Universidad Católica de Valencia
San Vicente Mártir

TFG

TRABAJO FIN DE GRADO

GRADO EN BIOTECNOLOGÍA

GENETIC CONTROL OF TRANSPOSABLE ELEMENTS IN NEUROLOGICAL DISEASE

Sara Ferrando Juan
Elisa Oltra
Curso 2019-2020

Facultad de Veterinaria y Ciencias Experimentales
Universidad Católica de Valencia
San Vicente Mártir

ACKNOWLEDGMENTS

I would like to gratefully thank the Catholic University of Valencia for giving me the opportunity to live through years of enriching experiences that I will remember all my life, and for always encouraging me to study and investigate everything that has sparked my interest.

I would like to express my gratitude to my supervisor Dr. Elisa Oltra for her guidance, energy and support throughout the year and more especially since the pandemic began.

Many thanks to all my friends, who supported me every day and who cared for every "misfortune" that has happened to me, which have not been few but thanks to them everything seemed little.

And finally, to all my family members, and especially my parents, thanks to who I am and to whom I can only express my sincere gratitude for always supporting me and even more during the academic stage that today culminates.

Thanks to all of you.

INDEX

TABLE INDEX

FIGURE INDEX

ABBREVIATIONS

TE(s)	Transposable element(s)
LTR	Long terminal repeats
DIRS	*Dictyostelium* intermediate repeat sequence
PLE	*Penelope* - like element
LINE	Long interspersed nuclear element
SINE	Short interspersed nuclear element
TIR	Terminal inverted repeats
ERV(s)	Endogenous retroviruses
HERV(s)	Human endogenous retrovirus(es)
MS	Multiple sclerosis
MSRV	Multiple sclerosis-associated retroviral virus
ALS	Amyotrophic lateral sclerosis
SCHZ	Schizophrenia
ME/CFS	Myalgic encephalomyelitis/chronic fatigue syndrome
PEM	Post-exertional malaise
PTSD	Posttraumatic Stress Disorder
PMC	Pubmed Central
WOS	Web of Science
NCBI	National center for biotechnology information
ISI	Institute for scientific information
FECYT	Spanish foundation for science and technology
GSE	General search equation

SPSE	Specific search equation
SNP(s)	Single polymorphism(s)
PBMC	Peripheral blood mononuclear cells

ABSTRACT

Transposable elements represent (TEs) 45% of the genome of all eukaryotic organisms, they are DNA sequences that are characterized by being able to move between and within the genome. Specifically, human endogenous retroviruses (HERVs), a superfamily of TEs involving 8% of the human genome, have previously been related to cancer and diseases of the nervous system, together with genetic and environmental factors. The objective of this study is to carry out a systematic review that analyzes the results obtained in the last 20 years regarding correlations between abnormal control of HERVs and neurological diseases, including multiple sclerosis, post-traumatic stress, amyotrophic lateral sclerosis, schizophrenia and chronic fatigue syndrome. The results obtained are very diverse, each family of HERVs behaves differently in each disease; being HERV-K and HERV-W the most prevalent and multiple sclerosis the disease with the highest level of evidence for correlation with HERVs, estimated by the number of reports. Future research is necessary to determine whether a cause-effect relationship between HERVs and these disease exists, as well as the exact role of the different families of HERVs in the development of each of these diseases. This knowledge should help, establish new diagnostic tools and perhaps more effective treatments for these diseases.

Keywords: *Transposable element, human endogenous retrovirus, genetic control, expression, neurological disease*

RESUMEN

Los elementos transponibles representan el 45% del genoma de todos los organismos eucariotas, son secuencias de DNA que se caracterizan por ser capaces de desplazarse por el genoma y entre ellos. En concreto los retrovirus endógenos humanos, una superfamilia de estos, han sido previamente relacionados con el desarrollo de enfermedades el cáncer o enfermedades del sistema nervioso; en conjunto con demás factores de carácter ambiental y genéticos. El objetivo de este estudio es realizar una revisión sistemática que analice los resultados obtenidos en los últimos 20 años acerca de dichas posibles correlaciones con enfermedades neurológicas como la esclerosis múltiple, el estrés postraumático, esclerosis lateral amiotrófica, esquizofrenia y síndrome de fatiga crónica. Los resultados obtenidos son muy diversos, cada familia de HERVs se comporta de manera diferente en cada enfermedad; siendo HERV-K y HERV-W los más estudiados y la esclerosis múltiple la enfermedad más revisada de las propuestas.Futura investigación es necesaria para determinar cuál es el papel exacto de las diferentes familias de HERVs en el genoma humano y el desarrollo de enfermedades, así como el posible establecimiento de nuevas técnicas diagnósticas y tratamientos para estas enfermedades.

Palabras clave: *Elementos transponibles, retrovirus endógenos humanos, control genético, expression, enfermedades neurológicas.*

1. INTRODUCTION

1.1. Transposable elements, scientific background

Transposable elements (TEs), commonly known as mobile genetic elements, transposons, jumping genes, junk DNA and many other synonyms, were discovered by Barbara McClintock (1956) and described for first time as “normal components of the chromosome complement and that they are responsible for controlling, differentially, the time and type of individual genes” (p. 70) while she was studying on maize cytogenetics.

Defining these terms clearly and concisely has been a great challenge for the scientific community, adding complexity to their classification and limiting the implementation of a unified classification system of TEs. All definitions included below, have been accepted in one way or another, and used indifferently in the literature leading to confusion and other aggravating consequences.

The first proposal for a concise definition of TEs was formulated by Haren et al. (1999) who stated that “Transposable elements are discrete segments of DNA capable of moving from one locus to another in their host genome or between different genomes” (p. 246). Later, Kidwell & Lisch, (2001) suggested that “TEs are DNA sequences that have the capacity to change genomic locations” (p. 2). In 2015, taking into account the knowledge acquired over time, the definition was reformulated and settled to include additional characteristics of TEs “TEs are discrete segments of DNA capable of moving within a host genome from one chromosome or plasmid location to another and which do not use a specific molecular machinery that they encode to infect the genome of new hosts by lateral transfer” (p. 91) (Piégu et al., 2015).

TEs have been found in the genome of all eukaryotic organisms, constituting the major component of the mobilome and being one of the main factors determining the genome´s size, reaching up to 45% in humans (International Human Genome Sequencing Consortium, 2001; Kidwell, 2002; Siefert, 2009; Sun et al., 2012; Elliott & Gregory, 2015; Mita & Boeke, 2016).

1.1.1. Classification of transposable elements

In the same way as with the definition, the classification of TEs has been a subject of conflict in the scientific community (Arkhipova, 2017). There are various classifications and naming systems depending on the categorizing factors chosen as criteria for their subgrouping and organization.

Finnegan (1989) came up with the first classification system based on their transposition mechanism used, dividing TEs into: Class I or retrotransposons, more colloquially known as "copy-and-paste" elements, are segments that transpose their genomic site by making use of an RNA intermediate that is converted into complementary DNA (cDNA) by reverse transcription; and Class II or transposons, commonly known as "cut-and-paste" elements, which transpose straightly from their loci as DNA sequences without requiring any use of intermediary molecules.

With the progress of investigation over time, different classification systems proposals were elucidated. Although aimed at becoming a unified consensus system, in some cases they have conducted to new independent classification methods. The latter has been the case for the reports of Wicker et al. (2007); Kapitonov & Jurka, (2008); Seberg & Petersen, (2009) and Curcio & Derbyshire, (2003) reviewed in Piégu et al. (2015).

Currently, despite intense debate, the most widely used classification and nomenclature system for TEs of eukaryotic organisms is the proposed by Wicker et al. (2007) based on the original suggestions by Finnegan (1989, 1992). This system takes into account TE structural similarities and functionality while applying mechanistic and enzymatic criteria.

Two types of TEs were described, Class I constituted by five orders: long terminal repeats (LTR), *Dictyostelium* intermediate repeat sequence (DIRS), *Penelope* - like element (PLE), long and short interspersed nuclear element (LINE and SINE); and Class II, constituted by two subclasses which differ in their replication potential: Subclass 1 that includes two orders: terminal inverted repeats (TIR) and *Crypton* TEs; and Subclass 2 which also includes two orders *Helitron* and *Maverick* (Table 1). Within the orders listed above, the TEs are subgrouped in superfamilies, among which the endogenous retroviruses (ERVs) belonging to the LTR order, focus of the present Final Degree Project are found.

Table 1. Proposed classification system for TEs. Own elaboration, adapted from Wicker et al. (2007).

Order			Superfamily
Class I - Retrotransposons		LTR	*Copia*
			Gypsy
			Bel-Pao
			Retrovirus
			ERV
		DIRS	*DIRS*
			Ngaro
			VIPER
		PLE	*Penelope*
		LINE	*R2*
			RTE
			Jockey
			L1
			I
Class II - DNA transposons	Subclass 1	TIR	*Tc1-Mariner*
			hAT
			Mutator
			Merlin
			Transib
			P
			PiggyBac
			PIF-Harbinger
			CACTA
			Crypton
	Subclass 2	Helitron	*Helitron*
		Maverick	*Maverick*

1.2. Human endogenous retroviruses

Firstly, it should be noted that in humans, retroviruses can be classified according to their source into: exogenous and endogenous, which were discovered in 1980. Exogenous retroviruses spread as any other virus, infecting their target cells through specific host receptors; while endogenous receive their name because they are part of the genome of the organism and therefore are located within each one of the cells that constitute the individual (Christensen, 2010, 2016).

Human endogenous retrovirus (HERVs) are the result of the endogenization process and fixation into the germline product of ancestral retroviral infections and are successfully inherited through generations following classical Mendelian rules. They were acquired along the primate evolution or even at a previous evolutionary point (Griffiths, 2001; International Human Genome Sequencing Consortium, 2004; Christensen, 2005; Bannert & Kurth, 2006; Christensen, 2010).

The basic HERV structure consists of the three major retroviral genes *gag*, *pol* and *env* encoding internal structural polyproteins, a reverse transcriptase and an envelope glycoprotein that determines viral tropism respectively. These sequences appear flanked on both 5' and 3' ends by inverted LTR sequences, corresponding to long sequences of DNA repeated numerous times (Figure 1). These genomic elements constitute about 8% of the human genome. However, most of them are presumably inactive under normal conditions due to negative selection and the accumulation of independent mutations such as deletions, insertions, addition of termination codons in their reading frames, nucleotide substitutions or frame shifts that were acquired during evolution and that sometimes difficult their identification (Mager & Goodchild, 1989; Rasmussen et al., 1997; Christensen, 2010; Tugnet et al., 2013; Christensen, 2016; Grandi & Tramontano, 2017).

