The Coevolution of LINE-1 Retroelements and the Immune Restriction Factor TEX19

Ricky Padilla Del Valle

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington, Seattle, WA

2024

Reading Committee:

Rick McLaughlin

Adam Geballe

Adam Steinbrenner

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

The Coevolution of LINE-1 Retroelements and the Immune Restriction Factor TEX19

Ricky Padilla Del Valle

Chair of the Supervisory Committee:

Rick McLaughlin

Pacific Northwest Research Institute, Molecular and Cellular Biology, and Genome Sciences

This thesis explores coevolutionary dynamics between genomes and intragenomic parasites, focusing on LINE-1 retrotransposons in humans. LINE-1s persist and evolve despite robust defense mechanisms, suggesting an ongoing genetic conflict. The study investigates how human genomes combat LINE-1 activity through restriction factor genes. Our research tested 34 diverse human LINE-1s from an individual genome against various restriction factors, finding that while some genes, like APOBEC3B, consistently restrict LINE-1s, TEX19 shows variable restriction, with 11 LINE-1s showing strong evasion. These evasive LINE-1s are also active *in vivo*, evidenced by their replication and 3' transductions in the human population. Additionally, variants of TEX19 in humans can lessen LINE-1 restriction, suggesting individual variability in LINE-1 regulation and potential implications for developmental disorders related to LINE-1s. The study further describes the dynamic evolution of *TEX19* across mammals, identifying multiple independent duplications, losses, and structural variants, especially in rodents, bats, and carnivores. Ultimately, this work illustrates the intricate coevolution between LINE-1s and TEX19, highlighting the evolutionary arms race that shapes innate immune defenses and providing insights into the mechanisms of genomic stability and the evolutionary pressures shaping LINE-1 and restriction factor interactions in mammals.

A mi abuela Carmen *Gloria* Morales Rodríguez, gracias. Me enseñaste a sembrar, a cocinar y a transformar mis ideas en manjares de sabiduría. Tu vasta experiencia tejió con amor y cuidado mi confianza a crecer, a hacer crucigramas a tinta al igual que se hace en la vida. Gracias por hablar de mi con rebosante orgullo en la antesala del doctor, en la fila del correo, en la plaza del mercado, en el más allá.

Gracias a mi madre Raquel Del Valle Martínez. Fuiste la musa, el coraje, y la valentía que capacitó y energizó mi motor por años. Gracias por enseñarme a seguir caminando, aunque llueva, truene o relampaguee, aunque no existan sendas diseñadas para mí. Aprendí de ti la escuela, el baile, y la pasión de oro que hoy decora mis atuendos, propulsa mis pasos y enriquece mis relaciones interpersonales. Gracias por tu ejemplo.

A mi padre Richard Padilla Morales, gracias. Fuiste la humildad, la serenidad y lo humano que invoco cuando todo es difícil. Aprendí de ti que ser fuerte es intentar la paciencia, a apreciar la manualidad, el trabajo y la construcción de mis esfuerzos. Aprendí de ti a siempre volver al origen, a los mangles de la playa, donde siempre hay inspiración. Gracias por tu insistencia.

A mis hermanes, Jorge Omar Serrano Del Valle y a Drachira Raquel Padilla Del Valle, gracias por acompañarme a creer en y a expresar quién soy, por interactuar con y desafiar mis intrigas, por enseñarme de la armonía que hay entre las olas del mar y las rocas de la orilla, a pesar de ser radicalmente distintas.

A mis amigues y familia, mil y unas gracias. A Berliza Marie Soriano, Kiara Minelly Rodríguez Núñez, Rossana Colón Thillet, Josef Cole, Santiago León Cole Colón, Erik Anderson Oliver y a mi tio *Tito* Víctor Manuel Padilla Morales, gracias por su crítica constructiva, apoyo sincondicional y gentil constancia.

A mis mentores, maestres y colegas, gracias. A Ms. Martí y Ms. Camen *Mandy* Matos (3r grado) de la escuela David G. Farragut. A Mr. Iván *Chijo* López (7mo grado), Mr. Otero y Ms. Mazda Soto (8vo) de la escuela José Gautier Benítez. A Ms. Brenda Cabrera (9no y 10mo grado) del Centro Residencial de Oportunidades Educativas en Mayagüez (CROEM). A Dr. Frank X. Ferrer

González (11mo y 12vo grado) y Dr. Carlos Ríos Velázquez de la Universidad de Puerto Rico en Mayagüez. Y al Dr. Rick N. McLaughlin, Jr. de la Universidad de Washington en Seattle. Gracias por formarse conmigo y por confiar en mi curiosidad. Su mentoría fue instrumental en mi desarrollo y práctica de las ciencias, siempre considerando el impacto de la investigación en mi comunidad. Gracias por escuchar, guiar y retar mis ideas. Gracias por ofrecerme y regarlame sus herramientas. Fue un honor instruirme con ustedes. Prometo ser para otros quienes ustedes fueron para mi.

A mi gatita Emma, gracias. Eres la mejor acompanante de mis días desde que abuela se fue. Has sido mi ancla y copirata en este viaje de océanos tormentosos y marejadas picadas. Gracias por traer novedad y calma a mis vivencias.

Y a mi madre patria Puerto Rico, Borikén. Soy bienaventurade y suertude de haber nacido en el desorden de tu realidad colonial mientras esperas tu gloriosa independencia. Heredo tu propósito y el privilegio de representar tu bandera a donde quiera que vaya. No hay nada como crecer en tus contextos; correr en tus montes, sembrar en tus sierras, caminar en tus valles, jugar en tus colinas, bañarme en tus riveras, tomar el sol en tus orillas, nadar en tus playas, pasear por tus bosques, dormir en tus montanas, explorar tus mangles, valorar tus arrecifes, reposar en tus arenas, saborear de tus frutos, escuchar tus coquies, sentir tus lluvias torreciales, declamar tus poemas, regocijarme al sonido de tus tambores, y temer tus tormentas. Eres y siempre serás el estímulo que me inspiró a entender la ciencia que apenas comienza a explicar tu natural, infinita y ancestral magia. Gracias a la ciudad del mangó. Gracias al pueblo de Betánces. Gracias al rio Yagüez, que me vio nacer. Gracias al rio Guanajibo, que me pidió ayuda. Gracias a Joyuda y al cerro Las Mesas. Armame de fe, valor y fuerza para levantarme con alevocia, táctica y estrategia en contra de tus opresores, para por fin vuelvan las riquesas a nuestra noble tierra. Gracias por persistir, que por ti resisto.

En pie de lucha,

Ricky

Table of Contents

Chapter 1: Background

In this section, I will start with an overview of coevolutionary processes on the scale of life, then present the interesting case of coevolution of genomes with intragenomic parasites called transposable elements (TEs). Next, I will review the diversity of TEs across life before focusing on the TEs present in human genomes. Using the human genome as a model, I will present the mechanisms used by the human genome to combat the deleterious effects of transposable elements on diverse aspects of human biology as well as the evolution of the central class of TEs in humans (LINE-1s) and how what we know about the coevolution of LINE-1s and innate immune restriction factors serves as motivation for the experimental and computational work presented in this thesis probing the coevolution of human LINE-1s and a restriction factor called TEX19.

Coevolution

Groups of organisms and other replicators must adapt to changes in their environment to optimize their fitness(*1*). Variation exists within a population among individuals, and when the environment changes, certain variants that encode traits may become advantageous. Individuals with these advantageous traits and their causative genetic variants are more likely to survive and, therefore, pass these genetic variants on to their offspring.

In addition to changes in the ‘non-living’ environment (i.e., meteor impact, earthquakes, which shape a species’ evolution, organisms are also subject to the pressures of coevolution – a process where one species affects the evolution of another. In coevolving species, a change in the genetic composition of one species drives a reciprocal change in the genetic composition of another species. Classic examples of coevolution include predator-prey interactions where an increase in the ability of the prey to evade or the predator to capture exerts pressure on the other party to adapt. Coevolution or covariation plays out at numerous scales of biological systems. For example, the amino acids of proteins coevolve over evolutionary time to create networks of covarying positions that underlie some key aspects of protein function(*2*). Similarly, genes coevolve to maintain signaling networks(*3*). In addition, some coevolutionary processes have driven deep intertwining of the genetics and physiology of organisms, like the coevolution of deep-sea tube worms and their resident bacteria. These tube worms have lost their digestive system and depend entirely on their endosymbiotic sulfur-fixing bacteria to chemosynthesize nutrients from the harsh environment of hydrothermal vents(*4*) Or the fascinating coevolution of flowering plants and their pollinators. Flowers have developed specific shapes, colors, and scents that attract their particular pollinators. In return, pollinators have evolved traits that enable them to access the nectar and pollen of those flowers(*5*).

Another central model system for studying coevolution has been the interaction of pathogens or parasites with their hosts, typified by extensive studies on phage-bacteria(*6*) and retrovirus-host systems(*7*, *8*). Less emphasis has been placed on understanding the coevolution of host genomes and their genome-bound endogenous elements, but several studies have focused on the theory and current evidence for this mode of coevolution.

Types of Coevolution: Symbiotic and Antagonistic

Symbiosis

Symbiotic relationships span a continuum from facultative, where the symbiosis is beneficial but not essential for survival, to obligate, where the survival of one or both replicators depends critically on the interaction. Understanding these relationships provides insight into the evolutionary forces and adaptations that arise from these complex interactions, driving the diversification and evolution of genomes and species. Based on the nature of the interaction, symbiotic relationships can be classified into three main types: mutualism, commensalism, and parasitism.

Mutualism

In mutualism, all partners benefit from the interaction. An example is the relationship between nitrogen-fixing bacteria and legumes, where bacteria convert nitrogen into a form usable by the plant while the plant provides carbohydrates to the bacteria(*9*). Although viruses are typically viewed as pathogenic, there are cases where viral infection can confer significant benefits to the host. An example of viral symbiosis involves the Glyptapanteles wasp, which utilizes a virus-like particle called the bracovirus to ensure its larval survival within the caterpillar host(*10*). The bracovirus is integrated into the wasp's genome and is expressed during reproduction. When the wasp injects its eggs into a caterpillar, it also delivers bracovirus particles, which manipulate the caterpillar's immune system to not only protect the developing wasp larvae from the host's immune response but also manipulate the caterpillar's behavior to protect the pupating larvae. This relationship is essential for wasp reproduction and survival, demonstrating a clear mutualistic relationship between the virus and its wasp host (*11*). Bacteriophages in the Human Gut infect bacteria in the human gastrointestinal tract, potentially regulating bacterial populations and contributing to gut health and stability(*12*). This balance can assist in digestion and protect against pathogenic bacterial overgrowth. Though primarily known as a bacterial endosymbiont, Wolbachia (gram-negative bacteria) can transfer virus-like elements to mosquito hosts, conferring resistance to RNA viral infections such as dengue and Zika. This phenomenon enhances the survival and reproductive success of the host mosquitoes(*13*, *14*). Parts of endogenous retroviruses in mammals have been co-opted to serve essential functions in placenta formation and function, a crucial aspect of reproductive success in placental mammals. That is the story of Syncytin, a remnant of ancient viral infection that integrated DNA into the genomes of their host organisms, facilitates the fusion of trophoblast cells to form this multinucleated cell layer, which is essential for the proper functioning of the placenta(*15*–*17*).

Commensalism

In commensalism, one species benefits without significantly harming or helping the other. An example includes epiphytic plants, which grow on other plants for physical support without affecting the host(*18*).

Parasitism

Parasites are organisms or replicators that benefit at the expense of the other (the host), often causing harm. For example, viruses benefit at their hosts' expense, often causing disease or death. Viral parasitism is commonly observed in all types of viruses affecting plants(*19*), animals , and bacteria(*20*, *21*). The influenza virus provides a clear example of parasitism. It infects host cells and hijacks the cellular machinery to replicate its genetic material and produce new virus particles. This replication often results in cell death and significant tissue damage, leading to flu symptoms. The virus spreads to new hosts while the original host suffers from the infection, illustrating a typical parasitic relationship(*22*). Rabies virus infects the nervous system of mammals, leading to severe neurological symptoms and almost invariably to death if untreated. The virus spreads through the saliva of infected animals, often via bites(*23*). The Ebola virus causes severe hemorrhagic fever in humans and primates with high mortality rates. The virus spreads rapidly through contact with body fluids, exploiting host systems while causing extensive tissue damage(*24*). Historically, the smallpox virus caused devastating epidemics with high mortality before its eradication. It used human hosts for replication, spreading efficiently and often fatally(*25*). Certain strains of Human Papillomavirus can cause cervical and other cancers, making it a parasitic virus that leads to significant host cellular transformation and tumorigenesis(*26*).

Evolutionary Forces of Symbiosis

Each party in a symbiotic relationship exerts selective pressures on the other, leading to adaptations that may enhance the symbiosis. For example, in the mutualistic relationship between bees and flowering plants, plants have evolved colorful flowers and nectar to attract bees, while bees have developed structures to collect and carry pollen. This mutualistic interaction has driven the evolutionary diversification of both groups(*27*).

In some symbiotic relationships, particularly among microbial communities, there is potential for genetic exchange or horizontal gene transfer, which can accelerate evolutionary changes by introducing novel genetic material. For example, symbiotic bacteria within the guts of some insects can exchange genes that confer abilities to digest certain foods or detoxify harmful compounds, thereby broadening the ecological niches the insects can exploit (*14*).

Symbiotic relationships can lead to speciation through processes such as symbiogenesis or by creating reproductive isolation. For instance, endosymbiotic events, where one organism lives inside another, can lead to entirely new organismal lineages, as seen in the origin of eukaryotic cells from prokaryotic ancestors. Moreover, organisms that become highly specialized and dependent on their symbiotic partners may diverge from closely related species that do not participate in such partnerships, leading to speciation(*28*).

Symbiosis often allows organisms to exploit ecological niches that would otherwise be inaccessible. For example, the symbiosis between certain fungi and plant roots (mycorrhiza) allows plants to absorb nutrients more efficiently from the soil, which can significantly influence the structure of ecological communities and the competitive dynamics among plant species(*29*).

Symbiotic relationships can provide a buffer against environmental stresses, promoting stability and survival in fluctuating environments. Coral reefs are a prime example, where the symbiotic relationship between corals and their photosynthetic algae is crucial for nutrient supply. However, environmental stressors like warming waters can disrupt this balance, leading to coral bleaching and the breakdown of the symbiotic relationship (*30*).

Antagonistic Coevolution

When two replicators interact in a way that an increase in the fitness of one necessarily results in a decrease in the fitness of the other, antagonistic coevolution may ensue characterized by reciprocal adaptation by each replicator over time.

Predator-Prey Interactions

The most recognizable form of antagonistic coevolution may be the interaction of predator and prey. Here, a predator evolves to become more efficient at capturing its prey, and the prey in turn evolves to evade' the predator's new, more efficient state.

Evolutionary forces of antagonistic coevolution

The Red Queen Hypothesis is often invoked to describe the evolutionary pressures and consequences resulting from antagonistic coevolution. In this theory, the 'Red Queen' comes from Lewis Carroll's "Through the Looking-Glass", where the Red Queen tells Alice that it takes all the running you can do to keep in the same place. This describes the concept of an evolutionary arms race, where species must continually evolve just to maintain their current level of fitness. The theory states that when two species compete for limited resources, in order for each party to persist, there must be ongoing and reciprocal adaptation on each side. If one species fails to adapt, it loses.

Host-virus interactions

Genetic conflict arises when the virus's and the host's evolutionary interests diverge, leading to an arms race of evolutionary adaptations. This conflict is a driving force for both viral innovation and host immune system development. HIV, the virus that causes AIDS, is a prime example of a pathogen engaged in genetic conflict with restriction factors (*31*–*33*). HIV continuously evolves mechanisms to evade the human immune system, such as rapidly changing its surface proteins (via high mutation rates), which makes it difficult for the immune system to detect and eradicate the virus. In response, the human immune system has evolved several mechanisms to recognize and attempt to control the virus, such as the expression of restriction factors like APOBEC3G (*34*), which can mutate and inactivate the viral genome. The ongoing adaptations between HIV and the human host represent a dynamic and continual genetic conflict.

Herpes Simplex Virus can evade the immune system by establishing latent infections in neuronal cells(*35*). The host immune system continuously attempts to suppress virus reactivation, leading to periodic clinical outbreaks. Hepatitis C Virus exhibits rapid mutation of its surface proteins to evade immune detection while the host evolves various immune responses to detect and eliminate the virus, often leading to chronic liver diseases and an ongoing state of immune activation (*36*). Influenza A virus strains undergo frequent antigenic shifts and drifts, modifying their hemagglutinin and neuraminidase proteins to escape pre-existing host immunity. This forces the host to constantly update its immune defenses, particularly through vaccination(*37*).

The interactions between viruses and their hosts, viewed as a spectrum with symbiosis and parasitism at opposite ends and genetic conflict occurring continually, are of paramount importance. Each party strives to optimize its evolutionary fitness, shaping the evolutionary trajectories of viruses and hosts. Understanding these dynamics not only provides profound insights into viral evolution and host resistance mechanisms but also plays a crucial role in developing more effective therapeutic strategies to manage viral infections and their impacts on ecosystems and human health.

Antagonistic Coevolution within Genomes

While viruses move among individuals, another evolutionary battle rages within each genome between genomic parasites called transposable elements and their host genomes. Transposable elements (TEs), also known as transposons, are mobile genetic elements found in the genomes of organisms ranging from bacteria to humans. These elements can move or transpose from one genomic location to another, influencing genome structure, function, and evolution. TEs are contained in a host's genome, and in order to survive the inevitable accumulation of mutations, they must replicate, bringing the cost of this genome instability to the host. However, the evolutionary pressures on intragenomic parasites are complicated by the fact that a TE must replicate, but if it replicates too successfully, it may decrease host fitness to such an extent that it decreases its own net fitness because TEs ultimately depend on the survival of their hosts. In this way, the inheritance of TEs, along with the rest of an organism's genes, is a distinctive characteristic of TE-host caoevolution compared to that of viruses. In addition, viruses too cannot survive if they eliminate their host.

Further complicating the complex coevolution of TEs and host genomes, their insertion into new genomic locations can lead to various genetic changes, including gene disruptions, duplications, rearrangements, and the generation of new genetic regulatory elements and genes(*38–41*). In some cases, these mutations have proven beneficial to the host genome, and domestication or co-option of sequence from TEs (and viruses) has enabled innovation of numerous aspects of an organism’s biology, including gene regulation, development, and morphology.

Domestication

One of the most prominent examples of viral domestication is the presence of endogenous viral elements (EVEs) in the human genome. These are sequences derived from ancient viral infections that have been integrated into the host genome and passed down through generations. While many of these viral sequences have lost their ability to produce infectious viruses, they have left behind remnants that can still be functional. Some HERVs have been co-opted by the human genome to play roles in normal cellular processes. For example, some HERV sequences have been repurposed as promoters or enhancers that regulate the transcription of nearby human genes. Domesticated viral elements can also modulate the immune system. Certain HERV proteins are involved in immune responses, including the modulation of immune signaling pathways. This can impact how the immune system responds to infections and how it distinguishes self from non-self.

Transposable elements

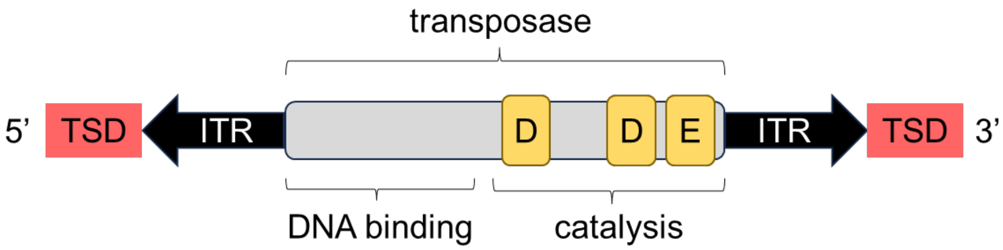
Barbara McClintock first discovered TEs in the 1940s through her groundbreaking work on *Zea mays* (maize) genetics(*42*, *43*) for which she was awarded the Nobel Prize in Physiology or Medicine in 1983. She discovered the Ac/Ds (Activator/Dissociation) transposon system, which comprises autonomous (Ac) and non-autonomous (Ds) DNA transposons, while studying the genetics of kernel pigmentation. Barbara McClintock’s contribution revolutionized how genetics was studied at the time despite the much later acknowledgment of her work by the scientific community. McClintock’s experiments, the development of high throughput sequencing technologies and bioinformatics has allowed the discovery and characterization of the vast diversity of TEs we know today.

TEs are a diverse group of genetic elements, but they can be broadly categorized into two main classes based on their mechanism of transposition: class II DNA transposons and class I retrotransposons.

DNA transposons

DNA transposons move directly as DNA segments through a 'cut-and-paste' mechanism mediated by a transposase enzyme. The key feature of DNA transposons is the presence of inverted terminal repeats (ITRs) flanking the transposon sequence. These ITRs contain recognition sites for the transposase enzyme, which catalyzes the excision and insertion of the transposon. The transposase recognizes the ITRs and mediates the excision of the transposon from its original genomic location. Following excision, the transposase facilitates the integration of the transposon into a new locus, often generating target site duplications (TSDs) flanking the insertion site.

DNA transposons exhibit remarkable diversity in their structure, organization, and transposition mechanisms. They can be classified into several families based on the presence of conserved protein domains within the transposase enzyme.



The **Tc1/mariner** Superfamily is one of the most widespread families of DNA transposons in eukaryotes (*44*, *45*) but is also present in prokaryotes (*46*). Members of the Tc1/mariner superfamily typically encode a single transposase enzyme characterized by a DDE motif, which is crucial for the DNA cutting and joining reactions of the transposition process (*47*). The transposase makes a staggered cut at the target site, resulting in the duplication of a short target sequence (TSD) upon insertion. Tc1/mariner elements range in size from 1.3 kb to 2.4 Kb. Tc1/mariner transposons help build synthetic *Sleeping Beauty* transposons that are used extensively as tools in genetic engineering and gene therapy due to their ability to transpose in a wide range of species (*48*). The figure was adapted from (*49*).

The **hAT (hobo/Ac/Tam3)** Superfamily includes the hobo elements in Drosophila, Activator (Ac) elements in maize, and Tam3 elements in Antirrhinum. Known for their broad host range, these transposons have played significant roles in studying genetic regulation and evolution in plants (*50*). They also utilize a transposase with a DDE/D motif; their typical length is 5 kb, and their transposition can cause a wide range of mutations, from simple insertions and deletions (*51*) to complex chromosomal rearrangements (*52*).

The **PiggyBac** superfamily was originally discovered in *Trichoplusia ni*, the cabbage looper moth (*53*). They are notable for their precise excision from the genome, which does not leave behind any extra DNA at the excision site. The PiggyBac transposon is used extensively in biotechnology (*54*) for gene transfer and expression studies because it can integrate efficiently into eukaryotic genomes and can be removed cleanly without altering the host DNA. It also uses a DDD motif in its transposase for catalyzing transposition.

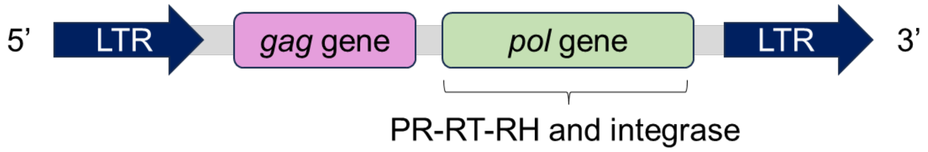
The **MuDR/Mutator** superfamily is found in all eukaryotes, but it has been studied predominantly in plants (*55*). These transposons are named after the Mutator elements in maize, which cause high mutation rates (*42*, *43*). MuDR/Mutator are characterized by their ability to acquire ORFs from the host while some of them have lost their transposase and only carry host genes (*56*, *57*). Like other DNA transposons, they rely on a transposase that recognizes TIRs, but they also contain a Zinc finger protein ORF.

The **Helitron** superfamily represents a unique class of DNA transposons that transpose via a rolling-circle replication mechanism (*58–60*), unlike the classic cut-and-paste mechanism. Instead of a transposase, Helitrons contain a protein named Rep/Helicase with two main functions, 1) HUH endonuclease that makes a single-stranded nick in the

DNA and 2) 5’ to 3’ helicase activity (*61*). They do not create target site duplications and often capture gene sequences from their host, contributing to gene shuffling (*62*) and evolution (*59*). Although less understood than other DNA transposons, Helitrons are abundant in many eukaryotic genomes and play significant roles in shaping genomic architecture (*63*).

Retrotransposons

The first elements recognized as retrotransposons were discovered in the yeast *Saccharomyces cerevisiae* in the late 1970s and early 1980s. These were the Ty (Transposon yeast) elements, specifically Ty1 and Ty2, which were found to transpose via an RNA intermediate in a manner analogous to retroviruses (*64*, *65*). Yeast geneticists observed unusual transcription of yeast genes that led to the identification of Ty elements, which were subsequently shown to use reverse transcription, akin to the process used by retroviruses, thus classifying them as retrotransposons (*65*, *66*). Retrotransposons (class I) mobilize through a 'copy-and-paste' mechanism involving reverse transcription of an RNA intermediate.



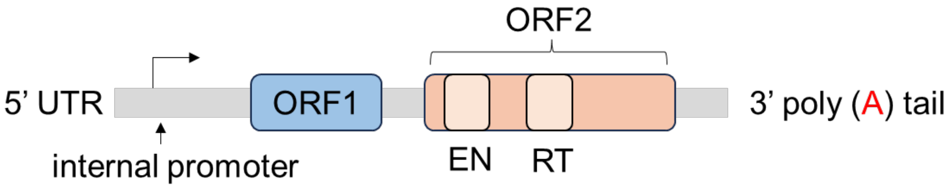
<u>LTR (Long Terminal Repeat) Retrotransposons.</u> LTR retrotransposons are structurally and functionally similar to retroviruses, from which they are thought to have originated (*67*, *68*). They are characterized by long terminal repeats (LTRs) at both ends of the DNA sequence, which play critical roles in transcription and integration. LTR

retrotransposons contain LTRs flanking internal genes that typically include `gag` (group-specific antigen; structural proteins) and `pol` (polymerase; includes reverse transcriptase, RNase H, integrase). They replicate through a process that involves transcription into RNA, translation of the necessary enzymes, reverse transcription of the RNA back into DNA, and then integration of this DNA into a new genomic location. Some LTR retrotransposons are Ty elements in yeast, Copia and Gypsy elements in plants and insects, and HERVs (Human Endogenous Retroviruses) in humans.

Ty elements are categorized into different families such as Ty1, Ty2, Ty3, Ty4, and Ty5, each characterized by distinct LTR sequences and slight variations in their integrase and other encoded proteins (*69*, *70*). Ty elements undergo transcription to produce RNA, which is then reverse-transcribed into cDNA by the element-encoded enzymes. The cDNA is integrated back into the yeast genome, often mediated by the preferences of the integrase enzyme. For instance, Ty1 and Ty2 prefer to integrate upstream of RNA polymerase III-transcribed genes like tRNA genes. Ty elements contribute to genetic diversity in yeast by generating mutations(*71*), altering gene expression, and creating new genetic material that can be subject to natural selection. They are also studied for their regulation and interactions with host factors, providing a model for understanding retrotransposition(*72*, *73*).

Copia and **Gypsy elements** are two of the major classes of LTR retrotransposons widely distributed in plants and insects(*74*–*76*). They are named based on their similarities to the respective retroviruses and are distinguished by the order of the encoded enzyme domains (reverse transcriptase, integrase). Frequently found in plant genomes, Copia elements contribute significantly to genome size and complexity by creating insertional mutations, altering gene regulatory networks, and contributing to genome rearrangements. Gypsy Elements are more common in insect genomes but are also prevalent in plants. Gypsy elements are particularly notable in the Drosophila genome, where they affect gene regulation and evolution. Both Copia and Gypsy elements are key players in shaping the architecture and evolution of their host genomes(*77–79*). They can influence gene expression, contribute to genomic instability, and act as sources of genetic innovation.

Human Endogenous Retroviruses (HERVs) are LTR retrotransposons in the human genome that are derived from ancient retroviral infections of germ cells. These elements have lost their infectious capacity but remain remnants that can be transcribed and influence human biology(*80*). HERVs are classified into several families based on sequence similarity to various parts of retroviral genomes, such as HERV-K, HERV-W, and HERV-H. They retain typical retroviral genes (gag, pol, and env), although many are truncated or mutated and have lost functional capability. While most HERVs are transcriptionally silent, due to host defense mechanisms like DNA methylation, some are transcribed and have been implicated in human physiology and pathology. For example, HERV-W expression is associated with multiple sclerosis, and HERV-K has been implicated in human cancer(*81*) and may have roles in early development(*82*). HERVs contribute regulatory elements to the human genome. For instance, some HERV LTRs act as promoters or enhancers for human genes, influencing gene expression patterns in development and disease(*83*, *84*).



Non-LTR Retrotransposons do not contain LTRs and are further classified into LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements). They integrate using a target-primed reverse transcription (TPRT) mechanism. The process begins with the transcription of the retrotransposon into RNA, which then associates with its ORF1p proteins and then uses the endonuclease and reverse transcriptase domains from ORF2p. This complex targets specific genomic sites (rich in As and Ts) and the endonuclease introduces a nick in the DNA, creating a free 3' hydroxyl group. This 3' end acts as a primer for reverse transcription, where the reverse transcriptase synthesizes complementary DNA (cDNA) from the retrotransposon RNA template. Following the degradation of the RNA template, a second strand of DNA is synthesized, resulting in the integration of the double-stranded retrotransposon DNA into the host genome.

DIRS (Dictyostelium Intermediate Repeat Sequence) retrotransposons, unlike typical LTR retrotransposons, contain split direct repeats and a tyrosine recombinase gene, which they use for integration. This distinguishes them from other retrotransposons that use integrase. They are found in the slime mold Dictyostelium and some fungi.

Penelope-like elements contain a reverse transcriptase closer to those found in telomerases and group II introns. They are named after the Penelope element of

Drosophila. Though not as widespread as other retrotransposons, they are particularly prevalent in certain invertebrate species.

LINEs (Long INterspersed Elements) encode the proteins necessary for their own transposition, typically including an RNA-binding protein and a protein with reverse transcriptase and endonuclease activity. LINEs use an internal promoter to transcribe RNA, which is then reverse transcribed and integrated at a new location, often generating target site duplications. LINE-1 and LINE-2 are found in humans, with LINE-1 being the only active LINE in the human genome capable of autonomous transposition(*85*). The promoter is also reversed transcribed as part of the entire retrotransposon.

SINEs (Short INterspersed Elements) do not encode proteins and rely on the machinery encoded by LINEs for their mobilization. SINEs are 100-700 basepairs long and use RNA polymerase III promoters (such as those for tRNA and 7SL RNA) for transcription. The LINE-encoded machinery facilitates their integration. Alu elements and MIRs (Mammalian-wide Interspersed Repeats) can be found in humans(*86*). SINEs are the most abundant TEs in human genomes with over 1 million copies present.

SVA (SINE-R/VNTR/Alu), like Alu elements, SVAs are non-autonomous and rely on L1 for mobility. SVAs are composite transposable elements that are composed of pieces of other TEs including, an Alu-like region, a GC-rich VNTR (variable number tandem repeat), and a SINE-R region related to HERVs. There are six subfamilies of SVAs (SVA_A to SVA_F), and approximatly half of the estimated 3,700 SVA copies in humans are unique to our species.

The detrimental effects of TEs in host genomes

While they have played significant roles in genomic evolution and innovation, TEs can also have detrimental effects on genome stability and organismal fitness. Here, we explore the various mechanisms by which transposable elements exert their negative impacts, the genomic defenses against these elements, and the broader implications for genome evolution and disease.

Transposable elements can cause significant genomic instability through several mechanisms. One major way is through insertional mutagenesis. When TEs insert themselves into new genomic locations, they can disrupt gene function if they insert within or near functional genes, leading to loss of function, altered gene expression, or the creation of aberrant proteins. Additionally, TE activity can lead to structural variations such as deletions, duplications, inversions, and translocations, which can disrupt chromosomal integrity and lead to gene dosage imbalances or the formation of fusion genes with oncogenic potential. The excision and insertion processes of TEs often involve the creation of double-strand breaks (DSBs). If these breaks are not properly repaired, they can result in genomic instability, cell cycle arrest, or apoptosis.

TEs can also affect the epigenetic landscape of the genome. The insertion of TEs can attract DNA methylation and repressive chromatin marks, leading to the silencing of nearby genes. This epigenetic disruption can have lasting effects on gene expression patterns and cellular function. Furthermore, TEs contain their own promoters and regulatory sequences, which can interfere with the normal regulation of adjacent genes.

This transcriptional interference can result in inappropriate gene activation or repression, contributing to developmental abnormalities and diseases(*39*, *41*, *87*, *88*).

The proliferation of TEs can lead to genome expansion, which has several potential disadvantages. An abundance of TEs can significantly increase genome size, leading to greater metabolic and energy costs associated with DNA replication, repair, and transcription. Moreover, a higher proportion of repetitive sequences can increase the mutation rate due to errors during DNA replication and repair. This increased mutation load can have deleterious effects on the fitness of the organism.

To mitigate the detrimental effects of TEs, organisms have evolved multiple defense mechanisms. RNA interference (RNAi) is a critical mechanism in silencing TEs at the post-transcriptional level. Small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs) guide the RNA-induced silencing complex (RISC) to degrade TE transcripts, preventing their translation and subsequent transposition. DNA methylation and histone modifications are key epigenetic mechanisms that silence TEs. The recruitment of repressive chromatin marks such as H3K9me3 and H3K27me3 can effectively shut down TE activity by condensing chromatin and preventing transcription. Additionally, specific proteins, such as APOBEC3 and TRIM5α in mammals, can directly inhibit TE mobilization by targeting TE components, such as reverse transcriptase or integrase, thereby preventing their replication and insertion. These mechanisms are presented in detail in a later section.

While TEs can be detrimental, their ability to induce genetic variation has also been a driving force in evolution. TEs have contributed to the development of new genes, regulatory networks, and genomic innovations. However, the balance between beneficial and detrimental effects is delicate, and the net impact on fitness can vary across species and environments. The reactivation of TEs has been implicated in various human diseases, particularly cancers and neurodegenerative disorders. For example, LINE-1 elements are active in many cancers, where their insertions can disrupt tumor suppressor genes or activate oncogenes(*89*). Additionally, TE activity has been associated with neurodevelopmental disorders such as autism and schizophrenia(*90*, *91*). There is evidence that TE activity increases with age, contributing to genomic instability in somatic cells. This increase in TE mobilization can exacerbate the decline in cellular function and increase the risk of age-related diseases.

