Regulatory networks of KRAB zinc finger genes and transposable elements changed during human brain evolution and disease

Yao-Chung Chen[1,2], Arnaud Maupas[3], Katja Nowick[1,2*]

[1] Human Biology and Primate Evolution, Institute for Biology, Freie Universität Berlin, Berlin, Germany

[2] Institute of Bioinformatics, Freie Universität Berlin, Berlin, Germany

[3] Institut de Biologie de l'École Normale Supérieure, Ecole Normale supérieure de Paris, Paris, France

[*]**Corresponding author/e-address**: katja.nowick@fu-berlin.de

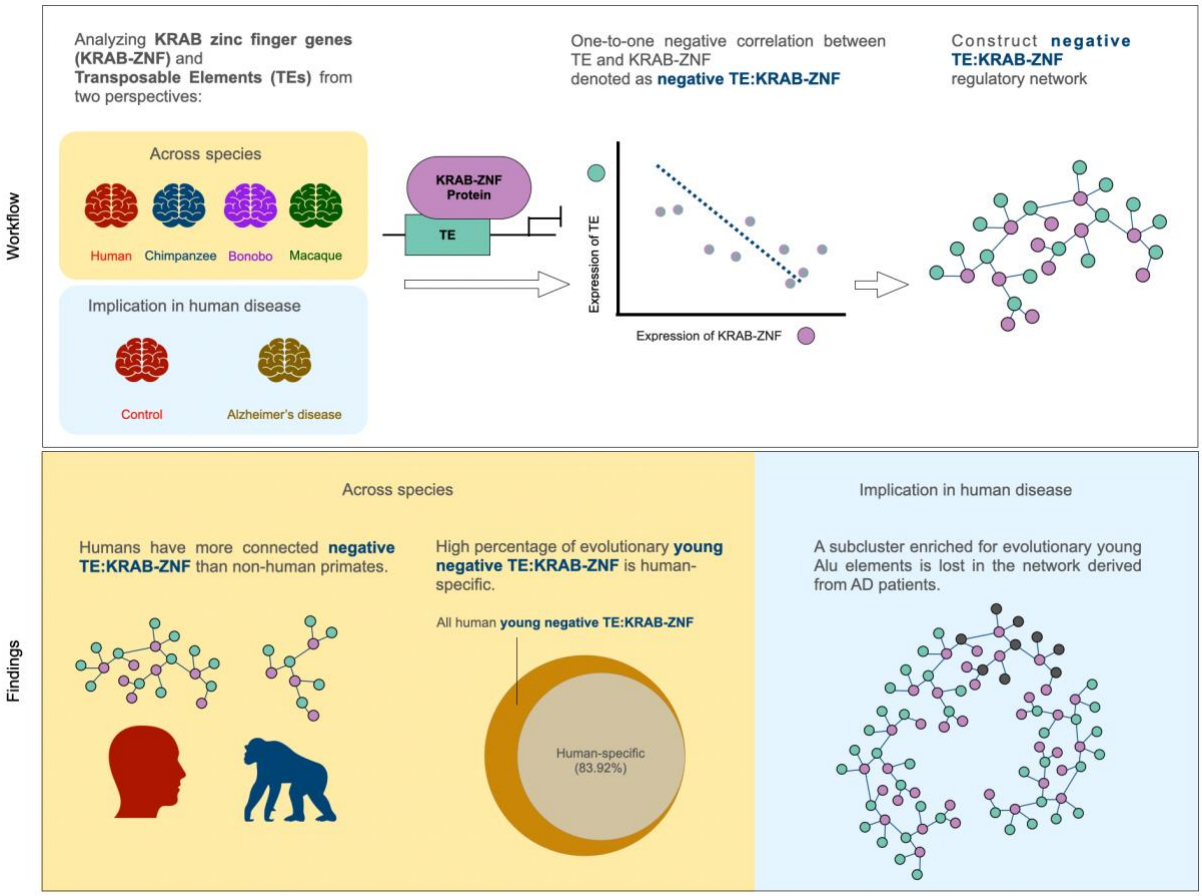
Abstract

Transposable elements (TEs), with their ability to change positions within a genome, have paradoxically been seen as both a potentially deleterious genomic phenomenon and a potent driving force behind evolution. The genome-protecting KRAB zinc finger (KRAB-ZNF) proteins play a critical role in repressing TE expression within mammalian genomes, engaging in a dynamic interplay. This interplay was suggested to evolve according to an arms-race model, wherein TEs strive to transpose within the genome and KRAB-ZNFs adapt to suppress them. Despite indications of the involvement of TEs and KRAB-ZNFs in brain evolution and disease, a systematic analysis of their interactions is still lacking. In this study, we explored the functional connections between KRAB-ZNFs and TEs in the context of human brain evolution and Alzheimer's disease (AD). We conducted an analysis of KRAB-ZNF genes and TEs expression patterns and networks using two independent RNA-seq datasets: (1) data from 33

human and multiple non-human primate brain regions, and (2) data from the temporal cortex and cerebellum of both healthy individuals and AD patients. To efficiently compare the regulatory networks across species, we developed the TEKRABber R package, which enables estimating TE and KRAB-ZNF expression levels and pairwise correlations between them, facilitating cross-species regulatory network comparisons. Our analysis highlighted species-specific expression variations, with many recently evolved TEs and KRAB-ZNF genes being differentially expressed between species, emphasizing their impact on evolution. Focusing on one-to-one negative correlations between TEs and KRAB-ZNF genes (negative TE:KRAB-ZNF), we found that many of those correlations engage recently evolved TEs and KRAB-ZNF genes and are specific to humans. Integrating findings from the second dataset, we identified 2492 human-specific negative TE:KRAB-ZNF pairs uniquely detectable in the healthy human temporal cortex, suggesting dysregulation in AD brains. A distinct subcluster in the network formed by these 2492 pairs consists of Alu elements, underscoring their unique role in human brain evolution and disease. Our results deepens insights into primate brain evolution and offer a new perspective on human neurodegenerative disease through the analysis of the negative TE:KRAB-ZNF regulatory network.

