Hengyuan, Valter Bergant, Goar Frishman, Andreas Ruepp, Andreas Pichlmair, Michelle Vincendeau, and Dmitrij Frishman. 2022. "Influenza A Virus Infection Reactivates Human Endogenous Retroviruses Associated with Modulation of Antiviral Immunity." *Viruses* 14(7). doi: 10.3390/v14071591.

Liu, Ting, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. 2017. "NF-KB Signaling in Inflammation." *Signal Transduction and Targeted Therapy* 2:17023-. doi: 10.1038/sigtrans.2017.23.

Liu, Yifeng, Yongjie Liang, and David Wishart. 2015. "PolySearch2: A Significantly Improved Text-Mining System for Discovering Associations

between Human Diseases, Genes, Drugs, Metabolites, Toxins and More." *Nucleic Acids Research* 43(W1):W535–42. doi: 10.1093/nar/gkv383.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12):550. doi: 10.1186/s13059-014-0550-8.

Lower, R., R. Tonjes, C. Korbmacher, R. Kurth, and J. Lower. 1995. "Identification of a Rev-Related Protein by Analysis of Spliced Transcripts of the Human Endogenous." *Microbiology* 69(1):141–49.

Maeder, Morgan L., Samantha J. Linder, Vincent M. Cascio, Yanfang Fu, Quan H. Ho, and J. Keith Joung. 2013. "CRISPR RNA-Guided Activation of Endogenous Human Genes." *Nature Methods* 10(10):977–79. doi: 10.1038/nmeth.2598.

Magin, Christine, Roswitha Löwer, and Johannes Löwer. 1999. "CORF and RcRE, the Rev/Rex and RRE/RxRE Homologues of the Human Endogenous Retrovirus Family HTDV/HERV-K." *Journal of Virology* 73(11):9496–9507. doi: 10.1128/jvi.73.11.9496-9507.1999.

Mangeney, M., N. De Parseval, G. Thomas, and T. Heidmann. 2001. "The Full-Length Envelope of an HERV-H Human Endogenous Retrovirus Has Immunosuppressive Properties." *Journal of General Virology* 82(10):2515–18. doi: 10.1099/0022-1317-82-10-2515.

Marazzi, Ivan, Jessica S. Y. Ho, Jaehoon Kim, Balaji Manicassamy, Scott Dewell, Randy A. Albrecht, Chris W. Seibert, Uwe Schaefer, Kate L. Jeffrey, Rab K. Prinjha, Kevin Lee, Adolfo García-Sastre, Robert G. Roeder, and Alexander Tarakhovsky. 2012. "Suppression of the Antiviral Response by an Influenza Histone Mimic." *Nature* 483(7390):428–33. doi: 10.1038/nature10892.

May, Jody C., Katrina L. Leaptrot, Bailey S. Rose, Kelly L. Wormwood Moser, Liulin Deng, Laura Maxon, Daniel DeBord, and John A. McLean. 2021. "Resolving Power and Collision Cross Section Measurement Accuracy of a Prototype High-Resolution Ion Mobility Platform Incorporating Structures for Lossless Ion Manipulation." *Journal of the American Society for Mass Spectrometry* 32(4):1126–37. doi: 10.1021/jasms.1c00056.

Mayer, J., M. Sauter, A. Rácz, D. Scherer, N. Mueller-Lantzsch, and E. Meese. 1999. "An Almost-Intact Human Endogenous Retrovirus K on Human Chromosome 7." *Nature Genetics* 21(3):257–58.

Mayer, Jens, Christian Harz, Laura Sanchez, Gavin C. Pereira, Esther Maldener, Sara R. Heras, Lyle W. Ostrow, John Ravits, Ranjan Batra, Eckart Meese, Jose Luis García-Pérez, and John L. Goodier. 2018. "Transcriptional Profiling of HERV-K(HML-2) in Amyotrophic Lateral Sclerosis and Potential Implications for Expression of HML-2 Proteins." *Molecular Neurodegeneration* 13(1):39. doi: 10.1186/s13024-018-0275-3.

Miyagawa, Shuji, Shino Nakatsu, Takatoshi Nakagawa, Akihiro Kondo, Katsuyoshi Matsunami, Kenji Hazama, Junko Yamada, Keizo Tomonaga, Takayuki Miyazawa, and Ryota Shirakura. 2005. "Prevention of PERV Infections in Pig to Human Xenotransplantation by the RNA Interference

Silences Gene." *Journal of Biochemistry* 137(4):503–8.