LINE-1s in the human genome

While the diversity of TEs and their effects across all genomes is staggering, the LINE-1 class of transposable elements underlies almost all of the TE activity in humans. This thesis explores the significant impact of LINE-1 retrotransposons on host genomes. It examines how their mutagenic actions alter genome evolution and organismal fitness.

LINE-1s constitute approximately 17% of the human genome(*92*). Despite the majority being truncated or mutated versions that are incapable of retrotransposition, a number of LINE-1 elements remain active in the human genome. LINE-1 elements are approximately 6 kb in length and are characterized by a 5’ untranslated region (UTR) that acts as an internal promoter, two non-overlapping open reading frames (ORF1 and ORF2), and a 3' UTR that ends in a poly(A) tail. ORF1 encodes an RNA-binding protein with nucleic acid chaperone activity, while ORF2 encodes endonuclease and reverse transcriptase activities. These proteins are essential for the retrotransposition process of LINE-1(*93–96*).

LINE-1 retrotransposes through an RNA intermediate through a "copy-and-paste" mechanism. This process involves transcription of the LINE-1 RNA, its export to the cytoplasm, translation of its ORFs, and the formation of a ribonucleoprotein (RNP) complex. The RNP is then imported back into the nucleus, where ORF2p endonuclease activity nicks the genomic DNA and uses the exposed 3' OH group to prime reverse transcription of LINE-1 RNA directly onto the genome. This results in the integration of a new LINE-1 copy, most likely in a new locus.

LINE-1 content, variation, and activity in human genomes

Studying LINE-1 elements poses significant challenges due to their repetitive nature and the difficulty in accurately attributing phenotypic outcomes to specific LINE-1 insertions. Advances in sequencing technology, like long-read sequencing and computational tools, continue to improve our ability to study these elements in finer detail. Understanding the precise mechanisms of LINE-1 regulation, the conditions under which they escape silencing, and their interactions with the host genome remains a critical area of ongoing research. This is particularly important in the context of human health and disease, where LINE-1 activity may have both deleterious and beneficial effects.

The highly repetitive nature of the human genome sequence has been evident since the early days of nucleic acid biochemistry, when high copy number repeats were inferred from cot curves of human DNA (*97*). The first identification of LINE-1 activity in a human occurred in 1988 when the Kazazian lab identified LINE-1 insertions in the Factor VIII genes which caused hemophilia (*98*). Based on two unrelated patients with such an insertion, a highly active source element was identified and cloned from human DNA and shown to be active in a cell-based retrotransposition assay (*99*).

With the first complete human genome sequence, the full extent of transposable element sequence started to become clear and motivated early studies to identify all active LINE-1s from an individual genome. This landmark study found only six highly active LINE-1s based on the draft version of the human reference genome (*100*). Subsequent studies focused mostly on using genome assemblies and sequencing reads to identify non-reference LINE-1 insertions from diverse human genomes (*101*–*105*). In addition, a few studies cloned LINE-1s from diverse genomes and tested their activity with in vitro retrotransposition assays (*85*).

Due to the length of LINE-1s, short read-based sequencing (where the read length was shorter than the repeat length) has been inadequate to resolve both the location and sequence of LINE-1 insertions that are not present in the human genome reference assembly. However, the broadening use of long read DNA sequencing has enabled unprecedented resolution of LINE-1s and other long repeats in diverse human genomes via the Human Pangenome Reference Consortium (HPRC) (*106*), the Telomere-to-Telomere (T2T) project and others (*107*). These studies showed that each individual has more than 1000 non-reference transposable element insertions (*103*, *104*, *108*). While most of these were Alu elements, each individual has around 250 non-reference LINE-1 insertions including around 50 full-length LINE-1s (*101*).

Recent work from our group used one genome that has been deeply sequenced with long reads (CHM1) to identify all full-length and intact (containing a complete ORF1 and ORF2 sequence) from this single genome (*109*). In this genome, we found 154 intact LINE-1s, of which we were able to clone 142 and test their ability to retrotranspose using an *in vitro* assay in HEK293T cells. With this approach, we found 35 significantly active LINE-1s in the CHM1 haploid genome, suggesting a given human may have around seventy (70) measurably active LINE-1s in their genome (diploid) – many more than previous estimates.

Evolutionary history of human LINE-1s

Using the sequences of the approximately 500,000 LINE-1s in the human genome, previous research has shown a clear succession of LINE-1 subfamilies during the evolutionary history of primates. These subfamilies of primate specific LINE-1s are named PA16 (oldest) to PA1 (youngest, aka L1Hs) (*110*–*112*). In addition to these primate-specific LINE-1s, remnants of even older LINE-1s can be found in the human genome, including elements that likely replicated more than 100 million years ago in the common ancestor of mammals (the L1M family) (*113*, *114*).

Within the human specific LINE-1s, a finer scale of age has been defined with subfamilies of the L1Hs family (*111*, *115*). The youngest subfamily is called ‘Ta’ (transcribed, active) which is defined by the presence of specific nucleotides in their 5’ UTRs ($A_{5930}C_{5931}A_{5932}$-G_{6015}) and arose around 4 million years ago (*111*). The Ta subfamily gave rise to two younger subfamilies called Ta-0 and the even younger Ta-1. These two subfamilies were proposed to encompass all of the active LINE-1s in modern humans (*89*), though we and others have found much older LINE-1s that still retain activity in some individuals (*85*, *109*).

While the large number of LINE-1 sequences in humans and other organisms has enabled the reconstruction of the evolutionary history of LINE-1s, a major unanswered question in the field is what drove the clear succession of LINE-1 families and subfamilies. Some possibilities include competition among LINE-1s, competition with non-autonomous elements like Alu, and pressure to evade innate immune restriction factors of host genomes (*116*).

Immune system and restriction factors

The innate immune system plays a fundamental role in the defense against pathogens, but it also extends its functions to include the surveillance and regulation of endogenous elements such as TEs. Understanding how cellular mechanisms evolved to control potentially deleterious transposition events is essential to unravel the intersection between innate immunity and genomic integrity. Despite their resilience, TEs are subject to various host mechanisms that regulate their activity and prevent genome instability. These include epigenetic silencing mechanisms, such as DNA methylation and histone modifications, and host surveillance systems that recognize and inhibit transposon mobility (*117*). However, dysregulation of TE activity can have severe consequences on genome stability and organismal fitness. In fact, TEs have been implicated in the development of diseases such as cancer and developmental disorders, underscoring the importance of understanding TE regulation. Different organisms contain different sets of transposable elements based on their evolutionary history, how permissive their genome is to mobile genetic elements, the fitness cost of TE replication, and other unexplored factors. Of particular interest to this thesis are the mechanisms used by host genomes to restrict the replication of TEs and the mechanisms by which TEs continue to replicate despite these numerous mechanisms.

The restriction of LINE-1 activity in the human genome occurs at three main stages of LINE-1 retrotransposition: silencing of LINE-1 transcription, targeting of LINE-1 components, blocking new integration of LINE-1 to the genome (*118*). Here, I provide an overview of examples of each class of restriction mechanisms.

1) Transcription repression (targets promoters and in-gene CpG),

> **DNA Methylation.** DNA methylation is a primary epigenetic mechanism that silences LINE-1 elements, particularly in somatic cells (*119–121*). Methylation of cytosine bases within LINE-1 promoters prevents transcriptional initiation by blocking transcription factor binding and recruiting methyl-CpG-binding domain proteins that promote a repressive chromatin state (*122*, *123*).
>
> **Histone Modifications**. Various histone modifications also play roles in LINE-1 silencing (*124*, *125*). For example, trimethylation of histone H3 at lysine 9 (H3K9me3) is associated with heterochromatin formation and is commonly found at LINE-1 elements, contributing to their repression. Histone deacetylases (HDACs) are also involved by removing acetyl groups from histones, leading to a more condensed chromatin state and reduced LINE-1 expression.

ZAP (Zinc-finger Antiviral Protein). ZAP restricts LINE-1 by recognizing specific sequences within LINE-1 RNA and recruiting the RNA exosome complex, which is responsible for RNA degradation (*126*, *127*). ZAP's role in targeting LINE-1 RNA emphasizes the adaptability of host defense mechanisms originally evolved to combat viral infections, which are also co-opted to control endogenous elements like LINE-1. This cross-functional activity of ZAP not only reduces LINE-1 retrotransposition but also mitigates the potential for LINE-1 to activate immune responses upon aberrant expression.

SAMHD1. SAMHD1 restricts LINE-1 activity through a mechanism that controls the cellular dNTP pool, which is essential for DNA synthesis, including that of LINE-1 via reverse transcription (*128*, *129*). By hydrolyzing dNTPs and maintaining their levels below the threshold necessary for efficient reverse transcription, SAMHD1 effectively starves LINE-1 elements of the building blocks required for their replication. This restriction mechanism is particularly effective in non-proliferating cells, where low dNTP levels are maintained to prevent unintended DNA synthesis and genomic instability.

2) Post-transcriptional modifications in the cytoplasm (targets LINE1 RNA, ORFs, or RNP formation and nuclear import)

PIWIL Proteins and piRNAs. PIWIL proteins, interacting with piRNAs, are pivotal in silencing LINE-1 elements, especially in the germline where LINE-1 activity could have heritable consequences (*125*, *130*). piRNAs are small RNAs that are complementary to transposon sequences, including LINE-1, and they guide PIWIL proteins to their targets. Upon binding, this complex can induce the cleavage of LINE-1 RNA, leading to its degradation and preventing translation. Additionally, the piRNA pathway contributes to the establishment of repressive chromatin marks at LINE-1 genomic loci (*131*). This transcriptional silencing is crucial for maintaining genomic integrity in reproductive cells, highlighting the evolutionary importance of this restriction pathway in controlling transposon activity.

TREX1. TREX1 protects genomic integrity by degrading cytoplasmic single-stranded DNA (ssDNA), which includes reverse-transcribed LINE-1 DNA intermediates (*132*). This activity prevents the potential integration of these ssDNA fragments into the genome, thwarting LINE-1 retrotransposition. The degradation of ssDNA by TREX1 also serves to prevent the activation of innate immune sensors that could detect this DNA as pathogenic, thereby avoiding inappropriate autoimmune responses (*133–135*). Dysfunction in TREX1 is associated with autoimmune diseases, such as Aicardi-Goutières syndrome, underscoring its importance in both genomic and immune system maintenance.

TEX19. TEX19 exerts restriction through at least two mechanisms. One region of the protein, the VPTEL domain, directly binds piRNAs (small RNAs involved in TE suppression) as well as PIWI proteins (*136*). The other domain of TEX19 directly binds the ORF1 protein of LINE-1 elements, acting as an adapter to bring the ubiquitin ligase UBR2 to ORF1p (*137*). This recruitment results in ubiquitination and subsequent degradation of LINE-1 ORF1p. These functions or other roles in development underlie the requirement of TEX19 for normal spermatogenesis and placental development (*138*–*144*).

3) blocking the integration of LINE-1 via RT (targets RNA intermediate, ORF2p EN, or RT domains).

APOBEC3. The APOBEC3 family, particularly APOBEC3A and APOBEC3B, is critical in inhibiting LINE-1 retrotransposition(*145*–*149*). These proteins execute their restriction functions through a direct mechanism involving the deamination of cytosine to uracil within LINE-1 RNA. This modification introduces mutations into the LINE-1 sequence during reverse transcription, often leading to detrimental mutations that can inhibit the functionality of the new LINE-1 insertions. Moreover, APOBEC3 proteins can bind directly to the LINE-1 RNP complex, potentially interfering with the reverse transcription process or the integration of LINE-1 DNA into the host genome. This dual action underscores their effectiveness as a defense mechanism against uncontrolled LINE-1 activity, which is particularly important given the potential for LINE-1 elements to contribute to genomic instability and disease.

Understanding and manipulating these restriction factors offer potential therapeutic avenues for conditions associated with dysregulated LINE-1 activity, including cancer and neurological disorders. Further research into these mechanisms will not only enhance our understanding of cellular defense systems but also inform the development of novel therapeutic strategies that exploit these natural defense mechanisms.

Chapter 2: Exploring Disruptions of Human Immune Regulation of LINE-1s

Summary

Because LINE-1s must jump in the germline and early embryo to survive as active TEs, our genomes have evolved restriction factor (RF) genes that control LINE-1 activity and maintain genome stability. Despite numerous defense mechanisms, LINE-1s continue replicating, as they have for over 100 million years(*150*, *151*). What is the reason for this? We know that the acquisition of mutations in the LINE-1 sequence can disrupt and circumvent host mechanisms of defense. LINE-1s in the ancestors of primates lost a bit of their 5' UTR, a 129-bp fragment in the internal promoter of the LINE-1, and that deletion led to the escape of a DNA-specific restriction factor called ZNF93(*152*, *153*). In this case, losing the DNA binding site for repression resulted in the evasion of host restriction and the expansion of LINE-1 families from PA4 to PA3. Other RFs in the human genome interact with LINE-1 RNA and proteins to inhibit retrotransposition, and simple changes can be completely disruptive to host regulation. For example, testis expressed 19 (TEX19), with the assistance of the ubiquitin ligase UBR2, binds LINE-1 ORF1p and tags it for proteasomal degradation, leading to a decrease of LINE-1 retrotransposition(*137*). Such protein-protein interaction interfaces are primed battlegrounds for host/pathogen coevolution(*31*). We hypothesized that LINE-1s and RFs are engaged in an ongoing genetic conflict, and the recent activity of some LINE-1s depends on their ability to evade directly binding post-transcriptional RFs. In addition,

we know that host proteins that engage against rapidly changing pathogens/parasites often accumulate signatures of positive selection and rapid evolution as a result of the fitness cost of infection/disease(*154*). Therefore, we think that 1) RFs of LINE-1 that are expressed during global de-repression of LINE-1s in early development and in the germline, and 2) RFs that present signatures of positive selection and rapid diversification are candidate RFs that *could* have driven LINE-1 evolution, promoting evasion mechanisms. To test this model, we measured the retrotransposition activity of thirty-four (34) diverse human LINE-1s from an individual human genome (*155*, *156*) in the presence/absence of several RFs. We reasoned that some restriction factors should restrict all of the tested LINE-1s because no data have shown otherwise. However, restriction factors showed variable patterns of restriction (restricting some but not all LINE-1s) and recently active human LINE-1s, including specific LINE-1s that may be able to evade parts of the innate immune defense system.

Introduction

The Long INterspersed Elements 1 (LINE-1s) continue to expand in the human genome(*157*) despite numerous immune system mechanisms targeting their retrotransposition. Is host regulation not good enough? Have new families of LINE-1s “figured out” how to overcome genome defenses? Other host genomes(*158*, *159*) exert tight epigenetic control of LINE-1s; what about some immune systems that are *more* permissive of LINE-1 retrotransposition than other hosts? Our research aims to question LINE-1’s recent retrotransposition activity and evolution in humans in relation to direct immune antagonists of LINE-1.

LINE-1s are transposable elements that replicate via a copy-paste mechanism. They use an RNA intermediate and reverse transcription to insert themselves into the host genome, which is why they are known as retrotransposons. LINE-1s themselves make up to 17% of the human genome, yet their retrotransposition also mobilizes non-autonomous retrotransposons like Alu(*86*), SVA(*160*), HAL1s, and host (retro)genes in *trans*(*161*) increasing the percentage of LINE-1-related genetic content to almost 30% of the genomic mass(*85*, *92*). Although most of the ~500,000 LINE-1 insertions are 5' truncations incapable of further retrotransposition(*162*), hundreds(*100*, *155*, *163*) of retrotransposition-competent LINE-1s are actively generating *de novo* insertions at an estimated rate of 1 per 20-200 births(*164*). While the human genome contains identifiable remnants of LINE-1s that replicated more than 100 million years ago in at least the last common ancestor of mammals, almost all of the currently active LINE-1s in humans belong to the human-specific subfamily called L1Hs (or L1PA1)(*165*, *166*).

Retrotransposition-competent LINE-1s are full-length, 6-7 kb long sequences that contain a 5' untranslated region (UTR) with an internal promoter, two open reading frames (ORF1 & ORF2), and a 3' UTR with a poly-A signal(*167*). ORF1p assembles as a homotrimer and binds to LINE-1 mRNA(*168*). ORF2p recognizes the 3' UTR of LINE-1 mRNA and binds to ORF1p-mRNA to form a ribonucleoprotein (RNP) complex(*94*, *169*).. LINE-1 RNP goes back into the nucleus, and ORF2p uses endonuclease and reverse transcriptase domains to insert LINE-1 into a new locus via target-primed reverse transcription (TPRT)(*96*, *170*, *171*). LINE-1 retrotransposition can be dangerously mutagenic(*172*), in many cases causing genomic instability(*173*, *174*) and debilitating disease(*41*, *175*).. LINE-1s can cause metabolic(*176*), neurological(*90*, *177*) and immune disorders (*178–185*). LINE-1 activation is linked to the aging (*186*) and progression/metastasis of human cancers (*187–191*). LINE-1 retrotransposition has become a huge genetic burden to the human genome due to the fitness cost of current and potential *de novo* mutagenic insertions (*192*, *193*).

The human genome has several genetic firewalls to protect itself from LINE-1s (*118*, *194*). Innate immune mechanisms target different vital parts of the LINE-1 retrotransposition cycle (*195*). For example, the HUSH complex along with TRIM28 and KRAB Zinc Finger Proteins (*120*, *196*) and associated RFs meet at the LINE-1 5’ UTR to modify and silence transcription via CpG methylation of the promoter and repressive H3K9 histone modifications (*119*, *131*, *197*, *198*). However, during early embryogenesis, access to essential genes for development requires global genomic demethylation, which also disables epigenetic repression of LINE-1 promoters (*199*). LINE-1 is then transcribed, and the genome relies on post-transcriptional mechanisms of defense to halt LINE-1 activity and avoid new potentially mutagenic insertions (*200*). RFs like Three-prime Repair EXonuclease 1 (TREX1) target LINE-1 via exonuclease-independent mechanism (thought to target the ORFs) to decrease LINE-1 retrotransposition *in vitro* (*134*)**.** Uridyltransferases TUT4/7 in concordance with helicase MOV10 modify LINE-1 via uridylation, the addition of non-template uridines to the 3’ end of the mRNA, permanently disrupting reverse transcription and integration (*201*). Potent mRNA editing enzymes like APOBEC3A (A3A), coupled with nuclear RNases, meet at the LINE-1 insertion site to modify LINE-1 mRNA, and stop insertion (*147–149*, *202*). A3B is expressed mostly in bone marrow cells and cancers, but also in germline cells. Testis Expressed 19 (TEX19) is specifically expressed in early embryonic development because it is activated by the same methylases that carry out global demethylation (*138*, *203*). With the help of ubiquitin ligase2 (UBR2) TEX19 tags ORF1p for proteasomal degradation resulting in decreased LINE-1 retrotransposition (*204*).

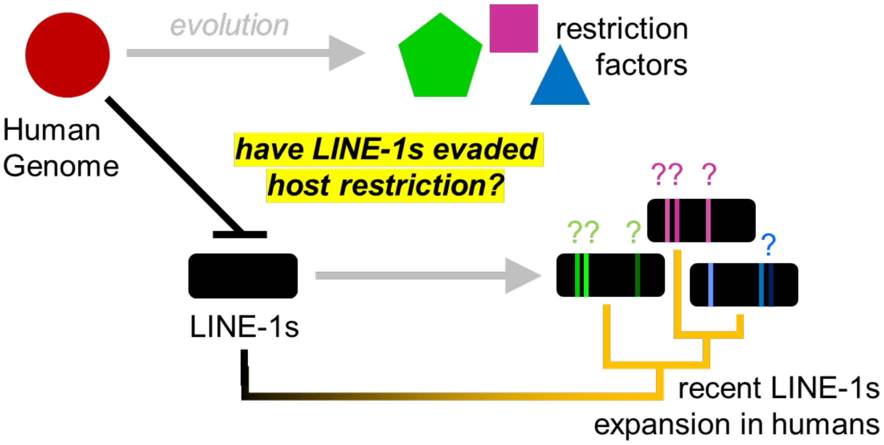


Figure 1. LINE-1s in the Human Genome (HG) are still active and making new copies despite multiple fronts of immune regulation; have retrocompetent LINE-1s in the HG evaded host restriction? We investigated this question by screening natural variation of LINE-1s in an individual HG against various restriction factors of LINE-1s expressed in the germline.

Are there retrocompetent LINE-1s in the human genome that have acquired mutations, evasive of RFs, which lead to increased retrotransposition in the germline? If so, which RFs and what are the LINE-1 sequence changes that have led to the continued expansion and persistence of LINE-1s in the human genome? We study this by screening LINE-1 natural variation in individual genomes against RFs that are most likely (known to restrict LINE-1, expressed in the germline, and dynamic evolution) to be in genetic conflict with LINE-1 (Fig 1). We used a recently described library of thirty-four (34) retrocompetent LINE-1s (relative to LINE1$_{RP}$) from the CHM1 individual haploid genome. We measure LINE-1 activity using a dual luciferase system while increasing amounts of restriction factors A3B, A3A, and TEX19 in a dose-response manner. We described the variation in restriction and found eleven (11) LINE-1s to be fully evasive to TEX19 relative to LINE-1_{RP}. In addition, we tested variants of human TEX19 found in clinical data repositories and found variations in restriction. We proposed mutations as potential determinants of evasion. We found that evasive LINE-1s *in vitro* are replicating *in vivo* (shown by 3' transduction identification in germline and cancer genomic datasets) and generating new copies in the human genome, consistent with the hypothesis that *evasive* mutations to RFs not only increase retrotransposition but also facilitate the expansion of LINE-1 in the human genome.

Results

Screening of retrocompetent LINE-1 library against restriction factors APOBEC3 and TEX19

We investigated specific RFs that may differ in effectively inhibiting young LINE-1s. We hypothesized that for young LINE-1s to evade host regulation, they need to overcome RFs expressed in the germline. We posit a scenario where polymorphisms present in young LINE-1s could disrupt the interaction with RFs, allowing for LINE-1 evasion. We chose LINE-1 RFs that 1) are expressed in the germline, 2) are known to inhibit LINE-1 retrotransposition *in vitro*, and 3) show signatures of rapid evolution. The most promising candidates are those that directly bind to LINE-1 protein machinery; such protein-protein interaction interfaces are primed battlegrounds for host/pathogen coevolution. We co-transfected a subset (7/34) of a library of retrocompetent LINE-1s from the individual human genome CHM1. Then we measured LINE-1 activity with and without RFs, using a dual luciferase reporter assay quantitative of LINE-1 retrotransposition. This assay uses the expression of Firefly luciferase as a proxy for LINE-1 retrotransposition (Fig 2, panel A). For RFs A3A, A3B, TREX1, and TEX19, we choose to transfect the amount of plasmid that restricts 50% (the IC50) for LINE-1$_{RP}$ activity to statistically compare to levels of activity without restriction (HEK-293Ts have low expression of these RFs). In this study, LINE-1s are named after their chromosome location, followed by the initial genomic coordinate. Most LINE-1s evaluated were targeted by A3B (50% decreased activity). However, several LINE-1s (L1_04_13638 and L1_14_58767) showed no statistical difference in activity with and without TEX19 (Fig, 2B), suggesting that those LINE-1s could be evading or differentially restricted by TEX19 *in vitro*.

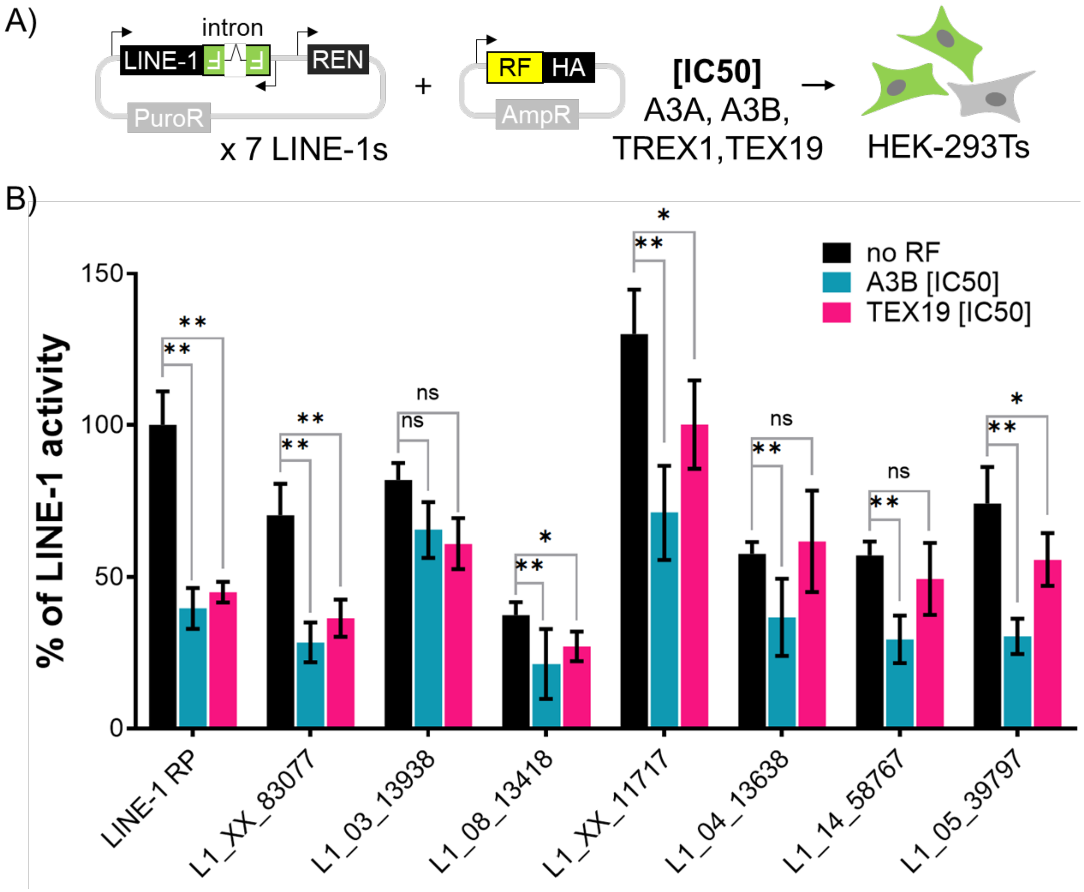


Figure 2. Several retrocompetent LINE-1s from CHM1 show potential evasion of restriction to TEX19. A) Cartoon diagram of the experimental setup. Several RFs were tested using their determined (not shown) IC50. B) One-point assay results showing the % of retrotransposition activity for seven (7) LINE-1s against A3B and TEX19. Columns were plotted using averages (n=8), and error bars show standard deviation values. Statistical significance analyses were also plotted, ** $p < 0.001$, * $p < 0.05$, ns $p > 0.05$.

Variable sensitivity of retrocompetent LINE-1s to the restriction factor TEX19

Next, we further explored the restriction of the library of thirty-four (34) active LINE-1s by transfecting each with increasing amounts (0-50 ng) of A3A, A3B, and TEX19 (Fig 3A). This library is a subset of LINE-1s from an individual human genome with >5% of LINE-1 activity relative to LINE-1_{RP}. We found that in the presence of A3A, all 34 LINE-1s were restricted as expected (Fig 3B for three of them, other data not shown). However, we found that 11/34 LINE-1s retained some activity even in the presence of high levels of TEX19. Further, some LINE-1s retained near-complete activity in the presence of TEX19 (Fig. 3.C). The figure shows dose-response data for three (3) LINE-1s (L1_08_72887, L1_15_70747, and L1_03_55767) and LINE-1_{RP} as a control. LINE-1_{RP} (black) and L1_15_70747 (red) are considered *TEX19-restricted* because they have similar response curves to increasing doses of TEX19. L1_08_72887 (blue) and L1_03_55767 (orange) are considered *TEX19-evasive* because retrotransposition levels show relative indifference of activity to high amounts of TEX19. We found 11/34 LINE-1s that we consider TEX19-evasive (binary evasion, almost to no hcange in % of activity); all of which are potently restricted by A3A. Residual activity of LINE-1 at 50 ng (>96% of TEX19 restriction of LINE-1_{RP}) of TEX19 show potency of TEX19-evasives in

Figure 3 C. We have also included dose-response curves for twenty-four (24) out of thirty-four (34) in Figure 4.