Keywords: primate evolution, Alzheimer's disease, KRAB zinc finger, transposable elements, regulatory network, bipartite network.

Graphical Abstract



Introduction

Transposable elements (TEs), also known as jumping genes, are mobile DNA sequences that can migrate and insert into new locations in the host genome and thereby result in phenotypic changes of the organism (Bourque et al. 2018; Schrader and Schmitz 2019). A variety of host factors are employed to regulate TE expression, such as small RNAs, chromatin and DNA modification pathways, and KRAB zinc finger (KRAB-ZNF) proteins (Colonna Romano and Fanti 2022). KRAB-ZNF proteins are the largest family of transcription factors in higher vertebrates, such as primates (Huntley et al. 2006; Ecco et al. 2017), characterized by fast evolution and contributing to gene expression differences between primates (Nowick et al. 2009; Nowick et al. 2013).

KRAB-ZNF proteins bind to the interspersed DNA sequences of TEs and repress their expression upon recruiting the cofactor KAP1 (Groner et al. 2010). The fast evolution of TEs and KRAB-ZNF genes lead to the hypothesis of an evolutionary arms-race model in which mutated TEs have the chance to escape the repression from KRAB-ZNF proteins until the KRAB-ZNF proteins evolve again a suitable recognition ability (Jacobs et al. 2014; Imbeault et al. 2017). For instance, in primates, coevolution between a family of TEs, retroelements in endogenous retroviruses (ERVs), and the evolving tandem repeats in KRAB-ZNF genes has been observed (Thomas and Schneider 2011).

Studying the expression of KRAB-ZNF genes and TEs in the evolution of the human brain has crucial relevance. The interaction between KRAB-ZNF proteins and TEs varies across tissue types and developmental stages. Typically, TEs are most active in the germline but become silenced in adult tissues. However, the adult human brain represents an exception where TEs remain actively transcribed (Bodea et al. 2018; Playfoot et al. 2021). For example, some TE subfamilies are still active in human brains, including L1 (Suarez et al. 2018), Alu elements (Larsen et al. 2018) and SINE-VNTR-Alu elements (SVAs) (Hancks and Kazazian 2010). Furthermore, TE-derived promoters play a role in gene transcription within the human brain, which in turn suggests the significance of TEs in gene regulation during neurodevelopment (Playfoot et al. 2021). Interestingly, a substantial number of differentially expressed primate-specific KRAB-ZNF genes were detected in the adult prefrontal cortex, comparing humans to chimpanzees (Nowick et al. 2009) and newly evolved KRAB-ZNF genes were predominantly detected in the developing human brain (Zhang et al. 2011). Therefore, analyzing the interplay between TEs and KRAB-ZNF genes in the context

of primate brain evolution could provide valuable insights into the complex mechanisms shaping human brain development and function.

Abnormalities in TE expression were linked to neural dysfunctions in neurodegenerative diseases, including Alzheimer's disease (AD), the primary cause of dementia (Ravel-Godreuil et al. 2021). A primary biomarker of AD, the Tau protein, which in its hyperphosphorylated form builds neurofibrillary tangles, seems to be able to induce the expression of at least some TEs (Guo et al. 2018). It has been proposed that the expression of TEs in the cytosol and endosomes induces neuroinflammation in AD patients (Evering et al. 2023). Additionally, a polymorphism found in the KRAB-ZNF gene *ZNF224* (rs3746319) demonstrated a notable association with both global AD neuropathology and cognitive functions, revealing a potential role in influencing intermediate phenotypes in AD progression (Shulman et al. 2010).

Taken together, the study of the regulatory networks involving KRAB-ZNF genes and TEs is expected to provide insights into the evolution of the human brain and the development of neurodegenerative diseases. Here, we present results from two independent RNA-seq datasets: one comparing different brain regions between humans and multiple non-human primates (NHPs) (**Primate Brain Data**), and the other comparing human control samples with AD samples (**Mayo Data**). These datasets collectively encompass a total of 514 samples from different species, brain regions, and disease states. We developed the TEKRABber R package for the systematic and comparative analysis of TE and KRAB-ZNF relationships. Our work reveals an intricate network of TEs and KRAB-ZNF genes in the brain that changed during evolution and is modified in AD brains.

Results

To investigate the regulatory network between KRAB-ZNFs and TEs during primate brain evolution and cognitive disorder, we analyzed two publicly available RNA-seq datasets, **Primate Brain Data** and **Mayo Data** (**Figure 1**, **Tables S1-S3**). The Primate Brain Data contains genome-wide expression information from 33 brain regions grouped into seven groups (Khrameeva et al. 2020) of four primate species, while the Mayo Data includes data from two brain regions of AD patients and controls. In brief, expression of all genes and TEs was quantified and normalized across samples of each dataset, and subsequently differential expression and correlations between KRAB-ZNFs and TEs were calculated (**Figure 1**).