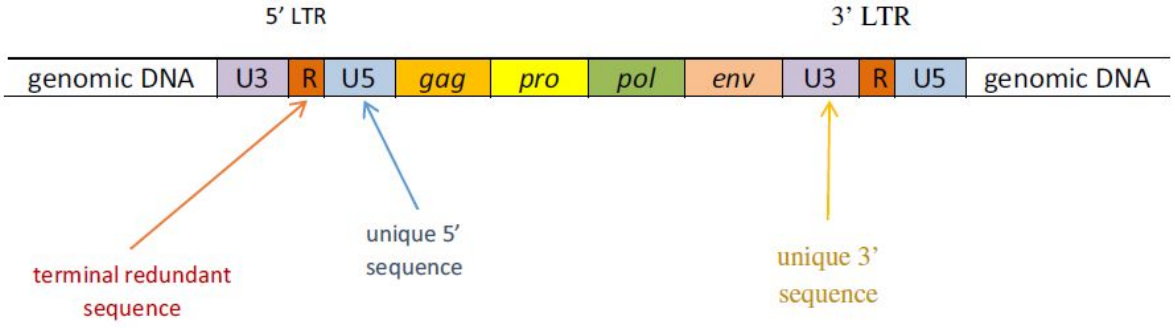

Figure 1: Diagrammatic general representation of HERVs structure (Morris et al., 2018).

1.2.1. Classification of Human endogenous retroviruses

To the date, about 31 to 40 independent families of HERVs have been described in the human genome, incorporated as result of an endogenization process. Some of these retroviruses have more than one identifying name; however as a general rule, a nomenclature based on the term "HERV" followed by the representative initial of the amino acid carried by the tRNA complementary to the primer binding site within the LTR region is used (Bannert & Kurth, 2006; Blikstad et al., 2008; Stocking & Kozack, 2008; Magiorkinis et al., 2013).

These families have traditionally been grouped into 3 classes, based on the similarity between their sequences: Class I or Gammaretrovirus, Class II or Betaretrovirus, and Class III or Spumavirus (Table 2) (Magiorkinis et al., 2013; Tugnet et al., 2013; Grandi & Tramontano, 2017). But in 2016, Vargiu et al. reorganized the classification based on phylogenetic similarities within in the *pol* gene across the different HERVs, by using an analysis method that they named “Simage”. This led to a total of 39 well-defined groups or HERV families and another 31 groups not considered family constituents because their origin was the result of a secondary integration of LTRs or other recombination events.

Table 2. Classification of HERVs. (Tugnet et al., 2013)

Class I	Class II	Class III
Group 1: HERV-HF HERV-H (RTVL-H, RGH) HERV-F	**Group 1: HML-1** HERV- K (HML-1.1)	HERV- L
Group 2: HERV-RW HERV-W HERV-R (ERV9) HERV-P (HuERS-P, HuRRS-P)	**Group 2: HML-2** HERV- K10 HERV-K-HTDV	HERV- S (HERV-18)
Group 3: HERV-ER1 HERV-E (4-1, ERVA, NP-2) 51-1 HERV-R (ERV3) RRHERV-1	**Group 3: HML-3** HERV- K (HML3.1)	HERV- U
Group 4: HERV-T HERV-T (S71, CRTK1,CRTK6)	**Group 4: HML-4** HERV- K-T47D	HERV- U3
Group 5: HERV-IP HERV-I (RTVL-I) HERV-IP-T47D (ERV-FTD)	**Group 5: HML-5** HERV- K-NMWV2	
Group 6: HERV-FRD ERV-FRD	**Group 6: HML-6** HERV- K (HML-6p)	
Others: HRES-1	**Group 7: HML-7** HERV- K-NMWV7	
	Group 8: HML-8 HERV- K-NMWV3	
	Group 9: HML-9 HERV- K-NMWV9	
	Group 10: HML-10 HERV-KC4	

1.3. Human endogenous retroviruses and disease

Currently, numerous are the evidences on the relationship between HERVs activations and various types of cancer such as breast cancer (Armbruester et al., 2002; Stauffer et al., 2004; Frank et al., 2008; Hohn et al., 2013; Magiorkinis et al., 2013; Kassiotis, 2014); autoimmune diseases such as rheumatoid arthritis (Sicat et al., 2005; Ehlhardt et al., 2006; Reynier et al., 2009) or systemic lupus erythematosus (Aldeman & Marchalonis, 2002; Tugnet et al., 2013); and neuropathological events such as those in multiple sclerosis (MS) or other immune-mediated diseases (Christensen, 2010; Magiorkinis et al., 2013; Grandi & Tramontano, 2017).

Although, the causal-effect relationship of retroviruses with some of the aforementioned diseases, such as MS, remains unclear; the alteration of the inactivated state of the HERVs and their deregulation, are believed to have pathogenic potential (Christensen, 2005, 2010, 2016; Hohn et al., 2013; Magiorkinis et al., 2013; Douville & Nath, 2014).

Family clustering of HERVs and their involvement with different autoimmune diseases supports the importance of their genetics in human pathogenesis. Although the mechanisms proposed are still under intense research, it is suspected that diverse environmental factors may explain why the expression of TEs varies within populations and perhaps triggers the development of autoimmune diseases (Kivity et al., 2009; H Perron et al., 2012; Tugnet et al., 2013).

1.3.1. Multiple Sclerosis

MS is a neurological and chronic inflammatory disease, leading to demyelination and axonal degeneration of the central nervous system (Keegan & Noseworthy, 2002; Christensen, 2016; Xiang et al., 2016; Morandi et al., 2017).

It is a disease that affects both men and women, although it appears more frequently in women, and is often diagnosed between the age of 20 to 40 years. The cause of MS has not yet been elucidated, but there is evidence pointing at stress caused by environmental factors as one of the main triggers. In recent years, the possible genetic variables involved, directly or indirectly, with this disease have generated interest, and the uncontrolled expression of HERVs has been proposed as a risk factor for its development (Prat & Antel, 2005; Ascherio & Munger, 2007; H. Perron et al., 2009; Christensen, 2016).

Specifically, research has focused primarily on two families of HERVs: HERV-H / F and HERV-W / multiple sclerosis-associated retroviral virus (MSRV) in blood, brain, and cerebrospinal fluid samples. The overexpression of these HERVs has also been detected in other diseases such as schizophrenia (Christensen, 2005; Prat & Antel, 2005; Antony et al., 2006; H. Perron et al., 2009; Laska et al., 2012; Douville & Nath, 2014; Hansen et al., 2017).

1.3.2. Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is an additional neurodegenerative disease affecting motor neurons, the cerebral cortex and spinal cord which means that it will lead to a progressive reduction of the individual's movement capacity. It is a variable disease with 90% of the cases diagnosed without a known genetic cause and with 10% of familiar cases (Alfahad & Nath, 2013; Renton et al., 2014; Brown & Phil, 2017; Mayer et al., 2018).

ALS has been associated to the overexpression of HERVs, some examples of identified evidences are: the *pol* activity level similar to that of human immunodeficiency virus patients (A. L. McCormick et al., 2008; Alfahad & Nath, 2013); and the overexpression of ERVK *env* protein promoting motor neuron death in a murine model of ALS (Li et al., 2015)

1.3.3. Schizophrenia

Schizophrenia (SCHZ) is a psychiatric disease including common symptoms such as hallucinations and the reduction and alteration of social capacities, including speech and affective relationships (L. M. McCormick & Flaum, 2005; Van Winkel et al., 2010).

It is a disease that affects both men and women, although it is generally diagnosed at earlier ages in men. The causes of the disease are thought to include a combination of genetic and environmental factors such as winter birth, maternal infection during pregnancy (by Influenza virus, Herpesvirus or *Toxoplasma gondii*), drug abuse during teenage years and the dysregulation of dopaminergic hormones, among other (Pearce et al., 2012; H Perron et al., 2012).

Nonetheless, genetic studies have revealed evidence of the relationship between HERVs overexpression and SCHZ. Similarly to what has been described for MS, MSRV components of the HERV-W family promotes the stimulation of the proinflammatory cascade of the immune system with neurotoxic potential also in SCHZ (Deb-Rinker et al., 1999; Hervé Perron et al., 2005; Huang et al., 2011; H Perron et al., 2012).

1.3.4. Chronic fatigue syndrome

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a complex long-lasting disease that causes the multisystemic weakening of the entire organism which consequently leads to: general pain in the joints and muscle exhaustion after physical exercise, an event described as PEM (post-exertional malaise), as well as headache, sore throat, cervical tenderness or axillary lymph node enlargement (Jason et al., 1999; Afari & Buchwald, 2003; Prins et al., 2006; Carruthers et al., 2011; Almenar-Pérez et al., 2019).

It is a disease that affects both men and women, although it is more prevalent in women. The cause or causes of the disease are still unknown since, as said above, its symptoms are varied and sometimes confused with other diseases. There is evidence to support that, at least for a subgroup of patients, it develops from an infection (Oakes et al., 2013; Scheibenbogen et al., 2017).

ME/CFS has been associated to the expression of HERVs by the proposal that their deregulation coincide with the symptomatic temporal outbreaks of symptoms (Oakes et al., 2013).

ME/CFS has been shown to appear comorbid to post-traumatic stress disorder (PTSD) (Dansie et al., 2012). Although no association between PTSD and HERVs available, the epigenetic transgenerational nature of PTSD and the fact that environmental stress triggers TE expression in rodents (Hunter et al., 2013), supports the search for a potential link between PTSD and HERV expression.

The aim of this study is to carry out a systematic review that analyzes the results of TE abnormal expression obtained in the last 20 years and the possible correlations between TE genetic control and different neurological diseases such as MS, PTSD, ALS, SCHZ, and ME/CFS.

2. OBJECTIVES

The main and specific objectives set for this Final Degree Project of bibliographic nature are listed below:

The main aim of this review is to gather available evidence supporting a role for aberrant silencing or activation of TEs in the development and/or maintenance of human neurological disease with a particular focus on MS, ALS, PTSD, SCHZ and ME/CFS.

Towards this end, the following specific aims were proposed:

1. Review the classification of transposable TEs.
2. Describe the different mechanisms of genetic control for main TE types in humans.
3. Explain, how the aberrant control of these genetic control mechanisms affects or is related to the development of certain neurological diseases.

3. MATERIAL & METHODS

3.1. Study search design

The documentary analysis method used to elaborate the hereby Final Degree Project was the classical systematic review method, the purpose of which is to obtain the "state of the art" information on the subject of study by rigorous and objective multiple searches. This method leads to the most complete and updated unbiased collection of documents in the topic of choice obtained by detecting and analyzing several rich and reliable sources of information.