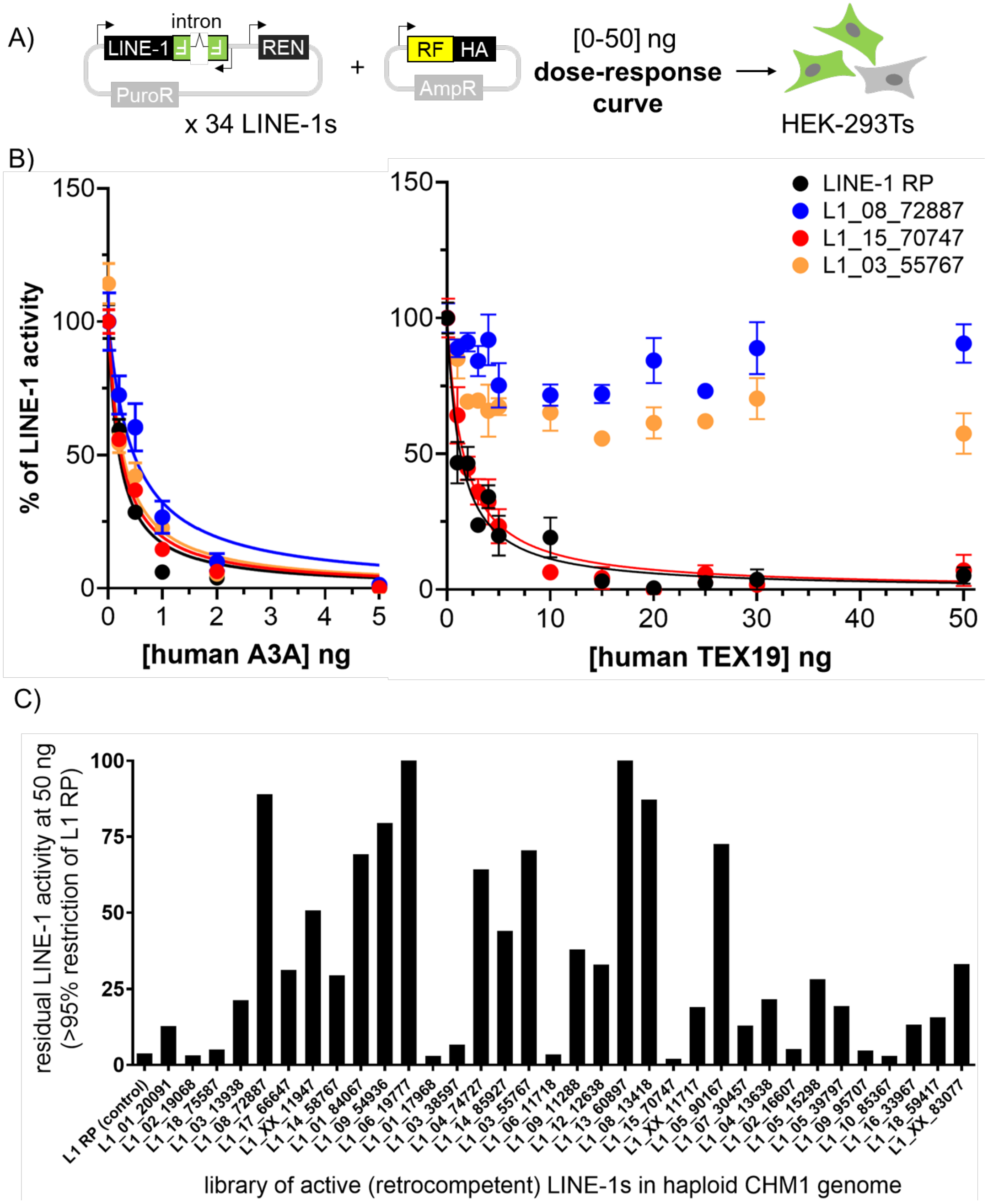
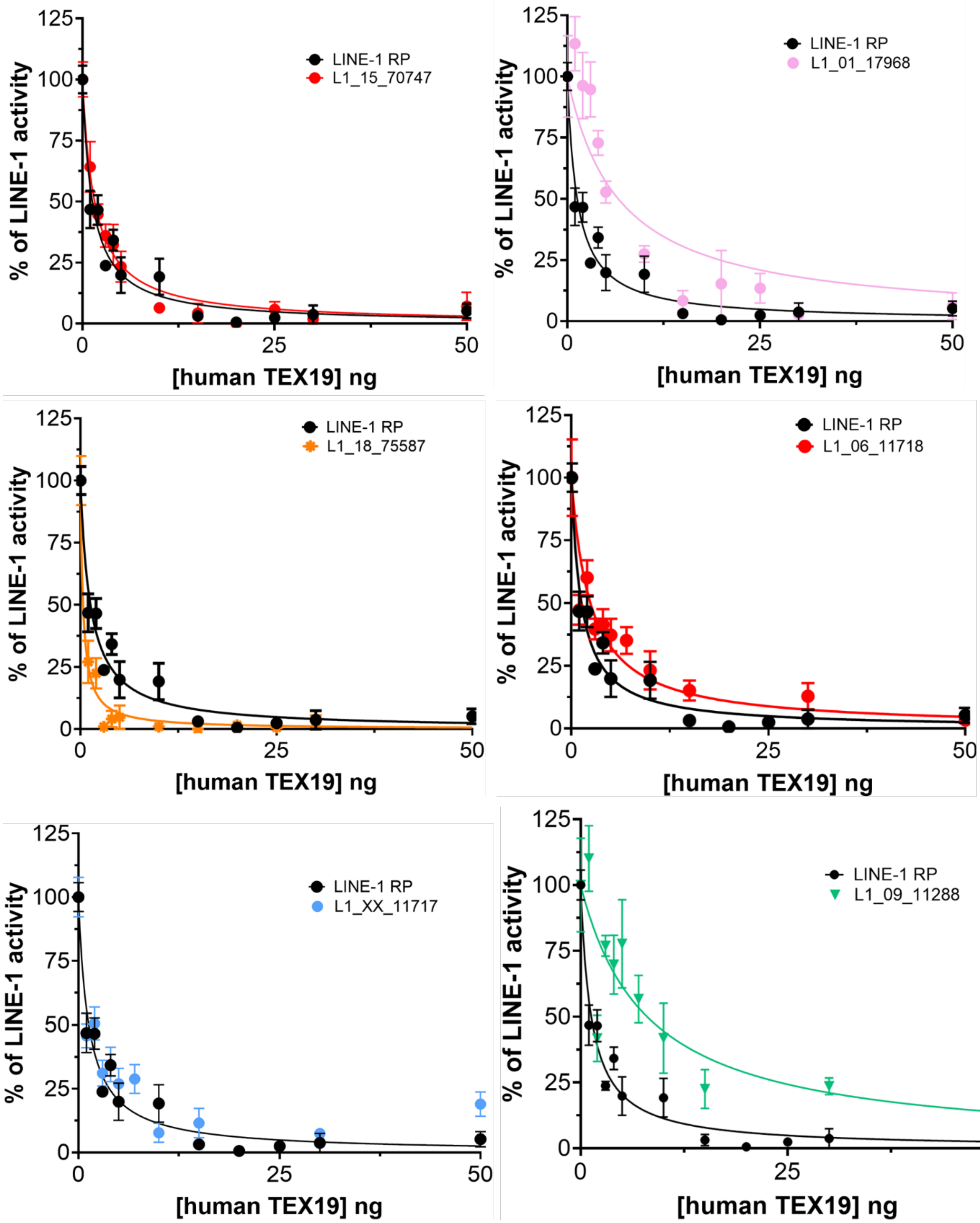
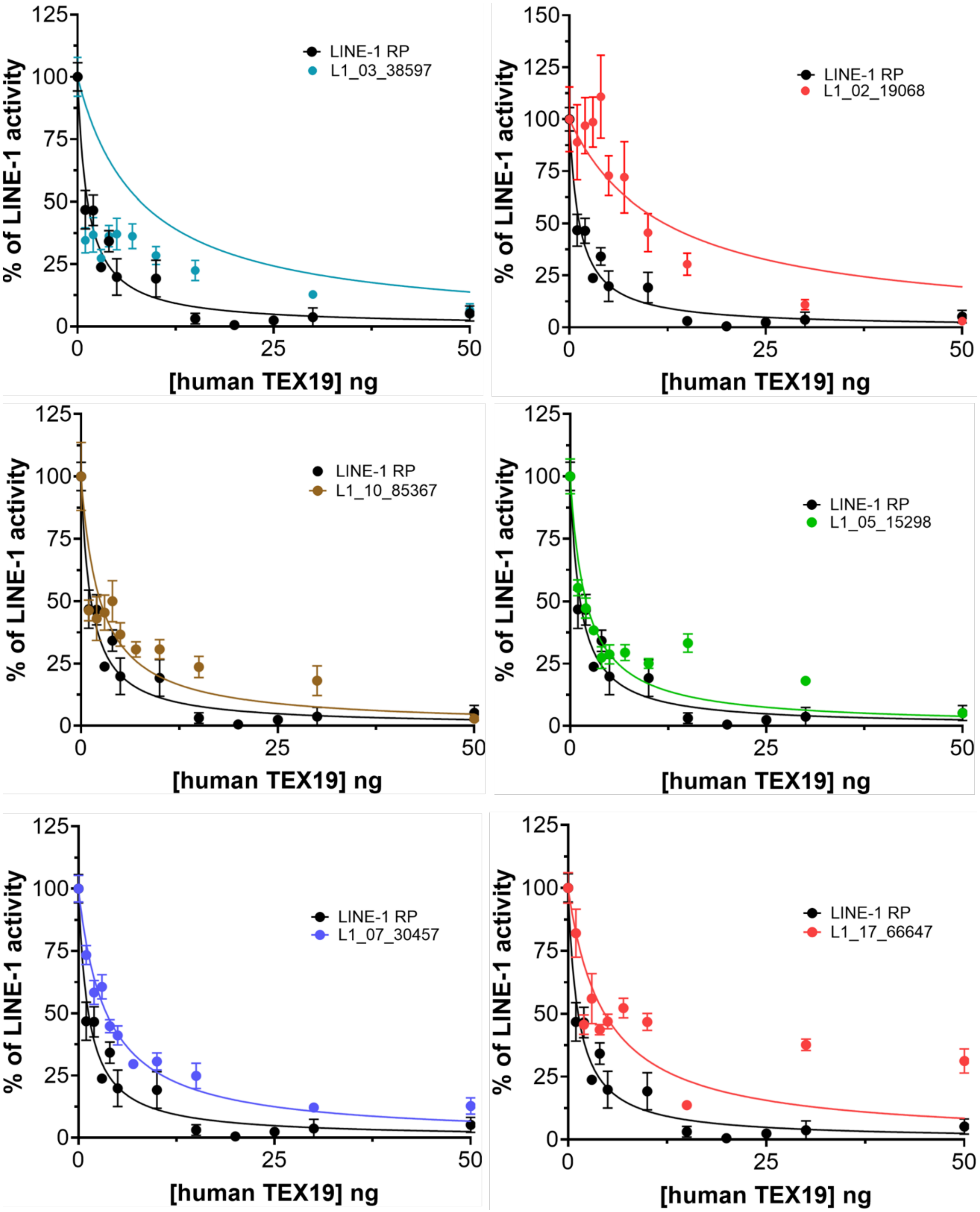
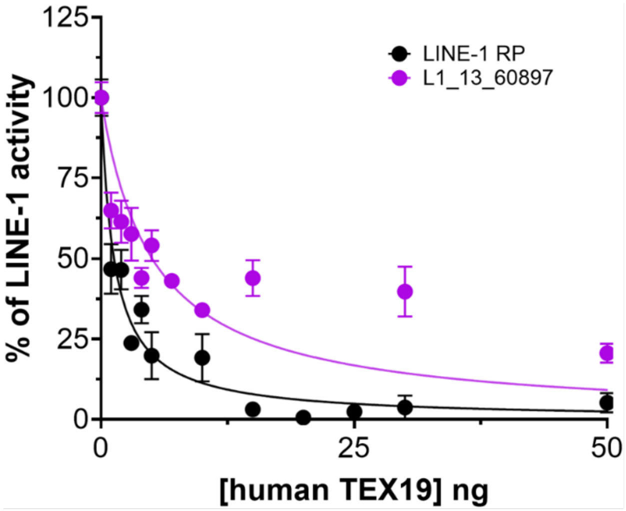


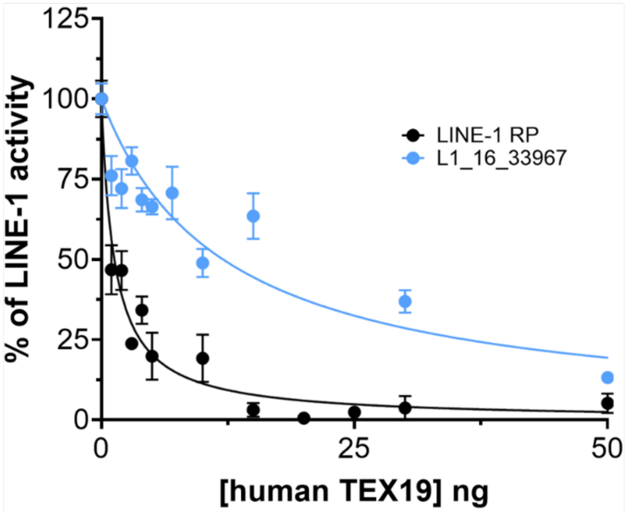
Figure 3. Some retrocompetent LINE-1s from a single genome (CHM1) evade restriction factor TEX19. A) cartoon diagram of the experimental setup, this assay differs from the one-point assay shown in Figure 2 in the transfection of multiple amounts of the RFs. Instead of transfecting one quantity we expanded the assay to a dose-response curve.

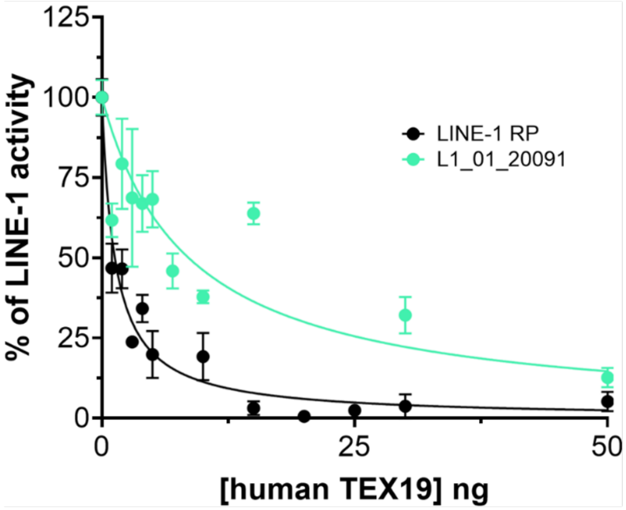
B) Dose-response curves showing examples of TEX19-evasive L1_08_72887 (blue) and L1_03_55767 (orange), and TEX19-restricted LINE-1RP (black) and L1_15_70747 (red). C) Summary of LINE % of activity at 50 ng of TEX19.

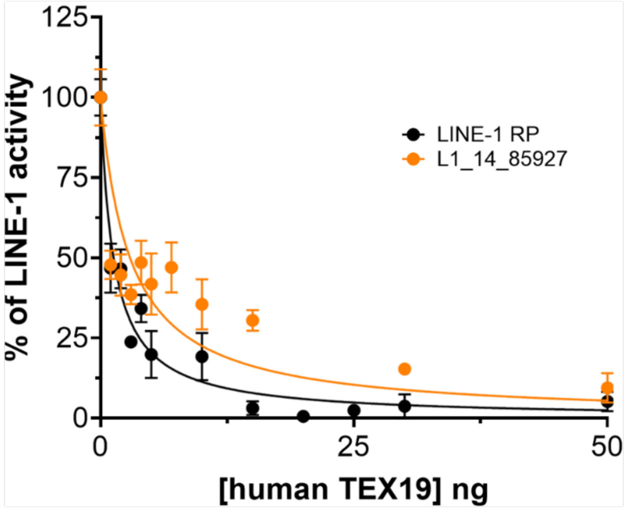


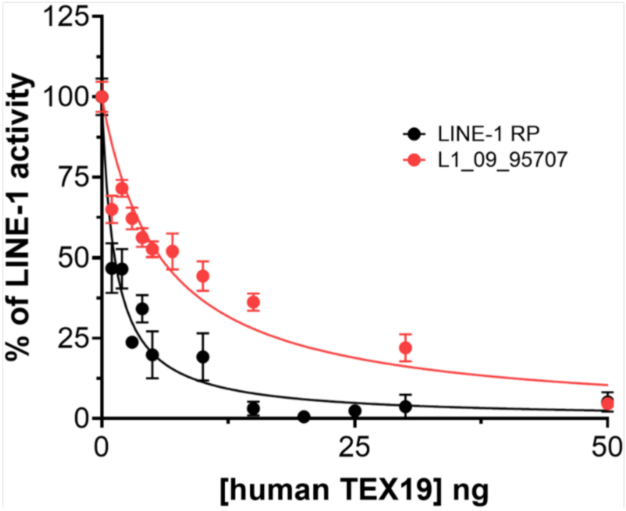

% of LINE-1 activity
[human TEX19] ng
125
100
75
50
25
0
0
25
50
LINE-1 RP
L1_03_38597
150
125
100
75
50
25
0
LINE-1 RP
L1_02_19068
LINE-1 RP
L1_10_85367
LINE-1 RP
L1_05_15298
LINE-1 RP
L1_07_30457
LINE-1 RP
L1_17_66647

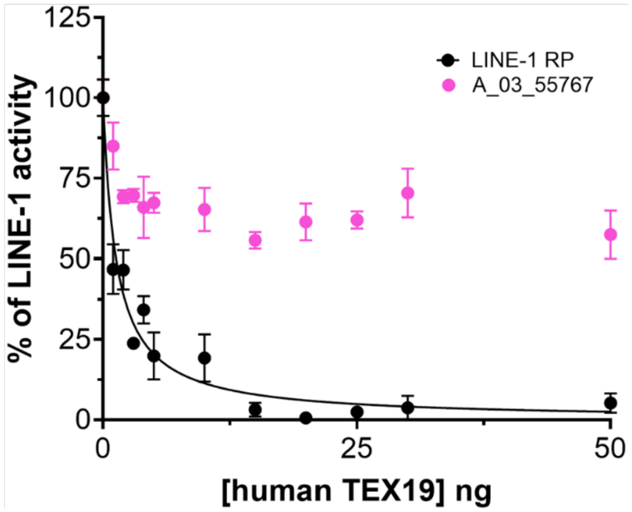

LINE-1 RP
L1_13_60897
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng


LINE-1 RP
L1_16_33967
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng


LINE-1 RP
L1_01_20091
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng


LINE-1 RP
L1_14_85927
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng


LINE-1 RP
L1_09_95707
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng


LINE-1 RP
A_03_55767
% of LINE-1 activity
125
100
75
50
25
0
0
25
50
[human TEX19] ng

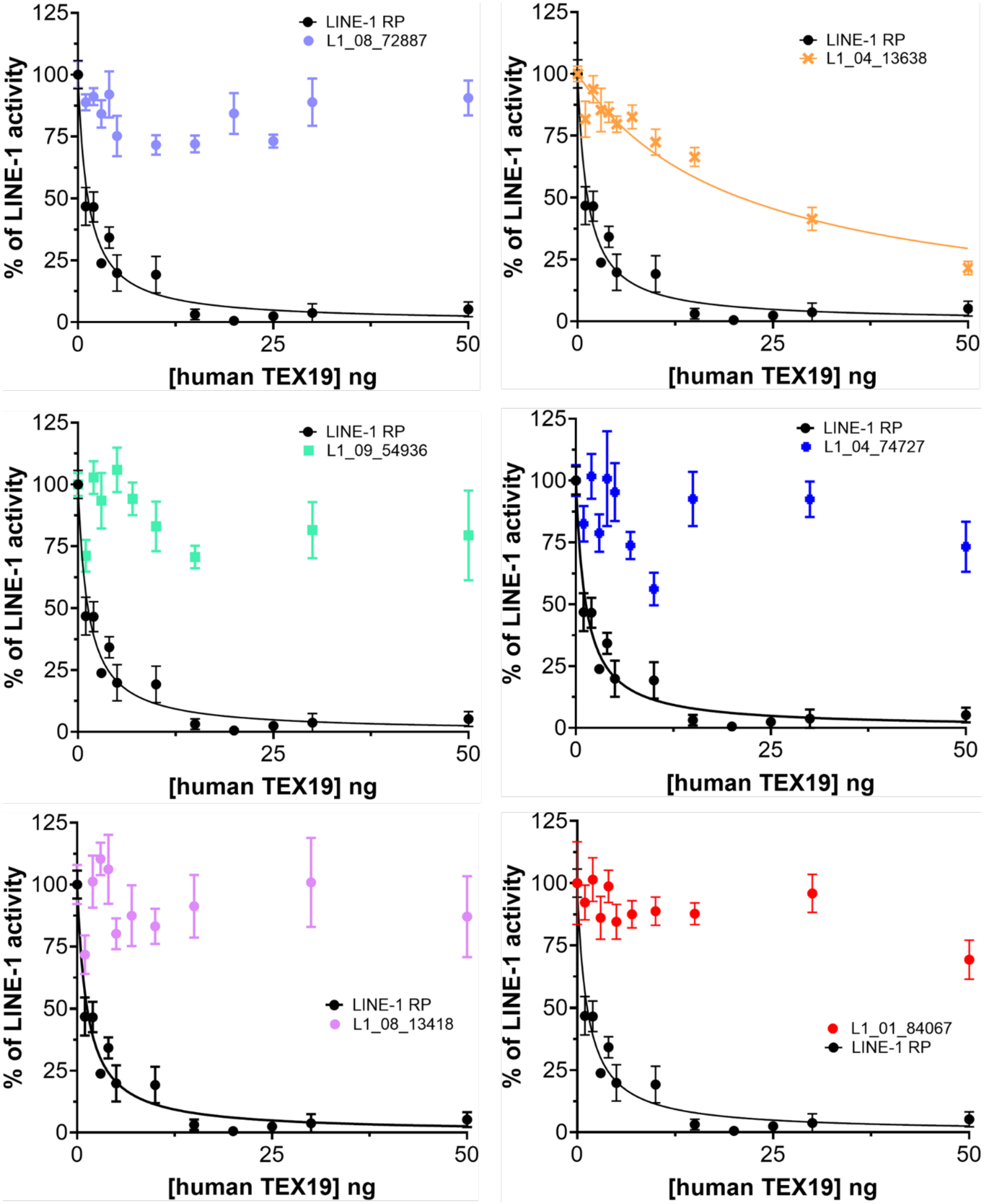

125
100
75
50
25
0
% of LINE-1 activity
[human TEX19] ng
0
25
50
LINE-1 RP
L1_08_72887
LINE-1 RP
L1_04_13638
LINE-1 RP
L1_09_54936
LINE-1 RP
L1_04_74727
LINE-1 RP
L1_08_13418
L1_01_84067
LINE-1 RP

Figure 4. Dose-response curves of retrocompetent LINE-1s from CHM1 library against TEX19. Each LINE-1 tested is compared with LINE-1_{RP}. Here, I show twenty-four (24) out of thirty-four (34) LINE-1s tested.

Sequence analyses of TEX19-evasive LINE-1s

We next made use of the sequence variation present in the CHM1 LINE-1s and tried to identify mutations as candidates of sequence determinants of evasion in those eleven (11) TEX19-evasive LINE-1s. We specifically searched for nonsynonymous mutations (amino acid change) because they have the potential to disrupt protein-protein interactions required for restriction. Translation-based sequence alignment shows nonsynonymous mutations located throughout ORF1 and ORF2, with the majority in ORF2 (Fig 6A). Initially, we did not find obvious differences shared among TEX19-evasive LINE-1s, suggesting that there might not be a direct correlation to a specific amino acid change but rather to an area/domain(s). Then, we asked if there were closely related LINE-1s that were differentially restricted by TEX19. This comparison decreases the number of changes being considered as potential determinants of TEX19 evasion. We found and described two interesting cases shown in Figure 5 and 6: 1) L1_10_85367 (grey) and L1_06_19777 (blue) and 2) L1_06_11718 (gray) and L1_05_90167 (red). There are only four (4) nonsynonymous changes in ORF1p (1) and ORF2p (3) sequence between L1_10_8536 (grey) and L1_06_19777 (blue) yet their activity levels to increasing amounts of TEX19 are completely different, L1_10_8536 (grey) is TEX19-restricted and L1_06_19777 (blue) is TEX19-evasive. In the case of L1_06_11718 (gray) and L1_05_90167 (red), we observe the same drastic differences at increasing amounts of TEX19, but they differ at nineteen (19) nonsynonymous changes (see amino acid differences in the cartoon of panels A, Fig. 5 for ORF1p and Fig. 6 for ORF2p). We annotated and visualized nonsynonymous changes for L1_06_11718 (gray) and L1_05_90167 (red), and we observed that the mutations are located in essential domains of ORF1p catalytic activity, like in the coil-coiled (CC) domain and the RNA recognition motif (RRM) highlighted in yellow in the monomer structure of ORF1p (Fig 5D). In addition, we show all the nonsynonymous mutations in all eleven (11) LINE-1s but compared to the consensus of LINE-1 Hs (Fig 5C, and Fig 6D). Some mutated positions unique to TEX19-evasives (shown in white spheres) are on the surface of ORF1p and exposed, yet the trimer interface is devoid of mutations, consistent with a role of some of these positions in protein-protein interactions (perhaps with TEX19) but not ORF1p trimer formation. Disrupting ORF1p trimerization would/could lead to killing LINE-1 activity completely.

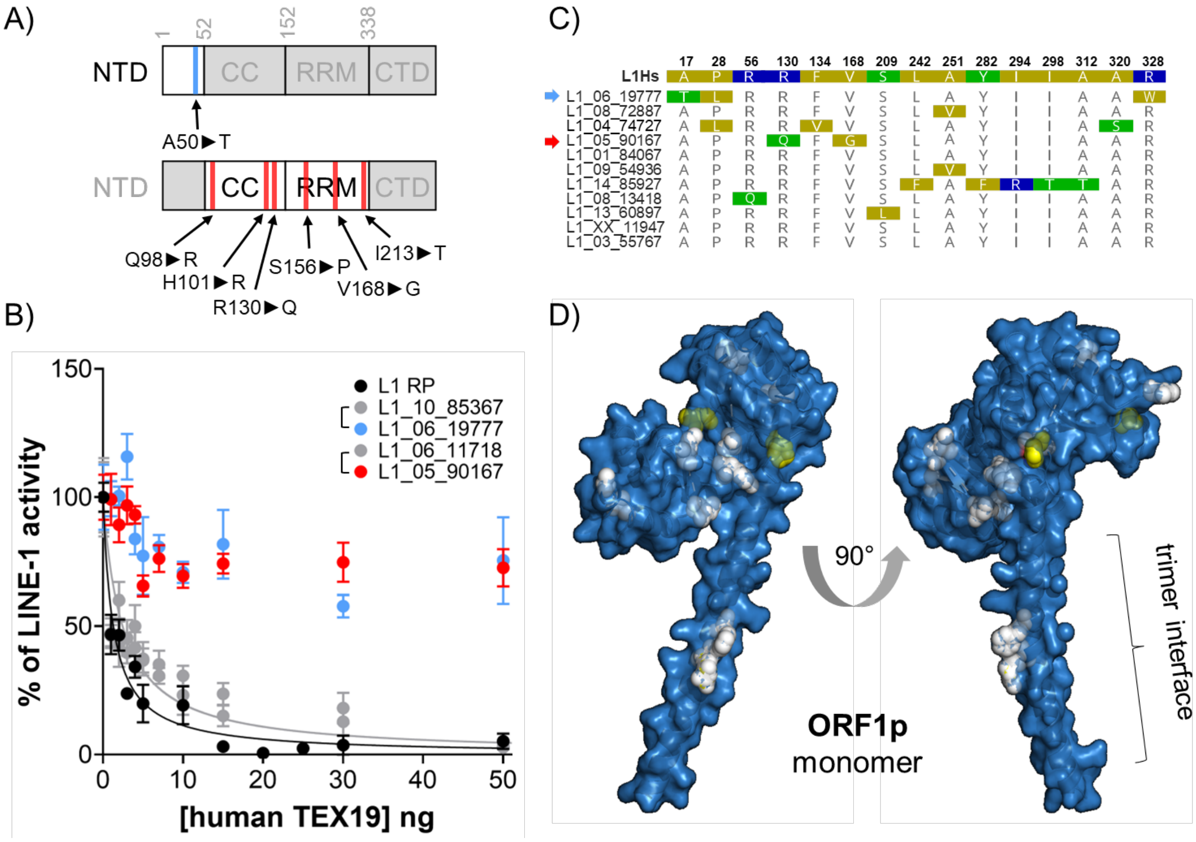
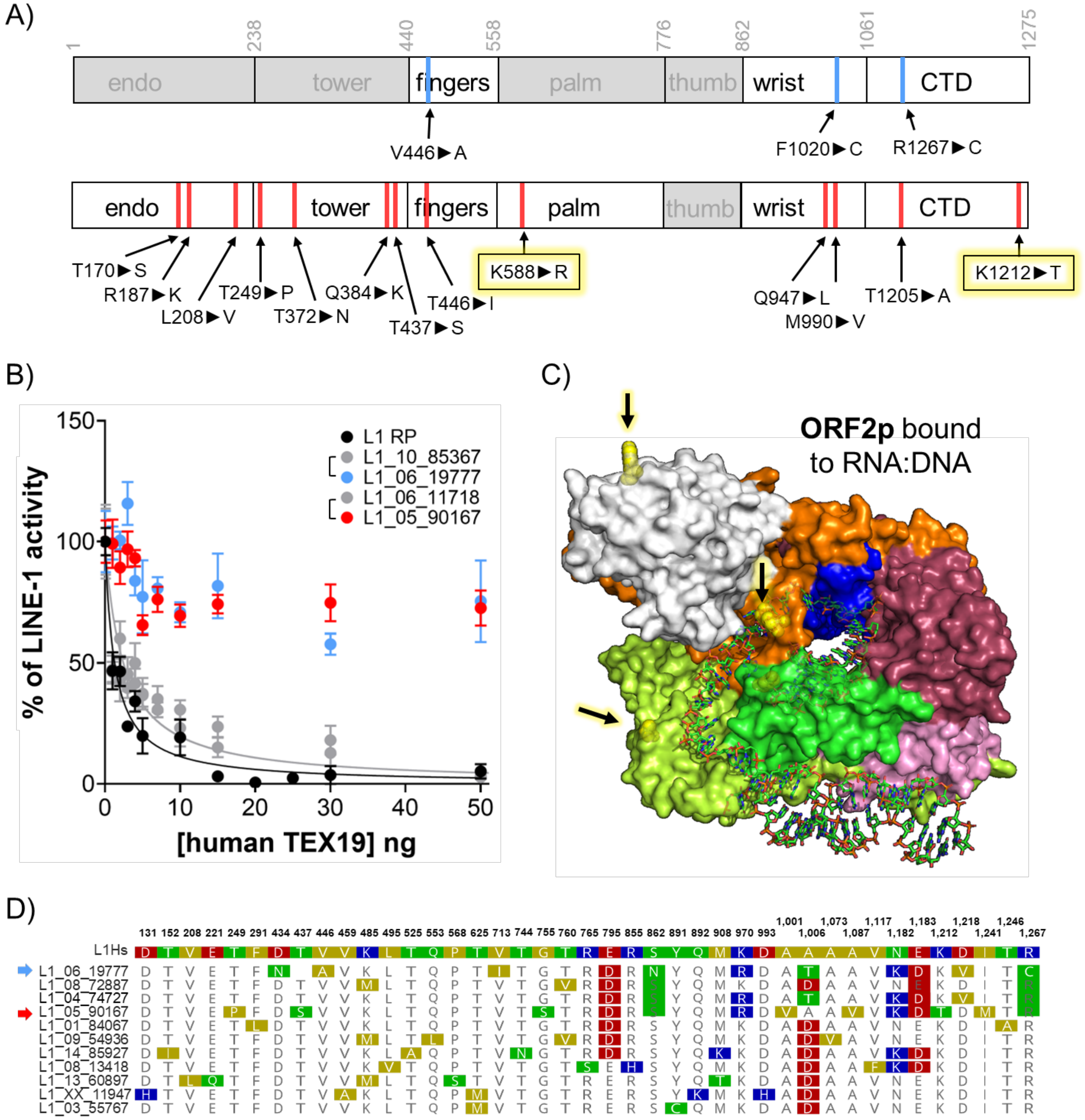


Figure 5. Sequence analysis of ORF1p in TEX19-evasive LINE-1s. A) cartoon of ORF1p from two evasive LINE-1s: blue, L1_06_19777, red, L1_05_990167. Vertical lines indicate unique nonsynonymous mutations when compared to a TEX19-restricted LINE-1 with highest sequence similarity. Domain boundaries are based on ORF1p structure (*205*). B) Dose-response curves showing LINE-1 percentage of activity against increasing amounts of transfected human TEX19. C) All the nonsynonymous changes relative to the L1Hs consensus sequence found in TEX19-evasive LINE-1s. D) Monomer structure of LINE-1 ORF1p (PDB 2YKP). Amino acid differences between evasive L1_05_990167 (red) and restricted L1_06_11718 (gray) are denoted by yellow spheres. Amino acid mutations in evasive LINE-1s compared to a consensus of L1Hs are denoted by white spheres.



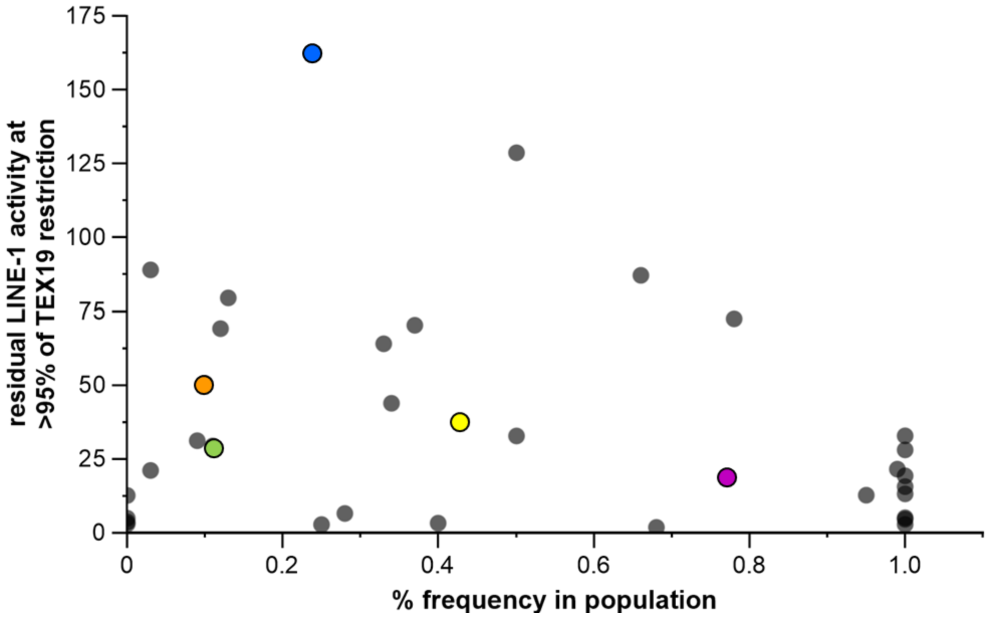
	131	152	208	221	249	291	434	437	446	459	485	495	525	553	568	625	713	744	755	760	765	795	855	862	891	892	908	970	993	1,001	1,006	1,073	1,087	1,117	1,182	1,183	1,212	1,218	1,241	1,246	1,267
L1Hs	D	T	V	E	T	F	D	T	V	V	K	L	T	Q	P	T	V	T	G	T	R	E	R	S	Y	Q	M	K	D	A	A	A	A	V	N	E	K	D	I	T	R
L1_06_19777	D	T	V	E	T	F	N		A	V	K	L	T	Q	P	T	I	T	G	T	R	D	R	N	Y	Q	M	R	D	A	T	A	A	V	K	D	K	V	I	T	C
L1_08_72887	D	T	V	E	T	F	D	T	V	V	M	L	T	Q	P	T	V	T	G	V	R	D	R	S	Y	Q	M	K	D	A	D	A	A	V	N	E	K	D	I	T	R
L1_04_74727	D	T	V	E	T	F	D	T	V	V	K	L	T	Q	P	T	V	T	G	T	R	D	R	S	Y	Q	M	R	D	A	T	A	A	V	K	D	K	V	I	T	R
L1_05_90167	D	T	V	E	P	F	D	S	V	V	K	L	T	Q	P	T	V	T	S	T	R	D	R	S	Y	Q	M	R	D	V	A	A	V	V	K	D	T	D	M	T	R
L1_01_84067	D	T	V	E	T	L	D	T	V	V	K	L	T	Q	P	T	V	T	G	T	R	D	R	S	Y	Q	M	K	D	A	D	A	A	V	N	E	K	D	I	A	R
L1_09_54936	D	T	V	E	T	F	D	T	V	V	M	L	T	L	P	T	V	T	G	V	R	D	R	S	Y	Q	M	K	D	A	D	V	A	V	N	E	K	D	I	T	R
L1_14_85927	D	I	V	E	T	F	D	T	V	V	K	L	A	Q	P	T	V	N	G	T	R	D	R	S	Y	Q	K	K	D	A	D	A	A	V	K	D	K	D	I	T	R
L1_08_13418	D	T	V	E	T	F	D	T	V	V	K	V	T	Q	P	T	V	T	G	T	S	E	H	S	Y	Q	M	K	D	A	D	A	A	F	K	D	K	D	I	T	R
L1_13_60897	D	T	L	Q	T	F	D	T	V	V	M	L	T	Q	S	T	V	T	G	T	R	E	R	S	Y	Q	T	K	D	A	D	A	A	V	N	E	K	D	I	T	R
L1_XX_11947	H	T	V	E	T	F	D	T	V	A	K	L	T	Q	P	M	V	T	G	T	R	E	R	S	Y	K	M	K	H	A	D	A	A	V	N	E	K	D	I	T	R
L1_03_55767	D	T	V	E	T	F	D	T	V	V	K	L	T	Q	P	M	V	T	G	T	R	E	R	S	C	Q	M	K	D	A	D	A	A	V	N	E	K	D	I	T	R

Figure 6. Sequence analysis of ORF2p in TEX19-evasive LINE-1s. A) cartoon of ORF2p from two evasive LINE-1s: blue, L1_06_19777, red, L1_05_990167. Vertical lines indicate unique nonsynonymous mutations when compared to a TEX19-restricted LINE-1 with highest sequence similarity. Domain boundaries are based on ORF2p structure *(95)*. B) Dose-response curves showing percentage of LINE-1 activity against increasing amounts of transfected human TEX19. C) Structure of LINE-1 ORF2p (PDB 8UW3). Mutated lysine residues are denoted by black-yellow arrows. D) All the nonsynonymous changes relative to the L1Hs consensus sequence found in TEX19-evasive LINE-1s.

We were surprised to find many changes in ORF2p (relative to ORF1p) because ORF1p is the known interactor with TEX19/UBR2 (*143*, *204*), and we expected that evasive LINE-1s would have mutations in their ORF1. The proposed mechanism for TEX19 restriction suggest model that may shows experimental evidence suggesting that TEX19/UBR2 ubiquitinates ORF1p and depletes ORF1p abundance via proteasomal degradation (*204*). We think that amino acids susceptible to protein ubiquitination are of interest because losing those could theoretically lead to LINE-1 evasion of TEX19/UBR2 restriction. We did not find mutated lysine (Lys) residues in ORF1p of TEX19-evasive LINE-1s, although other residues like cysteine (Cys), serine (Ser), and threonine (Thr) as well as N-terminal amino groups can be ubiquitinated and further modified like the lysine residues present in ubiquitin (*206*). In Fig 6C, we highlight in yellow two mutated lysines in the ORF2 p TEX19-evasive L1_05_90167. Other potentially relevant nonsynonymous changes are annotated in Figure 6 panel D, and they represent potential candidates for sequence determinants of evasion. Experimental data using site-directed mutagenesis will be essential for further understanding ORF mutations in TEX19-evasives.