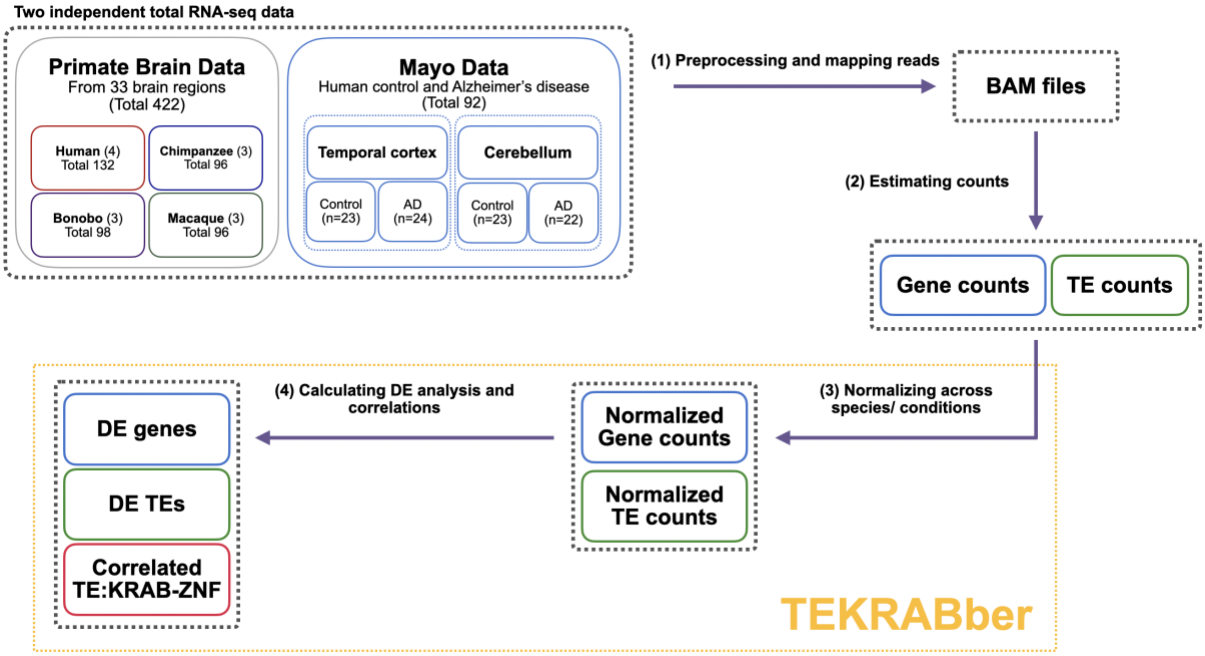


Figure 1. Overview of the analysis workflow. Two RNA-seq datasets, **Primate Brain Data** and **Mayo Data**, were analyzed in this study. **(1)** Transcriptomic data were first preprocessed with removing adapters and low-quality reads and then mapped to their reference genome using STAR to generate BAM files. **(2)** TEtranscript was used to quantify the expression of genes and TEs. **(3)** Expression profiles were normalized across different species. **(4)** Pairwise correlations and DE analysis were

calculated. Steps **(3)** and **(4)** were developed together in an R Bioconductor package, TEKRABber. (Workflow details can be found in Materials and Methods).

Primate Brain Data

1. Dynamics of expression across species and brain regions: Strong species differences especially of evolutionary young KRAB-ZNF genes and TEs

We obtained normalized expression values of KRAB-ZNF genes and TEs and assessed their variance across different species using t-SNE clustering (**Figure 2A**). The data labeled by species revealed distinct boundaries of variance, demonstrating clustering based on species. Specifically, humans and macaques formed their own clusters, while chimpanzees and bonobos were grouped in the same cluster, which agrees with phylogenetic distance. These findings indicated that clear differences in KRAB-ZNF genes and TE expression can be detected across species. However, the expression patterns differed less across brain regions within the same species. We further compared differentially expressed KRAB-ZNF genes and TEs. For example, there were 12 up-regulated and 42 down-regulated KRAB-ZNF genes, along with 31 up-regulated and 22 down-regulated TEs in humans compared to chimpanzees in the primary and secondary cortices (**Figure 2B;** see **Table S2** for information on Broadman areas included in the group “primary and secondary cortices”).

To investigate whether expression patterns differ between evolutionary old and young KRAB-ZNF genes and TEs, respectively, we dated these genomic elements and classified them into two groups based on their inferred evolutionary age. The first group consists of 237 KRAB-ZNF genes and 842 TEs that evolved prior to the emerge of Simiiformes (>44.2 million years ago (mya), denoted as **old** from here onward), while the second group consists of 100 KRAB-ZNF genes and 309 TEs that evolved less

than 44.2 mya (denoted as **young** from here onward) (**Figure S1-2**). We found that the young KRAB-ZNFs and young TEs exhibited significantly lower expression levels across all brain regions, regardless of species (**Figure 2C**). However, there were proportionally more young KRAB-ZNF genes and young TEs differentially expressed between humans and chimpanzees compared to old KRAB-ZNF genes and old TEs (**Figure 2D**).

We then focused on a subset of KRAB-ZNF genes and TEs that were human-specifically up-regulated or down-regulated compared to the three NHPs. There were 36 KRAB-ZNF genes and 18 TEs in total in this subset in the primary and secondary cortices (**Figure 2E**). For example, *ZNF267*, which shows higher expression in the human cortex, is involved in regulating cell proliferation (Yang et al. 2022) and was up-regulated in the prefrontal cortex of AD patients (Patel et al. 2021). It is used as one of the top 10 transcriptomic signatures for blood-based diagnosis of AD (Fehlbaum-Beurdeley et al. 2012). On the other hand, LTR12B was one of the most downregulated TEs in humans compared to NHPs. LTR12-related endogenous retroviruses had been reported to be repressed by ZNF676 and ZNF728 in early human development (Iouranova et al. 2022).