Currently, systematic reviews have been shown as the ideal tools to compile all the knowledge referent on a topic that can be used as justification for future research to be carried out. The Preferred Reporting Items for Systematic reviews and Meta-Analyzes (PRISMA) methodology has been used successfully in various areas of science including health care and biotechnology, for which reason it was selected for the present review.

For this particular study, the following systematic review was performed in the aspects of the genetic control mechanisms that direct the silencing or activation of TEs and their relationship with neurological diseases.

3.1.1. Databases

The databases consulted were: Pubmed, Pubmed Central (PMC), Web of Science (WOS), Google Scholar and Scopus, described below.

PubMed created in 1996, is a free web bibliographic platform which to date house around 30 million citations related to different areas of life sciences and the biomedical field. The citations of Pubmed published articles are included in the associated MEDLINE database and in scientific journals of different nature and relevance, as well as in books and publishers' websites.

PMC, a database created in the year 2000, is a search engine that offers free access to full text publications in scientific journals related to the area of biomedicine and life sciences. To date it houses around 6 million publications. Like PubMed, PubMed Central was developed by the United States National Library of Medicine and is managed by the National Center for Biotechnology Information (NCBI). However, these two databases, despite their resembling name similarities and although they archive knowledge from similar areas, they present notable differences. For example, PMC, in addition to archiving, generates metadata and medical ontology files that allow the data to be structured in XML format.

WOS is a free digital bibliographic repository of the company Institute for Scientific Information (ISI), accessible in Spain trough research centers, universities, hospitals and other associated professional centers thanks to a national subscription provided by the Ministry of Education and Science through the Spanish Foundation for Science and Technology (FECYT) since 2004. It allows access to a wide collection of multidisciplinary databases that collect millions of entries included numerous journals, books, reviews etc; offering bibliometric impact analysis resources among other available services.

Google Scholar was created in 2004. Constitutes a bibliographic search engine which in 2018 already contained around 389 million documents of a wide thematic range published in academic journals, books, conference papers, thesis and dissertations, preprints, abstracts, technical reports, and other scholarly literature, including court opinions and patents. It should be noted that sometimes the search results are files including free full texts and that in others, the web portal is redirected to other databases in which to consult the desired entry.

Scopus was created in 1996, being the oldest among the databases consulted together with Pubmed. It is a database of the Elsevier company which is accessed by subscription. However, in the same way as with WOS, Scopus is managed by FECYT at the national level in Spain. Like Science Direct and other databases, it is multidisciplinary and includes, in addition to the usual publications in scientific journals, books, bibliographic reviews, etc, patents and bibliometric impact factor analyzers.

3.2. Search equations

Firstly, in order to elaborate the search equations, a series of keywords or descriptor terms related to the previously established study aims were selected. Additionally, synonyms of these search terms and specific keywords were included in the equations guaranteeing maximum coverage of the reviewed topic (Table 3).

Table 3: Selected search terms and synonyms.

Descriptor	Synonymous terms
Transposable element	Mobile genetic element
	Interspersed repetitive sequence
	Endogenous retrovirus
Genetic control	Genetic expression regulation
Neurological disease	Nervous system disease
	Multiple Sclerosis
	Posttraumatic Stress Disorder
	Amyotrophic Lateral Sclerosis
	Schizophrenia
	Chronic fatigue syndrome

Secondly, Boolean operators were used. These are linker terms that allow combining and connecting the different descriptors or set of concepts and their synonyms, logically defining relationships between them. They allow expansion or limitation in the performed searches according to particular needs during the bibliographic review process, building search equations of a general or more specific type focus.

The Boolean operators used were the most basic: (AND) shows only the findings that contain all the defined search terms regardless of the order in which they have been described, so the more descriptors are combined with this operator, the smaller it will be the number of results. And (OR) that shows the results that contain any of the terms combined with this operator, therefore the greater the use, the greater the number of results. Being especially useful when the aim is to introduce synonyms in the search equations and when all possible combinations are pursued. In addition, the quotation mark symbol (“ ”) was used to retrieve or maintain adjacent terms in searches.

Finally, making use of the concepts cited on Table 3 and the Boolean operators previously described, the search equations were assembled. It should be noted at this point that the search equations included both: the descriptors in their singular and plural variants in order to cover the largest number of records in the databases consulted; and that the main field used was “subject”, except in some cases, such as in PubMed and PMC databases, which allow the search of terms in all fields to be simultaneously searched by default.

The final equations applied in the performed searches are shown in Table 4. At first instance, general search equations were defined with the purpose of offering an overview on the review topic in question.

- In the general search equation number 1 (GSE - 1) terms referring to TEs and their synonyms were used, including “mobile genetic element” and “interspersed repetitive sequence”, on which introductory information was searched, such as definitions of relevance concepts , classifications, abundance, distribution of transposable elements; or effect on the evolution and diversity of TEs in the genomes etc.

- The general search equation number 2 (GSE - 2) focused only on the descriptor “endogenous retrovirus” with the same purpose as the previous equation in terms of gathering information of the specific subgroup of transposable elements, which concentrated the majority of neurological syndromes in humans.

Subsequently, specific search equations were designed with the objective of carrying out a more individualized/specific search, narrowing potential hits around more specific topics.

- In the specific search equation number 1 (SPSE - 1), the same descriptors were used as in (GSE -1) but the "genetic control" terms and their synonyms were added to refine the more general findings towards defining transposon control mechanisms and genetic silencing.

- In the specific search equation number 2 (SPSE - 2) the chosen design and objective was the same as in (SPSE - 1) but using (GSE - 2) as a base; therefore, it is constituted by “endogenous retrovirus”, “genetic control” and their respective synonyms.

- In the specific search equation number 3 (SPSE - 3), it is the one with the highest specificity, all the neurological diseases descriptors proposed on Table1 were added to the (SPSE-2) search with the aim of narrowing previous searches as much as possible to the specific topic as an attempt to review each of the diseases formerly mentioned.

Table 4: General and specific search equations used in the above cited databases.

Code	Name	Search equation
(GSE - 1)	General search equation	("dna transposable elements" OR ("dna" AND "transposable" AND "elements") OR ("transposable" AND "element") OR "transposable element") OR ("interspersed repetitive sequences" OR ("interspersed" AND "repetitive" AND "sequences") OR ("mobile" AND "genetic" AND "element") OR "mobile genetic element")
(GSE - 2)	General search equation	("endogenous retroviruses" OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR "human endogenous retrovirus" OR "herv")
(SPSE - 1)	Specific search equation	(("dna transposable elements" OR ("dna" AND "transposable" AND "elements") OR ("transposable" AND "element") OR "transposable element") OR ("interspersed repetitive sequences" OR ("interspersed" AND "repetitive" AND "sequences") OR ("mobile" AND "genetic" AND "element") OR "mobile genetic element")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control"))
(SPSE - 2)	Specific search equation	(("endogenous retroviruses") OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR ("human endogenous retrovirus") OR ("herv")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control"))

(SPSE - 3)	Specific search equation	(("endogenous retroviruses") OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR ("human endogenous retrovirus") OR ("herv")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control")) AND (("nervous system diseases" OR ("nervous" AND "system" AND "diseases") OR ("nervous system diseases") OR ("neurological" AND "disease") OR "neurological disease")) AND (("multiple sclerosis") OR ("posttraumatic stress disorder") OR ("amyotrophic lateral sclerosis") OR (“Schizophrenia”) OR ("ME/CFS") OR ("chronic fatigue syndrome"))

3.3. Study selection criteria

For the correct selection and filtering of the articles obtained, a series of inclusion and exclusion criteria were defined that allowed the results initially obtained to be limited to only those articles of relevance that could contribute to answering and reaching the questions and objectives, respectively, presented in the present Final Degree Project objectives.

3.3.1. Inclusion criteria

- Only original publications in scientific journals, books, conference summaries and Web portals, verified and reliable were included. Review articles would artificially enlarge and bias level of evidence for topics reviewed, as they do not provide new data.
- Publications concerning to human beings, as relationship of TE expression and human disease was pursued.
- Publications after January 2000, to limit the large number of hits to recent findings.
- In English language.

3.3.2. Exclusion criteria

- Publications not concerning to human beings
- Publications before January 2000.
- Publications in a different language than that determined in the inclusion criteria.

Furthermore, manual curation was subsequently applied to discard redundant findings, unrelated hits to the purpose of the searches that had exceeded the applied filters or that did not agree with the previously established objectives and to add articles of interest that for some reason have not been allocated by designed systematic searches performed. The latter served of articles found cited on the hits obtained.

4. RESULTS

This section summarizes the findings obtained with the multiple Bibliographic searches performed, after listing the number of records obtained and the steps taken for their appropriate classification and analysis. Manuscript gathering details follows with summaries of the data extracted, synthesized and presented in a tabulated form, responding and covering the objectives set forth in this Final Degree Project.

4.1. Results of the bibliographic review

Tables from 5 to 9 show, the number of associated records obtained for each search equation carried out in each of the databases consulted and the cumulative total number of hits obtained; showing both the “raw” results without applying any type of restriction and those obtained after applying the inclusion and exclusion criteria detailed in the methodology, prior to manual curation.

And, table 10 illustrates the summary from of the cumulative total number of records obtained in each database, also before and after the filtering process. A total of 741.827 “raw” entries were identified at start of the search process, ending with 248.672 records after the study selection criteria filters were applied.

The term “entry” is referred to: research articles, previous reviews related to the topic, conference proceedings as well as book chapters or any other type of format for presenting information of a scientific and contrasted nature.

Table 5: Results summary of the searches carried out with the general search equation 1 (GSE-1).

(GSE-1)	("dna transposable elements" OR ("dna" AND "transposable" AND "elements") OR ("transposable" AND "element") OR "transposable element") OR ("interspersed repetitive sequences" OR ("interspersed" AND "repetitive" AND "sequences") OR ("mobile" AND "genetic" AND "element") OR "mobile genetic element")

Database	"raw" hits	After filtering
Pubmed	29.734	5.041
WOS	40.770	8.594
PMC	63.162	1.398
Scopus	32.843	7.671
Google Scholar	35.400	21.100
TOTAL	201.909	43.804

Table 6: Results summary of the searches carried out with the general search equation 2 (GSE-2).

(GSE-2)	("endogenous retroviruses" OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR "human endogenous retrovirus" OR "herv")

Database	"raw" hits	After filtering
Pubmed	4.454	2.211
WOS	6.998	4.723
PMC	39.706	573
Scopus	6.463	4.232
Google Scholar	91.000	51.500
TOTAL	148.621	63.239

Table 7: Results summary of the searches carried out with the specific search equation 1 (SPSE-1).