Population frequency of TEX19-evasive LINE-1s

Next, we show the calculated population frequency of all retrocompetent LINE-1s from CHM1 in other human genomes based on 1000-genome data project datasets and the human reference genome hg38 (*109*). Several of these retrocompetent LINE-1s are fixed (present in all humans). We found that TEX19-evasives range from unique to the CHM1 genome or <0.1% and not found in anyone else (sequenced) to present in 80% of humans (Fig 7).



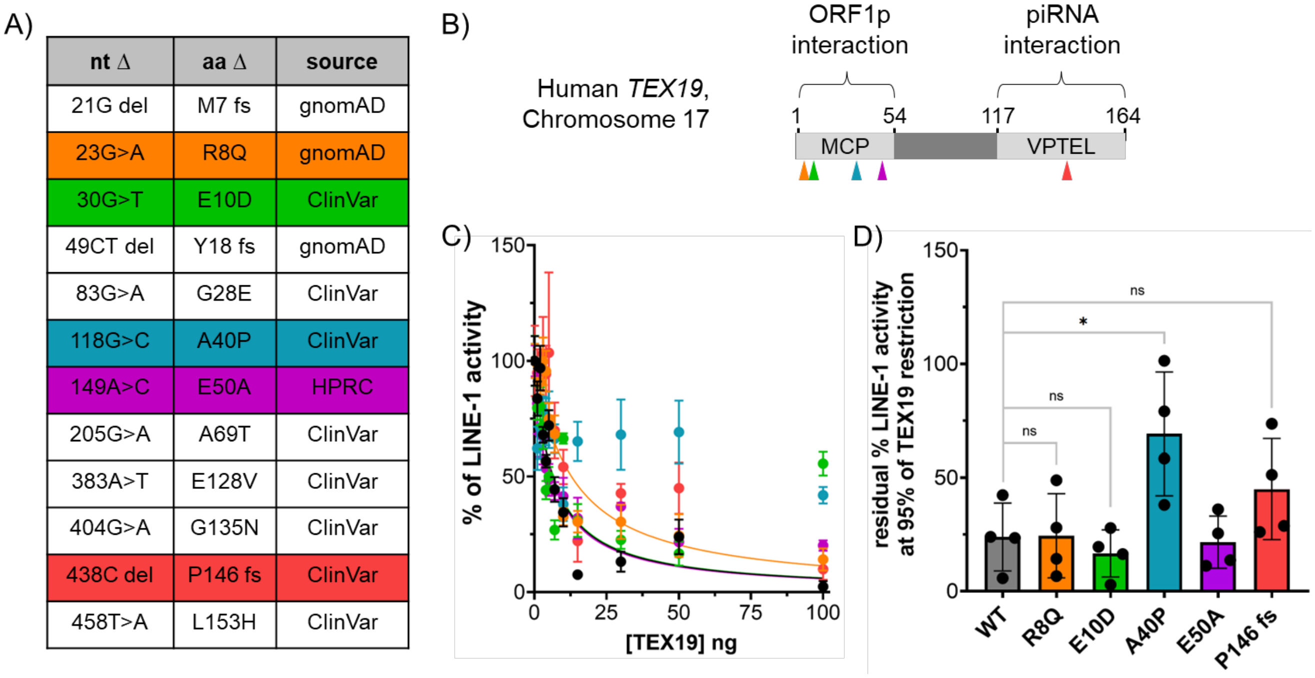
CHM1 name	Number of 3' transductions in Tubio et al. 2014 Table S5.	Number of 3' transductions in Rodriguez-Martin et al. 2020 Table S5.	Number of 3' transductions in Chuang et al. 2021 Table S3.
L1_01_84067	7	39	-
L1_03_55767	3	15	-
L1_04_74727	-	6	-
L1_06_19777	17	10	-
L1_08_13418	4	42	-
L1_08_72887	5	32	-
L1_09_54936		4	-
L1_13_60897	16	50	1
L1_14_85927	-	-	-
L1_XX_11947	20	112	-

L1_03_38597	-	-	-
L1_05_15298	-	-	-
L1_05_39797	17	25	-
L1_09_11288	6	128	-
L1_12_12638	-	-	-
L1_14_58767	40	347	2
L1_16_33967	-	-	-
L1_17_66647	2	3	-
L1_18_59417	10	35	-
L1_XX_11717	7	183	-
L1_XX_83077	-	-	-

Figure 7. TEX19 evasive LINE-1s from the CHM1 genome are replicating *in vivo*, as shown by the 3' transductions detected in cancer and germline datasets. The graph shows residual LINE-1 activity in relation to the frequency of that LINE-1 in the human population (*109*). Colored data points correspond to retrocompetent LINE-1s from CHm1 in the following table for which 3' transductions have been reported based on three different studies. " –" means either that LINE-1 is not present in that dataset, or absent from the considered genomes of the study.

3' transductions and in vivo activity of TEX19-evasive LINE-1s

Sometimes, during retrotransposition, LINE-1s can move bits of neighboring 3' sequences *in trans* when they integrate in a new locus; these are called 3' transductions. We searched for 3' transductions made by all retrocompetent LINE-1s in the CHM1 library and found that TEX19-evasive LINE-1s are actively making new copies of themselves based on previous whole-genome datasets for cancer and germline studies (Fig 7). We plotted residual LINE-1 activity at >95% of TEX19 restriction (relative to LINE-1RP) versus frequency in the population and included a table showing the values for 3' transductions found in three different studies (*101*, *207*, *208*). For example, L1_06_19777 (blue) is highly evasive of TEX19, present in less than 25% of the human population, and we detected seventeen (17) and ten (10) 3' transductions in two of the studies. We also found 3' transductions for other TEX19-evasives suggesting that our *in vitro* assay successfully detected TEX19-evasives LINE-1s that replicate *in vivo*. The colored circles in the graph match the colored rows in the table in Fig 7.



A)

nt Δ	aa Δ	source
21G del	M7 fs	gnomAD
23G>A	R8Q	gnomAD
30G>T	E10D	ClinVar
49CT del	Y18 fs	gnomAD
83G>A	G28E	ClinVar
118G>C	A40P	ClinVar
149A>C	E50A	HPRC
205G>A	A69T	ClinVar
383A>T	E128V	ClinVar
404G>A	G135N	ClinVar
438C del	P146 fs	ClinVar
458T>A	L153H	ClinVar

Figure 8. Variants of human TEX19 differ in their ability to restrict LINE-1 (RP) activity *in vitro*. A) Several human TEX19 variants were found in the human population and reported in ClinVar and gnomAD. B) TEX19 gene structure; triangles show placement for single mutants tested. C) Dose-response showing LINE-1 activity for five different TEX19 mutants in the human population. D) LINE-1 activity at 50ng of TEX19 (>95% restriction of LINE-1RP). Columns were plotted using averages (n=8), and error bars show standard deviation values. Statistical significance analyses were also plotted, ** $p < 0.001$, * $p < 0.05$, ns $p > 0.05$. Abbreviations: nt; nucleotide, aa; amino acid, del;

deletion, fs; frameshift, gnomAD; Genome Aggregation Database, HPRC; Human Pangenome Reference Consortium.

TEX19 clinical variants in the human genome

Loss of function mutations in host restriction factors of LINE-1 can also lead to increased retrotransposition, as is well established for TREX1 mutations that cause AGS(*183*). We searched for TEX19 variations in human genomes that could be associated with disease by collecting all TEX19 mutations found in two databases of candidates for disease-causing mutations (ClinVar and gnomAD). We found potentially “clinically relevant” mutations, yet no experimental data could attest to the “clinical’ consequences of these mutations on human health. We show (Fig 8A) single nucleotide TEX19 mutants and tested some of them based on their location in catalytic domains of TEX19. We found that the MCP region mutant varied in their restriction of LINE-1$_{RP}$. MCP is the region known to engage in protein-protein interaction with LINE-1 ORF1p. R8Q (orange), E10D (green), and E50A had no effect compared to the restriction of WT human TEX19 (Fig 8D; no statistical difference). However, A40P showed decreased WT TEX19 restriction capacity shown by (3-fold) increased LINE-1$_{RP}$ retrotransposition activity (statistically significant, upaired t test, $p = 0.0359$, $p<0.05$). P146 frameshift (fs) produces an early stop codon that shortens human TEX19. We expected this mutation to disrupt the piRNA interaction domain VPTEL. Although we did not see statistically significant differences compared to WT TEX19, we think that the slightly increased LINE-1$_{RP}$ retrotransposition activity seen in Figure 8, panel D, could mean that VPTEL mutations might disrupt LINE-1$_{RP}$ restriction. Lastly, E10D seems to increase LINE-1 activity in 50 ng (Fig 8, D), but this is a technical error; higher concentrations of TEX19 can lead to cell death, lowering the denominator value (REN luciferase) for calculating the % of activity, then appearing as an increase in LINE-1 activity.

Current and Future Work

Here I include experiments and ideas for characterizing sequence determinants of TEX19-evasion in relation to ORF1p abundance. We plan to work on finding LINE-1 evasion in other human genomes and ultimatly undestand more about LINE-1 expansion into new human-specific subfamilies.

Sequence determinants of LINE-1 evasion to TEX19

Currently, we are working on the cloning of single and multiple nucleotide mutants, using site-directed mutagenesis, for the reconstitution of TEX19-evasive phenotype. In Chapter 2, we identified two cases where only a few amino acid changes separated TEX19-evasive and restricted LINE-1s. For TEX19-evasive L1_06_1977 and TEX19-restricted L1_10_8536, there were only four (4) nonsynonymous changes. For TEX19-evasive L1_05_90167 and TEX19-restricted L1_06_11718, there are nineteen (19) nonsynonymous changes. We plan to measure the LINE activity of mutants against increasing amounts of human TEX19. We expect that some mutations could kill LINE-1 activity or be as restricted as WT, but we think others will show at least intermediate or full reconstitution of evasion.

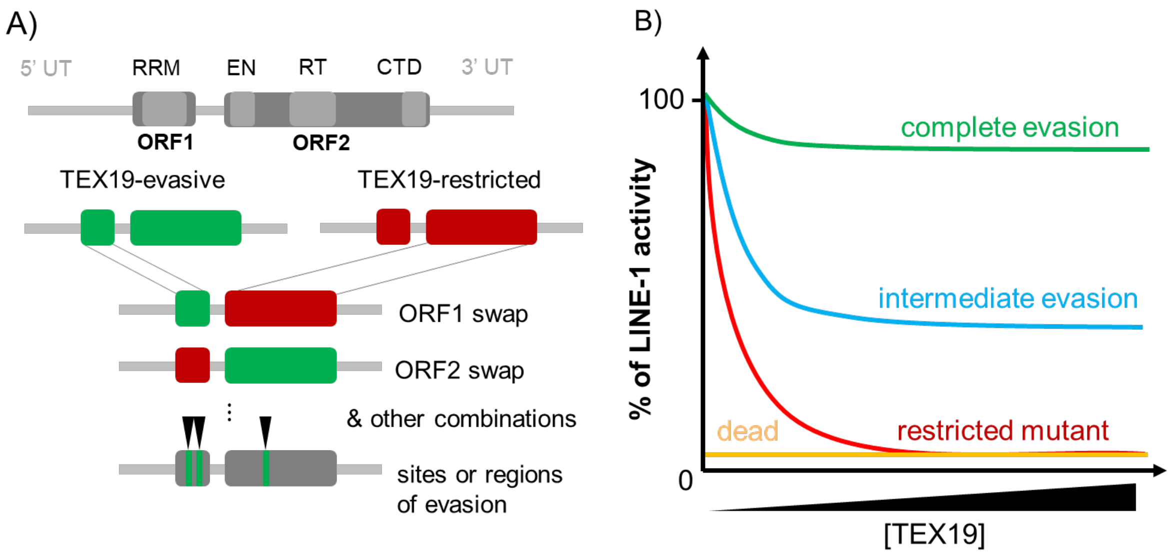


Figure 9. Cartoon of cloning scheme for identification and characterization of TEX9-evasive LINE-1s. A) Chimera cloning strategy. B) Dose-response curves of potential functional changes for mutants.

Another way we have approached identifying sequence determinants of evasion is by creating chimeras of TEX19-evasive and restricted LINE-1s (Fig 9).

If we assume that the proposed mechanism for TEX19 restriction of LINE-1 is accurate, that is, ORF1p abundance decreases when TEX19 is present, then we think that the ORF1p abundance from TEX19-evasive LINE-1s might be indifferent to TEX19 restriction (stays the same as without RF). Current work is focused on optimizing ORF1p detection in TEX1-evasive LINE-1s. For this, I have designed a shorter dose-response curve with four (4) points (0, 5, 15, 50 ng) instead of eleven (11), using 12-well plates (3 technical replicates per transfection). To measure the abundance of ORF1 protein, we transfected each LINE-1/restriction factor combination as described above except using 12-well tissue culture plates (0.3 x 10^6 cells, up to 1ug of pYX-LINE-1 per well), following the timeline used for dual-luciferase assay; five (5) days after puromycin (2.5 µg/uL) selection, we harvested cells from individual wells using trypsin 0.25% for cell dissociation. Then, we lysed the cells using radioimmunoprecipitation assay (RIPA) buffer and Protease Inhibitor Cocktail (1X). We quantified bulk protein content via the bicinchoninic acid (BCA) standard curve and calculated for loading equal amounts of total protein into the gel. We used beta-mercaptoethanol and heat (100C) to negatively charge and denature the proteins, respectively. We measured LINE-1 ORF1p abundance in relation to TEX19 (pCMV-TEX19-HA) expression using antibodies specific for ORF1p [EPR22227-6] (ab245249) and HA (clone 16B12). For protein loading control, we probed for protein abundance of COXIV (3E11), an isoform of the nuclear-encoded subunit IV of cytochrome c oxidase located in the inner membrane of the mitochondria. Bands were visualized on the ImageStudio™ Software in Odyssey® CLx imaging system using IR dye-coupled secondary antibodies IRDye® 800CW and IRDye® 680LT.

One limitation of using ORF1p abundance to further characterize LINE-1 evasion is that retrocompetent LINE-1s from the CHM1 genome have different levels of retrotransposition activity (20% to +100%) relative to LINE-1$_{RP}$, which means that ORF1p abundance among those LINE-1s might also differ. We plan to compare ORF1p abundance between no RF (pCMV-HA, empty) and RF (TEX19: 5, 15, 50 ng). If both TEX19-evasive and restricted LINE1s show the same rate of ORF1p depletion at increasing amounts of TEX19, this could suggest that evasion happens through a mechanism other than proteasomal degradation of ORF1p. In this direction, we will consider the potential role of ORF2p in the interaction of TEX19/UBR2-ORF1p.

Discussion

We found that while restriction factors like APOBEC3B and APOBEC3A potently restricted all LINE-1s, eleven (11) LINE-1s showed strong evasive phenotypes against TEX19 *in vitro*; they replicate at near-unrestricted levels, even in the presence of high concentrations of TEX19. This evasion is species-specific, as mouse TEX19.1 fully restricts these LINE-1s that evade human TEX19 (S2). This data completely contrasts the APOBEC3 restriction of all LINE-1s tested. APOBEC3 mode of restriction differs and involves DNA-protein interaction, and therefore harder to disrupt via nonsynonymous changes like protein-protein interactions.

We think this variability of restriction happened because of mutations specific to the youngest family of LINE-1s in the human genome, L1 Hs. Sequence analyses for identifying mutations as candidates for determinants of TEX19 evasion (Fig 5, 6) showed that most nonsynonymous changes between TEX19-evasive (L1_06_19777, L1_05_90167) and TEX19-restricted (L1_10_8536, L1_06_11718) were located throughout ORF1 and ORF2. We observed that mutations are located in essential domains of ORF1p catalytic activity, like in the coil-coiled (CC) domain and the RNA recognition motif (RRM). Preliminary conjectures point to the loss of ubiquitination sites (Lys) as a potential evasion mechanism, but functional assessment of evasive mutations will reveal which mutations are necessary for achieving evasion. Acquisition of mutations in the LINE-1 sequence can disrupt and circumvent host mechanisms of defense, leading to increased retrotransposition (evasion) and allowing for new LINE-1 expansions in the genome (*122*, *152*). *ZNF93* is a LINE-1 restriction factor expressed in the ancestor of primates, and it targeted the promoter of LINE-1s in the PA4 family via direct protein-DNA binding and subsequent recruitment of methylases for epigenetic silencing. LINE-1s from PA4 experienced a 120bp-deletion in the promoter region where ZNF93 used to bind, resulting in the transcriptional de-repression of LINE-1 and the continued expansion of PA4 to PA3(*123*).

Another case of evasion was found in brain LINE-1s that have acquired mutations (deletion) at the 5’ UTR binding site for a transcription factor Ying Yang 1 (YY1) responsible for the global silencing (methylation) of LINE-1s (*209*). Similarly to the ZNF93-LINE-1 story, we think that mutations that disrupt post-transcriptional silencing of LINE-1s can lead to increased LINE-1 retrotransposition and potentially help explain the most recent expansion of LINE-1s in the human genome. Mutations we pointed at could alter protein-protein interactions in TEX19-ORF1p, for example, are of special interest because they are relatively easy to acquire and a different residue could affect either binding affinity and/or spatial organization, which can be disruptive to function and lead to evasion. The fact that most mutations were found in ORF2p could be explained by ORF size; ORF2p is bigger than ORF1 and, therefore, accumulates more mutations. ORF2p could also be more mutationally tolerant compared to ORF1p. But also, it could mean that ORF2 plays a role in TEX19/UBR2-ORF1p interaction. If ORF2p is, in some way, a target of TEX19/UBR2 in humans, losses of lysine and other residues subject to ubiquitination could explain the evasion of proteasomal degradation as a mechanism that results in increased LINE-1 retrotransposition.

Further, we found that the TEX19-evasive LINE-1s have been active *in vivo*, making new copies of themselves as evidenced by 3’ transductions and clusters of related L1s in diverse human genomes. Most of the TEX19-evasive LINE-1s from the CHM1 genome are not fixed in the human population. Both 3’ transduction and population frequency of TEX19-evasive LINE-1s make a strong argument to consider the evasiveness of immune restriction factors as a mechanism for the most recent LINE-1 expansion and evolution in the human genome.

Lastly, we found that previously listed/annotated single-nucleotide variants of TEX19 in the human genome differently restrict LINE-1_{RP}. This suggests that individual genomes may not only harbor evasive LINE-1s to TEX19, but mutations in TEX19 can lead to upregulation of typically restricted LINE-1s. Knockout or knockdown of TEX19 paralogs in mice disrupts placenta/testis development. Yet, LINE-1's relation to this previously described phenotype remains elusive. This could explain the developmental disorder and disease phenotypes observed in mouse models lacking TEX19 (*144*, *203*) and contribute insights to the study of a similar LINE-1-TEX19 disease model in humans.

Ultimately, LINE-1 evasion to TEX19 could lead to *de novo* LINE-1 insertions with disease potential and the expansion of new families of LINE-1s in the human genome. We think that LINE-1 evolution could have happened through expansions that emerged as the result of bouts of intense selective pressure from mutations that evade host defense and damage host fitness.

Methodology

Plasmids and constructs

Active LINE-1s were cloned from BACs of the CHM1 genome as previously described (*109*). Briefly, LINE-1s were cloned into the pYX luciferase reporter construct (full LINE-1 sequence including 5’ UTR, two ORFs, and 3’ UTR, driven by a CAG promoter) in which firefly luciferase is appended to the 3’ UTR of the LINE-1, in the opposite direction, interrupted by an intron (Fig 2A) (*210*), and is a proxy for a complete retrotransposition cycle. Renilla luciferase is constitutively expressed from the plasmid backbone using an SHV-TK promoter and is used as a control for transfection efficiency and toxicity. LINE-1_{RP} (pYX017), a highly active LINE-1_{RP}, was used as a reference for LINE-1 activity, and an RT-mutant LINE-1 (pYX015/JM111) as a negative control for (zero) LINE-1 activity.

All restriction factors, human APOBEC3A (A3A), human APOBEC3B (A3B), human TEX19, and mouse TEX19.1, were cloned into pCMV-HA, which encodes an N-terminal HA tag and a cytomegalovirus (CMV) promoter. An empty pCMV-HA vector was used as a negative control.

Mutagenesis for clinical variants of human TEX19

We searched for variants of *TEX19* in the human population using the gnoMAD and ClinVar databases. Then, we listed single-nucleotide variants and selected five (5) mutations based on their location and relation to LINE-1 regulation in ORF1p(*204*). We designed oligos for single-mutagenesis of TEX19 (pCMV-TEX19-HA). Using Gibson assembly of TEX19 and pCMV-HA we generated constructs with desired mutations. Annotation is presented as follows: nucleotide position, nucleotide base in reference genome, nucleotide change or amino acid position in ORF1p, amino acid identity in reference genome, amino acid change. Constructs were made for the following variants (Fig 8A): 1) 23G>A, R8Q, 2) 30G>T, E10D, 3) 118G>C, A40P, 4) 149A>C, E50A, and 5) deletion of 438C, P146 to frameshift that leads to a premature stop codon.

Retrotransposition assays

To measure retrotransposition of LINE-1s, HEK-293T cells (ATCC CRL-3) were seeded in white, flat transparent bottom 96-well tissue culture plates (2.5 x 10^4 cells per well) and 24 hours later transfected with a pYX-LINE-1 (from library of retrocompetent LINE-1s in CHM1) and pCMV-restriction-factor or pCMV-empty using Trans®IT-LT1 transfection reagent from Mirus© at the recommended DNA:TransIT®-LT1 transfection ratio of 3:1. The amount of LINE-1 transfected (30 ng per well in 96-well plates) was maintained constant in all experiments. Twenty-four (24) hours after transfection, we selected for plasmid intake with puromycin [2.5 μg/uL].

Dual-luciferase reporter assay

The Dual-Luciferase® Reporter (DLR™) Assay System from Promega© was used following the manufacturer's instructions to assay firefly and renilla luciferase levels five (5) days after puromycin [2.5 μg/uL] selection.

Data normalization and manipulation

We collected firefly (FF) and renilla (REN) luciferase values using a UV-Vis spectrometer. Next, we calculated averages and standard deviations for eight (8) replicates per condition. We used FF values from LINE-1_{RP} as 100% of activity and RT-mutant LINE-1 (pYX015/JM111) as 0% of LINE-1 activity. The activity of all thirty-four (34) retrocompetent LINE-1s in CHM1 ranges from 20% to greater than100% of LINE-1_{RP}. In order to compare between LINE-1s, we normalized each to their activity without a restriction factor (pCMV-HA). REN values were used to control for differences in transcription but were not used to calculate a luciferase ratio to avoid incorrect visualization of LINE-1 activity at lower levels of plasmid transcription.

Determination of IC_{50} for LINE-1 Restriction Factors

We measured LINE-1_{RP} retrotransposition activity against a range of amounts ofTEX19 [0 – 50] ng. FF values were normalized to no RF (0 ng) condition. We plotted relative FF values as a percentage of LINE-1 retrotransposition activity, and modeled an inhibitory dose-response curve that fitted the experimental values for LINE-1_{RP} as a control for a TEX19-restricted LINE-1. Using that approach, we calculated the IC_{50} for A3A, A3B, TREX1, and TEX19.

Dose-response curves of restriction factors on LINE-1 activity

To measure the sensitivity of each LINE-1 to various restriction factors, we cotransfected increasing amounts of human APOBEC3A, human APOBEC3B, human TEX19, or mouse TEX19.1 with each LINE-1. Each condition was measured across at least 8 replicate wells of a 96-well plate and used to plot dose-response curves showing LINE-1 activity (relative to LINE-1_{RP}) versus the restriction factor of interest. For APOBEC3A and APOBEC3B, we used (0-5 ng); for hTEX19, (0-100 ng); and for mTEX19.1 (0 and 50 ng) per well.

LINE-1 sequence analysis

The amino acid sequences of the eleven (11) TEX19-evasive LINE-1s were aligned to the consensus amino acid sequence of the human-specific LINE-1 family (L1Hs). We also compared annotated mutations between closely related TEX19-evasive and

TEX19-restricted LINE-1s using a translation-based alignment algorithm in Genious®. Protein structures were visualized and annotated in PyMOL 2.6 software.

LINE-1 population frequency and 3' transductions

The inferred population frequency of each CHM1 LINE-1 was adapted from previously published work from our group (*109*). Briefly, population frequencies were calculated based on the 1000 Genomes Project, euL1db, and additional studies (*102*, *108*, *166*). The 3' transduction counts were based on the sum of the previously reported transduction counts from Tubio et al. 2014 (*207*), Rodriguez-Martin et al. 2020 (*208*), and Chuang et al. 2021 (*101*).

Acknowledgments

This work was supported by grants from the National Institutes of Health (R35GM142773 to RNM and T32GM007270 to RPDV). I also want to thank Mollie McGibbon for her work with humanTEX19 variants, Ernesto Torres for his engagement with the project and his interest in understanding the biochemical details of LINE-1 ORF1p and human TEX19 interaction, Emily Cross for her contribution working with TREX1, and Joseph McDonald for his work contributing to mapping sequence determinants of evasion. I thank the alumni and members of the McLaughlin lab for technical advice and critical comments on this manuscript.

Chapter 3: Decrypting the Evolution of *TEX19*

Summary

How have host genomes changed in response to a conflicting coevolutionary relationship with LINE-1 parasites? Most of what we know about *TEX19* has been studied in mice in relation to humans, so we searched for *TEX19* in other species. We proposed that by curating genomic variations (determining conservation, mapping nearby genes, identifying orthologs, describe genetic diversity of copy number, and structural variants) and comparing, sided by side, the patterns of variation to characteristics of LINE-1 content in those genomes, we can understand more about TEX19/LINE-1 coevolution in humans, and in other mammals. We identified extensive and recurrent signatures of *TEX19* rapid evolution across mammals, further supporting our model of TEX19/LINE-1 genetic conflict. The *TEX19* locus most commonly contains two (2) separate genes next to each other, although voles (in Rodentia) have up to fourteen (14) copies. Synteny and gene orientation suggest that TEX19 duplications were the result of different mechanisms in different species (i.e., local gene duplication with SECMT1, duplication of chromosomal segment, duplication of whole genome). In the entire order of Canidae (dog, wolf, fox) and in megabats, we cannot find traces of *TEX19*. The loss of *TEX19* in megabats is of particular interest because they also lack active LINE-1s. Lastly, structural variants show two main patterns: 1) *TEX19* is reduced to the beginning half in some species, which is the region responsible for restricting LINE-1 in mice and humans; 2) rodents and related groups show contraction and expansion of a glutamic acid (E) “loop” exactly where human TEX19 binds LINE-1

ORF1p for subsequent proteasomal degradation. The evolution of *TEX19* in mammals is dynamic and replete with signatures suggesting strong selection. Current work focuses on describing molecular evolution details and proposing functional assays to test natural variation of orthologs and the consequences for LINE-1 and TE restriction. Can hosts with duplications of *TEX19* restrict more LINE-1s? Is *TEX19* also important in the embryo/gametes/placenta of other mammals? How did *TEX19* end up in the genomes of placental mammals? We think the role of TEX19 in development of mice embryo and placenta could be tied to its role as a restriction factor of LINE-1.

Introduction

The germline houses a convergence of essential developmental programs and immune mechanisms that repress the potentially deleterious replication of transposable elements. Our perspective (and urgency to understand) as a host of TEs is biased yet is far from being the only (biological) story of host coexistence with TEs.

In 2008, Kuntz and colleagues successfully cloned the mouse Tex19.1 gene (Chromosome 7) and identified the only orthologous gene in humans, *TEX19* (Chromosome 11), through database analysis (*138*). There is a significant discrepancy in length between the two species, with the human protein being truncated due to the presence of a premature stop codon and consisting of only 164 amino acids, while the mouse (Chromosome 4) protein contains 351 amino acids. Nevertheless, both proteins have the same two conserved domains separated by a variable region. *TEX19* exhibits highly restricted expression patterns. Expression analysis using RT-PCR of various adult mouse tissues(*211*) found that mouse *Tex19*.1 was exclusively expressed in the testis, placenta, and ovary. Its expression is tightly regulated during spermatogenesis, with peak expression observed in spermatocytes and early spermatids. Further investigation detected its expression in unfertilized mouse eggs(*212*), early pluripotent mouse embryos, and later embryonic gonads(*213*). Similarly, human tissue analysis showed *TEX19* expression solely in the testis and not in the placenta(*214*). The expression of *TEX19* is tightly regulated at the transcriptional level, with specific transcription factors and cis-regulatory elements controlling its expression in the testis. SOX family transcription factors and testis-specific transcriptional regulators are key transcriptional regulators of *TEX19*, which bind to the promoter region of the *TEX19* gene and modulate its expression during spermatogenesis(*215*).

Studies in mice have shown that *TEX19* is essential for the normal development and maintenance of male germ cells (*138*, *141*, *142*, *216*). Mice that are genetically deficient in *TEX19* exhibit defects in spermatogenesis, including a reduction in the number of spermatogonia and spermatocytes, and a decrease in the production of mature spermatozoa (*140*, *141*, *143*). These defects are thought to be due to a failure of germ cells to progress through the meiotic prophase I stage, which is critical for the formation of haploid sperm cells. During meiosis, *TEX19* is required for the formation of synaptonemal complexes, which are structures that help align homologous chromosomes (*139*). Moreover, mice lacking *TEX19* exhibit underdeveloped embryos and placental defects (*141*, *144*).

This temporal and spatial specificity of *TEX19* expression coincides with the activation of genetic parasites of the genome. Studies employing gene knockout models in mice have provided compelling evidence that shows an increase in the rate of retrotransposition of LINE-1 transposable elements in the absence of *Tex19.1*(*137*). By regulating the activity of LINE-1s, *TEX19* plays a direct role in the maintenance of genomic stability. Dysregulation of retrotransposon-regulating proteins, such as Tex19.1, could lead to increased retrotransposon activity and potentially contribute to disease development(*217*). Mechanistically, TEX19 restricts LINE-1s by associating with a ubiquitin ligase (UBR2) in the cytoplasm and targeting LINE-1 ORF1p for proteasomal degradation (*218*).

TEX19 has undergone multiple duplication events during the evolution of mammals (mouse, rat, and cattle) (*219*, *220*), resulting in multiple paralogs with potentially distinct functions. Further, the protein is truncated in primates, suggesting evolutionary divergence between species. *TEX19* orthologs are present only in mammals. Previous analysis of *TEX19* orthologs across mammalian species shows TEX19 evolution in concordance with secreted and transmembrane protein 1 (SECTM1) because they are proximal to each other and found to be duplicated around the same time in rodents (*219*). The duplication event that gave rise to mouse Tex19.1 and Tex19.2 also duplicated the nearby gene *SECTM1* (secreted and transmembrane protein 1). Yet the same is not true for *TEX19* duplication in cattle (ruminants), where the *SECMT1* copy number has increased up to seven (7). In this study, we have found that there are many duplication events across mammals that do not involve *SECTM1*. In fact, we have found complete losses of *TEX19*, putting into question a conserved role of *TEX19* in organismal biology. Does the scale of evolution (gain, loss, structural variants) suggest that the ancestral function of *TEX19* is immunity, not development? Further research is needed to understand the consequences of *TEX19* duplication/loss events in placental mammals in relation to development and immunity against transposable elements.

We use comparative studies across mammalian species to 1) begin to understand the evolutionary dynamics and adaptive significance of *TEX19* restricting retrotransposons in the germline/testis/placenta and 2) to learn about the fundamental roles of *TEX19* in fertility, development, and reproduction across diverse mammalian taxa.

Results

We think that the acquisition of new restriction factors in the genome can be a mechanism by which host evolution can regain control of the persistent quorum of retrocompetent LINE-1s. Here, we show that LINE-1 restriction factor *TEX19* is subjected to constant changes driving its evolution in mammalian genomes. In almost every group of mammals, we observed several gene duplications, some shared among species (ancient) and others species-specific (recent). Figure 1 summarizes the profile of *TEX19* synteny in mammals.

Profile of TEX19 Content

We searched in publicly available genomes across all mammals (Monotremes, Marsupials, Afrotheria, Xenarthra, Glires, Euarchonta, Eulipotyphla, Chiroptera, Perissodactyla, Cetartiodactyla, Pholidota, And Carnivores) and found more than 570 orthologs of *TEX19* across species (Table 1).