Montesion, Meagan, Neeru Bhardwaj, Zachary H. Williams, Charlotte Kuperwasser, and John M. Coffin. 2018. "Mechanisms of HERV-K (HML-2) Transcription during Human Mammary Epithelial Cell Transformation" edited by V. Simon. *Journal of Virology* 92(1):e01258-17. doi: 10.1128/JVI.01258-17.

Morris, Gerwyn, Michael Maes, Marianna Murdjeva, and Basant K. Puri. 2019. "Do Human Endogenous Retroviruses Contribute to Multiple Sclerosis, and If So, How?" *Molecular Neurobiology* 56(4):2590–2605. doi: 10.1007/s12035-018-1255-x.

Mortelmans, Kristien, Feng Wang-Johanning, and Gary L. Johanning. 2016. "The Role of Human Endogenous Retroviruses in Brain Development and Function." *Apmis* 124(1–2):105–15. doi: 10.1111/apm.12495.

Mufson, Elliott J., Scott E. Counts, Stephen D. Ginsberg, Laura Mahady, Sylvia E. Perez, Stephen M. Massa, Frank M. Longo, and Milos D. Ikonomovic. 2019. "Nerve Growth Factor Pathobiology During the Progression of Alzheimer's Disease." *Frontiers in Neuroscience* 13:533. doi: 10.3389/fnins.2019.00533.

Nellåker, Christoffer, Yuanrong Yao, Lorraine Jones-Brando, François Mallet, Robert H. Yolken, and Håkan Karlsson. 2006. "Transactivation of Elements in the Human Endogenous Retrovirus W Family by Viral Infection." *Retrovirology* 3. doi: 10.1186/1742-4690-3-44.

Nevalainen, Tapio, Arttu Autio, Binisha Hamal Mishra, Saara Marttila, Marja Jylhä, and Mikko Hurme. 2018. "Aging-Associated Patterns in the Expression of Human Endogenous Retroviruses." *PLoS ONE* 13(12):1–11. doi: 10.1371/journal.pone.0207407.

Nikoletopoulou, Vassiliki, Heiko Lickert, José Maria Frade, Chantal Rencurel, Patrizia Giallonardo, Lixin Zhang, Miriam Bibel, and Yves-Alain Barde. 2010. "Neurotrophin Receptors TrkA and TrkC Cause Neuronal Death Whereas TrkB Does Not." *Nature* 467(7311):59–63. doi: 10.1038/nature09336.

Paces, Jan, Adam Pavlícek, and Václav Paces. 2002. "HERVd: Database of Human Endogenous Retroviruses." *Nucleic Acids Research* 30(1):205–6. doi: 10.1093/nar/30.1.205.

Paces, Jan, Adam Pavlícek, Radek Zika, Vladimir V Kapitonov, Jerzy Jurka, and Václav Paces. 2004. "HERVd: The Human Endogenous RetroViruses Database: Update." *Nucleic Acids Research* 32(Database issue):D50. doi: 10.1093/nar/gkh075.

Padmanabhan Nair, Vidya, Hengyuan Liu, Gabriele Ciceri, Johannes Jungverdorben, Goar Frishman, Jason Tchieu, Gustav Y. Cederquist, Ina Rothenaigner, Kenji Schorpp, Lena Klepper, Ryan M. Walsh, Tae Wan Kim, Daniela Cornacchia, Andreas Ruepp, Jens Mayer, Kamyar Hadian, Dmitrij Frishman, Lorenz Studer, and Michelle Vincendeau. 2021. "Activation of HERV-K(HML-2) Disrupts Cortical Patterning and Neuronal Differentiation by Increasing NTRK3." *Cell Stem Cell* 28(9):1566-1581.e8. doi: 10.1016/j.stem.2021.04.009.

Perez-Pinera, Pablo, D. Dewran Kocak, Christopher M. Vockley, Andrew F. Adler, Ami M. Kabadi, Lauren R. Polstein, Pratiksha I. Thakore, Katherine A. Glass, David G. Ousterout, Kam W. Leong, Farshid Guilak, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. 2013. "RNA-Guided Gene Activation by CRISPR-Cas9-Based Transcription Factors." *Nature Methods* 10(10):973–76. doi: 10.1038/nmeth.2600.