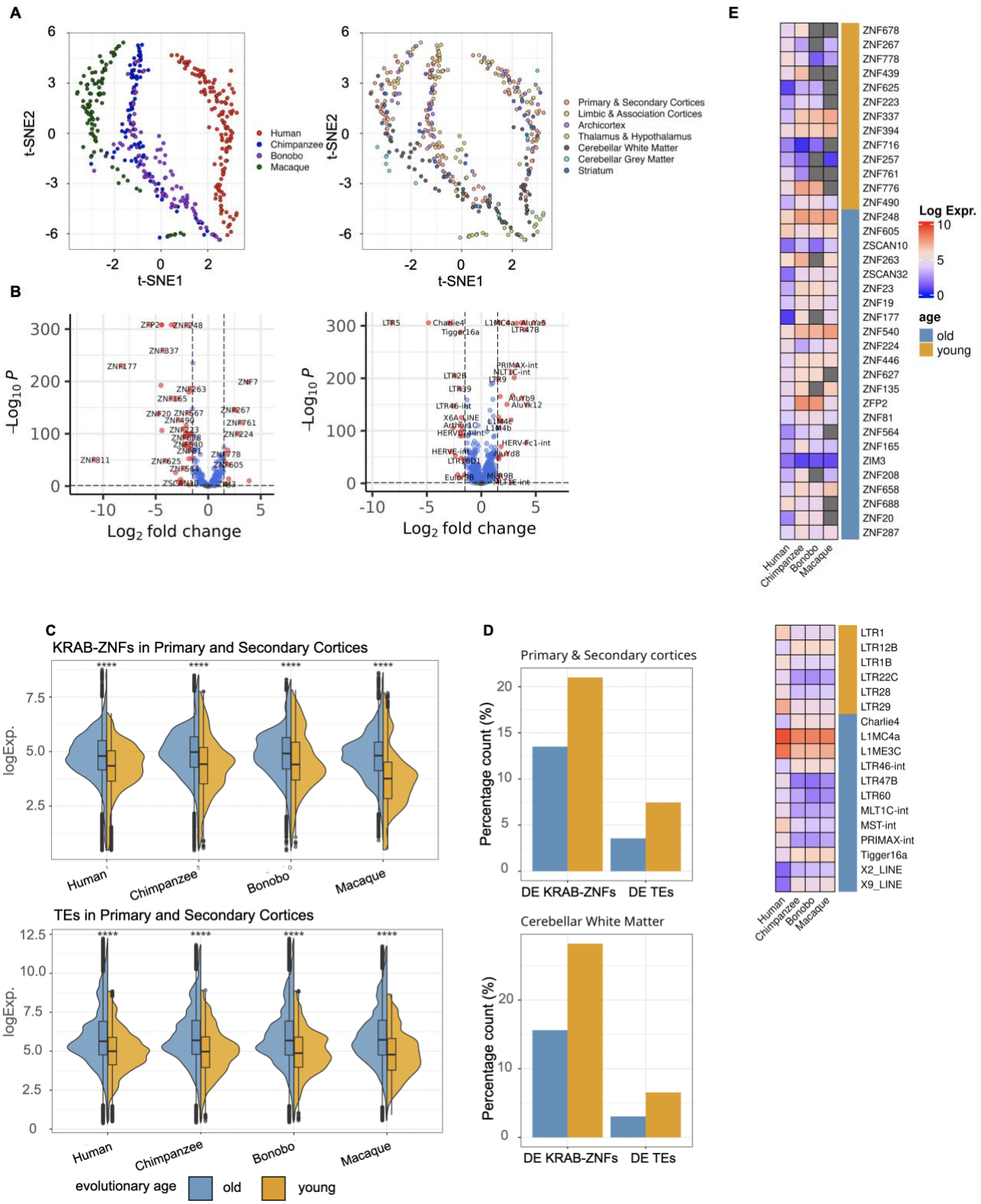


Figure 2. Expression of KRAB-ZNF genes and TEs in Primate Brain Data. (A) t-SNE plots of the expression of KRAB-ZNF genes and TEs from all 422 samples including human, chimpanzee, bonobos, and macaques labeled by species and different brain regions. **(B)** Differentially expressed KRAB-ZNF genes and TEs comparing human and chimpanzee in primary and secondary cortices. **(C)** Expression

of KRAB-ZNF genes and TEs in primary and secondary cortices across species. Both KRAB-ZNF genes and TEs were grouped in two groups based on their inferred evolutionary age, old (> 44.2 mya) and young (<=44.2mya). **(D)** Percentage of differentially expressed KRAB-ZNF genes and TEs in humans comparing to chimpanzees in primary and secondary cortices and cerebellar white matter. **(E)** Human-specific DE (differentially expressed) (human- specifically-changed) KRAB-ZNF genes and TEs in primary and secondary cortices compared to NHPs. Gray indicates no expression information. The colors for age inferences in **(C)**, **(D)**, and **(E)** are the same: blue for evolutionary old and yellow for evolutionary young KRAB-ZNFs and TEs, respectively.

2. Changes in correlations between TEs and KRAB-ZNF genes: Increased connectivity in the human brain with an enrichment for evolutionary young connections

To systematically analyze the functional relationships between TEs and KRAB-ZNF genes, we conducted pairwise correlation analysis using normalized expression levels within each group as defined in Khrameeva et al. (2020) and for each species (**Table S1-2**). Pearson's correlation coefficients were calculated between all pairs of expressed TE and KRAB-ZNF genes. In primary and secondary cortices, for example, we found between 48322, 19348, 26596, and 13305 negative and 50506, 16833, 26405, and 11384 positive correlations of TEs and KRAB-ZNF genes in humans, chimpanzees, bonobos, and macaques. In all brain regions we found approximately equal numbers of positive and negative correlations between TEs and KRAB-ZNF genes (**Table S4**). Negative correlations could indicate TE repression events by KRAB-ZNF proteins (keeping in mind that the direction of the interaction cannot be inferred from correlation analysis and the interaction could also be indirect). To cross-validate our findings, we compared our correlation results to results of a ChIP-Exo study of KRAB-ZNFs binding in TE regions (Imbeault et al. 2017) and confirmed a relationship for ZNF433, ZNF440, and ZNF468 binding to MER11A. Noteworthy, some of the ChIP-

Exo binding events appeared in positively correlated genes, such as ZNF141 binding to L1PAs, raising a question about their mode of interaction.