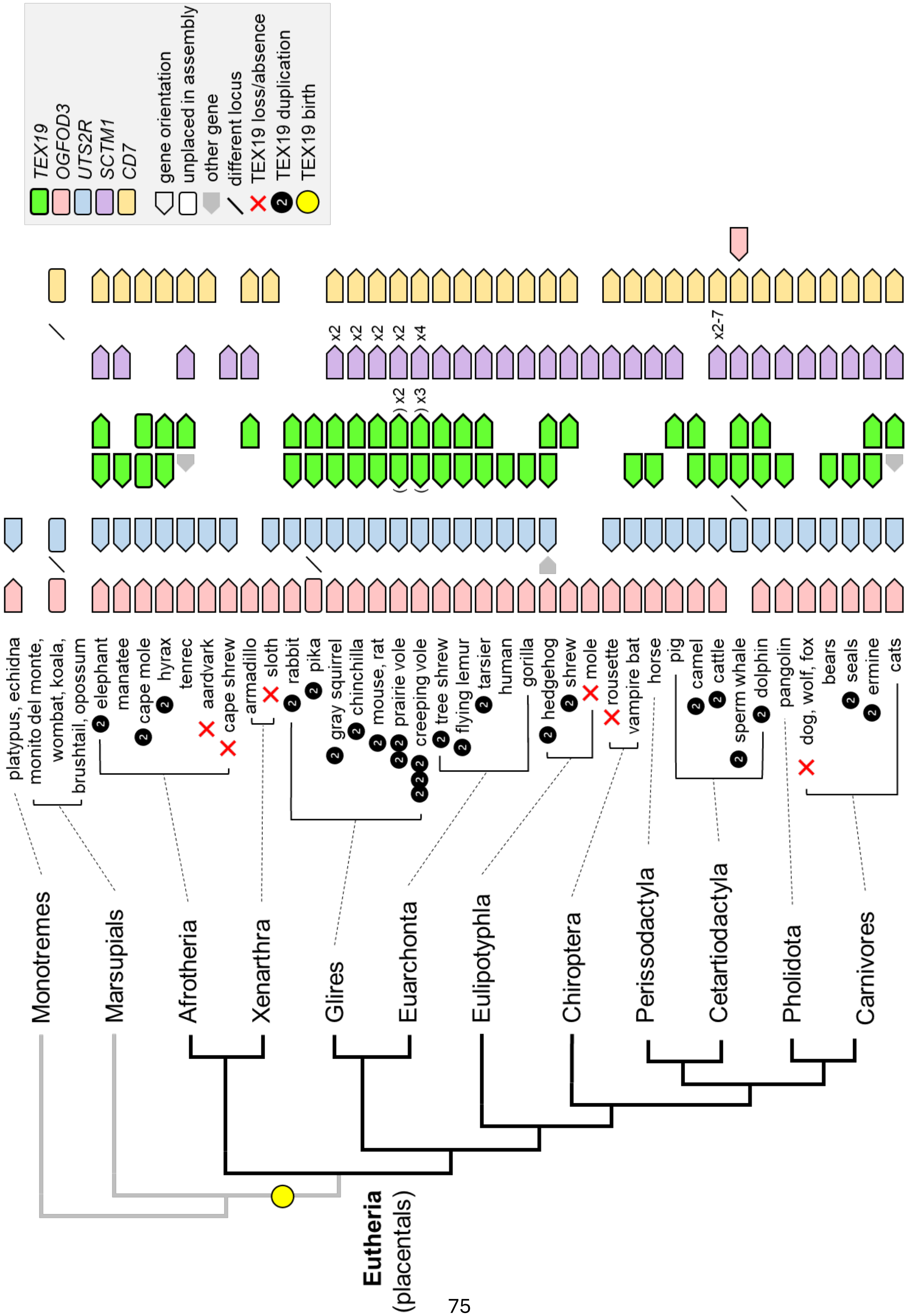

TEX19
OGFOD3
UTS2R
SCTM1
CD7
gene orientation
unplaced in assembly
other gene
different locus
TEX19 loss/absence
TEX19 duplication
TEX19 birth
Monotremes
Marsupials
Afrotheria
Xenarthra
Glires
Euarchonta
Eulipotyphla
Chiroptera
Perissodactyla
Cetartiodactyla
Pholidota
Carnivores
Eutheria
(placentals)
platypus, echidna
monito del monte, wombat, koala, brushtail, opossum
elephant
manatee
cape mole
hyrax
tenrec
aardvark
cape shrew
armadillo
sloth
rabbit
pika
gray squirrel
chinchilla
mouse, rat
prairie vole
creeping vole
tree shrew
flying lemur
tarsier
human
gorilla
hedgehog
shrew
mole
rousette
vampire bat
horse
pig
camel
cattle
sperm whale
dolphin
pangolin
dog, wolf, fox
bears
seals
ermine
cats
2
x2
x3
x4
x2-7

Figure 1. TEX19 evolution is highly dynamic, with independent duplications and multiple losses among different species of placental mammals. This mammalian cladogram shows TEX19 synteny and a summary of duplication events that led to increased copy number and apparent complete losses. For each group, I have annotated a few different representative species, but the sampling includes many more, as seen in the following figures.

Synteny, the preserved order of genes on chromosomes of related species that results from descent from a common ancestor, was helpful for mapping *TEX19*-containing loci. The orientation of *TEX19* relative to nearby genes, as well as the orientation of the open reading frames of nearby genes, allowed the separation of duplications into independent mutagenic events.

Table 1. Breakdown of *TEX19* found across placental mammals.

Group	***TEX19* genes in this study**
Afrotheria	14
Xenarthra	2
Primates	32
Dermoptera	3
Scandentia	3
Rodents	230
Lagomorph	19
Cetartiodactyla	188
Perissodactyla	4
Chiroptera	38
Eulipotyphla	4
Carnivorres	38

Pholidota	1

Copy Number Variants: single-copy, duplications, and losses

TEX19 duplication in mice into *Tex19.1* and *Tex19.2* also duplicated a nearby gene called *SECMT1* into *SECMT1a* and *SECMT1b.* But the same is not true for duplications outside of rodents. This duplication was local because other nearby genes (*OGFOD3, UTS2R*, and CD7) did not experience an increase in copy number. As for the mechanism that produced such duplications, an ancient segmental duplication that included *SECTM1* could explain *TEX19* duplication in the ancient ancestor of rodents, but the argument is not true for other rodents like voles or ruminants. However, it is still puzzling how *TEX19* duplicates are oriented in opposite directions, yet *SECTM1* duplicates point the same way. In contrast, another group of rodents, commonly known as voles, not only share that ancient duplication in the ancestor of mice but has undergone multiple duplication events at chromosome (*Microtus ochrogaster*, prairie vole) and whole genome level (*Microtus oregoni*, creeping vole), shown by the increased copy number of the entire *TEX19* locus to up to fourteen (14) *TEX19* paralogs.

Another example of ancient duplication is found in Cetartiodactyla (even-toe ungulates), specifically Ruminants (cattle, sheep, antelopes, deer, giraffes, etc), yet very few recent duplications (except in hybrid cow species). I have found that *SECTM1* in Ruminants increases in copy number from one (1) to seven (7) copies, while *TEX19* remains as a duo. Most *TEX19* duplicates point in opposite directions (mirrored), except for in the genome of rodent Jaculus jaculus (Lesser Egyptian jerboa), where there are three (3) TEX19 in tandem pointing in the same direction.

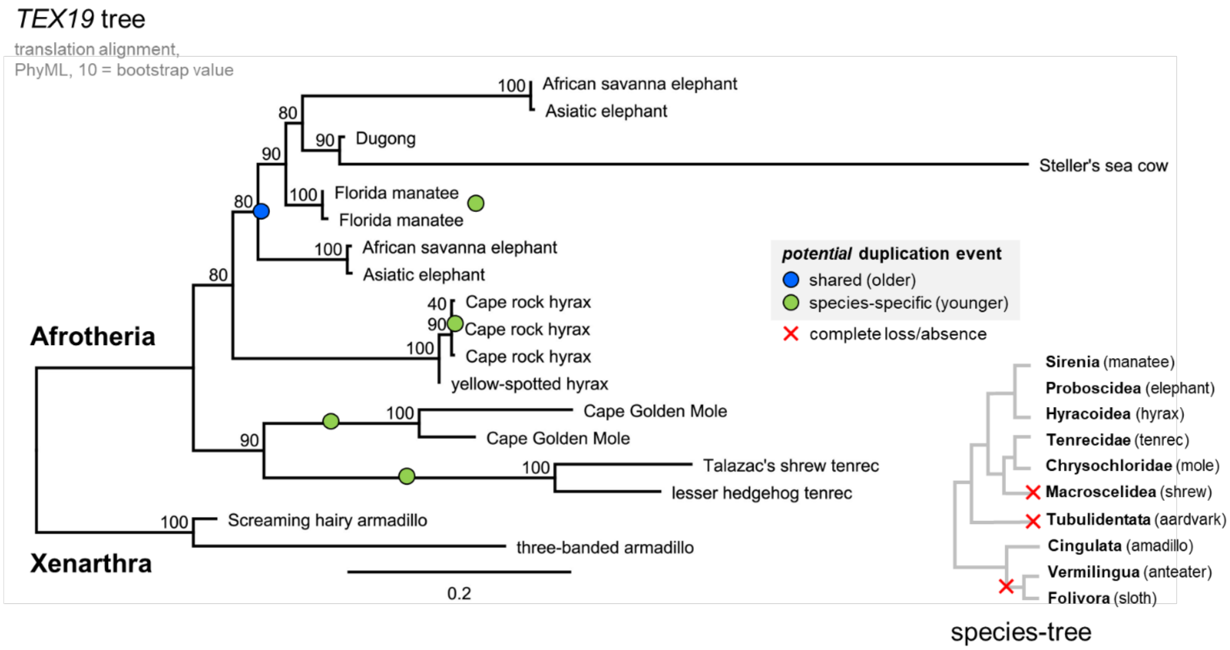


Figure 2. Phylogenetic tree showing the *TEX19* of Afrotherians and Xenarthrans (as root) mammals. Translation alignment, PhyML bootstrap = 10. We found complete losses in three groups. Macroscelidea, Tubilidentata and Vermilingua. There are both shared (ancient) mutations and species-specific (recent) duplications in Afrotherians.

Similarly, the orientation of further and more recent *TEX19* duplications in voles coincides (or arose from) the parental gene, in contrast to the ancient duplication in the common ancestor of mice and Ctenocystrica (gundis, porcupines, and guinea pigs). This suggests that different mechanisms of duplication were experienced by this region. This and other examples of duplications that appeared throughout the entire mammalian tax question the original state (one/two copies) of the first *TEX19* in a host genome.

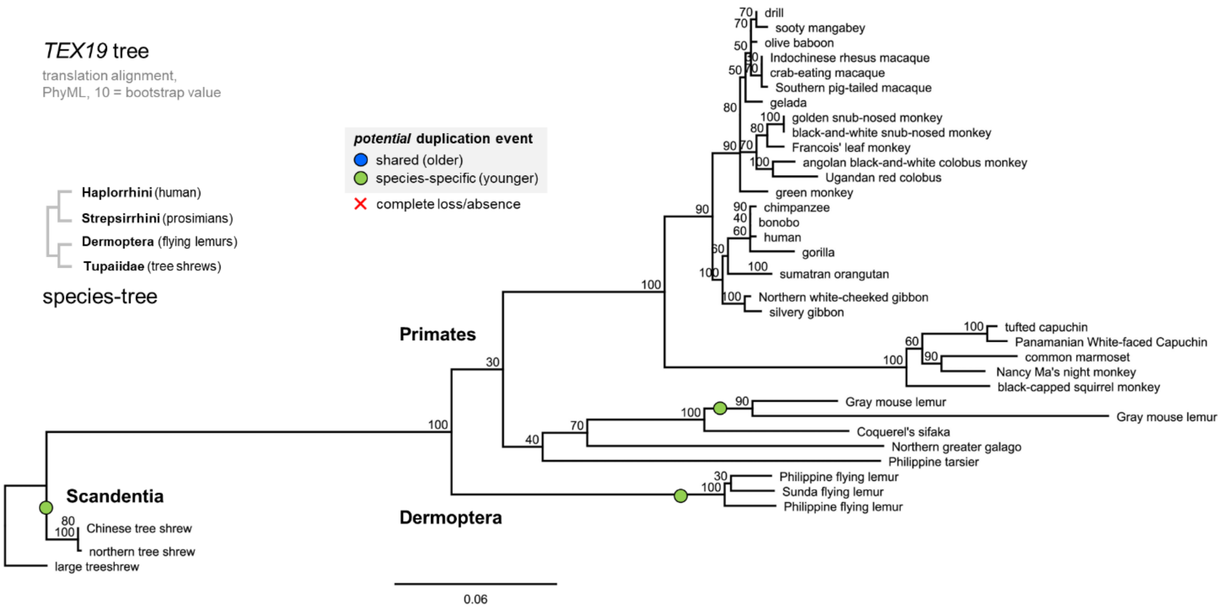
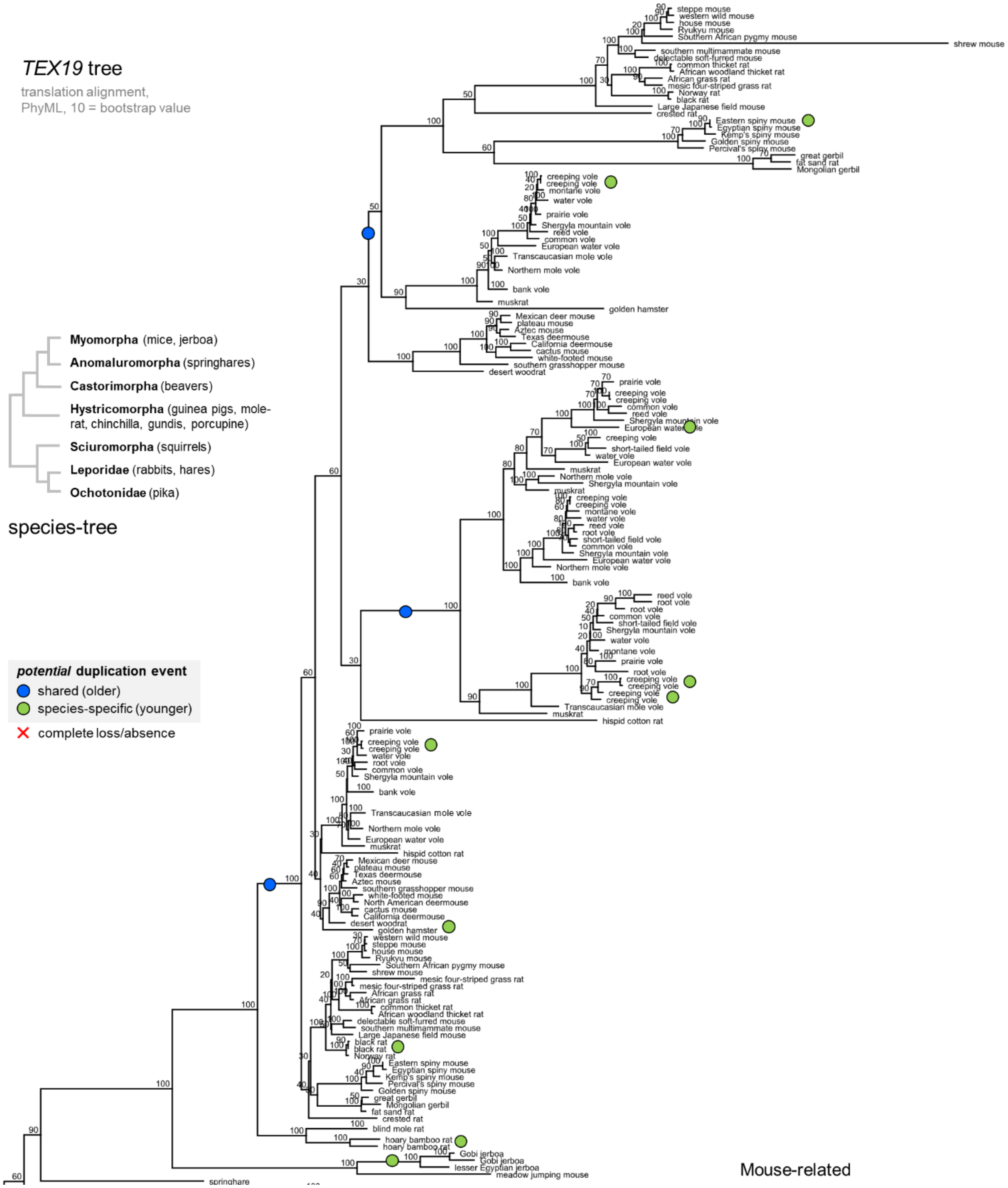


Figure 3. Scandentia, Dermoptera, and Primates *TEX19* gene tree resembles the species tree. Phylogenetic tree of *TEX19* showing Primates and Dermoptera, with Scandentia as root. Translation alignment, PhyML bootstrap = 10. No complete losses were found in these groups, all duplications appear to be recent.

In almost every group of mammals, we observed several gene duplications shared among species (ancient) and species-specific (recent). Comparisons between the gene and the species phylogenetic tree helped us determine if duplication events were ancient or recent. In Afrotherians (Fig 2), for example, we can detect ancient duplication in the ancestor of *Loxodonta africana* (African Savanna elephant) and *Elephas maximus indicus* (Asiatic elephant) because *the TEX19* phylogenetic tree shows them in separate clades each *TEX19* related among different species. This pattern repeats in Rodents (Fig 4), Ruminants and some Cetaceans (Fig 5), Chiroptera (Fig 6) and Carnivores (Fig 7). Groups with exclusively recent duplications were Primates (Strepsirrhini), Scandentia, and Dermoptera (Fig 3), Lagomorphs (Fig 4), Tylopoda (Fig 5), and Eulipotyphla (Fig 6). Those were also the same groups with most of the single-copy *TEX19*, and the ones with closely related species where *TEX19* has been lost.


TEX19 tree
translation alignment,
PhyML, 10 = bootstrap value
Myomorpha (mice, jerboa)
Anomaluromorpha (springhares)
Castorimorpha (beavers)
Hystricomorpha (guinea pigs, mole-rat, chinchilla, gundis, porcupine)
Sciuromorpha (squirrels)
Leporidae (rabbits, hares)
Ochotonidae (pika)
species-tree
potential duplication event
● shared (older)
● species-specific (younger)
✕ complete loss/absence
steppe mouse
western wild mouse
house mouse
Ryukyu mouse
Southern African pygmy mouse
shrew mouse
southern multimammate mouse
delectable soft-furred mouse
common thicket rat
African woodland thicket rat
African grass rat
mesic four-striped grass rat
Norway rat
black rat
Large Japanese field mouse
crested rat
Eastern spiny mouse
Egyptian spiny mouse
Kemp's spiny mouse
Golden spiny mouse
Percival's spiny mouse
great gerbil
fat sand rat
Mongolian gerbil
creeping vole
creeping vole
montane vole
water vole
prairie vole
Shergyla mountain vole
reed vole
common vole
European water vole
Transcaucasian mole vole
Northern mole vole
bank vole
muskrat
golden hamster
Mexican deer mouse
plateau mouse
Aztec mouse
Texas deermouse
California deermouse
cactus mouse
white-footed mouse
southern grasshopper mouse
desert woodrat
prairie vole
creeping vole
creeping vole
common vole
reed vole
Shergyla mountain vole
European water vole
creeping vole
short-tailed field vole
water vole
European water vole
muskrat
Northern mole vole
Shergyla mountain vole
muskrat
creeping vole
creeping vole
montane vole
water vole
reed vole
root vole
short-tailed field vole
common vole
Shergyla mountain vole
European water vole
Northern mole vole
bank vole
reed vole
root vole
root vole
common vole
short-tailed field vole
Shergyla mountain vole
water vole
montane vole
prairie vole
root vole
creeping vole
creeping vole
creeping vole
creeping vole
Transcaucasian mole vole
muskrat
hispid cotton rat
prairie vole
creeping vole
creeping vole
water vole
root vole
common vole
Shergyla mountain vole
bank vole
Transcaucasian mole vole
Northern mole vole
European water vole
muskrat
hispid cotton rat
Mexican deer mouse
plateau mouse
Texas deermouse
Aztec mouse
southern grasshopper mouse
white-footed mouse
North American deermouse
cactus mouse
California deermouse
desert woodrat
golden hamster
western wild mouse
steppe mouse
house mouse
Ryukyu mouse
Southern African pygmy mouse
shrew mouse
mesic four-striped grass rat
mesic four-striped grass rat
African grass rat
African grass rat
common thicket rat
African woodland thicket rat
delectable soft-furred mouse
southern multimammate mouse
Large Japanese field mouse
black rat
black rat
Norway rat
Eastern spiny mouse
Egyptian spiny mouse
Kemp's spiny mouse
Percival's spiny mouse
Golden spiny mouse
great gerbil
Mongolian gerbil
fat sand rat
crested rat
blind mole rat
hoary bamboo rat
hoary bamboo rat
Gobi jerboa
Gobi jerboa
lesser Egyptian jerboa
meadow jumping mouse
springhare
Mouse-related

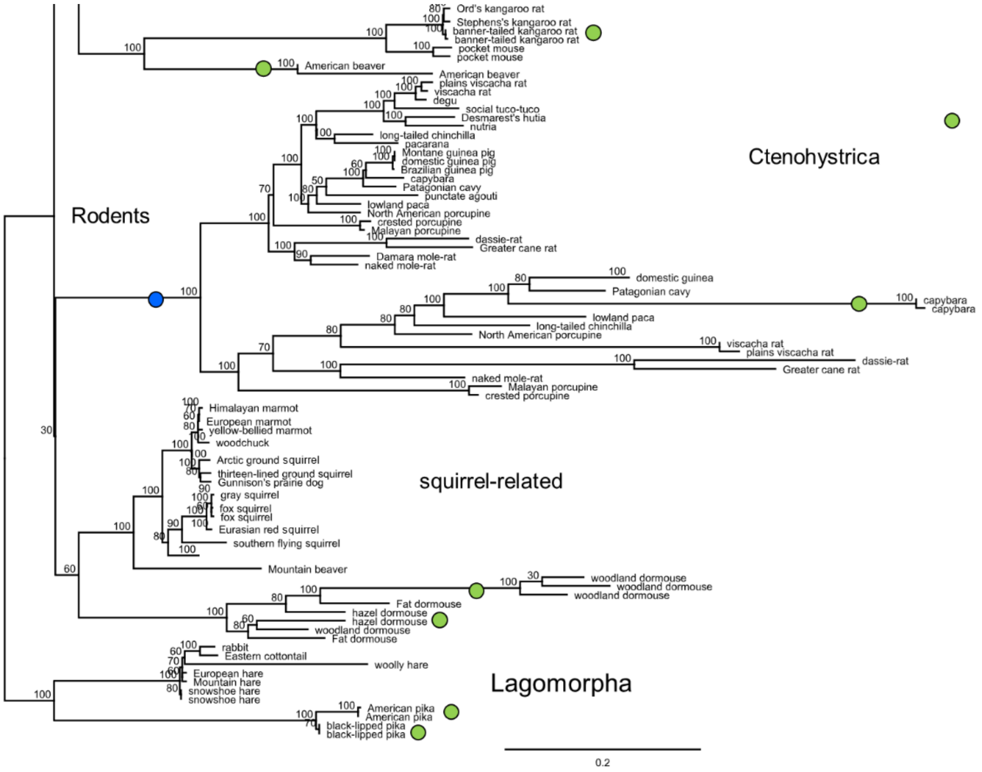
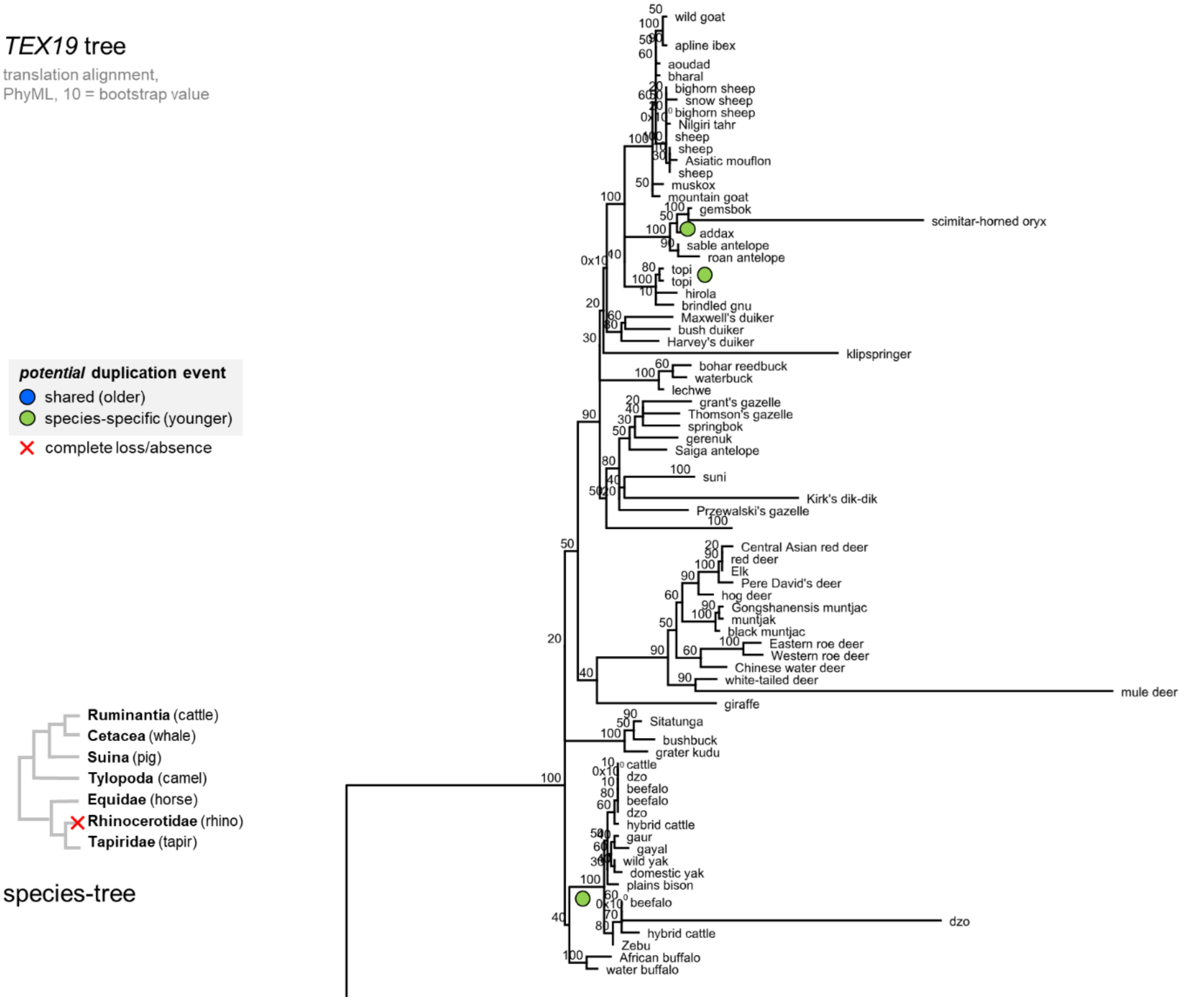


Figure 4. Phylogenetic tree showing *TEX19* from Rodents and Lagomorphs (as root) mammals. Translation alignment, PhyML bootstrap = 10. No complete losses were found in these groups. There are many shared (ancient) and recent duplications. Rodents have the fastest-growing copy number of *TEX19* of all groups.

We found complete losses of *TEX19* in the genomes of Canidae (dogs, wolves, foxes, coyotes) and Megabats (Pteropodinae). We also found other isolated species with no trace of *TEX19*: Afrotherian groups Macroscelidea (shrews) and Tubulidentata (aardvark), Xenarthran groups Vermilingua (anteater) and Folivora (sloths), Perissodactyla group Rinocerotidae (rhino), and Eulipotyphla group Talpidae (mole, shrew moles, desmans). How do we know that the absence of *TEX19* is not a consequence of limited sampling and not a loss? In theory, and because some of these groups are not comprehensively studied at the genomic level, it is possible that the absence of *TEX19* is a factor of limited sampling and data access. For both Chiroptera and Canidae, I have done an extensive search using not only *TEX19* from other bats but also from outgroups. None of these genomes have but fragmented traces of low to no similarity to *TEX19* from microbats: *Rousettus aegyptiacus* (Egyptian rousette), *Pteropus alecto* (black flying fox), *Pteropus vampyrus* (large flying fox), *Pteropus giganteus* (Indian flying fox), *Pteropus pselaphon* (Bonin flying fox), *Cynopterus brachyotis* (lesser short-nosed fruit bat), *Eidolon helvum* (straw-colored fruit bat), *Macroglossus sobrinus* (Long-tongued fruit bat), and *Rousettus leschenaultii* (Leschenault's rousette).



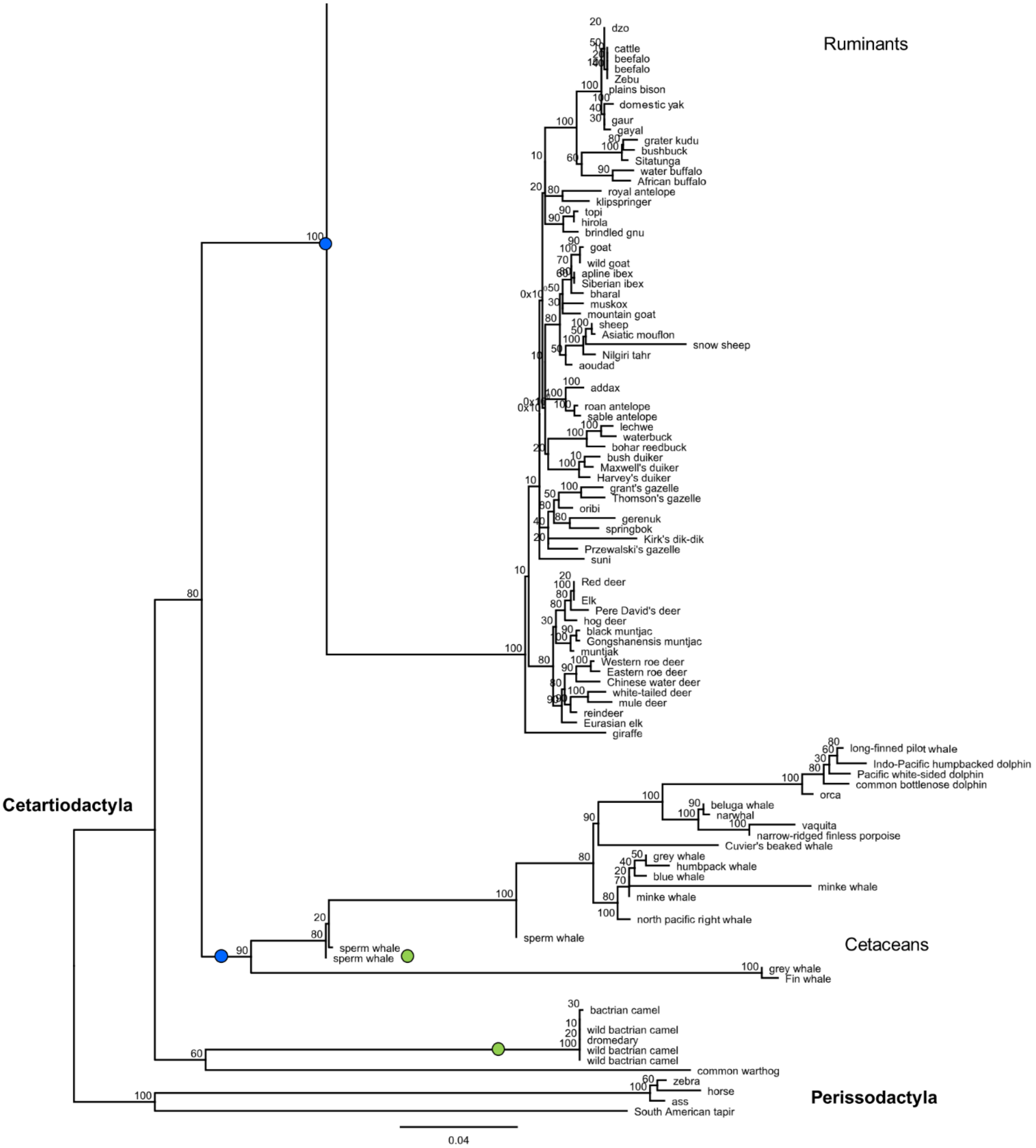


Figure 5. Phylogenetic tree showing *TEX19* in Cetratiodactyla and Perissodactyla (as root) mammals. Translation alignment, PhyML bootstrap = 10. No complete losses were found, except for in Rhinocerotidae (rhino). In this case, the *TEX19* open reading frame is fragmented. In addition, there is one shared (ancient) and several (recent) duplications of *TEX19*.

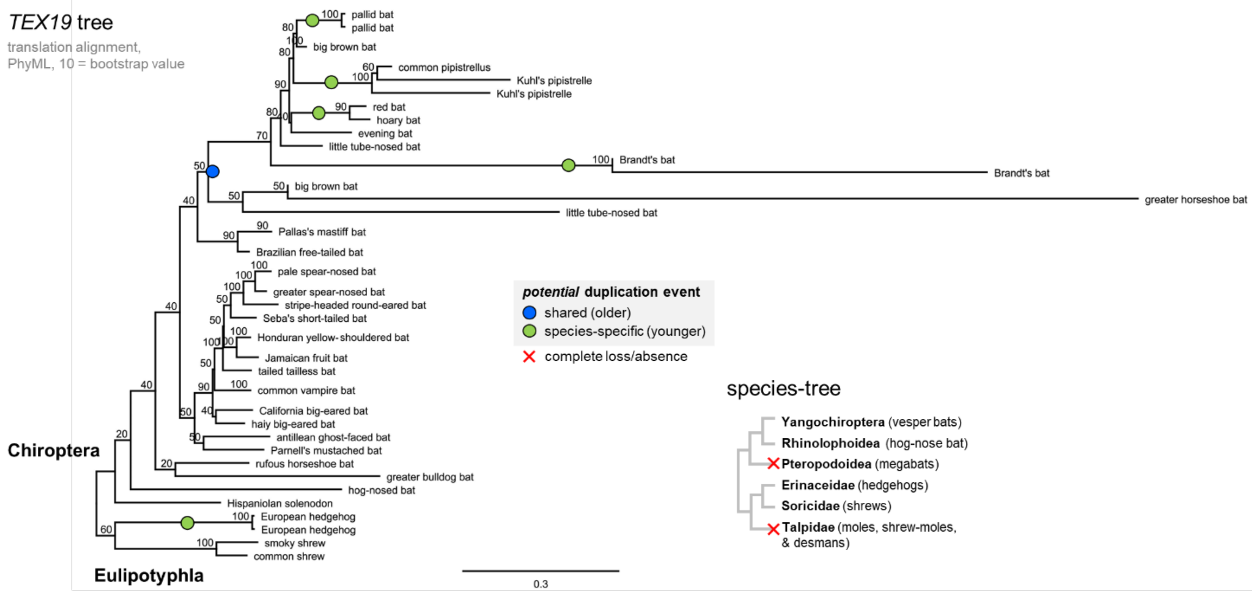
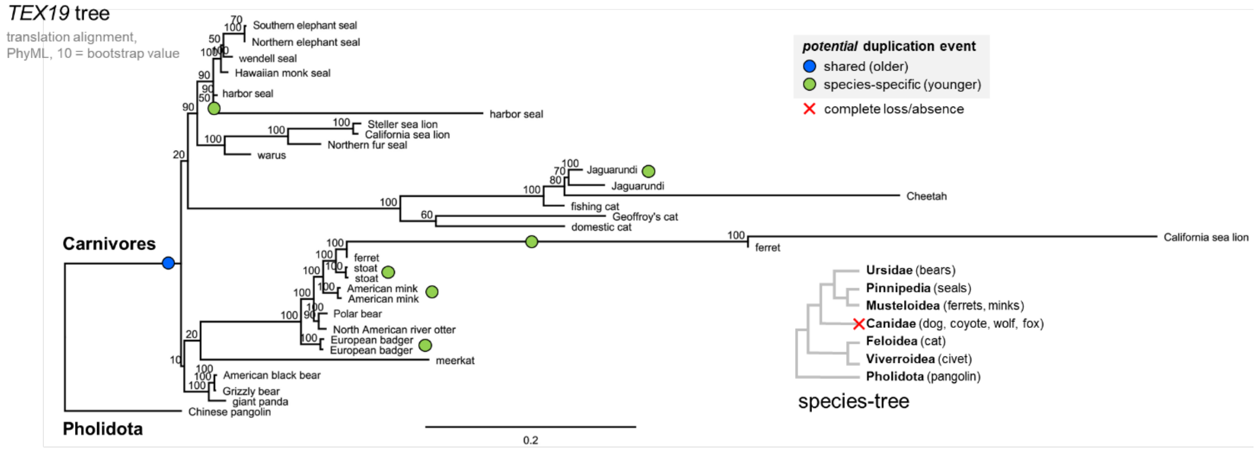


Figure 6. Phylogenetic tree of *TEX19* showing Chiroptera and Eulipotyphla (as root). Translation alignment, PhyML bootstrap = 10. Complete losses were found in the Chiroptera group Pteropodinae (Megabats) and in the Eulipotyphla group of Talpidae. There is one ancient *TEX19* duplication present in Yangchiroptera (vesper bats) and Rhinolophoidea (hog-nose bats). And a few (recent) duplications of *TEX19* specific to species in Yangochiroptera.