Ponferrada, V. G., B. S. Mauck, and Dawn P. Wooley. 2003. "The Envelope Glycoprotein of Human Endogenous Retrovirus HERV-W Induces Cellular Resistance to Spleen Necrosis Virus." *Archives of Virology* 148(4):659–75. doi: 10.1007/s00705-002-0960-x.

Rehwinkel, Jan, and Michaela U. Gack. 2020. "RIG-I-like Receptors: Their Regulation and Roles in RNA Sensing." *Nature Reviews. Immunology* 20(9):537–51. doi: 10.1038/s41577-020-0288-3.

Rolland, Alexandre, Evelyne Jouvin-Marche, Christophe Viret, Mathias Faure, Hervé Perron, and Patrice N. Marche. 2006. "The Envelope Protein of a Human Endogenous Retrovirus-W Family Activates Innate Immunity through CD14/TLR4 and Promotes Th1-like Responses." *Journal of Immunology (Baltimore, Md. : 1950)* 176(12):7636–44. doi: 10.4049/jimmunol.176.12.7636.

Römer, Christine. 2021. "Viruses and Endogenous Retroviruses as Roots for Neuroinflammation and Neurodegenerative Diseases." *Frontiers in Neuroscience* 15:648629. doi: 10.3389/fnins.2021.648629.

Roulois, David, Helen Loo Yau, Rajat Singhania, Yadong Wang, Arnavaz Danesh, Shu Yi Shen, Han Han, Gangning Liang, Peter A. Jones, Trevor J. Pugh, Catherine O'Brien, and Daniel D. De Carvalho. 2015. "DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts." *Cell* 162(5):961–73. doi: 10.1016/j.cell.2015.07.056.

Saini, Sunil Kumar, Andreas Due Ørskov, Anne-Mette Bjerregaard, Ashwin Unnikrishnan, Staffan Holmberg-Thydén, Annie Borch, Kathrine Valentini Jensen, Govardhan Anande, Amalie Kai Bentzen, Andrea Marion Marquard, Tripti Tamhane, Marianne Bach Treppendahl, Anne Ortved Gang, Inge Høgh Dufva, Zoltan Szallasi, Nicola Ternette, Anders Gorm Pedersen, Aron Charles Eklund, John Pimanda, Kirsten Grønbæk, and Sine Reker Hadrup. 2020. "Human Endogenous Retroviruses Form a Reservoir of T Cell Targets in Hematological Cancers." *Nature Communications* 11(1):5660. doi: 10.1038/s41467-020-19464-8.

Saxena, Mansi, and Garabet Yeretssian. 2014. "NOD-Like Receptors: Master Regulators of Inflammation and Cancer." *Frontiers in Immunology* 5:327. doi: 10.3389/fimmu.2014.00327.

Scheiblauer, H., A. P. Kendal, and R. Rott. 1995. "Pathogenicity of Influenza A/Seal/Mass/1/80 Virus Mutants for Mammalian Species." *Archives of Virology* 140(2):341–48. doi: 10.1007/BF01309867.

Schmidt, Nora, Patricia Domingues, Filip Golebiowski, Corinna Patzina, Michael

H. Tatham, Ronald T. Hay, and Benjamin G. Hale. 2019. “An Influenza Virus-Triggered SUMO Switch Orchestrates Co-Opted Endogenous Retroviruses to Stimulate Host Antiviral Immunity.” *Proceedings of the National Academy of Sciences of the United States of America* 116(35):17399–408. doi: 10.1073/pnas.1907031116.

Schmitt, Katja, Kristina Heyne, Klaus Roemer, Eckart Meese, and Jens Mayer. 2015. “HERV-K(HML-2) Rec and Np9 Transcripts Not Restricted to Disease but Present in Many Normal Human Tissues.” *Mobile DNA* 6:4. doi: 10.1186/s13100-015-0035-7.

Schön, Ulrike, Olivia Diem, Laura Leitner, Walter H. Günzburg, Dixie L. Mager, Brian Salmons, and Christine Leib-Mösch. 2009. “Human Endogenous Retroviral Long Terminal Repeat Sequences as Cell Type-Specific Promoters in Retroviral Vectors.” *Journal of Virology* 83(23):12643–50. doi: 10.1128/JVI.00858-09.

Schwarz, Robert, Philipp Koch, Jeanne Wilbrandt, and Steve Hoffmann. 2022. “Locus-Specific Expression Analysis of Transposable Elements.” *Briefings in Bioinformatics* 23(1):1–10. doi: 10.1093/bib/bbab417.