Given that KRAB-ZNFs are usually considered to act as repressors, we focused our subsequent analyses only the significant negative correlations between the TE and KRAB-ZNF gene (adjusted *p-value* < 0.05, *R* < 0) and denoted the correlated pairs as **negative TE:KRAB-ZNF** in the following content. In the human primary and secondary cortices, we found 48322 negative TE:KRAB-ZNF. There were more negative TE:KRAB-ZNF in humans than in any NHPs in all investigated brain areas, except in the hypothalamus and thalamus (**Table S4)**. We next normalized the number of negative TE:KRAB-ZNF by dividing by the count of all possible correlations based on the number of expressed TEs and KRAB-ZNF genes. The excess of correlations in humans compared to NHP was most pronounced in the primary and secondary cortices (**Figure 3A**). Notably, the sample size for humans was larger, consisting of four biological replicates, while all NHPs had only three replicates available. To exclude the possibility that the higher number of negative TE:KRAB-ZNF was due to the difference in sample size, we employed leave-one-out testing, for which we always excluded one of the human samples and repeated the analysis with the remaining three human samples (**Figure S3**). Regardless of which human sample was excluded, humans still had clearly more negative TE:KRAB-ZNF compared to NHPs. These correlations included an average of 284 KRAB-ZNF genes and 861 TEs compared to NHPs (219-285 KRAB-ZNFs and 820-875 TEs in NHPs). Since the number of KRAB-ZNF genes and TEs involved in negative TE:KRAB-ZNF was very consistent across the four leave-one-out re-analyses and the full analysis with four human samples, we decided to perform all our subsequent analyses with all four human samples to keep

as much information as possible.

To investigate the evolution of negative TE:KRAB-ZNF interactions, we utilized the same age classification as above, i.e., KRAB-ZNFs and TEs that evolved after the emergence of Simiiformes were denoted as young. Accordingly, we dated the interaction as “young” if both the TE and the KRAB-ZNF were young. Most young negative TE:KRAB-ZNF were found in the primary and secondary cortices. We found that young negative TE:KRAB-ZNF were far more numerous in the human primary and secondary cortices compared to the three other species (**Figure 3B** and **Figure S4**). Interestingly, a large percentage of these young negative TE:KRAB-ZNF were human-specific (**Figure 3C**). For instance, in the primary and secondary cortices, 83.92% of the young negative TE:KRAB-ZNF were only found in humans. To exclude the possibility that these results were caused by KRAB-ZNF genes and TEs expressed only in humans, we also accounted for KRAB-ZNF genes and TEs that were expressed in humans and at least one NHP. Under this condition, human primary and secondary cortices still exhibit the highest proportion of young negative TE:KRAB-ZNF (76.83%), compared to the percentage in chimpanzees (63.13%), bonobos (59.03%), and macaques (72.21%). Thus, we conclude that young TEs and young KRAB-ZNF genes seem to negatively interact more often in humans than in NHP, at least in the brain. An example of a human-specific negative TE:KRAB-ZNF is AluYa5:*ZNF658* (**Figure 3D**). *ZNF658* has been reported to regulate gene expression in response to cellular zinc in humans (Ogo et al. 2015; Francis et al. 2019) and AluYa5 mediates genetic disorders (Kim et al. 2016) such as via *ACE* in Alzheimer’s Disease (Wu et al. 2013) and via *ALMS1* in Alström Syndrome (Taşkesen et al. 2012).

In all brain regions, the majority of TEs that were differentially expressed were negatively correlated with at least one KRAB-ZNF protein (over 80% in all brain regions with the exception of the thalamus and hypothalamus, where it was only over 55%; **Figure S5**). In the human primary and secondary cortices, almost all (92.45%) TEs that were differentially expressed and correlated with KRAB-ZNF genes were expressed at lower levels than non-differentially expressed TEs (**Figure 3E**). The pair LTR3B:*ZNF224* is an illustrative example (**Figure 3F**). They are negatively correlated with each other in both humans and chimpanzees. While *ZNF224* is up-regulated in humans compared to chimpanzees, LTR3B shows down-regulation in humans relative to chimpanzees.

Bringing it all together, the human brain had more negative TE:KRAB-ZNF than the brain of NHPs, and most of the TEs of those pairs were significantly lower expressed in humans, suggesting more repressing control from human KRAB-ZNFs compared to KRAB-ZNFs of NHPs.