In the case of Canidae, we have even more genomes that lack *TEX19* despite having nearby genes conserved. For other local complete losses, like in Rhinocerotidea (rhinos), I found fragments of shortened open reading frames (coding for less than a hundred amino acids), indicating the possibility of alternative splicing isoforms of *TEX19*. This also happens in microbats and feliformes, both of which are directly related to groups that have lost their *TEX19* from their genomes. This finding is more a description of a pattern than a hypothesis based on preliminary data.

Lastly, orthologs of *TEX19* are present only in placental mammals, not in Monotremes or Marsupials. The current understanding of *TEX19*'s origin is still unclear. We think that *TEX19* was born in the last common ancestors of placentals. There are no hits when doing remote homology searches (except for genomes contaminated with mammalian

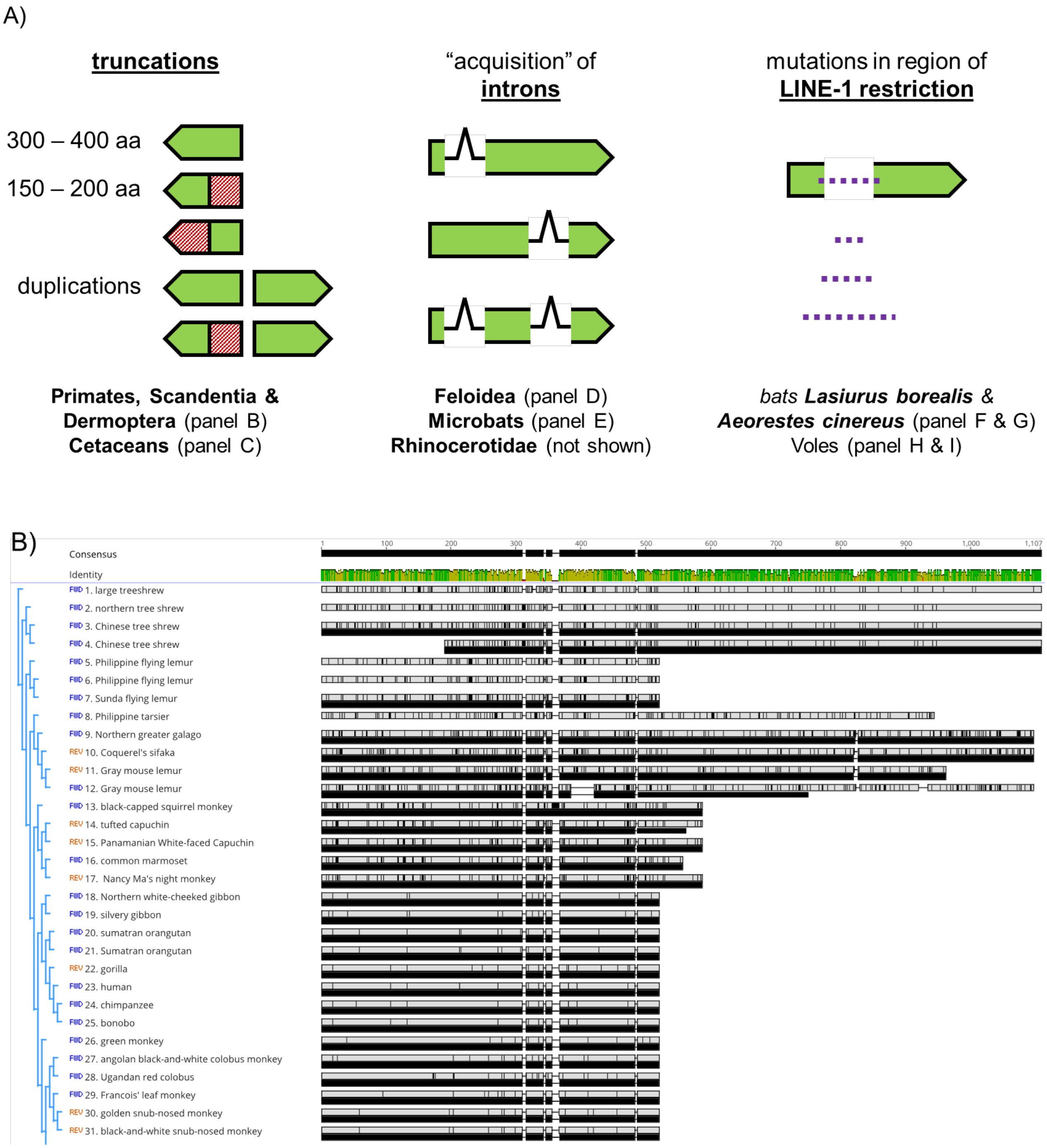
DNA, i.e., mouse). At the end of this chapter, I expand more on the origin of *TEX19* and the challenges to finding an answer.



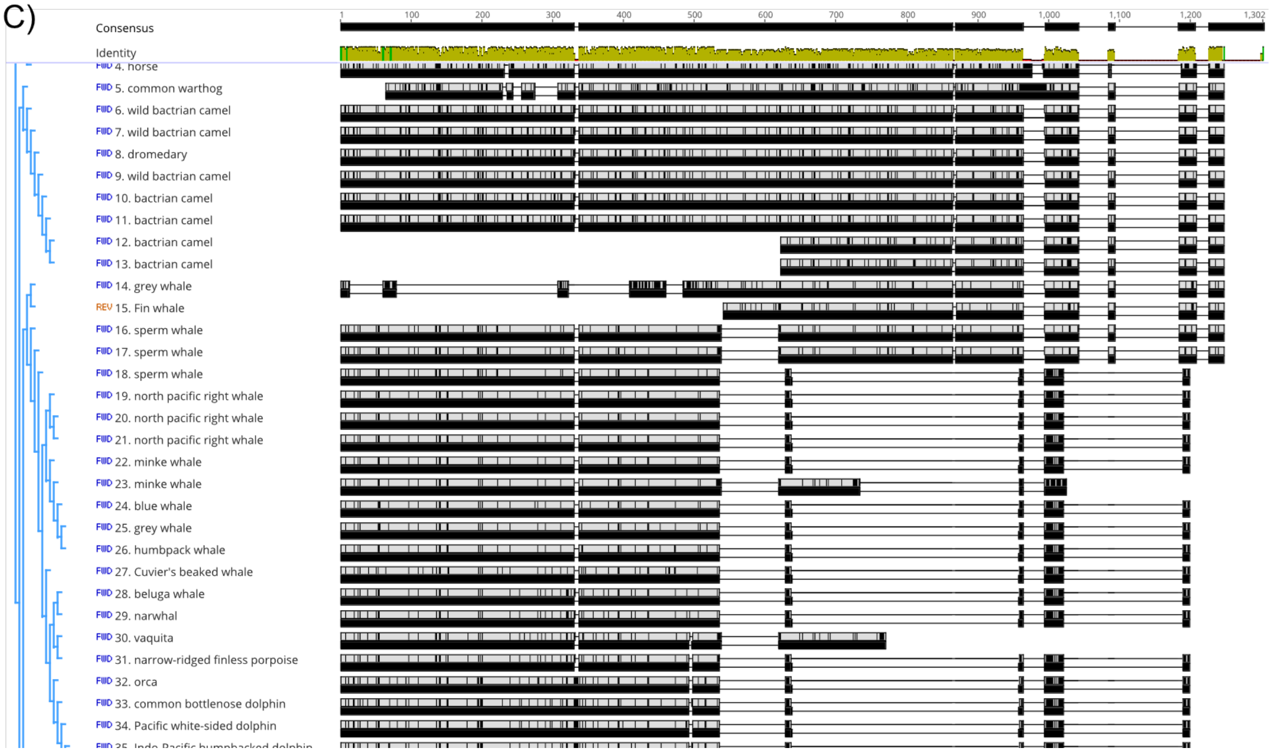
ure 7. Phylogenetic tree of *TEX19* showing Carnivores and Phalloidota (as root). Translation alignment, PhyML bootstrap = 10. Complete losses were found in Canidae (dogs, coyotes, wolves, and foxes). This is the major loss of TEX19 in all mammals. There is one ancient TEX19 duplication located in different clades for *Zalophus californianus* (sea lion), yet many recent duplications are specific to Musteloidea (ferrets and minks).

Structural Variants: truncations, introns, and repetitive loops

We found two (2) main gene lengths: human TEX19 with 164 amino acids (aa) aligns with the beginning half of the mouse Tex19.1 and Tex19.2 with 317 aa and 349 aa, respectively. There are species in most groups with a short (likely half) version of TEX19, but next, I mention the group where I detect this pattern consistently. This shorter form of TEX19 is present in other primates (Haplorinni) also with 164 – 186 aa, compared to TEX19 of Scandentia and Dermoptera with 309 – 362 aa (Fig 8B). Both elephants in Afrotherians have a short TEX19, 129 aa compared to 359 aa. Similarly, TEX19 in Cetaceans is mostly found as a short version with 192 aa and *Physeter catodon* (sperm whales) also has the TEX19 long version containing 351 aa and (Fig 8C).

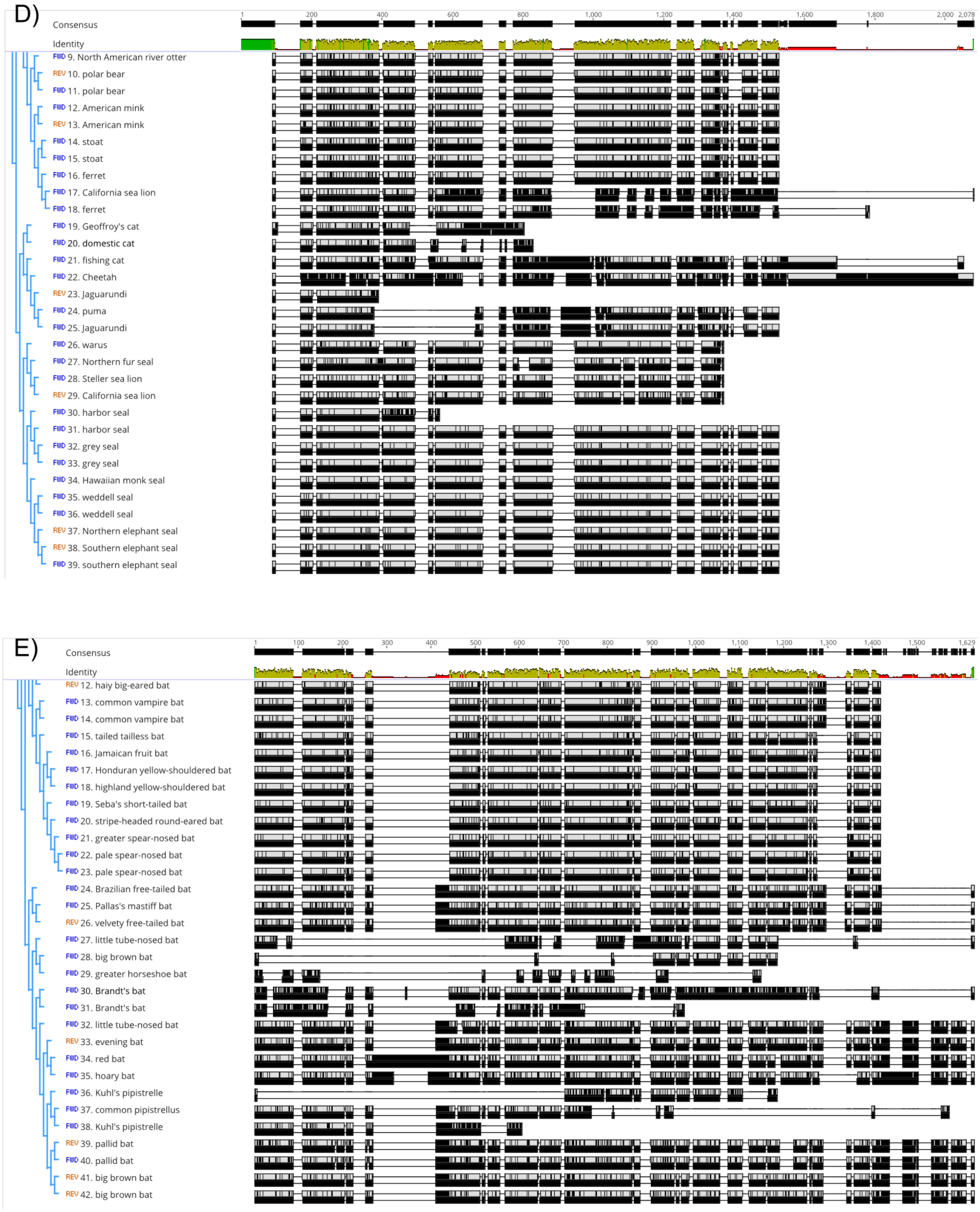

A)
truncations
300 – 400 aa
150 – 200 aa
duplications
"acquisition" of introns
mutations in region of LINE-1 restriction
Primates, Scandentia & Dermoptera (panel B)
Cetaceans (panel C)
Feloidea (panel D)
Microbats (panel E)
Rhinocerotidae (not shown)
bats Lasiurus borealis & Aeorestes cinereus (panel F & G)
Voles (panel H & I)
B)
Consensus
Identity
1
100
200
300
400
500
600
700
800
900
1,000
1,107
FWD 1. large treeshrew
FWD 2. northern tree shrew
FWD 3. Chinese tree shrew
FWD 4. Chinese tree shrew
FWD 5. Philippine flying lemur
FWD 6. Philippine flying lemur
FWD 7. Sunda flying lemur
FWD 8. Philippine tarsier
FWD 9. Northern greater galago
REV 10. Coquerel's sifaka
REV 11. Gray mouse lemur
FWD 12. Gray mouse lemur
FWD 13. black-capped squirrel monkey
REV 14. tufted capuchin
REV 15. Panamanian White-faced Capuchin
FWD 16. common marmoset
REV 17. Nancy Ma's night monkey
FWD 18. Northern white-cheeked gibbon
FWD 19. silvery gibbon
FWD 20. sumatran orangutan
FWD 21. Sumatran orangutan
REV 22. gorilla
FWD 23. human
FWD 24. chimpanzee
FWD 25. bonobo
FWD 26. green monkey
FWD 27. angolan black-and-white colobus monkey
FWD 28. Ugandan red colobus
FWD 29. Francois' leaf monkey
REV 30. golden snub-nosed monkey
REV 31. black-and-white snub-nosed monkey

C)

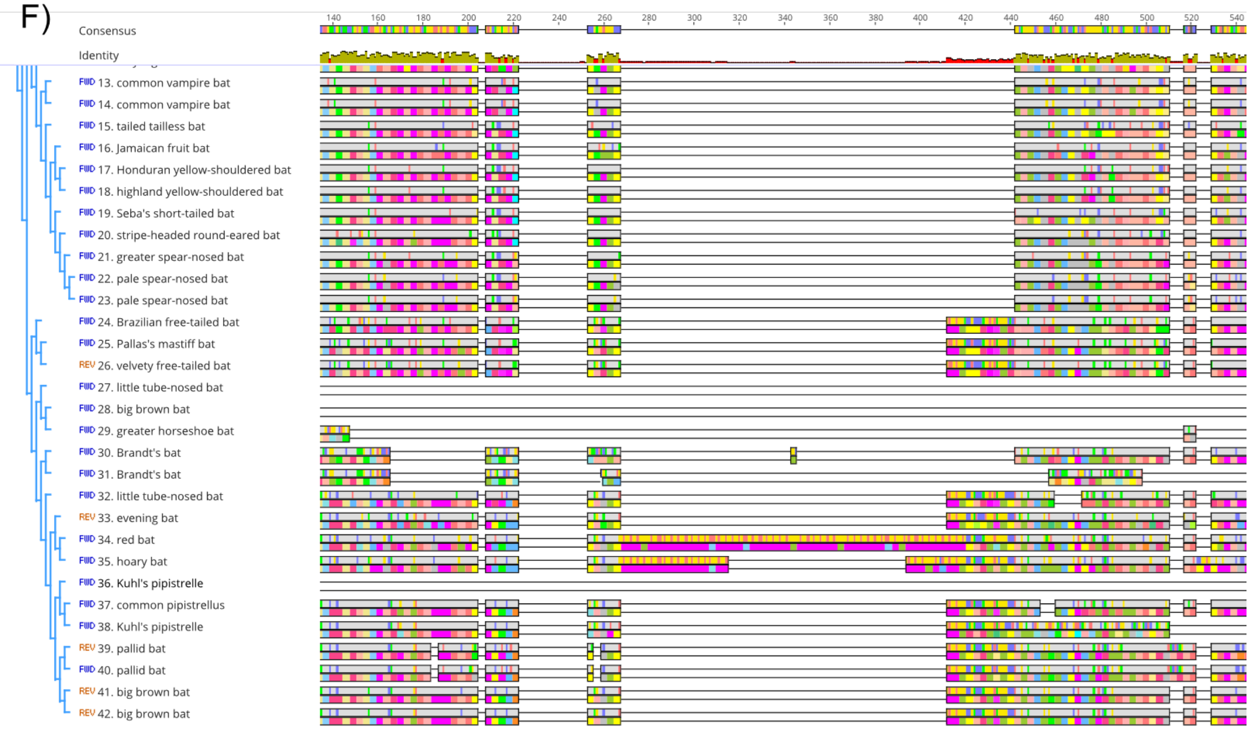


We also found TEX19s with only the C-terminal half of in two (2) different species in Cetartiodactyla (Fig 8C), specifically *Camelus bactrianus* (camels) and *Alces alces* (fin whales). Rodents and Ruminants conserve long TEX19 for duplications. Whereas *Physeter catodon* (sperm whales) have duplications composed of a mixture of short and long TEX19 (not shown).

The second main structural variant that I identified was the presence of introns unique to two groups, Felidae and Microbats. *TEX19* exists mostly intron-less in all other groups. Still, I was able to find the coding sequences, and we annotated isoforms that align in pieces to the closest *TEX19* in carnivores (Fig 8D). For Felidae and Chiroptera, all TEX19s analyzed were found in unique locations. These isoforms are truncated versions of TEX19. Similarly, I found introns in *TEX19* of Chiroptera (Fig 8E). How do these introns end up in between TEX19?

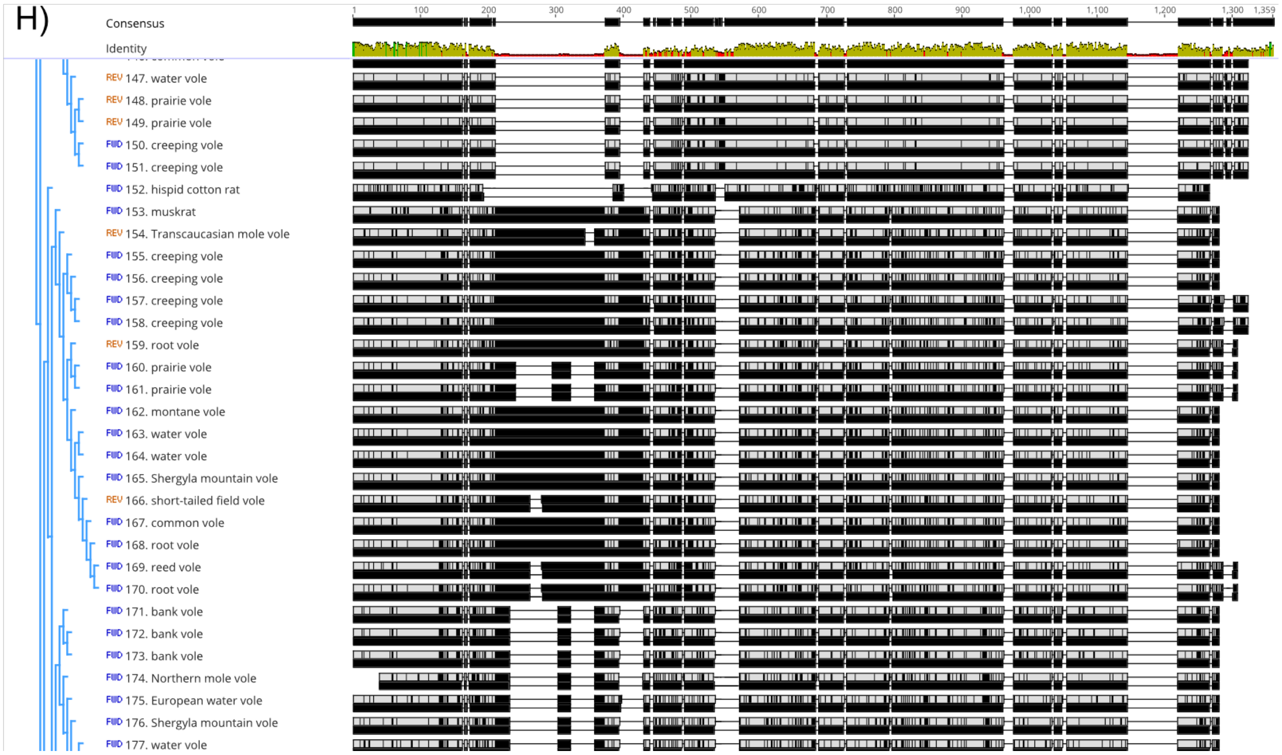

D)
Consensus
Identity
1
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
2,078
FWD 9. North American river otter
REV 10. polar bear
FWD 11. polar bear
FWD 12. American mink
REV 13. American mink
FWD 14. stoat
FWD 15. stoat
FWD 16. ferret
FWD 17. California sea lion
FWD 18. ferret
FWD 19. Geoffroy's cat
FWD 20. domestic cat
FWD 21. fishing cat
FWD 22. Cheetah
REV 23. Jaguarundi
FWD 24. puma
FWD 25. Jaguarundi
FWD 26. warus
FWD 27. Northern fur seal
FWD 28. Steller sea lion
REV 29. California sea lion
FWD 30. harbor seal
FWD 31. harbor seal
FWD 32. grey seal
FWD 33. grey seal
FWD 34. Hawaiian monk seal
FWD 35. weddell seal
FWD 36. weddell seal
REV 37. Northern elephant seal
REV 38. Southern elephant seal
FWD 39. southern elephant seal
E)
Consensus
Identity
1
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,300
1,400
1,500
1,629
REV 12. haiy big-eared bat
FWD 13. common vampire bat
FWD 14. common vampire bat
FWD 15. tailed tailless bat
FWD 16. Jamaican fruit bat
FWD 17. Honduran yellow-shouldered bat
FWD 18. highland yellow-shouldered bat
FWD 19. Seba's short-tailed bat
FWD 20. stripe-headed round-eared bat
FWD 21. greater spear-nosed bat
FWD 22. pale spear-nosed bat
FWD 23. pale spear-nosed bat
FWD 24. Brazilian free-tailed bat
FWD 25. Pallas's mastiff bat
REV 26. velvety free-tailed bat
FWD 27. little tube-nosed bat
FWD 28. big brown bat
FWD 29. greater horseshoe bat
FWD 30. Brandt's bat
FWD 31. Brandt's bat
FWD 32. little tube-nosed bat
REV 33. evening bat
FWD 34. red bat
FWD 35. hoary bat
FWD 36. Kuhl's pipistrelle
FWD 37. common pipistrellus
FWD 38. Kuhl's pipistrelle
REV 39. pallid bat
FWD 40. pallid bat
REV 41. big brown bat
REV 42. big brown bat

F)



G)

Acquisition of introns, specifically at the beginning of the gene, right where the TEX19 is thought to interact with LINE-1s, is of particular interest because that mutagenic event could directly alter the state of genetic conflict between TEX19 and LINE-1 ORF1p. In fact, in *Lasiurus borealis* (red bat) and *Aeorestes cinereus* (hoary bat), we observe the last structural variant of notice found, a segment of repetitive nature that encodes mostly for a poly glutamic acid (Glu, E) amino acid stretch interrupted by a few glycines (G) and lysines (K) (Fig 8F). This Glu repetitive segment is fifty-one (51) aa long for *Lasiurus borealis* and twenty-four (24) aa long for *Aeorestes cinereus* (Fig 8G). Intrerestly, the other group of mammals that also contains such a structural variant are the voles, known for harboring the highest amount of *TEX19* paralogs in all mammals (Fig 8H).



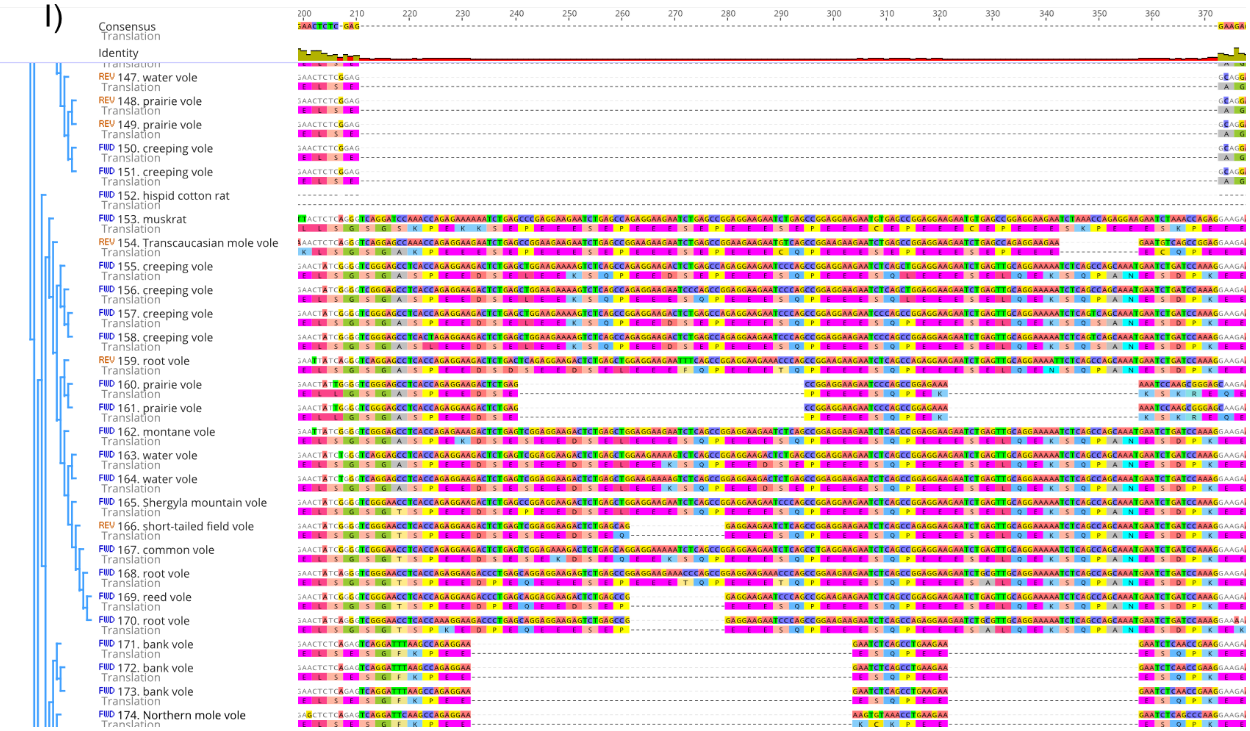


Figure 8. Structural Variants of TEX19 in mammals. A) Cartoon diagram of the three main structural variants of *TEX19*. B) Aligment (of phylogenetic tree) of Primate, Scandentia and Dermoptera *TEX19*. C) Aligment (of phylogenetic tree) of Cetartiodactyla, zooming in to Tylopoda and Cetaceans TEX19. D) Aligment (of phylogenetic tree) of Carnivores, zooming in to Felidae *TEX19*. E) Aligment (of phylogenetic tree) of Chiroptera, showing Microbats *TEX19*. F&G) Zoom in into Chiroptera showing beginning half of *TEX19*. H&I) Zoom in into Rodent, Voles showing beginning half of *TEX19*.

TEX19 versus the content of TEs and LINE-1s in mammals

What could *TEX19* duplication or complete loss mean for regulation of TEs? We know TEX19 targets LINE-1s, therefore gaining more copies or losing it could make the gene subject to strong selection. Of all groups of mammals that have completely lost *TEX19,* only one group, Megabats, has been shown to have also lost LINE-1 activity (*221*, *222*). Other groups like Rhinocerotidae and Felidae, where I found potentially dysfunctional isoforms of TEX19, have also lost LINE-1 activity (no recent expansions) in their genomes(*223*). However, this correlation does not hold for Canidae, where *TEX19* cannot be found but LINE-1s are still expanding. The relationship between hosts' TEX19 and their respective LINE-1s (and TEs in general) might differ depending on what other regulatory mechanisms are also restricting LINE-1 retrotransposition activity.

Current and Future Work

More on the origin of TEX19

Identifying so many events of gene duplication in *TEX19* evolution ignited the question of how many TEX19 genes there were in the common ancestor of placental mammals around 100 Mya ago(*224*, *225*).. Did the TEX19 genomic journey start as a duo? Even when adopting this perspective, many recent duplications and complete losses shape the evolution of this gene. I used the extensive number of TEX19s and their sequences to search for traces of this gene outside of placental mammals, and I did not find evidence to propose a credible origin story. The fact that the gene is mostly found intron-less could suggest an RT-mediated duplication event (from an unknown source gene), like via LINE-1 retrotransposition of host genes, endogenous retrovirus activity, or retroviral infection. Alternatively, TEX19 could have formed *de novo* from existing non-coding sequences in the placental mammal's last common ancestor, or it could have been horizontally transferred from an unknown source genome.

Functional and Molecular Evolution of TEX19 in Rodents

More work is being done to understand the signatures of selection and molecular evolution in TEX19. Structural variants like truncations, insertions, and introns have been a technical challenge to work around for methods detecting selection. However, we do see some homologous positions with high dN/dS ratios that could indicate positive selection at specific positions in TEX19. In addition, the rate of structural variation is incredibly high and likely reflects recurrent selection for functional variation in the *TEX19* gene.

In addition, we plan to test the natural variation of TEX19 changes across mammals against the most recent expansion of human LINE-1s described in Chapter 2. We think that structural and nonsynonymous changes found in mammalian TEX19 could be a result of each species' genetic conflict against LINE-1 retrotransposition and could help us understand better TEX19-evasive phenotypes in humans.

Discussion

The molecular evolution of the *TEX19* gene family in mammalian genomes is characterized by multiple duplication events, high sequence divergence, and complete losses, all suggesting this gene has evolved rapidly. Previous reports have suggested that *TEX19* evolution is correlated with the evolution of the nearby immune gene *SECMT1* because of the duplication of both genes in the mice locus (*226*). However, in other rodents and other mammals, TEX19 has undergone exclusive duplication events that leave the *SECTM1* copy number intact. In ruminants, specifically hybrid species of cattle, *SECTM1* has increased in copy number up to seven (7) copies without any change to the nearby *TEX19* copy number. Synteny and gene orientation in *TEX19*-containing loci across mammals suggest that *TEX19* has experienced multiple independent multiplication events. Ancient duplications are shared among different species of the same group, and I observed they also cluster together in phylogenetic clades, whereas species-specific duplications are recent.

We found that *TEX19* increased in copy number in the same groups while it was also lost/deleted from the genome (Chiroptera, Carnivores, Perissodactyla). Carnivores is an example of a group that has all copy number variant forms of *TEX19*: single copy (bears), duplications (seals and leopards), and complete losses (dogs, foxes, wolves, coyotes). Is the development of the embryo and placenta, the healthy progression of spermatogenesis, and the fertility of gametes not seriously impaired for *TEX19*-deficient genomes? Losses of *TEX19* might not be expected since *TEX19*-deficient mice suffer from serious pathologies like reduction of perinatal survival (*211*) and increased activity of endogenous retroelements like ERVs and LINE-1s(*142*).

Human and mouse TEX19 have two known functions: restriction of LINE-1 retrotransposons (mediated by the N-terminus) and TE regulation via recruitment of piRNAs and PIWI-interacting proteins (mediated by the C-terminus). Most truncations of *TEX19* are a short version consisting of only the beginning half of mouse *Tex19.1* and *Tex19.2*. Since this region is essential for LINE-1 restriction in vitro, those mammals with short TEX19 would still be able to restrict LINE-1s. However, we noticed that the domain responsible for piRNA silencing of TEs changed due to the deletion of this region. Therefore, mammals with a single copy of the short version of *TEX19* (humans) might not be as potent regulators of TEs because of this mutation. This short version also exists in whales (similar to primate TEX19 in length and domains, but formed via an independent truncation event), but some also have the long version of *TEX19* and a second copy. There are copy number gains of *TEX19* in humans, but they are the result of segmental and whole chromosome duplications, not local duplications, as we see across mammals.