(SPSE-1)	(("dna transposable elements" OR ("dna" AND "transposable" AND "elements") OR ("transposable" AND "element") OR "transposable element") OR ("interspersed repetitive sequences" OR ("interspersed" AND "repetitive" AND "sequences") OR ("mobile" AND "genetic" AND "element") OR "mobile genetic element")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control"))

Database	"raw" hits	After filtering
Pubmed	5.802	919
WOS	37.134	9.246
PMC	53.418	1.216
Scopus	7.926	3.715
Google Scholar	35.400	21.300
TOTAL	139.680	36.396

Table 8: Results summary of the searches carried out with the specific search equation 2 (SPSE-2).

(SPSE - 2)	(("endogenous retroviruses") OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR ("human endogenous retrovirus") OR ("herv")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control"))

Database	"raw" hits	After filtering
Pubmed	860	499
WOS	1.499	1.108
PMC	34.387	505
Scopus	25.274	19.625
Google Scholar	93.000	40.900
TOTAL	155.020	62.637

Table 9: Results summary of the searches carried out with the specific search equation 3 (GSE-3).

(SPSE - 3)	((("endogenous retroviruses") OR ("endogenous" AND "retroviruses") OR ("human" AND "endogenous" AND "retrovirus") OR ("human endogenous retrovirus") OR ("herv")) AND (("gene expression regulation" OR ("gene" AND "expression" AND "regulation") OR ("genetic" AND "control") OR "genetic control")) AND (("nervous system diseases" OR ("nervous" AND "system" AND "diseases") OR ("nervous system diseases") OR ("neurological" AND "disease") OR "neurological disease")) AND ((("multiple sclerosis") OR ("posttraumatic stress disorder") OR ("amyotrophic lateral sclerosis") OR (“Schizophrenia”) OR ("ME/CFS") OR ("chronic fatigue syndrome"))

Database	“raw” hits	After filtering
Pubmed	33	30
WOS	46	42
PMC	1.808	89
Scopus	1.710	1.535
Google Scholar	93.000	40.900
TOTAL	96.597	42.596

Table 10. Number of “raw” records per database and the number of records after applying the inclusion and exclusion criteria per database.

	PubMed	**WOS**	**PMC**	**Scopus**	**Google Scholar**	**# Total records**
“Raw” hits	40.883	86.447	192.481	74.216	347.800	**741.827**
After filtering	8.700	23.713	3.781	36.778	175.700	**248.672**

4.1.1. PRISMA flow diagram

As mentioned previously, 741.827 “raw” records were identified through database searching, screening those following the study selection criteria, were reduced to 248,672 hits, a significantly lower quantity than those obtained in first instance (Figure 2).

Figure 2: Number of records obtained for each search equation before and after applying study selection criteria.

However, 248,672 publications were still too high a number, so prior to continuing with manuscript eligibility process, it should be noted that at this point, with the information collected in tables from 5 to 10, it was decided to focus the bibliographic review on the SPSE-3 due to it was the most complete search equation since it encompasses all the key terms to be addressed. And more specifically in the databases: PubMed, WOS and PMC; due to the impossibility of reviewing the large number of records obtained. Therefore, the number of results at this point is 161 in the set of the three databases mentioned.

After manual curation, 98 unrelated hits to the previously established objectives and 14 duplicated records were discarded, (Table 11 shows, the rest of the selection process in detail).

Table 11. Detailed selection process for SPES-3 equation in each of the selected databases.

Database	Nº records	Nº Selected	Nº Duplicated	Relevant studies for analysis	Removed reviews
Pubmed	30	28	-	28	8
WOS	42	28	13	15	6
PMC	89	7	1	6	6
Total	161	63	14	49	20

Also, review articles were excluded as the final goal was to obtain empirical evidence of the relationship of particular TEs with the selected neurologic and autoimmune diseases. Nevertheless, they were screened for potential additional articles missed by the searches performed.

At the end, 29 records was the final number of relevant original articles obtained to be included for the detailed analysis in the present Final Degree Project.

A flow chart representing the information gathered and the screening process applied is shown underneath (Figure 3).

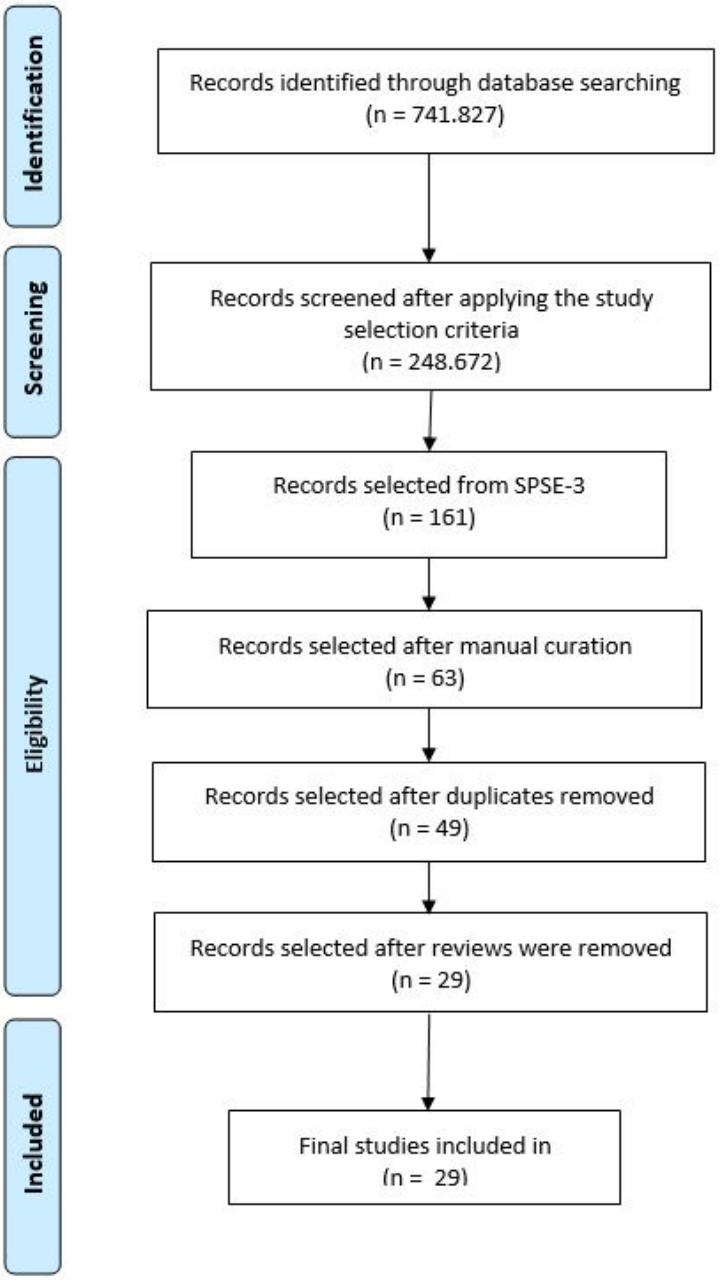

Figure 3. PRISMA flow diagram. PRISMA flow chart for search and record screening process.

4.1.2. Productivity per year

Next, Figure 4 shows a graph with the annual production of scientific articles published in the PubMed database, during the study period containing information about the control of gene expression in both the TEs and the HERVs in humans.

It can be appreciated how the number of published articles referring to TEs is significantly higher; the number of publications increases progressively with some variability over time until reaching 2013 when the maximum number is found with 288 publications, as of this year the cipher stabilizes with a small repeak in 2017, year from which productivity begins to decrease. Meanwhile HERVs productivity in the last 20 years has been maintained with an average of about 32 publications/year and always below 64 which constitutes the maximum number of articles, reached in 2012.

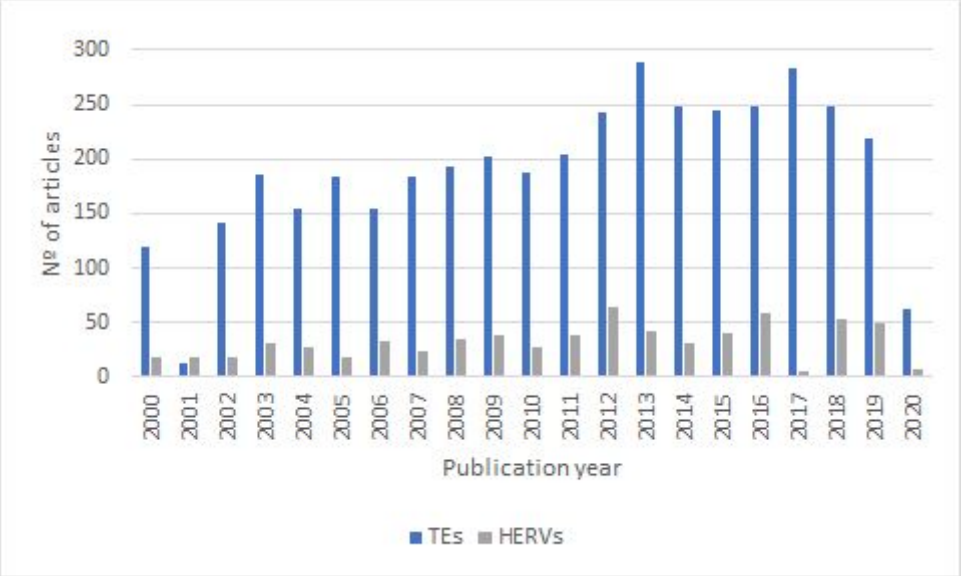

Figure 4: Productivity per year.

4.2. Genetic control mechanisms and disease

Of the 29 articles selected, 34.5% focused on the study of HERV-K elements, coinciding with the group of HERV-W; 13.8% studied HERV-H; 6.9% of the articles focused on the study of HERV-FRD; and finally the remaining 10.3% of the articles focused on other HERVs such as HERV-R or HERV-T that were only studied in 1 of the articles analyzed (Figure 5). It should be noted that 3 articles studied more than one HERV and therefore they were counted more than once.