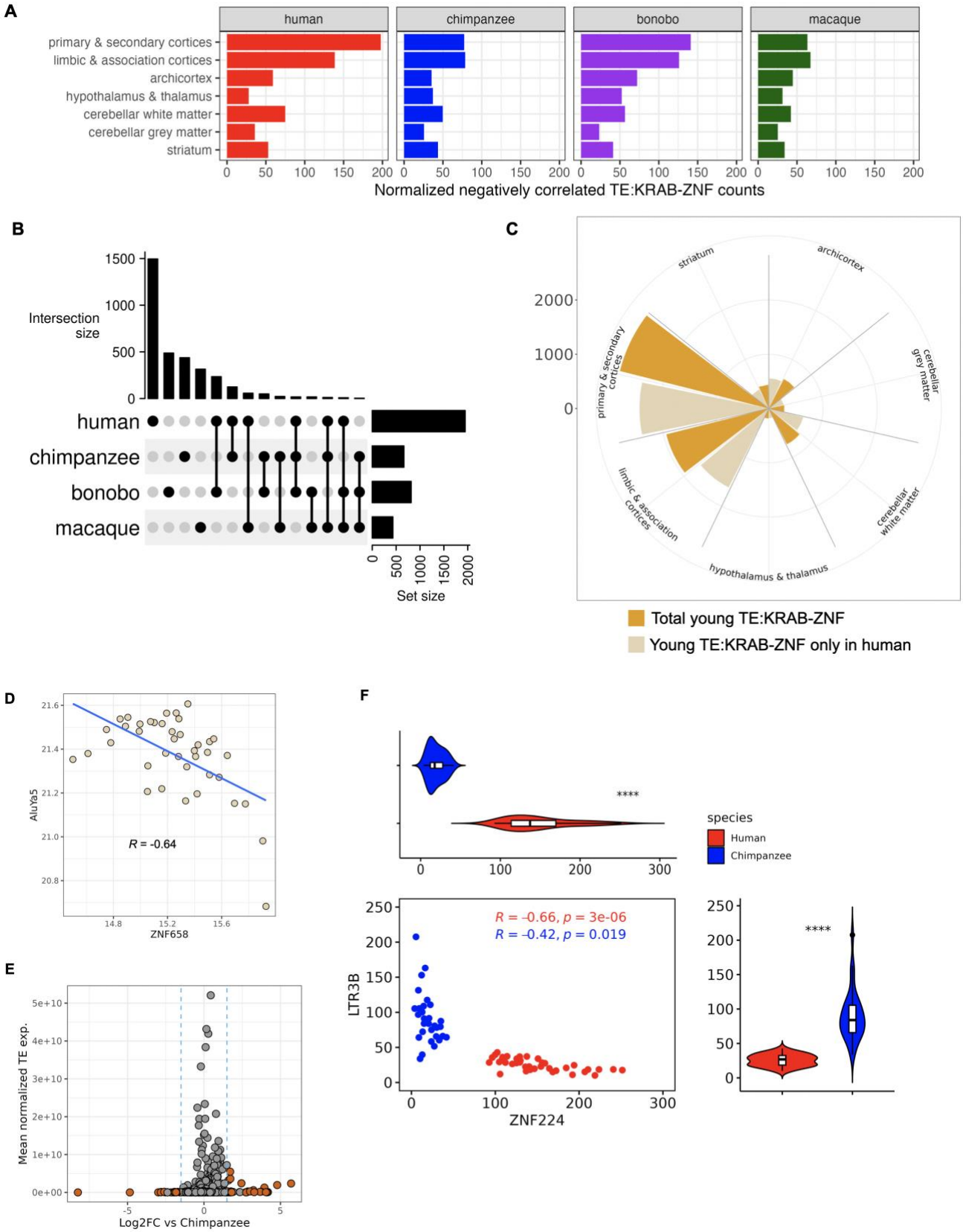


Figure 3. Negative TE:KRAB-ZNF in Primate Brain Data. (A) Negative TE:KRAB-ZNF counts were labeled by each brain region across species. Counts were normalized by all the possible combinations of KRAB-ZNF genes and TEs. **(B)** Overlap percentage and count of young negative TE:KRAB-ZNF

comparing human and NHPs in primary and secondary cortices. Full comparisons including all negative TE:KRAB-ZNF in **Figure S4. (C)** Distribution of young negative TE:KRAB-ZNF across different brain regions in humans. A relatively high percentage of young negative TE:KRAB-ZNF was only detected in humans but not in other NHPs. **(D)** An example of the negative TE:KRAB-ZNF pairs (AluYa5: *ZNF658*) showing the negative correlation in human primary and secondary cortices. **(E)** Humans TEs that were differentially expressed and correlated compared to chimpanzees with at least one KRAB-ZNF gene show lower expression levels. **(F)** Gene and TE expression difference between humans and chimpanzees. The upper panel shows the gene expression of *ZNF224*, and the right panel shows the expression of LTR3B, which are differing between two species. The main panel displays the merged gene expression and correlation in LTR3B:*ZNF224*.

To gain a more systematic understanding of the relationship between TEs and KRAB-ZNF genes, we constructed bipartite networks of negative TE:KRAB-ZNF from the expression data of the primary and secondary cortices for each species. One partition was the TEs while the other partition was the KRAB-ZNF genes. The resulting network thus only consists of links between TEs and KRAB-ZNF genes but not of links connecting TEs with TEs or KRAB-ZNF genes with KRAB-ZNF genes. Comparing those networks across species, we revealed that the human and chimpanzee networks were the most similar to each other (**Figure S6**). We proceeded to compare the negative TE:KRAB-ZNF networks of the primary and secondary cortices between humans and chimpanzees. For that, we first analyzed the subset of KRAB-ZNF genes and TEs, which were human-specifically changed (subset in **Figure 2E**) to construct the negative TE:KRAB-ZNF network for both humans and chimpanzees (**Figure 4A** and **4B**). Interestingly, connectivity was much higher in humans compared to the chimpanzees network. We questioned whether this was a species difference or related to the fact that we selected the KRAB-ZNF genes and TEs that had changed human-specifically. Thus, we created a second subset of chimpanzee-specifically changed