Another aspect that may have made it simpler for duplications of *TEX19* to be useful is the lack of introns in the *TEX19* gene of most species. This also points to *de novo* gene birth of TEX19 or TE-mediated birth of the gene. In fact, species where I observed introns in *TEX19* also have immediately related species that have completely lost *TEX19.* For example, in the genome of Microbats (Vesper bats), it was particularly challenging to find *TEX19* because of the presence of introns and the shortening of isoforms of *TEX19.* This is interesting because we also found *TEX19* interrupted by introns in Felidae, which, like Microbats to Megabats, are directly related to a group that has lost *TEX19*, Canidae. I think that the acquisition of introns that results in the fragmentation of the regions with protein activity could help drive the genomic breakdown of *TEX19* in those genomes. In fact, we see that the coding sequence of *TEX19* in Rhinocerotidae is fragmented like the previous examples, and it is one of the few mammals that have also lost LINE-1 activity. We think that the loss of a parasite terminates the coevolution dynamic, and the immune gene is not needed. Megabats have lost LINE-1 activity, and we found that there is no *TEX19* present in their genomes. Yet, *TEX19* is only one of a collection of immune genes that work in concert to restrict LINE-1s. In other species like Canidae, *TEX19* has been lost, but LINE-1s continue to replicate and expand in their genomes. Presumably, a number of other hosts continue to be arrested into a similar coevolution dynamic.

Our work suggests the interaction of LINE-1 and TEX19 has been a recurrent interface for the ancient and ongoing evolutionary battle between LINE-1s and the restriction factors of the mammalian immune system. These evolutionary patterns underscore the importance of TEX19 in mammalian innate immunity and suggest selective pressures acting to maintain and diversify its functions in some mammals. It has been lost in other mammals, presumably with other genes playing a redundant function or via TEX19 function becoming obsolete.

Materials and Methods

TEX19 ortholog/paralog identification in mammalian genomes

I searched for mammalian genomes in NCBI Genome Datasets (https://www.ncbi.nlm.nih.gov/data-hub/genome/)and, used nucleotide BLAST (blastn and megablast optimization algorithms) to search for *TEX19*. There are well-annotated genomes in each group of mammals, and using their *TEX19* made it easier to find other *TEX19*s in genomes without annotations. We filtered for reference genomes assembled at the scaffold level, then collected TEX19 coding (nt) and protein (aa) sequences in Geneious Prime ® for further genomic analysis. For reference genomes, we made note of the nearby genes (synteny) *OGFOD3, UTS2R, SECMT1*, and *CD7,* and their orientation in relation to *TEX19*. There are genomes where it is hard to find *TEX19* because of the high sequence divergence, so I used the closest neighboring *TEX19* to blast.

Detection of independent gene duplication events and mechanisms

Gene orientation and, in some cases, the copy number variation of nearby genes helped gain more resolution about *TEX19* duplication events. I collected copies of *TEX19* in a reference genome with different coordinates to avoid exact duplicates. Then, I recorded the content of *TEX19* and the orientation of the gene in relation to the rest of the synteny (annotated in Fig 1). More work needs to be done to understand the mechanisms of duplication, and one way we started approaching the question was by comparing the TEX19-containing locus between species, using dot plots as a method for visualization.

Genetic comparisons and phylogenetic analyses

I extracted/collected coding sequences for *TEX19* in each group. Outgroups were chosen based on the field's most recent understanding and use of species phylogenetic relationships of mammals(*227*). Then, I ran a translation alignment and built a maximum likelihood phylogenetic tree with10 bootstraps (we plan to increase this number to >100 for publication) using the HKY85 substitution model(*228*).

Alignments and identification of structural variants

Phylogenetic trees and alignments were used to search for structural differences in gene length and conservation of characterized domains. I focused on the insertions or deletions of segments that disrupt the long (mouse) or short (human) versions of *TEX19*.

Remote homology search for *TEX19* origin

I ran preliminary searches using protein sequences of one species from each group of mammals (i.e., mouse, human, elephant, cattle, etc.). We used remote blast DELTA-BLAST and HMMER web server (http://hmmer.org/).

Acknowledgments

This work was supported by grants from the National Institutes of Health (R35GM142733 to RNM and T32GM007270 to RPDV). I thank the alumni and members of the McLaughlin lab for technical advice and critical comments on this manuscript.

References

1. D. H. Janzen, WHEN IS IT COEVOLUTION? *Evolution* **34**, 611–612 (1980).
2. S. W. Lockless, R. Ranganathan, Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
3. M. E. Salazar, M. T. Laub, Temporal and evolutionary dynamics of two-component signaling pathways. *Curr Opin Microbiol* **24**, 7–14 (2015).
4. T. Lösekann, A. Robador, H. Niemann, K. Knittel, A. Boetius, N. Dubilier, Endosymbioses between bacteria and deep-sea siboglinid tubeworms from an Arctic Cold Seep (Haakon Mosby Mud Volcano, Barents Sea). *Environ Microbiol* **10**, 3237–3254 (2008).
5. E. Barreto, M. M. A. Boehm, E. Ogutcen, S. Abrahamczyk, M. Kessler, J. Bascompte, A. S. Dellinger, C. Bello, D. M. Dehling, F. Duchenne, M. Kaehler, L. P. Lagomarsino, L. G. Lohmann, M. A. Maglianesi, H. Morlon, N. Muchhala, J. F. Ornelas, M. Perret, N. R. Salinas, S. D. Smith, J. C. Vamosi, I. G. Varassin, C. H. Graham, Macroevolution of the plant-hummingbird pollination system. *Biol Rev Camb Philos Soc*, doi: 10.1111/BRV.13094 (2024).
6. B. Koskella, M. A. Brockhurst, Bacteria–phage coevolution

as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* **38**, 916 (2014).

7. J. Marchi, M. Lässig, A. M. Walczak, T. Mora, Antigenic waves of virus-immune coevolution. *Proc Natl Acad Sci U S A* **118** (2021).
8. A. Hayward, M. Grabherr, P. Jern, Broad-scale phylogenomics provides insights into retrovirus - Host evolution. *Proc Natl Acad Sci U S A* **110**, 20146–20151 (2013).
9. T. C. De La Peña, E. Fedorova, J. J. Pueyo, M. Mercedes Lucas, The Symbiosome: Legume and Rhizobia Co-evolution toward a Nitrogen-Fixing Organelle? *Front Plant Sci* **8**, 2229 (2017).
10. C. A. Desjardins, D. E. Gundersen-Rindal, J. B. Hostetler, L. J. Tallon, D. W. Fadrosh, R. W. Fuester, M. J. Pedroni, B. J. Haas, M. C. Schatz, K. M. Jones, J. Crabtree, H. Forberger, V. Nene, Comparative genomics of mutualistic viruses of Glyptapanteles parasitic wasps. *Genome Biol* **9**, R183 (2008).
11. J. Gauthier, H. Boulain, J. J. F. A. van Vugt, L. Baudry, E. Persyn, J. M. Aury, B. Noel, A. Bretaudeau, F. Legeai, S. Warris, M. A. Chebbi, G. Dubreuil, B. Duvic, N. Kremer, P. Gayral, K. Musset, T. Josse, D. Bigot, C. Bressac,

S. Moreau, G. Periquet, M. Harry, N. Montagné, I. Boulogne, M. Sabeti-Azad, M. Maïbèche, T. Chertemps, F. Hilliou, D. Siaussat, J. Amselem, I. Luyten, C. Capdevielle-Dulac, K. Labadie, B. L. Merlin, V. Barbe, J. G. de Boer, M. Marbouty, F. L. Cônsoli, S. Dupas, A. Hua-Van, G. Le Goff, A. Bézier, E. Jacquin-Joly, J. B. Whitfield, L. E. M. Vet, H. M. Smid, L. Kaiser, R. Koszul, E. Huguet, E. A. Herniou, J. M. Drezen, Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Communications Biology 2021 4:1* **4**, 1–15 (2021).

12. M. Shahab, N. Shahab, Coevolution of the Human Host and Gut Microbiome: Metagenomics of Microbiota. *Cureus* **14** (2022).

13. H. Ding, H. Yeo, N. Puniamoorthy, Wolbachia infection in wild mosquitoes (Diptera: Culicidae): implications for transmission modes and host-endosymbiont associations in Singapore. *Parasites & Vectors 2020 13:1* **13**, 1–16 (2020).

14. N. M. Guruprasad, S. K. Jalali, H. P. Puttaraju, Wolbachia-a foe for mosquitoes. *Asian Pac J Trop Dis* **4**, 78 (2014).

15. A. Dupressoir, C. Lavialle, T. Heidmann, From ancestral infectious retroviruses to bona

fide cellular genes: role of the captured syncytins in placentation. *Placenta* **33**, 663–671 (2012).

16. M. Sha, X. Lee, X. ping Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, J. C. Keith, J. M. McCoy, Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature 2000 403:6771* **403**, 785–789 (2000).

17. S. Mi, X. Lee, X. Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, Jr. Keith J. C., J. M. McCoy, Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).

18. G. Chomicki, S. S. Renner, Partner abundance controls mutualism stability and the pace of morphological change over geologic time. *Proc Natl Acad Sci U S A* **114**, 3951–3956 (2017).

19. A. Fraile, F. García-Arenal, The coevolution of plants and viruses: resistance and pathogenicity. *Adv Virus Res* **76**, 1–32 (2010).

20. H. Brüssow, C. Canchaya, W.-D. Hardt, Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and*

Molecular Biology Reviews **68**, 560 (2004).

21. M. A. B. Shabbir, H. Hao, M. Z. Shabbir, Q. Wu, A. Sattar, Z. Yuan, Bacteria vs. bacteriophages: Parallel evolution of immune arsenals. *Front Microbiol* **7**, 213809 (2016).

22. K. Voskarides, E. Christaki, G. K. Nikolopoulos, Influenza Virus—Host Co-evolution. A Predator-Prey Relationship? *Front Immunol* **9** (2018).

23. A. R. Khairullah, S. C. Kurniawan, A. Hasib, O. S. M. Silaen, A. Widodo, M. H. Effendi, S. C. Ramandinianto, I. B. Moses, K. H. P. Riwu, S. M. Yanestria, Tracking lethal threat: in-depth review of rabies. *Open Vet J* **13**, 1385–1399 (2023).

24. S. T. Jacob, I. Crozier, W. A. Fischer, A. Hewlett, C. S. Kraft, M. A. de La Vega, M. J. Soka, V. Wahl, A. Griffiths, L. Bollinger, J. H. Kuhn, Ebola virus disease. *Nature Reviews Disease Primers 2020 6:1* **6**, 1–31 (2020).

25. N. Kaur, J. Dabar, P. Bassi, Monkeypox: A re-emerging disease. *Indian J Pharmacol* **56**, 129–135 (2024).

26. E. A. Alrefai, R. T. Alhejaili, S. A. Haddad, Human Papillomavirus and Its Association With Cervical Cancer: A Review. *Cureus* **16** (2024).

27. A. Shimizu, I. Dohzono, M. Nakaji, D. A. Roff, D. G. Miller,

S. Osato, T. Yajima, S. Niitsu, N. Utsugi, T. Sugawara, J. Yoshimura, Fine-tuned Bee-Flower Coevolutionary State Hidden within Multiple Pollination Interactions. *Scientific Reports 2014 4:1* **4**, 1–9 (2014).

28. T. H. Coale, V. Loconte, K. A. Turk-Kubo, B. Vanslembrouck, W. K. E. Mak, S. Cheung, A. Ekman, J. H. Chen, K. Hagino, Y. Takano, T. Nishimura, M. Adachi, M. Le Gros, C. Larabell, J. P. Zehr, Nitrogen-fixing organelle in a marine alga. *Science* **384**, 217–222 (2024).

29. P. Bonfante, A. Genre, Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications 2010 1:1* **1**, 1–11 (2010).

30. K. Kawamura, S. Sekida, K. Nishitsuji, E. Shoguchi, K. Hisata, S. Fujiwara, N. Satoh, In vitro Symbiosis of Reef-Building Coral Cells With Photosynthetic Dinoflagellates. *Front Mar Sci* **8**, 706308 (2021).

31. M. D. Daugherty, H. S. Malik, Rules of Engagement: Molecular Insights from Host-Virus Arms Races. *Annu Rev Genet* **46**, 677–700 (2012).

32. K. Uriu, Y. Kosugi, J. Ito, K. Sato, The Battle between Retroviruses and APOBEC3 Genes: Its Past and Present. *Viruses* **13** (2021).

33. Y. Nakano, H. Aso, A. Soper, E. Yamada, M. Moriwaki, G. Juarez-Fernandez, Y. Koyanagi, K. Sato, A conflict of interest: The evolutionary arms race between mammalian APOBEC3 and lentiviral Vif. *Retrovirology* **14** (2017).

34. A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).

35. M. Miranda-Saksena, C. E. Denes, R. J. Diefenbach, A. L. Cunningham, Infection and Transport of Herpes Simplex Virus Type 1 in Neurons: Role of the Cytoskeleton. *Viruses* **10** (2018).

36. D. Lapa, A. R. Garbuglia, M. R. Capobianchi, P. Del Porto, Hepatitis C Virus Genetic Variability, Human Immune Response, and Genome Polymorphisms: Which Is the Interplay? *Cells* **8** (2019).

37. M. Wille, E. C. Holmes, The Ecology and Evolution of Influenza Viruses. *Cold Spring Harb Perspect Med* **10**, 1–19 (2020).

38. R. Wijngaard, G. Demidov, L. O’Gorman, J. Corominas-Galbany, B. Yaldiz, W. Steyaert, E. de Boer, L. E. L. M. Vissers, E. J. Kamsteeg, R. Pfundt, H. Swinkels, A. den Ouden, I. B. A. W. te Paske, R. M. de Voer, L. Faivre, A. S. Denommé-Pichon, Y.

Duffourd, A. Vitobello, M. Chevarin, V. Straub, A. Töpf, A. J. van der Kooi, F. Magrinelli, C. Rocca, M. G. Hanna, J. Vandrovcova, S. Ossowski, S. Laurie, C. Gilissen, Mobile element insertions in rare diseases: a comparative benchmark and reanalysis of 60,000 exome samples. *European Journal of Human Genetics 2023 32:2* **32**, 200–208 (2023).

39. K. H. Burns, Transposable elements in cancer. *Nature Reviews Cancer 2017 17:7* **17**, 415–424 (2017).

40. K. Kaer, M. Speek, Retroelements in human disease. *Gene* **518**, 231–241 (2013).

41. L. M. Payer, K. H. Burns, Transposable elements in human genetic disease. *Nat Rev Genet* **20**, 760–772 (2019).

42. H. B. Creighton, B. McClintock, A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. *Proc Natl Acad Sci U S A* **17**, 492 (1931).

43. B. McClintock, The Order of the Genes C, Sh and Wx in Zea Mays with Reference to a Cytologically Known Point in the Chromosome. *Proc Natl Acad Sci U S A* **17**, 485 (1931).

44. M. Dupeyron, T. Baril, C. Bass, A. Hayward, Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored

diversity of pogo-like elements. *Mob DNA* **11** (2020).

45. H. M. Robertson, The Tcl-mariner superfamily of transposons in animals. [Preprint] (1995). https://doi.org/10.1016/0022-1910(94)00082-R.

46. C. Cassier-Chauvat, M. Poncelet, F. Chauvat, Three insertion sequences from the cyanobacterium Synechocystis PCC6803 support the occurrence of horizontal DNA transfer among bacteria. *Gene* **195**, 257–266 (1997).

47. J. W. Jacobson, M. M. Medhora, D. L. Hartl, Molecular structure of a somatically unstable transposable element in Drosophila. *Proc Natl Acad Sci U S A* **83**, 8684–8688 (1986).

48. S. A. Narayanavari, S. S. Chilkunda, Z. Ivics, Z. Izsvák, Sleeping Beauty transposition: from biology to applications. *Crit Rev Biochem Mol Biol* **52**, 18–44 (2017).

49. H. Quesneville, Twenty years of transposable element analysis in the Arabidopsis thaliana genome. *Mob DNA* **11** (2020).

50. K. R. Oliver, J. A. McComb, W. K. Greene, Transposable elements: powerful contributors to angiosperm evolution and diversity.

Genome Biol Evol **5**, 1886–1901 (2013).

51. A. D. Hudson, R. Carpenter, E. S. Coen, Phenotypic effects of short-range and aberrant transposition in Antirrhinum majus. *Plant Mol Biol* **14**, 835–844 (1990).

52. W. McGinnis, A. W. Shermoen, S. K. Beckendorf, A transposable element inserted just 5' to a Drosophila glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75–84 (1983).

53. Hwei-gene Heidi Wang, M. J. Fraser, L. C. Cary, Transposon mutagenesis of baculoviruses: analysis of TFP3 lepidopteran transposon insertions at the FP locus of nuclear polyhedrosis viruses. *Gene* **81**, 97–108 (1989).

54. Q. Chen, W. Luo, R. A. Veach, A. B. Hickman, M. H. Wilson, F. Dyda, Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nat Commun* **11** (2020).

55. M. Dupeyron, K. S. Singh, C. Bass, A. Hayward, Evolution of Mutator transposable elements across eukaryotic diversity. *Mob DNA* **10** (2019).

56. N. Juretic, D. R. Hoen, M. L. Huynh, P. M. Harrison, T. E. Bureau, The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res* **15**, 1292–1297 (2005).

57. C. Feschotte, E. J. Pritham, DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331–368 (2007).

58. V. V Kapitonov, J. Jurka, Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8714–8719 (2001).

59. J. Thomas, E. J. Pritham, Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* **3** (2015).

60. C. Feschotte, S. R. Wessler, Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8923–8924 (2001).

61. M. Chandler, F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalian, B. Ton-Hoang, Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol.* **11**, 525–538 (2013).

62. I. Grabundzija, S. A. Messing, J. Thomas, R. L. Cosby, I. Bilic, C. Miskey, A. Gogol-Döring, V. Kapitonov, T. Diem, A. Dalda, J. Jurka, E. J. Pritham, F. Dyda, Z. Izsvák, Z. Ivics, A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **7**, 10716 (2016).

63. Z. Wang, G. Zhao, Q. Yang, L. Gao, C. Liu, Z. Ru, D. Wang, J. Jia, D. Cui, Helitron and

CACTA DNA transposons actively reshape the common wheat - AK58 genome. *Genomics* **114**, 110288 (2022).

64. D. J. Garfinkel, J. D. Boeke, G. R. Fink, Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42**, 507–517 (1985).

65. J. Mellor, M. H. Malim, K. Gull, M. F. Tuite, S. McCready, T. Dibbayawan, S. M. Kingsman, A. J. Kingsman, Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature* **318**, 583–586 (1985).

66. J. Clare, P. Farabaugh, Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2829–2833 (1985).

67. H. S. Malik, T. H. Eickbush, Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res* **11**, 1187–1197 (2001).

68. H. S. Malik, W. D. Burke, T. H. Eickbush, The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16**, 793–805 (1999).

69. M. Carr, D. Bensasson, C. M. Bergman, Evolutionary Genomics of Transposable Elements in Saccharomyces

cerevisiae. *PLoS One* **7** (2012).

70. J. M. Kim, S. Vanguri, J. D. Boeke, A. Gabriel, D. F. Voytas, Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res* **8**, 464–478 (1998).

71. C. Smukowski Heil, K. Patterson, A. S. M. Hickey, E. Alcantara, M. J. Dunham, Transposable Element Mobilization in Interspecific Yeast Hybrids. *Genome Biol Evol* **13** (2021).

72. A. Beauregard, M. J. Curcio, M. Belfort, The take and give between retrotransposable elements and their hosts. *Annu Rev Genet* **42**, 587 (2008).

73. P. Lesage, A. L. Todeschini, Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenet Genome Res* **110**, 70–90 (2005).

74. N. Friesen, A. Brandes, J. S. Heslop-Harrison, Diversity, Origin, and Distribution of Retrotransposons (gypsy and copia) in Conifers. *Mol Biol Evol* **18**, 1176–1188 (2001).

75. R. C. Paz, M. E. Kozaczek, H. G. Rosli, N. P. Andino, M. V. Sanchez-Puerta, Diversity, distribution and dynamics of full-length Copia and Gypsy LTR retroelements in

Solanum lycopersicum. *Genetica* **145**, 417–430 (2017).

76. M. Vukich, T. Giordani, L. Natali, A. Cavallini, Copia and Gypsy retrotransposons activity in sunflower (Helianthus annuus L.). *BMC Plant Biol* **9**, 1–12 (2009).

77. S. J. Zhang, L. Liu, R. Yang, X. Wang, Genome Size Evolution Mediated by Gypsy Retrotransposons in Brassicaceae. *Genomics Proteomics Bioinformatics* **18**, 321–332 (2020).

78. J. M. Park, G. M. Schneeweiss, H. Weiss-Schneeweiss, Diversity and evolution of Ty1-copia and Ty3-gypsy retroelements in the non-photosynthetic flowering plants Orobanche and Phelipanche (Orobanchaceae). *Gene* **387**, 75–86 (2007).

79. X. Zhang, Y. Qi, The Landscape of Copia and Gypsy Retrotransposon During Maize Domestication and Improvement. *Front Plant Sci* **10**, 475423 (2019).

80. J.-L. Blond, D. Lavillette, V. Cheynet, O. Bouton, G. Oriol, S. Chapel-Fernandes, B. Mandrand, F. Mallet, F.-L. Cosset, An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus

Receptor. *J Virol* **74**, 3321–3329 (2000).

81. G. Kassiotis, Endogenous retroviruses and the development of cancer. *J Immunol* **192**, 1343 (2014).
82. Y. Xiang, H. Liang, The Regulation and Functions of Endogenous Retrovirus in Embryo Development and Stem Cell Differentiation. *Stem Cells Int* **2021** (2021).
83. T. Kaneko-Ishino, F. Ishino, The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front Microbiol* **3**, 27133 (2012).
84. K. Imakawa, K. Kusama, T. Kaneko-Ishino, S. Nakagawa, K. Kitao, T. Miyazawa, F. Ishino, Endogenous Retroviruses and Placental Evolution, Development, and Diversity. *Cells* **11** (2022).
85. C. R. Beck, P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge, J. V. Moran, LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170 (2010).
86. M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41–48 (2003).
87. J. ME, G. R, J. PA, J. J, Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders.

Trends Genet **36**, 610–623 (2020).

88. A. A, D. T. I, V.-T. N, E. G. MR, S. I, Transposable elements in brain health and disease. *Ageing Res Rev* **64** (2020).

89. C. R. Beck, J. L. Garcia-Perez, R. M. Badge, J. V Moran, LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 187–215 (2011).

90. N. A. Suarez, A. Macia, A. R. Muotri, LINE-1 retrotransposons in healthy and diseased human brain. *Dev Neurobiol* **78**, 434–455 (2018).

91. B. Zhao, Q. Wu, A. Y. Ye, J. Guo, X. Zheng, X. Yang, L. Yan, Q.-R. Liu, T. M. Hyde, L. Wei, A. Y. Huang, Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLoS Genet* **15**, e1008043 (2019).

92. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J.

Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, others, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

93. S. L. Martin, The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J Biomed Biotechnol* **2006**, 45621 (2006).

94. L. Dai, J. LaCava, M. S. Taylor, J. D. Boeke, Expression and detection of LINE-1 ORF-encoded proteins. *Mob Genet*

Elements **4**, e29319–e29319 (2014).

95. E. T. Baldwin, T. van Eeuwen, D. Hoyos, A. Zalevsky, E. P. Tchesnokov, R. Sánchez, B. D. Miller, L. H. Di Stefano, F. X. Ruiz, M. Hancock, E. Işik, C. Mendez-Dorantes, T. Walpole, C. Nichols, P. Wan, K. Riento, R. Halls-Kass, M. Augustin, A. Lammens, A. Jestel, P. Upla, K. Xibinaku, S. Congreve, M. Hennink, K. B. Rogala, A. M. Schneider, J. E. Fairman, S. M. Christensen, B. Desrosiers, G. S. Bisacchi, O. L. Saunders, N. Hafeez, W. Miao, R. Kapeller, D. M. Zaller, A. Sali, O. Weichenrieder, K. H. Burns, M. Götte, M. P. Rout, E. Arnold, B. D. Greenbaum, D. L. Romero, J. LaCava, M. S. Taylor, Structures, functions and adaptations of the human LINE-1 ORF2 protein. *Nature* **626**, 194–206 (2024).

96. P. Mita, A. Wudzinska, X. Sun, J. Andrade, S. Nayak, D. J. Kahler, S. Badri, J. LaCava, B. Ueberheide, C. Y. Yun, D. Fenyö, J. D. Boeke, LINE-1 protein localization and functional dynamics during the cell cycle. *Elife* **7**, e30058 (2018).

97. R. J. Britten, D. E. Kohne, Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science (1979)* **161**, 529–540 (1968).

98. H. H. Kazazian, C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, S. E. Antonarakis, Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).

99. B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, Jr. Kazazian H. H., Isolation of an active human transposable element. *Science (1979)* **254**, 1805–1808 (1991).

100. B. Brouha, J. Schustak, R. M. Badge, S. Lutz-Prigge, A. H. Farley, J. V Moran, H. H. Kazazian Jr, Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280–5285 (2003).

101. N. T. Chuang, E. J. Gardner, D. M. Terry, J. Crabtree, A. A. Mahurkar, G. L. Rivell, C. C. Hong, J. A. Perry, S. E. Devine, Mutagenesis of human genomes by endogenous mobile elements on a population scale. *Genome Res.* **31**, 2225–2235 (2021).

102. R. C. Iskow, M. T. McCabe, R. E. Mills, S. Torene, W. S. Pittard, A. F. Neuwald, E. G. Van Meir, P. M. Vertino, S. E. Devine, Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, doi: 10.1016/j.cell.2010.05.020 (2010).

103. P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H.-Y. Fritz, M. K. Konkel, A. Malhotra, A. M. Stütz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E.-W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, 1000 Genomes Project Consortium, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, J. O. Korbel, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).

104. E. J. Gardner, V. K. Lam, D. N. Harris, N. T. Chuang, E. C. Scott, W. S. Pittard, R. E. Mills, 1000 Genomes Project Consortium, S. E. Devine,

The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res.* **27**, 1916–1929 (2017).

105. S. E. Devine, Emerging Opportunities to Study Mobile Element Insertions and Their Source Elements in an Expanding Universe of Sequenced Human Genomes. *Genes 2023, Vol. 14, Page 1923* **14**, 1923 (2023).

106. W.-W. Liao, M. Asri, J. Ebler, D. Doerr, M. Haukness, G. Hickey, S. Lu, J. K. Lucas, J. Monlong, H. J. Abel, S. Buonaiuto, X. H. Chang, H. Cheng, J. Chu, V. Colonna, J. M. Eizenga, X. Feng, C. Fischer, R. S. Fulton, S. Garg, C. Groza, A. Guarracino, W. T. Harvey, S. Heumos, K. Howe, M. Jain, T.-Y. Lu, C. Markello, F. J. Martin, M. W. Mitchell, K. M. Munson, M. N. Mwaniki, A. M. Novak, H. E. Olsen, T. Pesout, D. Porubsky, P. Prins, J. A. Sibbesen, J. Sirén, C. Tomlinson, F. Villani, M. R. Vollger, L. L. Antonacci-Fulton, G. Baid, C. A. Baker, A. Belyaeva, K. Billis, A. Carroll, P.-C. Chang, S. Cody, D. E. Cook, R. M. Cook-Deegan, O. E. Cornejo, M. Diekhans, P. Ebert, S. Fairley, O. Fedrigo, A. L. Felsenfeld, G. Formenti, A. Frankish, Y. Gao, N. A. Garrison, C. G. Giron, R. E. Green, L. Haggerty, K. Hoekzema, T. Hourlier, H. P. Ji, E. E. Kenny, B. A. Koenig, A. Kolesnikov, J. O. Korbel, J. Kordosky, S. Koren, H. Lee, A. P. Lewis, H.

Magalhães, S. Marco-Sola, P. Marijon, A. McCartney, J. McDaniel, J. Mountcastle, M. Nattestad, S. Nurk, N. D. Olson, A. B. Popejoy, D. Puiu, M. Rautiainen, A. A. Regier, A. Rhie, S. Sacco, A. D. Sanders, V. A. Schneider, B. I. Schultz, K. Shafin, M. W. Smith, H. J. Sofia, A. N. Abou Tayoun, F. Thibaud-Nissen, F. F. Tricomi, J. Wagner, B. Walenz, J. M. D. Wood, A. V Zimin, G. Bourque, M. J. P. Chaisson, P. Flicek, A. M. Phillippy, J. M. Zook, E. E. Eichler, D. Haussler, T. Wang, E. D. Jarvis, K. H. Miga, E. Garrison, T. Marschall, I. M. Hall, H. Li, B. Paten, A draft human pangenome reference. *Nature* **617**, 312–324 (2023).

107. S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A. V Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V Maduro, T. Marschall, A.

M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sović, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, The complete sequence of a human genome. *Science (1979)* **376**, 44–53 (2022).

108. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

109. L. Yang, G. A. Metzger, R. Padilla Del Valle, D. Delgadillo Rubalcaba, R. N. McLaughlin Jr, Evolutionary insights from profiling LINE-1 activity at allelic resolution in a single human genome. *EMBO J.* **43**, 112–131 (2024).

110. H. Khan, A. Smit, S. Boissinot, Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).

111. S. Boissinot, P. Chevret, A. V Furano, L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* **17**, 915–928 (2000).

112. A. F. Smit, G. Tóth, A. D. Riggs, J. Jurka, Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401–417 (1995).

113. A. F. Smit, G. Tóth, A. D. Riggs, J. Jurka, Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401–417 (1995).

114. F. H. Burton, D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell, C. A. Hutchison, Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J Mol Biol* **187**, 291–304 (1986).

115. J. Skowronski, T. G. Fanning, M. F. Singer, Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385–1397 (1988).

116. A. J. Betancourt, K. H.-C. Wei, Y. Huang, Y. C. G. Lee, Causes and Consequences of

Varying Transposable Element Activity: An Evolutionary Perspective. *Annu Rev Genomics Hum Genet*, doi: 10.1146/ANNUREV-GENOM-120822-105708 (2024).

117. R. L. Cosby, N. C. Chang, C. Feschotte, Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev* **33**, 1098–1116 (2019).
118. J. L. Goodier, Restricting retrotransposons: a review. *Mob DNA* **7**, 16 (2016).
119. X. G. Pappalardo, V. Barra, Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics Chromatin* **14**, 25 (2021).
120. P. Turelli, N. Castro-Diaz, F. Marzetta, A. Kapopoulou, C. Raclot, J. Duc, V. Tieng, S. Quenneville, D. Trono, Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements . *Genome Res.* **24**, 1260–1270 (2014).
121. J. A. Hackett, J. P. Reddington, C. E. Nestor, D. S. Dunican, M. R. Branco, J. Reichmann, W. Reik, M. A. Surani, I. R. Adams, R. R. Meehan, Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623–3632 (2012).

122. M. Bruno, M. Mahgoub, T. S. Macfarlan, The arms race between KRAB-zinc finger proteins and endogenous retroelements and its impact on mammals. *Annu. Rev. Genet.* **53**, 393–416 (2019).

123. F. M. J. Jacobs, D. Greenberg, N. Nguyen, M. Haeussler, A. D. Ewing, S. Katzman, B. Paten, S. R. Salama, D. Haussler, An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–245 (2014).

124. K. Zheng, J. Xiol, M. Reuter, S. Eckardt, N. L. Leu, K. J. McLaughlin, A. Stark, R. Sachidanandam, R. S. Pillai, P. J. Wang, Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc Natl Acad Sci U S A* **107**, 11841–11846 (2010).

125. M. DiGiacomo, S. Comazzetto, H. Saini, S. DeFazio, C. Carrieri, M. Morgan, L. Vasiliauskaite, V. Benes, A. J. Enright, D. O’Carroll, Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. *Mol Cell* **50**, 601–608 (2013).

126. J. L. Goodier, G. Pereira, L. E. Cheung, R. Rose, Jr. Kazazian H. H., The broad-

spectrum antiviral protein ZAP restricts human retrotransposition. *PLoS Genet* (2015).

127. J. B. Moldovan, J. V. Moran, The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. *PLoS Genet* **11** (2015).

128. K. Zhao, J. Du, X. Han, J. L. Goodier, P. Li, X. Zhou, W. Wei, S. L. Evans, L. Li, W. Zhang, L. E. Cheung, G. Wang, Jr. Kazazian H. H., X. F. Yu, Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep* **4**, 1108–1115 (2013).

129. S. Hu, J. Li, F. Xu, S. Mei, Y. Le Duff, L. Yin, X. Pang, S. Cen, Q. Jin, C. Liang, F. Guo, SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. *PLoS Genet* **11**, e1005367 (2015).

130. J. Castaneda, P. Genzor, G. W. van der Heijden, A. Sarkeshik, 3rd Yates J. R., N. T. Ingolia, A. Bortvin, Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *EMBO J* **33**, 1999–2019 (2014).

131. D. Pezic, S. A. Manakov, R. Sachidanandam, A. A. Aravin, piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3

mark in germ cells. *Genes Dev.* **28**, 1410–1428 (2014).

132. D. B. Stetson, J. S. Ko, T. Heidmann, R. Medzhitov, Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).

133. M. Morita, G. Stamp, P. Robins, A. Dulic, I. Rosewell, G. Hrivnak, G. Daly, T. Lindahl, D. E. Barnes, Gene-targeted mice lacking the Trex1 (DNase III) 3'-->5' DNA exonuclease develop inflammatory myocarditis. *Mol Cell Biol* **24**, 6719–6727 (2004).

134. P. Li, J. Du, J. L. Goodier, J. Hou, J. Kang, H. H. Kazazian Jr, K. Zhao, X.-F. Yu, Aicardi-Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. *Nucleic Acids Res.* **45**, 4619–4631 (2017).

135. M. A. Lee-Kirsch, M. Gong, D. Chowdhury, L. Senenko, K. Engel, Y. A. Lee, U. De Silva, S. L. Bailey, T. Witte, T. J. Vyse, J. Kere, C. Pfeiffer, S. Harvey, A. Wong, S. Koskenmies, O. Hummel, K. Rohde, R. E. Schmidt, A. F. Dominiczak, M. Gahr, T. Hollis, F. W. Perrino, J. Lieberman, N. Hübner, Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet* **39**, 1065–1067 (2007).

136. Y. Tarabay, M. Achour, M. Teletin, T. Ye, A. Teissandier, M. Mark, D. Bourc'his, S. Viville, Tex19 paralogs are new members of the piRNA pathway controlling retrotransposon suppression. *J Cell Sci* **130**, 1463–1474 (2017).