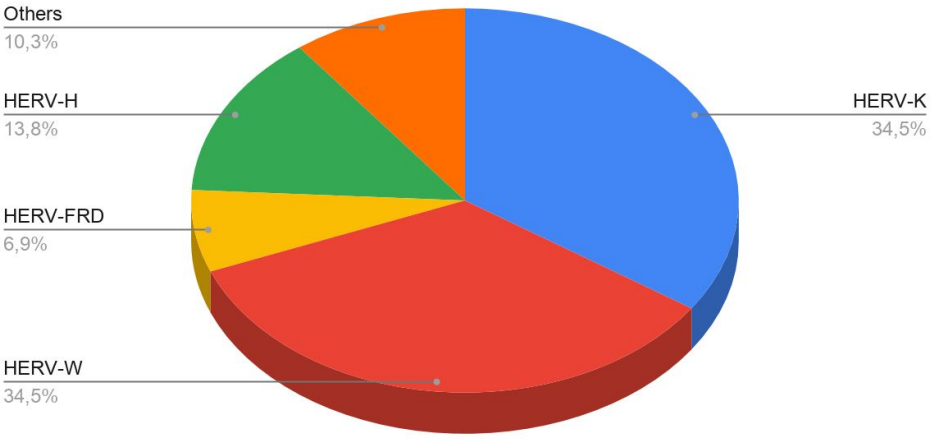

Figure 5: Percentage representation of the HERVs studied in the articles reviewed.

Also, of the diseases proposed for this review, MS is the most studied with a clear prevalence above all of them, being the focus of research in 60% of the publications reviewed; followed by ALS with 20% of publications, and SCHZ and CFS, with 10% for both diseases (Figure 6). About the additional pathology of interest PTSD, no publication was found among the selected articles as expected.

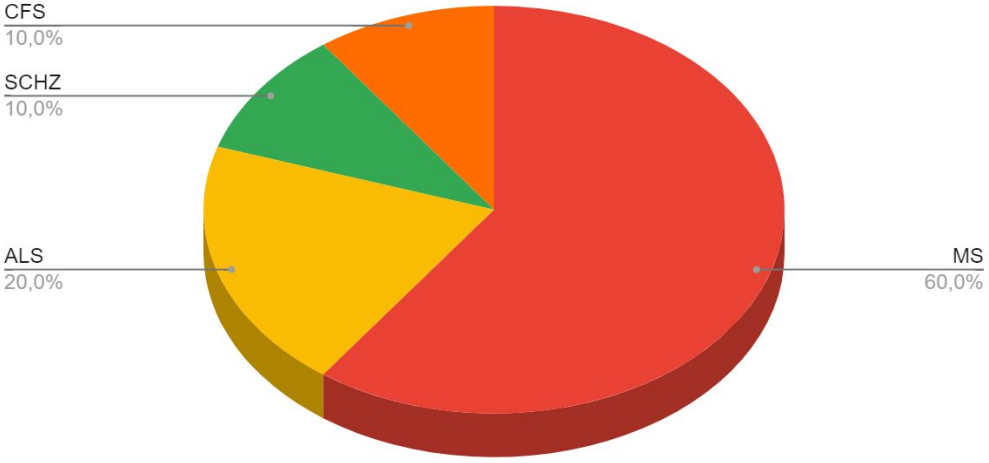

Figure 6: Percentage representation of the neurological diseases studied in the articles reviewed.

Table 12 presents a summary of the HERVs studied in the 29 selected articles, indicating the neurological disease and HERV element evaluated, the technique used by the authors and the protein or gene (nucleic acid) studied in each publication. Citing references are provided for each subtype of study.

Table 12: Summary of the HERVs studied

Disease	HERVs	Protein / nucleic acid investigated	Used technique	Reference
MS	HERV-FRD	*env*	qRT-PCR	(Antony et al., 2006; Brütting et al., 2018)
	HERV-R	*env*	RT-PCR	(Antony et al., 2006)
	HERV-T	*env*	RT-PCR	(Antony et al., 2006)
	HERV-H	*env*	RT-PCR	(Nissen et al., 2012; Antony et al., 2006; Brütting et al., 2016)
		gag	Flow cytometry	(Laska et al., 2012)
	HERV-K	*env*	RT-PCR	(Antony et al., 2006; Brütting et al., 2016)
	HERV-E	*env*	RT-PCR	(Antony et al., 2006)
	HERV-W	*env*	RT-PCR	(Hervé Perron et al., 2005; Antony et al., 2006; Mameli et al., 2013; Schmitt et al., 2013; Morandi et al., 2019)
		gag		(Hervé Perron et al., 2005)
		pol	RT-PCR	(Nowak et al., 2003)
ALS	HERV-K	*env*	qRT-PCR, q-PCR	(Li et al., 2015; ; Mayer et al., 2018)
		pol	q-PCR	(Manghera et al., 2016)
SCHZ	HERV-W	*env*	RT-PCR	(Perron et al., 2012; Xiao et al., 2017)
CFS	HERV-K	*env*	RT-PCR	(Oakes et al., 2013)

4.2.1. Multiple sclerosis

A total of nine articles studied the correlation between HERVs and MS. More specifically, most of the studies focussed on the HERV-W (6 articles) (Hervé Perron et al., 2005; Nowak et al., 2003; Antony et al., 2006; Mameli et al., 2013; Schmitt et al., 2013; Morandi et al., 2019); followed by studies of HERV-H elements (4 articles) (Antony et al., 2006; Laska et al., 2012; Nissen et al., 2012; Brütting et al., 2016), HERV-FRD (2 articles) (Antony et al., 2006; Brütting et al., 2016, HERV-K (2 articles) (Antony et al., 2006; Brütting et al., 2016), and HERV-R, HERV-T, HERV-E (1 article) (Antony et al., 2006).

Antony et al. (2006) analyzed the role of the *env* encoded syncytin-1 protein expression in brain, cerebrospinal fluid and plasma samples from 20 MS patients versus 19 non-MS subjects, and showed that the levels of expression of this protein-encoded transcript, measured using RT-PCR were higher in the brains of patients with MS. In the case of cerebrospinal fluid and brain samples, no significant differences were found for its expression between patients with MS and non-MS controls; and in the same way it happened in the case of plasma samples for HERV-H, -K (HML-2), -E, and -W.

Also, Brütting et al. (2016) studied the possible correlation between HERV-H and HERV-K *env* encoded protein with MS, by searching for putative open reading frames in the chromosomal regions surrounding single nucleotide polymorphisms (SNPs) present in patients, finding homologies of 80%; however this was not enough evidence to determine the specific role of these HERVs in the development of MS. It might be important to evaluate whether these SNPs in fact translate into induced transcription from these regions of the genome

Previously, analyzing the expression of HERV-H in the germ line of MS patients, no significant increase in the number of copies was detected with respect to the controls (Nissen et al., 2012), indicating no transpositions were associated to MS. By contrast, when flow cytometry was used to determine, the levels of expression of the *gag* protein in blood samples a greater presence in patients with MS evidenced, importantly suggesting that increased transcription from this element associates with MS (Laska et al., 2012).

HERV-W was the most studied TE, therefore including the largest number of related articles found in the systematic search, a total of six. They include analysis of the three main proteins that make HERV-W up: *gag, pol* and *env*. Confirmation of expression by more than one part of the element provides more solid evidence for the relationship between this element and MS.

Env protein correlation with MS seems to be unclear because it was overexpressed in macrophages during early MS lesions (Hervé Perron et al., 2005); but it does not show significant differences in expression when studying samples of white matter brain tissue of patients with MS (Schmitt et al., 2013). However, the differential or aberrant expression might affect specific cell types and/or stages of the disease.

Gag protein pathophysiological role in MS was more directly demonstrated, as its overexpression was observed in the endothelium and accumulated in the axonal structures of the white matter of the brain of patients with MS (Hervé Perron et al., 2005).

Pol protein overexpression in MS patients when compared to levels present in healthy control group, in both serum and PBMC (peripheral blood mononuclear cells) samples, suggested its implication in the pathogenesis of the disease (Nowak et al., 2003).

4.2.2. Amyotrophic lateral sclerosis

Of the articles reviewed, three relate to ALS and in all of them the relationship of HERV-K with this disease was studied in postmortem brain samples by different techniques including qRT-PCR and q-PCR. Specifically, the expression of the HERV-K viral RNA *env* and pol.

Two studies analyzed the role of the *env* protein in the development of ALS. Li et al. (2015) showed that the expression within neurons of patients with ALS may contribute to neurodegeneration and disease pathogenesis; while Mayer et al. (2018) specifically focused on the HML-2 group, finding no significant differences in expression levels between healthy controls and ALS patients, so, further research is necessary to determine their role in the triggering and/or progression of the disease.

And Manghera et al. (2016) evidenced the reactivation of HERV-K by the transcription factors NF-.kB / IRF1 as result of the inflammation present in the tissues of patients with ALS, for which they propose that HERV-K are involved in the development of the disease and that these transcription factors could be used as biomarkers for ALS diagnosis.

4.2.3. Schizophrenia

In the case of SCHZ, two articles were studied. Both evaluated how the development of the disease is potentiated by the presence of the *env* protein encoded by the HERV-W family of TEs.

On one hand, the role of *env* was analyzed from the study of microglial cells from brain samples; showing that the overexpression of the protein associated with increased levels of nitric oxide on the analyzed cells, which is linked to the inflammation in the disease (Xiao et al., 2017).

On another hand, it was shown that the levels of *env* transcription were increased in patients with SCHZ than in healthy controls in blood samples; while the opposite occurred with the number of copies of DNA in leukocytes (H Perron et al., 2012).

Meaning that although a reduced number of copies of HERVs is present in the genome of patients with SCHZ, they lead to higher levels of *env* transcripts. Therefore, the role of this part of HERV elements in the pathogenesis and the development of neurological diseases, including SCHZ, seems confirmed.

4.2.4. Chronic fatigue syndrome

Only one article was found that studied the involvement of HERVs in CFS. Oakes et al. (2013) using blood and saliva samples, the authors tried to demonstrate a correlation between the levels of transcription of the *env* protein in HERV-K18 and the number of viral copies of HHV-6 and HH-7; by qPCR technique; with the aim of using them as biomarkers for the detection of the disease. However, no significant difference was observed between CFS patients and the healthy control groups.

5. DISCUSSION

In first place, before starting the discussion, it must be taken into account that due to the large number of articles obtained through the search equations, it was decided to only analyze those belonging to the specific search equation number 3; since it is the most complete of all and allowed to reduce the number of hits obtained. Therefore, the review does not cover bibliography or findings on the initially set objective one, and only indirectly, some information on objective two.

SPE-3 search equation, identified as mostly containing information related to secondary objectives number two and three: write the different mechanisms of genetic control for main transposable element types in humans; and explain, how the aberrant control of these genetic control mechanisms affect or are related to the development of certain neurological diseases. To at least partially cover objective one: the review of the classification of transposable elements, a non-systematic review was carried out in the introduction section.

Then, for the results section, only the 29 articles obtained after filtering the SPE-3 were used, following the chosen inclusion and exclusion criteria. And, to carry out the discussion, the discussion section of the articles obtained in the SPE-3 systematic search were used. In addition, a quick parallel search of reviews was carried out to include specific topics not covered in depth by any of the 20 reviews obtained in the systematic search. This extra effort was directed to cover all the available and most current information about the present Final Degree Project topic.