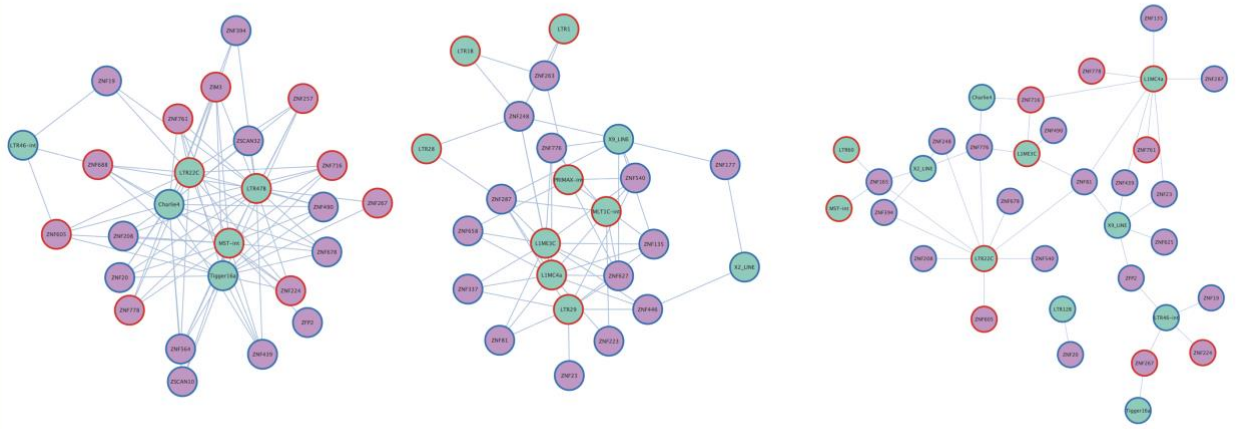
KRAB-ZNF genes and TEs to construct their networks (**Figure 4C** and **4D**). Even though this network is very small, because not many KRAB-ZNF genes and TEs are chimpanzee specifically expressed, we do not observe higher connectivity in the chimpanzee network (**Figure 4C** and **Figure 4D**), indicating that the higher connectivity in humans as observed in **Figure 4A** is a human specific feature.

We further investigated the sub-networks of primate-specific Alu elements and SVAs with their correlated KRAB-ZNF genes in humans by only utilizing human-specific negative TE:KRAB-ZNF. We found that Alu elements and SVAs have more links than KRAB-ZNF genes, meaning that they serve as hubs and that, on average, these TEs interact with more KRAB-ZNFs than KRAB-ZNFs interact with TEs (**Figure 4E** and **4F**).

Part I: Human-specifically-changed subset

A. Human negative TE:KRAB-ZNF network

B. Chimpanzee negative TE:KRAB-ZNF network



Part II: Chimpanzee-specifically-changed subset

C. Human negative TE:KRAB-ZNF network

D. Chimpanzee negative TE:KRAB-ZNF network

Part III: Human-specific negative TE:KRAB-ZNF network

E. Alu

F. SVAs

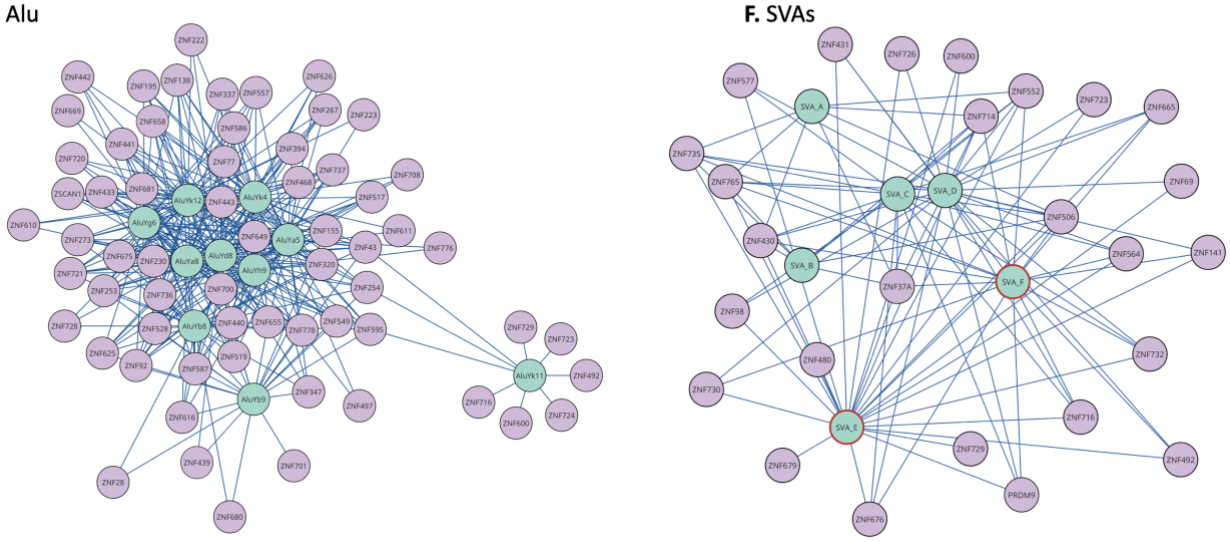


Figure 4. Negative TE:KRAB-ZNF regulatory network in primary and secondary cortices. Part I: A human-specifically-changed subset of KRAB-ZNF genes and TEs was used (subset in **Figure 2E**) to create the regulatory network in **(A)** and **(B)**. **(A)** Human-specifically-changed subset in humans and **(B)** human-specifically-changed subset in chimpanzees. KRAB-ZNF nodes were colored purple, and TE nodes were green. The red border indicated this gene or TE was up-regulated in humans, and the blue was down-regulated in humans. **Part II:** A chimpanzee-specifically-changed subset of KRAB-ZNF genes and TEs was used to create a regulatory network in **(C)** human samples and **(D)** chimpanzee samples. The red border was for up-regulated genes and TEs in chimpanzees, and the blue border was indicating down-regulation in chimpanzees. **Part III: (E)** Alu:KRAB-ZNF network with only human-specific negative TE:KRAB-ZNF **(F)** SVA:KRAB-ZNF network with only human-specific negative TE:KRAB-ZNF. Human-specific SVA_E and SVA_F were colored with a red border.