Sheng, Peike, Krystal A. Flood, and Mingyi Xie. 2020. “Short Hairpin RNAs for Strand-Specific Small Interfering RNA Production .” *Frontiers in Bioengineering and Biotechnology* 8.

Smit, A. F. 1999. “Interspersed Repeats and Other Mementos of Transposable Elements in Mammalian Genomes.” *Current Opinion in Genetics & Development* 9(6):657–63. doi: 10.1016/s0959-437x(99)00031-3.

Smith, Christof C., Kathryn E. Beckermann, Dante S. Bortone, Aguirre A. Cubas, Lisa M. Bixby, Samuel J. Lee, Anshuman Panda, Shridar Ganesan, Gyan Bhanot, Eric M. Wallen, Matthew I. Milowsky, William Y. Kim, Kimryn Rathmell, Ronald Swanstrom, Joel S. Parker, Jonathan S. Serody, Sara R. Selitsky, and Benjamin G. Vincent. 2018. “Endogenous Retroviral Signatures Predict Immunotherapy Response in Clear Cell Renal Cell Carcinoma.” *Journal of Clinical Investigation* 128(11):4804–20. doi: 10.1172/JCI121476.

Smith, Cynthia L., and Janan T. Eppig. 2009. “The Mammalian Phenotype Ontology: Enabling Robust Annotation and Comparative Analysis.” *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* 1(3):390–99. doi: 10.1002/wsbm.44.

Smith, T. F., and M. S. Waterman. 1981. “Identification of Common Molecular Subsequences.” *Journal of Molecular Biology* 147(1):195–97. doi: 10.1016/0022-2836(81)90087-5.

Soltani, Mohammad H., Rita Pichardo, Ziqui Song, Namrata Sangha, Fabian Camacho, Kapaettu Satyamoorthy, Omar P. Sangueza, and Vijayasaradhi Setaluri. 2005. “Microtubule-Associated Protein 2, a Marker of Neuronal Differentiation, Induces Mitotic Defects, Inhibits Growth of Melanoma Cells, and Predicts Metastatic Potential of Cutaneous Melanoma.” *The American Journal of Pathology* 166(6):1841–50. doi: 10.1016/S0002-9440(10)62493-5.

Sperber, Göran O., Tove Airola, Patric Jern, and Jonas Blomberg. 2007.

"Automated Recognition of Retroviral Sequences in Genomic Data - RetroTector©." *Nucleic Acids Research* 35(15):4964–76. doi: 10.1093/nar/gkm515.

Srinivasachar Badarinarayan, Smitha, and Daniel Sauter. 2021. "Switching Sides: How Endogenous Retroviruses Protect Us from Viral Infections." *Journal of Virology* 95(12). doi: 10.1128/jvi.02299-20.

Srinivasachar Badarinarayan, Smitha, Irina Shcherbakova, Simon Langer, Lennart Koepke, Andrea Preising, Dominik Hotter, Frank Kirchhoff, Konstantin M. J. Sparrer, Gunnar Schotta, and Daniel Sauter. 2020. "HIV-1 Infection Activates Endogenous Retroviral Promoters Regulating Antiviral Gene Expression." *Nucleic Acids Research* 48(19):10890–908. doi: 10.1093/nar/gkaa832.

Subramanian, Ravi P., Julia H. Wildschutte, Crystal Russo, and John M. Coffin. 2011. "Identification, Characterization, and Comparative Genomic Distribution of the HERV-K (HML-2) Group of Human Endogenous Retroviruses." *Retrovirology* 8:90. doi: 10.1186/1742-4690-8-90.

Szobota, Stephanie, Pranav D. Mathur, Sairey Siegel, KristenAnn Black, H. Uri Saragovi, and Alan C. Foster. 2019. "BDNF, NT-3 and Trk Receptor Agonist Monoclonal Antibodies Promote Neuron Survival, Neurite Extension, and Synapse Restoration in Rat Cochlea Ex Vivo Models Relevant for Hidden Hearing Loss." *PLOS ONE* 14(10):e0224022.

Tai, Hao-Chih, Mohamed Ezzelarab, Hidetaka Hara, David Ayares, and David K. C. Cooper. 2007. "Progress in Xenotransplantation Following the Introduction of Gene-knockout Technology." *Transplant International* 20(2):107–17.