137. M. MacLennan, M. García-Cañadas, J. Reichmann, E. Khazina, G. Wagner, C. J. Playfoot, C. Salvador-Palomeque, A. R. Mann, P. Peressini, L. Sanchez, K. Dobie, D. Read, C.-C. Hung, R. Eskeland, R. R. Meehan, O. Weichenrieder, J. L. García-Pérez, I. R. Adams, Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *Elife* **6**, e26152 (2017).

138. S. Kuntz, E. Kieffer, L. Bianchetti, N. Lamoureux, G. Fuhrmann, S. Viville, Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and Germ Line. *Stem Cells* **26**, 734–744 (2008).

139. J. H. Crichton, C. J. Playfoot, M. MacLennan, D. Read, H. J. Cooke, I. R. Adams, Tex19.1 promotes Spo11-dependent meiotic recombination in mouse spermatocytes. *PLoS Genet* **13** (2017).

140. J. H. Crichton, D. Read, I. R. Adams, Defects in meiotic recombination delay progression through pachytene in Tex19.1 −/− mouse spermatocytes.

Chromosoma **127**, 437–459 (2018).

141. Y. Tarabay, E. Kieffer, M. Teletin, C. Celebi, A. Van Montfoort, N. Zamudio, M. Achour, R. El Ramy, E. Gazdag, P. Tropel, M. Mark, D. Bourc'his, S. Viville, The mammalian-specific Tex19.1 gene plays an essential role in spermatogenesis and placenta-supported development. *Hum Reprod* **28**, 2201–14 (2013).

142. R. Öllinger, A. J. Childs, H. M. Burgess, R. M. Speed, P. R. Lundegaard, N. Reynolds, N. K. Gray, H. J. Cooke, I. R. Adams, Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet* **4** (2008).

143. F. Yang, Y. Cheng, J. Y. An, Y. T. Kwon, S. Eckardt, N. A. Leu, K. J. Mclaughlin, P. J. Wang, The ubiquitin ligase Ubr2, a recognition E3 component of the N-end rule pathway, stabilizes Tex19.1 during spermatogenesis. *PLoS One* **5** (2010).

144. J. Reichmann, J. P. Reddington, D. Best, D. Read, R. Öllinger, R. R. Meehan, I. R. Adams, The genome-defence gene tex19.1 suppresses line-1 retrotransposons in the placenta and prevents intra-uterine growth retardation in mice. *Hum Mol Genet* **22**, 1791–1806 (2013).

145. M. D. Stenglein, R. S. Harris, APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**, 16837–16841 (2006).

146. H. Muckenfuss, M. Hamdorf, U. Held, M. Perkovic, J. Lower, K. Cichutek, E. Flory, G. G. Schumann, C. Munk, APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* **281**, 22161–22172 (2006).

147. S. Wissing, M. Montano, J. L. Garcia-Perez, J. V Moran, W. C. Greene, Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* **286**, 36427–36437 (2011).

148. M. Kinomoto, T. Kanno, M. Shimura, Y. Ishizaka, A. Kojima, T. Kurata, T. Sata, K. Tokunaga, All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* **35**, 2955–2964 (2007).

149. S. R. Richardson, I. Narvaiza, R. A. Planegger, M. D. Weitzman, J. V Moran, APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**, e02008 (2014).

150. H. H. Kazazian, Mobile elements: drivers of genome

evolution. *Science* **303**, 1626–1632 (2004).

151. A. F. A. Smit, The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* **6**, 743–748 (1996).

152. F. M. Jacobs, D. Greenberg, N. Nguyen, M. Haeussler, A. D. Ewing, S. Katzman, B. Paten, S. R. Salama, D. Haussler, An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242–245 (2014).

153. M. Bruno, M. Mahgoub, T. S. Macfarlan, The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* **53**, 393–416 (2019).

154. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687–695 (2012).

155. L. Yang, G. A. Metzger, R. Padilla Del Valle, D. Delgadillo Rubalcaba, R. N. McLaughlin Jr, Evolutionary insights from profiling LINE-1 activity at allelic resolution in a single human genome. *EMBO J* **43**, 112–131 (2024).

156. K. M. Steinberg, V. A. Schneider, T. A. Graves-Lindsay, R. S. Fulton, R. Agarwala, J. Huddleston, S. A. Shiryev, A. Morgulis, U. Surti, W. C. Warren, D. M. Church, E. E. Eichler, R. K. Wilson, Single haplotype

assembly of the human genome from a hydatidiform mole. *Genome Res* **24**, 2066 (2014).

157. C. R. Beck, P. Collier, C. Macfarlane, M. Malig, J. M. Kidd, E. E. Eichler, R. M. Badge, J. V Moran, LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170 (2010).

158. A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res* **16**, 203–215 (2008).

159. J. N. Wells, C. Feschotte, A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet* **54**, 539–561 (2020).

160. D. C. Hancks, J. L. Goodier, P. K. Mandal, L. E. Cheung, H. H. Kazazian, Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* **20**, 3386 (2011).

161. A. Buzdin, S. Ustyugova, E. Gogvadze, T. Vinogradova, Y. Lebedev, E. Sverdlov, A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of L1. *Genomics* **80**, 402–406 (2002).

162. G. Grimaldi, M. F. Singer, Members of the KpnI family of long interspersed repeated sequences join and interrupt

alpha-satellite in the monkey genome. *Nucleic Acids Res* **11**, 321 (1983).

163. R. M. Badge, R. S. Alisch, J. V Moran, ATLAS: a system to selectively identify human-specific L1 insertions. *Am J Hum Genet* **72**, 823–838 (2003).

164. J. Xing, Y. Zhang, K. Han, A. H. Salem, S. K. Sen, C. D. Huff, Q. Zhou, E. F. Kirkness, S. Levy, M. A. Batzer, L. B. Jorde, Mobile elements create structural variation: analysis of a complete human genome. *Genome Res* **19**, 1516–1526 (2009).

165. J. P. McMillan, M. F. Singer, Translation of the human LINE-1 element, L1Hs. *Proc Natl Acad Sci U S A* **90**, 11533–11537 (1993).

166. A. A. Mir, C. Philippe, G. Cristofari, euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* **43**, D43–D47 (2015).

167. A. F. Scott, B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, L. Margolet, Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* **1**, 113–125 (1987).

168. S. L. Martin, D. Branciforte, D. Keller, D. L. Bain, Trimeric structure for an essential protein in L1

retrotransposition. *Proc Natl Acad Sci U S A* **100**, 13815–13820 (2003).

169. M. S. Taylor, J. LaCava, P. Mita, K. R. Molloy, C. R. L. Huang, D. Li, E. M. Adney, H. Jiang, K. H. Burns, B. T. Chait, M. P. Rout, J. D. Boeke, L. Dai, Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**, 1034–1048 (2013).

170. Q. Feng, J. V Moran, H. H. Kazazian, J. D. Boeke, Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. *Cell* **87**, 905–916 (1996).

171. G. J. Cost, Q. Feng, A. Jacquier, J. D. Boeke, Human L1 element target-primed reverse transcription in vitro. *EMBO J* **21**, 5899–5910 (2002).

172. S. J. Wheelan, Y. Aizawa, J. S. Han, J. D. Boeke, Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* **15**, 1073–1078 (2005).

173. D. E. Symer, C. Connelly, S. T. Szak, E. M. Caputo, G. J. Cost, G. Parmigiani, J. D. Boeke, Human L1 Retrotransposition Is Associated with Genetic Instability In Vivo. *Cell* **110**, 327–338 (2002).

174. R. Cordaux, M. A. Batzer, The impact of retrotransposons on

human genome evolution. *Nat Rev Genet* **10**, 691 (2009).

175. D. Zhou, S. Qi, W. Zhang, L. Wu, A. Xu, X. Li, B. Zhang, Y. Li, S. Jia, H. Wang, J. Jia, X. Ou, J. Huang, H. You, Insertion of LINE-1 Retrotransposon Inducing Exon Inversion Causes a Rotor Syndrome Phenotype. *Front Genet* **10**, 1399 (2020).

176. V. Turcot, A. Tchernof, Y. Deshaies, L. Pérusse, A. Bélisle, S. Marceau, S. Biron, O. Lescelleur, L. Biertho, M.-C. Vohl, LINE-1 methylation in visceral adipose tissue of severely obese individuals is associated with metabolic syndrome status and related phenotypes. *Clin Epigenetics* **4**, 10 (2012).

177. A. R. Muotri, M. C. N. Marchetto, N. G. Coufal, R. Oefner, G. Yeo, K. Nakashima, F. H. Gage, L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).

178. H. E. Volkman, D. B. Stetson, The enemy within: endogenous retroelements and autoimmune disease. *Nat Immunol* **15**, 415–422 (2014).

179. M. K. Crow, Long interspersed nuclear elements (LINE-1): potential triggers of systemic autoimmune disease. *Autoimmunity* **43**, 7–16 (2010).

180. C. A. Thomas, L. Tejwani, C. A. Trujillo, P. D. Negraes, R. H. Herai, P. Mesci, A. Macia, Y. J. Crow, A. R. Muotri,

Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell Stem Cell* **21**, 319-331.e8 (2017).

181. C. P. Mavragani, I. Sagalovskiy, Q. Guo, A. Nezos, E. K. Kapsogeorgou, P. Lu, J. Liang Zhou, K. A. Kirou, S. V. Seshan, H. M. Moutsopoulos, M. K. Crow, Expression of Long Interspersed Nuclear Element 1 Retroelements and Induction of Type I Interferon in Patients With Systemic Autoimmune Disease. *Arthritis and Rheumatology* **68**, 2686–2696 (2016).

182. X. Zhang, R. Zhang, J. Yu, New understanding of the relevant role of LINE-1 retrotransposition in human disease and immune modulation. *Front. Cell Dev. Biol.* **8**, 657 (2020).

183. Y. J. Crow, B. E. Hayward, R. Parmar, P. Robins, A. Leitch, M. Ali, D. N. Black, H. van Bokhoven, H. G. Brunner, B. C. Hamel, P. C. Corry, F. M. Cowan, S. G. Frints, J. Klepper, J. H. Livingston, S. A. Lynch, R. F. Massey, J. F. Meritet, J. L. Michaud, G. Ponsot, T. Voit, P. Lebon, D. T. Bonthron, A. P. Jackson, D. E. Barnes, T. Lindahl, Mutations in the gene encoding the 3’-5’ DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus.

Nat Genet **38**, 917–920 (2006).

184. Z. Cutts, S. Patterson, L. Maliskova, K. E. Taylor, C. Ye, M. Dall'Era, J. Yazdany, L. Criswell, G. K. Fragiadakis, C. Langelier, J. A. Capra, M. Sirota, C. M. Lanata, Cell-Specific Transposable Element Gene Expression Analysis Identifies Associations with Systemic Lupus Erythematosus Phenotypes. *bioRxiv*, 2023.11.27.567477 (2023).

185. J. Nakkuntod, Y. Avihingsanon, A. Mutirangura, N. Hirankarn, Hypomethylation of LINE-1 but not Alu in lymphocyte subsets of systemic lupus erythematosus patients. *Clin Chim Acta* **412**, 1457–1461 (2011).

186. M. De Cecco, T. Ito, A. P. Petrashen, A. E. Elias, N. J. Skvir, S. W. Criscione, A. Caligiana, G. Brocculi, E. M. Adney, J. D. Boeke, O. Le, C. Beauséjour, J. Ambati, K. Ambati, M. Simon, A. Seluanov, V. Gorbunova, P. E. Slagboom, S. L. Helfand, N. Neretti, J. M. Sedivy, L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).

187. E. Lee, R. Iskow, L. Yang, O. Gokcumen, P. Haseley, L. J. Luquette, J. G. Lohr, C. C. Harris, L. Ding, R. K. Wilson, D. A. Wheeler, R. A. Gibbs, R. Kucherlapati, C. Lee, P. V. Kharchenko, P. J. Park,

Landscape of somatic retrotransposition in human cancers. *Science (1979)* **337**, 967–971 (2012).

188. K. H. Burns, Transposable elements in cancer. *Nat. Rev. Cancer* **17**, 415–424 (2017).

189. G. L. Bratthauer, T. G. Fanning, Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* **7**, 507–510 (1992).

190. Y. Miki, I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K. W. Kinzler, B. Vogelstein, Y. Nakamura, Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* **52**, 643–645 (1992).

191. Z. Tang, J. P. Steranka, S. Ma, M. Grivainis, N. Rodić, C. R. L. Huang, I.-M. Shih, T.-L. Wang, J. D. Boeke, D. Fenyö, K. H. Burns, Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E733–E740 (2017).

192. S. Boissinot, A. Entezam, A. V Furano, Selection against deleterious LINE-1-containing loci in the human lineage. *Mol Biol Evol* **18**, 926–935 (2001).

193. S. Boissinot, J. Davis, A. Entezam, D. Petrov, A. V Furano, Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9590–9594 (2006).

194. Y. Ariumi, Guardian of the human genome: Host defense mechanisms against LINE-1 retrotransposition. *Front. Chem.* **4**, 28 (2016).

195. M. S. Protasova, T. V. Andreeva, E. I. Rogaev, Factors regulating the activity of LINE1 retrotransposons. *Genes (Basel)* **12**, 1562 (2021).

196. L. Robbez-Masson, C. H. C. Tie, L. Conde, H. Tunbak, C. Husovsky, I. A. Tchasovnikarova, R. T. Timms, J. Herrero, P. J. Lehner, H. M. Rowe, The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* **28**, 836–845 (2018).

197. J. A. Hackett, J. P. Reddington, C. E. Nestor, D. S. Dunican, M. R. Branco, J. Reichmann, W. Reik, M. A. Surani, I. R. Adams, R. R. Meehan, Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623–3632 (2012).

198. A. Molaro, H. S. Malik, Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Curr. Opin. Genet. Dev.* **37**, 51–58 (2016).

199. H. Guo, P. Zhu, L. Yan, R. Li, B. Hu, Y. Lian, J. Yan, X. Ren, S. Lin, J. Li, X. Jin, X. Shi, P.

Liu, X. Wang, W. Wang, Y. Wei, X. Li, F. Guo, X. Wu, X. Fan, J. Yong, L. Wen, S. X. Xie, F. Tang, J. Qiao, The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).

200. J. G. Pizarro, G. Cristofari, Post-transcriptional control of LINE-1 retrotransposition by cellular host factors in somatic cells. *Front. Cell Dev. Biol.* **4**, 14 (2016).

201. Z. Warkocki, P. S. Krawczyk, D. Adamska, K. Bijata, J. L. Garcia-Perez, A. Dziembowski, Uridylation by TUT4/7 restricts retrotransposition of human LINE-1s. *Cell* **174**, 1537-1548.e29 (2018).

202. H. P. Bogerd, H. L. Wiegand, A. E. Hulme, J. L. Garcia-Perez, K. S. O'Shea, J. V Moran, B. R. Cullen, Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8780–8785 (2006).

203. C. Celebi, A. van Montfoort, V. Skory, E. Kieffer, S. Kuntz, M. Mark, S. Viville, Tex 19 paralogs exhibit a gonad and placenta-specific expression in the mouse. *Journal of Reproduction and Development* **58**, 360–365 (2012).

204. M. MacLennan, M. García-Cañadas, J. Reichmann, E. Khazina, G. Wagner, C. J. Playfoot, C. Salvador-Palomeque, A. R. Mann, P.

Peressini, L. Sanchez, K. Dobie, D. Read, C.-C. Hung, R. Eskeland, R. R. Meehan, O. Weichenrieder, J. L. García-Pérez, I. R. Adams, Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *Elife* **6** (2017).

205. E. Khazina, O. Weichenrieder, Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *Elife* **7** (2018).

206. I. R. Kelsall, Non-lysine ubiquitylation: Doing things differently. *Front. Mol. Biosci.* **9**, 1008175 (2022).

207. J. M. C. Tubio, Y. Li, Y. S. Ju, I. Martincorena, S. L. Cooke, M. Tojo, G. Gundem, C. P. Pipinikas, J. Zamora, K. Raine, A. Menzies, P. Roman-Garcia, A. Fullam, M. Gerstung, A. Shlien, P. S. Tarpey, E. Papaemmanuil, S. Knappskog, P. Van Loo, M. Ramakrishna, H. R. Davies, J. Marshall, D. C. Wedge, J. W. Teague, A. P. Butler, S. Nik-Zainal, L. Alexandrov, S. Behjati, L. R. Yates, N. Bolli, L. Mudie, C. Hardy, S. Martin, S. McLaren, S. O'Meara, E. Anderson, M. Maddison, S. Gamble, C. Foster, A. Y. Warren, H. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, A. G. Lynch, T. Visakorpi, W. B. Isaacs, L. Van't Veer, C. Caldas, C. Desmedt, C. Sotiriou, S.

Aparicio, J. A. Foekens, J. E. Eyfjörd, S. R. Lakhani, G. Thomas, O. Myklebost, P. N. Span, A. L. Børresen-Dale, A. L. Richardson, M. Van De Vijver, A. Vincent-Salomon, G. G. Van Den Eynden, A. M. Flanagan, P. A. Futreal, S. M. Janes, G. S. Bova, M. R. Stratton, U. McDermott, P. J. Campbell, Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science (1979)* **345** (2014).

208. B. Rodriguez-Martin, E. G. Alvarez, A. Baez-Ortega, J. Zamora, F. Supek, J. Demeulemeester, M. Santamarina, Y. S. Ju, J. Temes, D. Garcia-Souto, H. Detering, Y. Li, J. Rodriguez-Castro, A. Dueso-Barroso, A. L. Bruzos, S. C. Dentro, M. G. Blanco, G. Contino, D. Ardeljan, M. Tojo, N. D. Roberts, S. Zumalave, P. A. Edwards, J. Weischenfeldt, M. Puiggròs, Z. Chong, K. Chen, E. A. Lee, J. A. Wala, K. M. Raine, A. Butler, S. M. Waszak, F. C. P. Navarro, S. E. Schumacher, J. Monlong, F. Maura, N. Bolli, G. Bourque, M. Gerstein, P. J. Park, D. C. Wedge, R. Beroukhim, D. Torrents, J. O. Korbel, I. Martincorena, R. C. Fitzgerald, P. Van Loo, H. H. Kazazian, K. H. Burns, P. J. Campbell, J. M. C. Tubio, K. C. Akdemir, P. C. Boutros, D. D. L. Bowtell, B. Brors, K. Chan, I. Cortés-Ciriano, A. J. Dunford, X. Estivill, D.

Etemadmoghadam, L. Feuerbach, J. L. Fink, M. Frenkel-Morgenstern, D. W. Garsed, D. A. Gordenin, D. Haan, J. E. Haber, J. M. Hess, B. Hutter, M. Imielinski, D. T. W. Jones, M. D. Kazanov, L. J. Klimczak, Y. Koh, K. Kumar, J. J. K. Lee, A. G. Lynch, G. Macintyre, F. Markowetz, A. Martinez-Fundichely, M. Meyerson, S. Miyano, H. Nakagawa, S. Ossowski, J. V. Pearson, K. Rippe, S. A. Roberts, R. Scully, M. Shackleton, N. Sidiropoulos, L. Sieverling, C. Stewart, I. Villasante, N. Waddell, L. Yang, X. Yao, S. S. Yoon, C. Z. Zhang, Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet* **52**, 306–319 (2020).

209. F. J. Sanchez-Luque, M.-J. H. C. Kempen, P. Gerdes, D. B. Vargas-Landin, S. R. Richardson, R.-L. Troskie, J. S. Jesuadian, S. W. Cheetham, P. E. Carreira, C. Salvador-Palomeque, M. García-Cañadas, M. Muñoz-Lopez, L. Sanchez, M. Lundberg, A. Macia, S. R. Heras, P. M. Brennan, R. Lister, J. L. Garcia-Perez, A. D. Ewing, G. J. Faulkner, LINE-1 evasion of epigenetic repression in humans. *Mol. Cell* **75**, 590-604.e12 (2019).

210. Y. Xie, J. M. Rosser, T. L. Thompson, J. D. Boeke, W. An, Characterization of L1 retrotransposition with high-throughput dual-luciferase

assays. *Nucleic Acids Res* **39**, e16 (2011).

211. C. Celebi, A. van Montfoort, V. Skory, E. Kieffer, S. Kuntz, M. Mark, S. Viville, Tex 19 paralogs exhibit a gonad and placenta-specific expression in the mouse. *J Reprod Dev* **58**, 360–365 (2012).

212. S. Kuntz, E. Kieffer, L. Bianchetti, N. Lamoureux, G. Fuhrmann, S. Viville, Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* **26**, 734–744 (2008).

213. R. Ollinger, A. J. Childs, H. M. Burgess, R. M. Speed, P. R. Lundegaard, N. Reynolds, N. K. Gray, H. J. Cooke, I. R. Adams, Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet* **4**, e1000199 (2008).

214. S. Kuntz, E. Kieffer, L. Bianchetti, N. Lamoureux, G. Fuhrmann, S. Viville, Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and Germ Line. *Stem Cells* **26**, 734–744 (2007).

215. M. Zeng, Y. Lu, X. Liao, D. Li, H. Sun, S. Liang, S. Zhang, Y. Ma, Z. Yang, DAZL binds to 3′UTR of Tex19.1 mRNAs and regulates Tex19.1 expression. *Mol Biol Rep* **36**, 2399–2403 (2009).

216. V. Planells-Palop, A. Hazazi, J. Feichtinger, J. Jezkova, G. Thallinger, N. O. Alsiwiehri, M. Almutairi, L. Parry, J. A. Wakeman, R. J. McFarlane, Human germ/stem cell-specific gene TEX19 influences cancer cell proliferation and cancer prognosis. *Mol Cancer* **16** (2017).

217. V. Planells-Palop, A. Hazazi, J. Feichtinger, J. Jezkova, G. Thallinger, N. O. Alsiwiehri, M. Almutairi, L. Parry, J. A. Wakeman, R. J. McFarlane, Human germ/stem cell-specific gene TEX19 influences cancer cell proliferation and cancer prognosis. *Mol Cancer* **16** (2017).

218. M. MacLennan, M. García-Cañadas, J. Reichmann, E. Khazina, G. Wagner, C. J. Playfoot, C. Salvador-Palomeque, A. R. Mann, P. Peressini, L. Sanchez, K. Dobie, D. Read, C.-C. Hung, R. Eskeland, R. R. Meehan, O. Weichenrieder, J. L. García-Pérez, I. R. Adams, Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *Elife* **6** (2017).

219. L. Bianchetti, Y. Tarabay, O. Lecompte, R. Stote, O. Poch, A. Dejaegere, S. Viville, Tex19 and Sectm1 concordant molecular phylogenies support co-evolution of both eutherian-specific genes. *BMC Evol Biol* **15** (2015).

220. S. Kuntz, E. Kieffer, L. Bianchetti, N. Lamoureux, G. Fuhrmann, S. Viville, Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* **26**, 734–744 (2008).

221. R. N. McLaughlin, J. M. Young, L. Yang, R. Neme, H. A. Wichman, H. S. Malik, Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency. *PLoS Genet* **10** (2014).

222. M. A. Cantrell, L. A. Scott, C. J. Brown, A. R. Martinez, H. A. Wichman, Loss of LINE-1 activity in the megabats. *Genetics* **178**, 393–404 (2008).

223. A. Sookdeo, C. M. Hepp, S. Boissinot, Contrasted patterns of evolution of the LINE-1 retrotransposon in perissodactyls: the history of a LINE-1 extinction. *Mob DNA* **9** (2018).

224. M. A. O’Leary, J. I. Bloch, J. J. Flynn, T. J. Gaudin, A. Giallombardo, N. P. Giannini, S. L. Goldberg, B. P. Kraatz, Z. X. Luo, J. Meng, X. Ni, M. J. Novacek, F. A. Perini, Z. S. Randall, G. W. Rougier, E. J. Sargis, M. T. Silcox, N. B. Simmons, M. Spaulding, P. M. Velazco, M. Weksler, J. R. Wible, A. L. Cirranello, The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667 (2013).

225. M. Dos Reis, P. C. J. Donoghue, Z. Yang, Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biol Lett* **10** (2014).

226. L. Bianchetti, Y. Tarabay, O. Lecompte, R. Stote, O. Poch, A. Dejaegere, S. Viville, Tex19 and Sectm1 concordant molecular phylogenies support co-evolution of both eutherian-specific genes. *BMC Evol Biol* **15** (2015).

227. N. S. Upham, J. A. Esselstyn, W. Jetz, Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol* **17**, e3000494 (2019).

228. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–21 (2010).

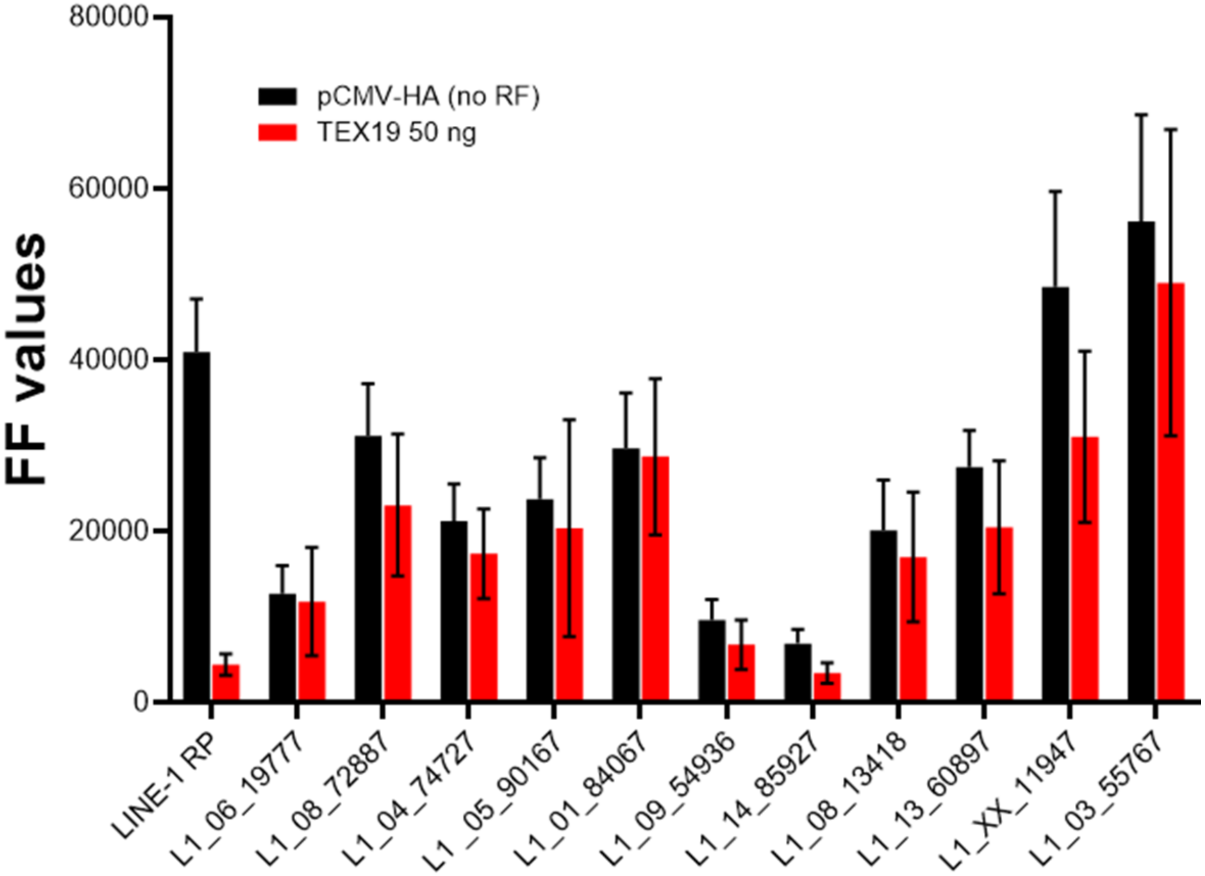
Supplementary Materials

Supplementary Table 1. List of nomenclature used for identifying and referring to LINE-1s from the CHM1 genome. We named them by indicating the chromosome number where they are located and then the first five (5) numbers of the genomic coordinates of Hg38 genome reference (10^8).

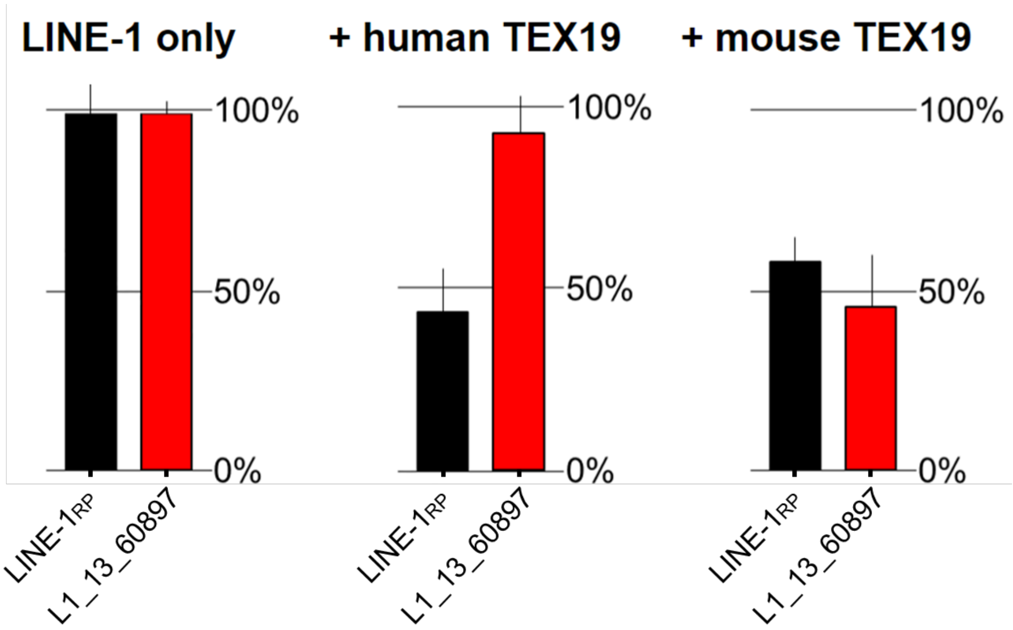
CHM1 name	CHM1 contig/scaffold number
L1_02_19068	LJII01000002.1_49345876_49351905_+_6029_L1HS
L1_18_75587	LJII01000023.1_28589228_28595280_-_6052_L1HS
L1_01_20091	LJII01000015.1_39372723_39378740_-_6017_L1HS
L1_03_13938	LJII01000008.1_9140235_9146251_+_6016_L1HS
L1_08_72887	LJII01000006.1_12767654_12773704_+_6050_L1HS
L1_17_66647	LJII01000087.1_1071256_1077285_+_6029_L1HS
L1_XX_11947	LJII01000068.1_3777361_3783378_-_6017_L1HS
L1_14_58767	LJII01000004.1_41603277_41609287_-_6010_L1HS
L1_01_84067	LJII01000042.1_579595_585611_-_6016_L1HS
L1_09_54936	LJII01000016.1_5326820_5332854_+_6034_L1HS
L1_06_19777	LJII01000031.1_6909916_6915928_-_6012_L1HS
L1_01_17968	LJII01000015.1_18071811_18077859_-_6048_L1HS
L1_03_38597	LJII01000005.1_38023157_38029186_+_6029_L1HS
L1_04_74727	LJII01000011.1_28169369_28175400_+_6031_L1HS
L1_14_85927	LJII01000004.1_14418042_14423764_+_5722_L1HS
L1_03_55767	LJII01000005.1_55224943_55231003_+_6060_L1HS

L1_06_11718	LJII01000001.1_55856666_55862697_+_6031_L1HS
L1_09_11288	LJII01000020.1_15791033_15797083_-_6050_L1HS
L1_12_12638	LJII01000007.1_6996638_7002665_+_6027_L1HS
L1_13_60897	LJII01000012.1_8309279_8315328_-_6049_L1HS
L1_08_13418	LJII01000010.1_6943196_6949225_+_6029_L1HS
L1_15_70747	LJII01000058.1_1912005_1917420_-_5415_L1HS
L1_XX_11717	LJII01000068.1_3549304_3555334_+_6030_L1HS
L1_05_90167	LJII01000003.1_85827333_85833369_-_6036_L1HS
L1_07_30457	LJII01000034.1_23870329_23876360_+_6031_L1HS
L1_04_13638	LJII01000061.1_7645740_7651794_+_6054_L1HS
L1_10_85367	LJII01000089.1_1817529_1823560_-_6031_L1HS
L1_09_95707	LJII01000153.1_1205393_1211411_-_6018_L1HS
L1_02_16607	LJII01000009.1_6866544_6872576_-_6032_L1HS
L1_16_33967	chr16_33952564_33958612_-_6048_L1HS
L1_18_59417	LJII01000023.1_12381651_12387681_-_6030_L1PA2
L1_05_39797	LJII01000059.1_6809248_6815277_-_6029_L1HS
L1_05_15298	LJII01000003.1_23061067_23067085_+_6018_L1HS
L1_XX_83077	LJII01000027.1_19179605_19185659_+_6054_L1HS

Supplementary Figure 1: Eleven (11) LINE-1s with and without TEX19.



Supplementary Figure 2: Mouse and Human TEX19 versus LINE-1$_{RP}$ and L1_13_60897.



Supplementary Table 2. 3' Transductions of retrocompetent LINE-1s from CHM1 present in germline and cancer genome datasets from three separate studies.

CHM1 name	Number of 3' transductions in Tubio et al. 2014 Table S5.	Number of 3' transductions in Rodriguez-Martin et al. 2020 Table S5.	Number of 3' transductions in Chuang et al. 2021 Table S3.
L1_01_17968		6	
L1_01_20091			
L1_01_84067	7	39	
L1_02_16607			
L1_02_19068			
L1_03_13938			0

L1_03_38597			0
L1_03_55767	3	15	0
L1_04_13638		22	
L1_04_74727		6	
L1_05_11018	3	5	
L1_05_15298			
L1_05_39797	17	25	
L1_05_90167	19	35	0
L1_06_11718			
L1_06_19777	17	10	
L1_07_30457	1	36	
L1_08_13418	4	42	
L1_08_72887	5	32	
L1_09_11288	6	128	
L1_09_54936		4	0
L1_09_95707			
L1_10_85367		8	
L1_12_12638			
L1_13_60897	16	50	1
L1_14_58767	40	347	2
L1_14_85927			
L1_15_70747			
L1_16_33967			

L1_17_66647	2	3	0
L1_18_59417	10	35	
L1_18_75587			
L1_XX_11717	7	183	
L1_XX_11947	20	112	
L1_XX_83077			