5.1. Multiple sclerosis

In first place, regarding the results obtained in the review for the studied MS association with HERV-H and HERV-K elements, the results obtained were found somewhat diverse: Antony et al. (2006) and Nissen et al. (2012) agreed that there is no significant difference in HERV-H concentration between MS patients and healthy controls, while Laska et al. (2012) evidenced the opposite.

The differences between these studies are several: the sequences analyzed, the methodology used and the type of samples. In the first two studies: *env*, RT-qPCR as the target gene and brain, serum or cerebrospinal fluid mass the samples evaluated; while for the third article flow cytometry, directly assaying on the protein product of the *gag* gene on PBMC samples were analyzed.

And with HERV-W, the authors evaluated the gene encoding the *gag* and *pol* proteins demonstrating to have a role in MS development (Nowak et al., 2003; Hervé Perron et al., 2005); while the role of the *env* encoding protein could not be confirmed due to the controversial evidence found (Hervé Perron et al., 2005; Antony et al., 2006; Schmitt et al., 2013).

The heterogeneity associated to the methods may explain the apparent controversy found. The discrepancies are expected to the covered by future research studies.

5.2. Amyotrophic lateral sclerosis

Secondly, very diverse results were provided in the studies by Li et al. (2015) and Manghera et al. (2016), presenting evidence that, in the case of HERV-K, it plays a role in the pathogenesis of ALS; while other failed to reach the same conclusion (Mayer et al., 2018).

The reasons for these controversial conclusions seem to be that the results obtained by Li et al. (2015) and Manghera et al. (2016) may seem initially valid and demonstrate the relationship between HERV-K and the disease. However, because overexpression of a non-disease-causing HERV may mask another lesser expressed HERV (Dolei et al., 2019), the next step was to check whether *env* proteins inducing toxicity to the identified HERV-K (HML-2) loci were transcribed and this was not the case, as later studies showed (Mayer et al., 2018).

Furthermore, it is possible that disparate results were obtained due to the choice of primers used for the RT-qPCR in each investigation, since the more specific they are, the more they will filter the results obtained. To cover larger number of copies, some authors make use of degenerate primers at the expense of limiting the information on the loci being affected.

Lastly, it should be noted that in different ALS studies, it has not been possible to simulate 100% the cell culture conditions of the samples, that would occur *in vivo*. For example, it was not possible to simulate the loss of motor neurons, this detail is a limitation and partly alters the results obtained, as would the growth of undesirable cells, considered "foreign cells" in the culture. To avoid this problem, (Mayer et al., 2018) proposes the prior measurement of HERV concentrations and subsequent classification of the foreign cells for culture. This could constitute an approach, to normalize the results obtained.

5.3. Schizophrenia

In third place, Perron et al. (2012) and Xiao et al. (2017) demonstrated that HERV-W play a role in the progression of SCHZ.

The difference between the two studies is that Xiao et al. (2017) focuses on the role of nitrogen oxide in the development of inflammation, this being a signaling factor or a response when HERV-W is overexpressed. The authors propose that the overexpression of this TE element could be used as a biomarker of SCHZ, but its specificity remains to be established, as its overexpression has also been shown in other neurological diseases progressing with inflammation and abnormal immune responses.

And Perron et al., (2005), show that the transcription levels of *env* HERV-W transcripts which they studied in both SCHZ and patients with bipolarity with respect to healthy controls; confirming their suspicions that individuals with SCHZ or bipolar disorder present with abnormal HERV-W expression levels (Frank et al., 2005; Huang et al., 2011) and so do patients suffering from both diseases simultaneously.

This would support the potential lack of specificity of these TEs and perhaps end up explaining the overlapping neurological symptoms present in some of these patients.

5.4. Chronic fatigue syndrome

And lastly, in fourth place, the lack of correlation between HERV-K and ME/CFS reported by Oakes et al. (2013) is in disagreement with the more recent results obtained by Rodrigues et al. (2019), also in blood samples, who showed able to demonstrate HERV-K overexpression in moderately affected patients. The latter authors also informed that in the cohort studied HERV-W, in contrast with the HERV-K elements do not show any differential expression; which in turn is consistent with the results obtained by (Montoya et al., 2017) in the study including evaluation of cytokine signatures reflecting an inflammatory process associated to ME/CFS.

The difference between the results by Oakes et al. (2013) and Rodrigues et al. (2019) could be explained because the first one only focuses on the expression of the HERV-K18 *env* protein, using specific primers for the analysis of this sequence. While in Rodrigues' study, the *env* HERV-K uses generic, degenerate primers to detect abnormal expression of as many subfamilies as possible (Rodrigues et al., 2019).

Now, taking into account the results obtained by Rodrigues et al. (2019), it can be interpreted that different HERVs have different roles in developing ME/CFS; although knowing what pattern of expression for each patient is somewhat unexplored at present, it may turn to end up being the diagnosis or treatment approach for this disease, as individualized medicine may be necessary for an effective treatment of complex diseases.

The reason for the predicted individual variability is unknown, but everything indicates that the aberrant expression of these TEs is due, at least partly, to the environmental stress factors that each person in their environment is subjected to throughout its particular life, a process known as the individual exposome (Andreoli & Tincani, 2017; Almenar-Pérez et al., 2019).

Latter in 2020, Ovejero et al. were the first to observe high levels of HERV-H, -K and -W in patients of fibromyalgia, a disease highly related to ME/MFS. Interestingly enough the levels of some of these TEs correlated with the levels of inflammation associated factors interferon beta and gamma and not with other, including tumor necrosis alpha. As they used degenerate primers, nonspecific, to screen by RT-qPCR most of the elements present in the genome no specificity to the disease can be attributed, and thus, the potential use as biomarkers of the disease remains to be determined. Future research will be necessary to identify specific genomic loci differentially affected in FM and / or ME/CFS and their potential association with the inflammation markers of the individuals.

6. CONCLUSIONS

From the bibliographic review accomplished by this Final Degree Project is finished, the following conclusions can be drawn:

1. The absence of a consensus for the nomenclature and classification of both the transposable elements and the human endogenous retroviruses in particular; can lead to confusion and make difficult to collect and interpret significant data.

2. Among the different neurological diseases, multiple sclerosis is the most studied by far.

3. The controversial findings may associate to methodological heterogeneity and the lack of studies using individual loci interrogation.

4. Knowing what elements and how genetic control over the expression of HERVs occurs, allows them to be used as biomarkers or biosensors and to develop possible diagnostic guides and treatments for these diseases.

5. Future research is necessary to determine the exact role of the different families of HERVs in the human genome and the regulation of the expression to be able to predict development of the associated diseases

7. FUTURE LINES OF RESEARCH

The review carried out in this Final Degree Project has allowed the compilation of interesting aspects, specifically the potential roles played by the HERVs with regard to their regulation in the human genome and the involvement of their expression with different neurological diseases.

This, in turn has allowed raising new questions as well as being able to define possible future lines of research, which are listed below:

First of all, regarding the study of TEs in general, the findings stress the need of establishing fluid communication in the scientific community. Improved collaborative circles in the field could provide significant information about how TEs are classified and controlled towards the establishment of consensus classifications and nomenclatures that will in turn contribute to filling the knowledge gaps currently present in the area.

Secondly, continuation of the investigations of HERV families as well as their complementary sequences in various organisms seems not only granted but necessary, being the establishment of model organisms of great interest, as they may allow unravelling the function of TEs and to establish cause-effect relationships with physiology and pathology.

Thirdly, through genetic engineering, taking into account the knowledge about the expression and regulation of HERVs, their regulatory mechanisms could be used to induce resistance to exogenous or harmful infections.

And lastly, highlighting the importance of continuing the study of HERVs and their relationships with different diseases, to be used as biomarkers to improve current diagnosis methods, establish prognosis predictions and perhaps allow treatment response monitoriezation in individualized medical programs.

What environmental factors might trigger unproper control of HERVs is still an enigmatic issue. As exogenous viral infections might be one of the triggers, it might be of interest to evaluate what are the effects of the recent SARS-CoV-2 pandemic on the control of these endogenous sequences.

8. REFERENCES

Afari, N., & Buchwald, D. (2003). Chronic Fatigue Syndrome: A Review. *Am J Psychiatry*, 16.

Aldeman, M. K., & Marchalonis, J. J. (2002). Endogenous retroviruses in systemic Lupus Erythematosus: Candidate Lupus Viruses. *Clinical Immunology*, *102*(2), 107–116.

Alfahad, T., & Nath, A. (2013). Retroviruses and amyotrophic lateral sclerosis. *Antiviral Research*, *99*(2), 180–187. https://doi.org/10.1016/j.antiviral.2013.05.006

Almenar-Pérez, E., Ovejero, T., Sánchez-Fito, T., Espejo, J. A., Nathanson, L., & Oltra, E. (2019, April). Epigenetic Components of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Uncover Potential Transposable Element Activation. In *Clinical Therapeutics*. https://linkinghub.elsevier.com/retrieve/pii/S0149291819300724

Andreoli, L., & Tincani, A. (2017). Undifferentiated connective tissue disease, fibromyalgia and the environmental factors: *Current Opinion in Rheumatology*, *29*(4), 355–360. https://doi.org/10.1097/BOR.0000000000000392

Antony, J. M., Izad, M., Bar-Or, A., Warren, K. G., Vodjgani, M., Mallet, F., & Power, C. (2006). Quantitative Analysis of Human Endogenous Retrovirus-W *env* in Neuroinflammatory Diseases. *AIDS Research and Human Retroviruses*, *22*(12), 1253–1259. https://doi.org/10.1089/aid.2006.22.1253

Arkhipova, I. R. (2017). Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mobile DNA*, *8*(1), 19. https://doi.org/10.1186/s13100-017-0103-2

Armbruester, V., Sauter, M., Krautkraemer, E., Meese, E., Kleiman, A., Best, B., Roemer, K., & Mueller-Lantzsch, N. (2002). A Novel Gene from the Human Endogenous Retrovirus K Expressed in Transformed Cells1. *Clinical Cancer Research*, *8*, 1800–1807.

Ascherio, A., & Munger, K. L. (2007). Environmental Risk Factors for Multiple Sclerosis. Part I: the role of infection. *Ann. Neurol.*, *61*, 288–299.

Bannert, N., & Kurth, R. (2006). The Evolutionary Dynamics of Human Endogenous Retroviral Families. *Annual Review of Genomics and Human Genetics*, *7*(1), 149–173. https://doi.org/10.1146/annurev.genom.7.080505.115700

Blikstad, V., Benachenhou, F., Sperber, G. O., & Blomberg, J. (2008). Evolution of human endogenous retroviral sequences:a conceptual account. *Cell. Mol. Life Sci.*, *65*, 3348–3365.