Tavares, Luciana P., Mauro M. Teixeira, and Cristiana C. Garcia. 2017. "The Inflammatory Response Triggered by Influenza Virus: A Two Edged Sword." *Inflammation Research : Official Journal of the European Histamine Research Society ... [et Al.]* 66(4):283–302. doi: 10.1007/s00011-016-0996-0.

The Gene Ontology Consortium. 2018. "The Gene Ontology Resource: 20 Years and Still GOing Strong." *Nucleic Acids Research* 47(D1):D330–38. doi: 10.1093/nar/gky1055.

Tönjes, R. R., R. Löwer, K. Boller, J. Denner, B. Hasenmaier, H. Kirsch, H. König, C. Korbmacher, C. Limbach, R. Lugert, R. C. Phelps, J. Scherer, K. Thelen, J. Löwer, and R. Kurth. 1996. "HERV-K: The Biologically Most Active Human Endogenous Retrovirus Family." *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology : Official Publication of the International Retrovirology Association* 13 Suppl 1:S261-7. doi: 10.1097/00042560-199600001-00039.

Toufaily, Chirine, Sebastien Landry, Christine Leib-Mosch, Eric Rassart, and Benoit Barbeau. 2011. "Activation of LTRs from Different Human Endogenous Retrovirus (HERV) Families by the HTLV-1 Tax Protein and T-Cell Activators." *Viruses* 3(11):2146–59. doi: 10.3390/v3112146.

Vargiu, Laura, Patricia Rodriguez-Tomé, Göran O. Sperber, Marta Cadeddu, Nicole Grandi, Vidar Blikstad, Enzo Tramontano, and Jonas Blomberg. 2016. "Classification and Characterization of Human Endogenous Retroviruses

Mosaic Forms Are Common." *Retrovirology* 13(1):1–29. doi: 10.1186/s12977-015-0232-y.

Vincendeau, Michelle, Ingmar Göttesdorfer, Julia M. H. Schreml, Armand G. Ngounou Wetie, Jens Mayer, Alex D. Greenwood, Markus Helfer, Susanne Kramer, Wolfgang Seifarth, Kamyar Hadian, Ruth Brack-Werner, and Christine Leib-Mösch. 2015. "Modulation of Human Endogenous Retrovirus (HERV) Transcription during Persistent and de Novo HIV-1 Infection." *Retrovirology* 12(1):27. doi: 10.1186/s12977-015-0156-6.

Wang, L., X. Wang, and M. L. Adamo. 2000. "Two Putative GATA Motifs in the Proximal Exon 1 Promoter of the Rat Insulin-like Growth Factor I Gene Regulate Basal Promoter Activity." *Endocrinology* 141(3):1118–26. doi: 10.1210/endo.141.3.7344.

Wang, Miao, Yang Qiu, Haizhou Liu, Bilin Liang, Baofeng Fan, Xi Zhou, and Di Liu. 2020. "Transcription Profile of Human Endogenous Retroviruses in Response to Dengue Virus Serotype 2 Infection." *Virology* 544(January):21–30. doi: 10.1016/j.virol.2020.01.014.

Wang, Miao, Liying Wang, Haizhou Liu, Jianjun Chen, and Di Liu. 2021. "Transcriptome Analyses Implicate Endogenous Retroviruses Involved in the Host Antiviral Immune System through the Interferon Pathway." *Virologica Sinica* 36(6):1315–26. doi: 10.1007/s12250-021-00370-2.

Wang, Tongguang, Marie Medynets, Kory R. Johnson, Tara T. Doucet-O'Hare, Brianna DiSanza, Wenxue Li, Yadi Xu, Anna Bagnell, Richa Tyagi, Kevon Sampson, Nasir Malik, Joseph Steiner, Alina Hadegan, Jeffrey Kowalak, James O'Malley, Dragan Maric, and Avindra Nath. 2020. "Regulation of Stem Cell Function and Neuronal Differentiation by HERV-K via MTOR Pathway." *Proceedings of the National Academy of Sciences of the United States of America* 117(30):17842–53. doi: 10.1073/pnas.2002427117.

Wang, Xiuling, Xiulin Wu, Jin Huang, Haiyan Li, Qiujin Yan, and Fan Zhu. 2021. "Human Endogenous Retrovirus W Family Envelope Protein (HERV-W Env) Facilitates the Production of TNF-α and IL-10 by Inhibiting MyD88s in Glial Cells." *Archives of Virology* 166(4):1035–45. doi: 10.1007/s00705-020-04933-8.