Brown, R. H., & Phil, D. (2017). Amyotrophic Lateral Sclerosis. *N Engl j Med*, 11.

Brütting, C., Emmer, A., Kornhuber, M., & Staege, M. S. (2016). A survey of endogenous retrovirus (ERV) sequences in the vicinity of multiple sclerosis (MS)-associated single nucleotide polymorphisms (SNPs). *Molecular Biology Reports*, *43*(8), 827–836. https://doi.org/10.1007/s11033-016-4004-0

Brütting, C., Narasimhan, H., Hoffmann, F., Kornhuber, M. E., Staege, M. S., & Emmer, A. (2018). Investigation of Endogenous Retrovirus Sequences in the Neighborhood of Genes Up-regulated in a Neuroblastoma Model after Treatment with Hypoxia-Mimetic Cobalt Chloride. *Frontiers in Microbiology*, *9*, 287. https://doi.org/10.3389/fmicb.2018.00287

Carruthers, B. M., van de Sande, M. I., De Meirleir, K. L., Klimas, N. G., Broderick, G., Mitchell, T., Staines, D., Powles, A. C. P., Speight, N., Vallings, R., Bateman, L., Baumgarten-Austrheim, B., Bell, D. S., Carlo-Stella, N., Chia, J., Darragh, A., Jo, D., Lewis, D., Light, A. R., … Stevens, S. (2011). Myalgic encephalomyelitis: International Consensus Criteria: Review: ME: Intl. Consensus Criteria. *Journal of Internal Medicine*, *270*(4), 327–338. https://doi.org/10.1111/j.1365-2796.2011.02428.x

Christensen, T. (2005). Association of human endogenous retroviruses with multiple sclerosis and possible interactions with herpes viruses. *Reviews in Medical Virology*, *15*(3), 179–211. https://doi.org/10.1002/rmv.465

Christensen, T. (2010). HERVs in Neuropathogenesis. *Journal of Neuroimmune Pharmacology*, *5*(3), 326–335. https://doi.org/10.1007/s11481-010-9214-y

Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. *APMIS*, *124*(1–2), 116–126. https://doi.org/10.1111/apm.12486

Curcio, M. J., & Derbyshire, K. M. (2003). The outs and ins of transposition: From Mu to Kangaroo. *Nature Reviews Molecular Cell Biology*, *4*(11), 865–877. https://doi.org/10.1038/nrm1241

Dansie, E. J., Heppner, P., Furberg, H., Goldberg, J., Buchwald, D., & Afari, N. (2012). The Comorbidity of Self-Reported Chronic Fatigue Syndrome, Post-Traumatic Stress Disorder, and Traumatic Symptoms. *Psychosomatics*, *53*(3), 250–257. https://doi.org/10.1016/j.psym.2011.08.007

Deb-Rinker, P., Klempan, T. A., O'Reilly, R. L., Torrey, E. F., & Singh, S. M. (1999). Molecular Characterization of a MSRV-like Sequence Identified by RDA from Monozygotic Twin Pairs Discordant for Schizophrenia. *Genomics*, *61*(2), 133–144. https://doi.org/10.1006/geno.1999.5946

Dolei, A., Ibba, G., Piu, C., & Serra, C. (2019). Expression of HERV Genes as Possible Biomarker and Target in Neurodegenerative Diseases. *International Journal of Molecular Sciences*, *20*(15), 3706. https://doi.org/10.3390/ijms20153706

Douville, R. N., & Nath, A. (2014). Huamn endogenous retroviruses and the nervous system. *Handboof of Clinical Neurobiology*, *123*.

Ehlhardt, S., Seifert, M., Schneider, J., Ojak, A., Zang, K., & Mehraein, Y. (2006). Human

endogenous retrovirus HERV-K (HML-2) rec expression and transcriptional activities in nirmal and rheumatoid arthritis synovia. *Journal Rheumatology*, *1*(33), 16–23.

Elliott, T. A., & Gregory, T. R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1678), 20140331. https://doi.org/10.1098/rstb.2014.0331

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Elevier Science Publishers Ltd.*, *5*(4), 103–107.

Finnegan, D. J. (1992). Transposable elements. *Current Opinion in Genetics & Development*, *2*(6), 861´ – 867.

Frank, O., Giehl, M., Zheng, C., Hehlmann, R., Leib-Mösch, C., & Seifarth, W. (2005). Human Endogenous Retrovirus Expression Profiles in Samples from Brains of Patients with Schizophrenia and Bipolar Disorders. *Journal of Virology*, *79*(17), 10890–10901. https://doi.org/10.1128/JVI.79.17.10890-10901.2005

Frank, O., Verbeke, C., Schwarz, N., Mayer, J., Fabarius, A., Hehlmann, R., Leib-Mösch, C., & Seifarth, W. (2008). Variable Transcriptional Activity of Endogenous Retroviruses in Human Breast Cancer. *Journal of Virology*, *82*(4), 1808–1818. https://doi.org/10.1128/JVI.02115-07

Grandi, N., & Tramontano, E. (2017). Type W Human Endogenous Retrovirus (HERV-W) Integrations and Their Mobilization by L1 Machinery: Contribution to the Human Transcriptome and Impact on the Host Physiopathology. *Viruses*, *9*(7), 162. https://doi.org/10.3390/v9070162

Griffiths, D. J. (2001). *Endogenous retroviruses in the human genome sequence*. 5.

Hansen, D. T., Petersen, T., & Christensen, T. (2017). Retroviral envelope proteins: Involvement in neuropathogenesis. *Journal of the Neurological Sciences*, *380*, 151–163. https://doi.org/10.1016/j.jns.2017.07.027

Haren, L., Ton-Hoang, B., & Chandler, M. (1999). Integrating DNA: Transposases and Retroviral Integrases. *Annual Review of Microbiology*, *53*(1), 245–281. https://doi.org/10.1146/annurev.micro.53.1.245

Hohn, O., Hanke, K., & Bannert, N. (2013). HERV-K (HML-2), the best preserved family of HERVs: Endogenization, expression, and implications in health and disease. *Frontiers in Oncology*, *3*(246).

Huang, W., Li, S., Hu, Y., Yu, H., Luo, F., Zhang, Q., & Zhu, F. (2011). Implication of the env Gene of the Human Endogenous Retrovirus W Family in the Expression of BDNF and DRD3 and Development of Recent-Onset Schizophrenia. *Schizophrenia Bulletin*, *37*(5), 988–1000. https://doi.org/10.1093/schbul/sbp166

Hunter, R. G., McEwen, B. S., & Pfaff, D. W. (2013). Environmental stress and transposon transcription in the mammalian brain. *Mobile Genetic Elements*, *3*(2), e24555. https://doi.org/10.4161/mge.24555

Ibba, G., Piu, C., Uleri, E., Serra, C., & Dolei, A. (2018). Disruption by SaCas9 Endonuclease of HERV-Kenv, a Retroviral Gene with Oncogenic and Neuropathogenic Potential, Inhibits Molecules Involved in Cancer and Amyotrophic Lateral Sclerosis. *Viruses*, *10*(8), 412. https://doi.org/10.3390/v10080412

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/nature03001

Jason, L. A., Richman, J. A., Rademaker, A. W., Jordan, K. M., Plioplys, A. V., Taylor, R. R.,

McCready, W., Huang, C.-F., & Plioplys, S. (1999). A Community-Based Study of Chronic Fatigue Syndrome. *Archives of Internal Medicine*, *159*(18), 2129.

Kapitonov, V. V., & Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *G e n e Ti c s*, 2.

Kassiotis, G. (2014). Endogenous Retroviruses and the Development of Cancer. *The Journal of Immunology*, *192*(4), 1343–1349. https://doi.org/10.4049/jimmunol.1302972

Keegan, B. M., & Noseworthy, J. H. (2002). Multiple Sclerosis. *Annual Review of Medicine*, *53*, 285–302.

Kidwell, M. G. (2002). *Transposable elements and the evolution of genome size in eukaryotes*. 15.

Kidwell, M. G., & Lisch, D. R. (2001). PERSPECTIVE: TRANSPOSABLE ELEMENTS, PARASITIC DNA, AND GENOME EVOLUTION. *Evolution*, *55*(1), 1–24. https://doi.org/10.1111/j.0014-3820.2001.tb01268.x

Kivity, S., Agmon-Levin, N., Blank, M., & Shoenfeld, Y. (2009). Infections and autoimmunity – friends or foes? *Trends in Immunology*, *30*(8), 409–414. https://doi.org/10.1016/j.it.2009.05.005

Laska, M. J., Brudek, T., Nissen, K. K., Christensen, T., Moller-Larsen, A., Petersen, T., & Nexo, B. A. (2012). Expression of HERV-Fc1, a Human Endogenous Retrovirus, Is Increased in Patients with Active Multiple Sclerosis. *Journal of Virology*, *86*(7), 3713–3722. https://doi.org/10.1128/JVI.06723-11

Li, W., Lee, M.-H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., Campanac, E., Hoffman, D. A., von Geldern, G., Johnson, K., Maric, D., Morris, H. D., Lentz, M., Pak, K., Mammen, A., Ostrow, L., Rothstein, J., & Nath, A. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Science Translational Medicine*, *7*(307), 307ra153-307ra153. https://doi.org/10.1126/scitranslmed.aac8201

Mager, D. L., & Goodchild, N. L. (1989). *Homologous recombination between the LTRs of human retrovirus-like element causes a 5-kb deletion in two siblings*. *45 (6)*, 848–854.

Magiorkinis, G., Belshaw, R., & Katzourakis, A. (2013). There and back again: Revisiting the pathophysiological roles of human endogenous retroviruses i the post-genomic era. *Phil Trans R Soc B*, 368.

Mameli, G., Madeddu, G., Mei, A., Uleri, E., Poddighe, L., Delogu, L. G., Maida, I., Babudieri, S., Serra, C., Manetti, R., Mura, M. S., & Dolei, A. (2013). Activation of MSRV-Type Endogenous Retroviruses during Infectious Mononucleosis and Epstein-Barr Virus Latency: The Missing Link with Multiple Sclerosis? *PLoS ONE*, *8*(11), e78474. https://doi.org/10.1371/journal.pone.0078474

Manghera, M., Ferguson-parry, J., Lin, R., & Douville, R. N. (2016). NF-kB and IRF1 induce endogenous retrovirus-K expression via interferon stimulated response elments in its 5' long terminal repeat. *Journal of Virology*.

Mayer, J., Harz, C., Sanchez, L., Pereira, G. C., Maldener, E., Heras, S. R., Ostrow, L. W., Ravits, J., Batra, R., Meese, E., García-Pérez, J. L., & Goodier, J. L. (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. https://doi.org/10.1186/s13024-018-0275-3

McClintock, B. (1956). Intranuclear systems controlling gene action and mutation. *Brookhaven Symp Biol.*, *8*, 58–74.

McCormick, A. L., Brown, R. H., Cudkowicz, M. E., Al-Chalabi, A., & Garson, J. A. (2008). Quantification of reverse transcriptase in ALS and elimination of a novel retroviral candidate. *Neurology*, *70*(4), 278–283. https://doi.org/10.1212/01.wnl.0000297552.13219.b4

McCormick, L. M., & Flaum, M. (2005). *Diagnosing schizophrenia circa 2005: How and why?*

5.

Mita, P., & Boeke, J. D. (2016). How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development*, *37*, 90–100.

Montoya, J. G., Holmes, T. H., Anderson, J. N., Maecker, H. T., Rosenberg-Hasson, Y., Valencia, I. J., Chu, L., Younger, J. W., Tato, C. M., & Davis, M. M. (2017). Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proceedings of the National Academy of Sciences*, *114*(34), E7150–E7158. https://doi.org/10.1073/pnas.1710519114

Morandi, E., Tanasescu, R., Tarlinton, R. E., Constantinescu, C. S., Zhang, W., Tench, C., & Gran, B. (2017). The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. *PLoS ONE*, *12*(2).

Morandi, E., Tanasescu, R., Tarlinton, R. E., Constantin-Teodosiu, D., & Gran, B. (2019). Do Antiretroviral Drugs Protect From Multiple Sclerosis by Inhibiting Expression of MS-Associated Retrovirus? *Frontiers in Immunology*, *9*, 3092. https://doi.org/10.3389/fimmu.2018.03092

Morris, G., Maes, M., Murdjeva, M., & Puri, B. K. (2018). Do human endogenous retrviruses contribute to multiple sclerosis and if so, how? *Molecular Neurobiology*.

Nissen, K. K., Laska, M. J., Hansen, B., Pedersen, F. S., & Nexø, B. A. (2012). No additional copies of HERV-Fc1 in the germ line of multiple sclerosis patients. *Virology Journal*, *9*(1), 188. https://doi.org/10.1186/1743-422X-9-188

Nowak, J., Januszkiewicz, D., Pernak, M., Liwen, I., Zawada, M., Rembowska, J., Nowicka, K., Lewandowski, K., Hertmanoswska, H., & Wender, M. (2003). Multiple sclerosis-asscociated virus -related pol sequences found both in multiple sclerosis and healthy donors are more frequentlu expressed in multiple sclerosis patientes. *Journal of NeuroVirology*, *9*, 112–117.

Oakes, B., Hoagland-Henefield, M., Komaroff, A. L., Erickson, J. L., & Huber, B. T. (2013).

Human Endogenous Retrovirus-K18 Superantigen Expression and Human Herpesvirus-6 and Human Herpesvirus-7 Viral Loads in Chronic Fatigue Patients. *Clinical Infectious Diseases*, *56*(10), 1394–1400. https://doi.org/10.1093/cid/cit086

Ovejero, T., Sadones, O., Sánchez-Fito, T., Almenar-Pérez, E., Espejo, J. A., Martín-Martínez, E., Nathanson, L., & Oltra, E. (2020). Activation of Transposable Elements in Immune Cells of Fibromyalgia Patients. *International Journal of Molecular Sciences*, *21*(4), 1366. https://doi.org/10.3390/ijms21041366

Pearce, B. D., Kruszon-Moran, D., & Jones, J. L. (2012). The Relationship Between Toxoplasma Gondii Infection and Mood Disorders in the Third National Health and Nutrition Survey. *Biological Psychiatry*, *72*(4), 290–295. https://doi.org/10.1016/j.biopsych.2012.01.003

Perron, H., Bernard, C., Bertrand, J.-B., Lang, A. B., & Popa, I. (2009). Endogenous retroviral genes, Herpesviruses and gender in Multiple Sclerosis. *Journal of the Neurological Science*, *286*, 65–72.

Perron, H, Hamdani, N., Faucard, R., Lajnef, M., Jamain, S., Daban-Huard, C., Sarrazin, S., LeGuen, E., Houenou, J., Delavest, M., Moins-Teisserenc, H., Bengoufa, D., Yolken, R., Madeira, A., Garcia-Montojo, M., Gehin, N., Burgelin, I., Ollagnier, G., Bernard, C., … Leboyer, M. (2012). Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. *Translational Psychiatry*, *2*(12), e201–e201. https://doi.org/10.1038/tp.2012.125

Perron, Hervé, Lazarini[23], F., Ruprecht[4], K., Péchoux-Longin[6], C., Seilhean[2], D., Sazdovitch[2], V., Créange[7], A., Battail-Poirot[1], N., Sibai[1], G., Santoro[1], L., Jolivet[1], M., Darlix[6], J.-L., Rieckmann[4], P., Arzberger[5], T., Hauw[2], J.-J., & Lassmann[8], H. (2005). Human endogenous retrovirus (HERV)-W ENV and GAG proteins: Physiological expression in human brain and pathophysiological modulation in multiple sclerosis lesions. *Journal of NeuroVirology*, *11*(1), 23–33. https://doi.org/10.1080/13550280590901741

Piégu, B., Bire, S., Arensburger, P., & Bigot, Y. (2015). A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution*, *86*, 90–109. https://doi.org/10.1016/j.ympev.2015.03.009

Prat, A., & Antel, J. (2005). Pathogenesis of multiple sclerosis. *Current Opinion in Neurology*, *18*, 225–230.

Prins, J. A., Van der Meer, J. W. M., & Bleijenberg, G. (2006). Chronic fatigue syndrome. *The Lancet*, *367*, 346–355.

Rasmussen, H. B., Geny, C., Deforges, L., Perron, H., Tourtelotte, W., Heltberg, A., & Clausen, J. (1997). Expression of endogenous retroviruses in blood mononuclear-cells and brain tissue from multiple sclerosis patients. *Acta Neurologica Scandinavica*, *95*, 38–44. https://doi.org/10.1111/j.1600-0404.1997.tb08148.x

Renton, A. E., Chiò, A., & Traynor, B. J. (2014). State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, *17*(1), 17–23. https://doi.org/10.1038/nn.3584

Reynier, F., Verjat, T., Turrel, F., Imbert, P. E., Marotte, H., Mougin, B., & Miossec, P. (2009). Increase in Human Endogenous Retrovirus HERV-K (HML-2) Viral Load in Active Rheumatoid Arthritis. *Scandinavian Journal of Immunology*, *70*(3), 295–299. https://doi.org/10.1111/j.1365-3083.2009.02271.x

Rodrigues, L. S., da Silva Nali, L. H., Leal, C. O. D., Sabino, E. C., Lacerda, E. M., Kingdon, C. C., Nacul, L., & Romano, C. M. (2019). HERV-K and HERV-W transcriptional activity in myalgic encephalomyelitis/chronic fatigue syndrome. *Autoimmunity Highlights*, *10*(1), 12. https://doi.org/10.1186/s13317-019-0122-8

Scheibenbogen, C., Freitag, H., Blanco, J., Capelli, E., Lacerda, E., Authier, J., Meeus, M., Castro Marrero, J., Nora-Krukle, Z., Oltra, E., Strand, E. B., Shikova, E., Sekulic, S., & Murovska, M. (2017). The European ME/CFS Biomarker Landscape project: An initiative of the European network EUROMENE. *Journal of Translational Medicine*,

15(1), 162. https://doi.org/10.1186/s12967-017-1263-z

Schmitt, K., Richter, C., Backes, C., Eckart, M., Ruprecht, K., & Mayer, J. (2013). Comprehensive analysis of human endogenous retrovirus group HERV-W locus transcription in multiple sclerosis brain lesions by high-throughput amplicon sequencing. *Journal of Virology*, *87*(24), 13837–13852.

Seberg, O., & Petersen, G. (2009). A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature Reviews Genetics*, *10*(4), 276–276. https://doi.org/10.1038/nrg2165-c3

Sicat, J., Sutkowski, N., & Huber, B. T. (2005). Expression of Human Endogenous Retrovirus HERV-K18 superantigen is elevated in juvenile Rheumatoid Arthritis. *Journal Rheumatology*, *9*(32), 1821–1831.

Siefert, J. L. (2009). Defining the Mobilome. In M. B. Gogarten, J. P. Gogarten, & L. C. Olendzenski (Eds.), *Horizontal Gene Transfer* (Vol. 532, pp. 13–27). Humana Press. https://doi.org/10.1007/978-1-60327-853-9_2

Stauffer, Y., Thelier, G., Sperisen, P., Lebedev, Y., & Jongeneel, V. (2004). Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immunity*, *4*, 19.

Stocking, C., & Kozack, C. A. (2008). Murine endogenous retroviruses. *Cell. Mol. Life Sci.*, *65*, 3383–3398.

Sun, C., López Arriaza, J. R., & Mueller, R. L. (2012). Slow DNA Loss in the Gigantic Genomes of Salamanders. *Genome Biology and Evolution*, *4*(12), 1340–1348. https://doi.org/10.1093/gbe/evs103

Tugnet, N., Rylance, P., Roden, D., Trela, M., & Nelson, P. (2013). Human endogenous retroviruses (HERVs) and autoimmune Rheumatic disease: Is there a link? *The Open*

Rheumatology Journal, *7*, 13–21.

Van Winkel, R., Esquivel, G., Kenis, G., Wichers, M., Collip, D., Peerboms, O., Rutten, B., Myin-Germeys, I., & Van Os, J. (2010). Genome-wide findings in schizophrenia and the role of gene-environment interplay. *CNS Neuroscience & Therapeutics*, *16*, 185–192.

Vargiu, L., Rodriguez-Tomez, P., Cadeddu, M., Grandi, N., Blikstad, V., Tramontano, E., & Blomberg, J. (2016). Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology*, *13*.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Xiang, L., Xiao, B., & Chen, X.-S. (2016). DNA methylation: A new player in Multiple Sclerosis. *Molecular Neurobiology*.

Xiao, R., Li, S., Cao, Q., Wang, X., Yan, Q., Tu, X., Zhu, Y., & Zhu, F. (2017). Human endogenous retrovirus W env increases nitric oxide production and enhances the migration ability of microglia by regulating the expression of inducible nitric oxide synthase. *Virologica Sinica*, *32*(3), 216–225. https://doi.org/10.1007/s12250-017-3997-4