Weber, Melanie, Vidya Padmanabhan Nair, Tanja Bauer, Martin F. Sprinzl, Ulrike Protzer, and Michelle Vincendeau. 2021. "Increased HERV-K(HML-2) Transcript Levels Correlate with Clinical Parameters of Liver Damage in Hepatitis C Patients." *Cells* 10(4). doi: 10.3390/cells10040774.

Wilhelm, M., T. Heyman, M. Boutabout, and F. X. Wilhelm. 1999. "A Sequence Immediately Upstream of the Plus-Strand Primer Is Essential for plus-Strand DNA Synthesis of the Saccharomyces Cerevisiae Ty1 Retrotransposon." *Nucleic Acids Research* 27(23):4547–52. doi: 10.1093/nar/27.23.4547.

Yamashita, Masahiro, and Michael Emerman. 2004. "Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells." *Journal of Virology* 78(11):5670–78. doi: 10.1128/JVI.78.11.5670-5678.2004.

Yang, Jin, Hal P. Bogerd, Sheila Peng, Heather Wiegand, Ray Truant, and Bryan

R. Cullen. 1999. “An Ancient Family of Human Endogenous Retroviruses Encodes a Functional Homolog of the HIV-1 Rev Protein.” *Proceedings of the National Academy of Sciences of the United States of America* 96(23):13404–8. doi: 10.1073/pnas.96.23.13404.

Yates, Andrew D., Premanand Achuthan, Wasiu Akanni, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, Irina M. Armean, Andrey G. Azov, Ruth Bennett, Jyothish Bhai, Konstantinos Billis, Sanjay Boddu, José Carlos Marugán, Carla Cummins, Claire Davidson, Kamalkumar Dodiya, Reham Fatima, Astrid Gall, Carlos Garcia Giron, Laurent Gil, Tiago Grego, Leanne Haggerty, Erin Haskell, Thibaut Hourlier, Osagie G. Izuogu, Sophie H. Janacek, Thomas Juettemann, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, Jose Gonzalez Martinez, Thomas Maurel, Mark McDowall, Aoife McMahon, Shamika Mohanan, Benjamin Moore, Michael Nuhn, Denye N. Oheh, Anne Parker, Andrew Parton, Mateus Patricio, Manoj Pandian Sakthivel, Ahamed Imran Abdul Salam, Bianca M. Schmitt, Helen Schuilenburg, Dan Sheppard, Mira Sycheva, Marek Szuba, Kieron Taylor, Anja Thormann, Glen Threadgold, Alessandro Vullo, Brandon Walts, Andrea Winterbottom, Amonida Zadissa, Marc Chakiachvili, Bethany Flint, Adam Frankish, Sarah E. Hunt, Garth IIsley, Myrto Kostadima, Nick Langridge, Jane E. Loveland, Fergal J. Martin, Joannella Morales, Jonathan M. Mudge, Matthieu Muffato, Emily Perry, Magali Ruffier, Stephen J. Trevanion, Fiona Cunningham, Kevin L. Howe, Daniel R. Zerbino, and Paul Flicek. 2020. “Ensembl 2020.” *Nucleic Acids Research* 48(D1):D682–88. doi: 10.1093/nar/gkz966.

Young, George R., Jonathan P. Stoye, and George Kassiotis. 2013. “Are Human Endogenous Retroviruses Pathogenic? An Approach to Testing the Hypothesis.” *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 35(9):794–803. doi: 10.1002/bies.201300049.

Zhang, Yanxiao, Ting Li, Sebastian Preissl, Maria Luisa Amaral, Jonathan D. Grinstein, Elie N. Farah, Eugin Destici, Yunjiang Qiu, Rong Hu, Ah Young Lee, Sora Chee, Kaiyue Ma, Zhen Ye, Quan Zhu, Hui Huang, Rongxin Fang, Leqian Yu, Juan Carlos Izpisua Belmonte, Jun Wu, Sylvia M. Evans, Neil C. Chi, and Bing Ren. 2019. “Transcriptionally Active HERV-H Retrotransposons Demarcate Topologically Associating Domains in Human Pluripotent Stem Cells.” *Nature Genetics* 51(9):1380–88. doi: 10.1038/s41588-019-0479-7.

Zhu, Anqi, Joseph G. Ibrahim, and Michael I. Love. 2019. “Gene Expression Heavy-Tailed Prior Distributions for Sequence Count Data : Removing the Noise and Preserving Large Differences.” 35(November 2018):2084–92. doi: 10.1093/bioinformatics/bty895.