Mayo Data

1. Variance in TE and KRAB-ZNF gene expression: Lower variances observed in disease conditions than in brain regions among control and AD patients

Several KRAB-ZNFs and TEs with human-specific expression patterns or correlations have been associated with AD. For example, *ZNF267*, which was up-regulated in the human cortex compared to NHPs (**Figure 2E**), is a clear transcriptomic signature for the diagnosis of AD (Fehlbaum-Beurdeley et al. 2012) and AluYa5 expression leads to genetic dysregulation in AD (Kim et al. 2016). We thus hypothesized that the network of KRAB-ZNFs and TEs might be severely altered in the brains of AD patients. We utilized the Mayo Data to test this hypothesis.

To first observe the effect on expression variances, we conducted t-SNE analysis using the expression levels of KRAB-ZNF genes and TEs of the temporal cortex and cerebellum of human control individuals and AD patients. Results showed that variances were primarily influenced by brain regions rather than conditions (**Figure 5A**). Similar to what was mentioned for the Primate Brain Data (**Figure 2C**), we found

that evolutionary young KRAB-ZNF genes and young TEs were expressed at lower levels than their older counterparts (**Figure 5B**). Comparing expression levels between AD and control samples, we obtained for the temporal cortex six KRAB-ZNF genes and four TEs that were up-regulated in AD and six TEs that were down-regulated in AD (**Figure 5C**). In the cerebellum, there were six up-regulated and one down-regulated TEs, and four down-regulated and ten up-regulated KRAB-ZNF genes (**Figure 5D**).

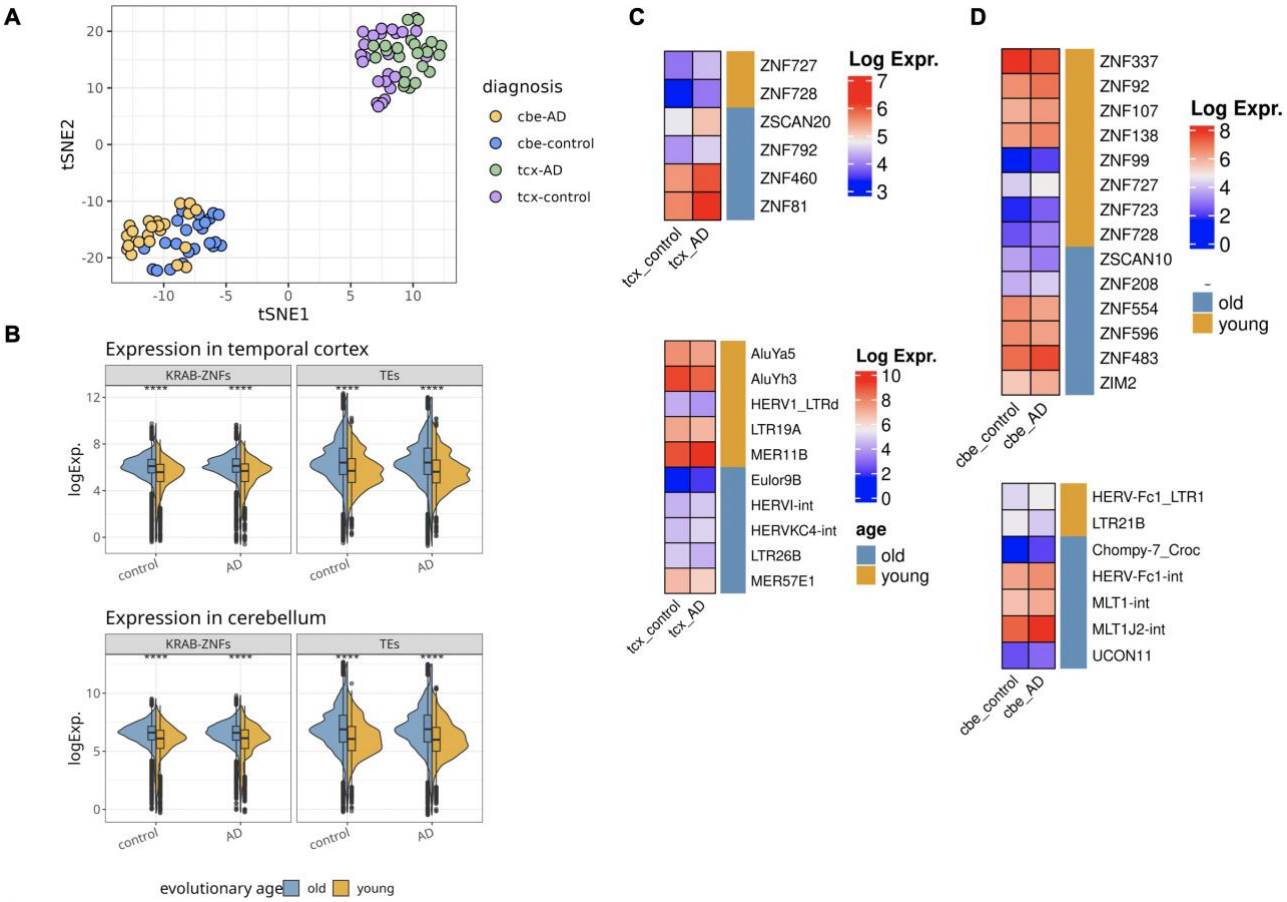


Figure 5. Expression of TEs and KRAB-ZNF genes in Mayo Data. (A) Variations in the expression of KRAB-ZNF genes and TEs using t-SNE analysis. **(B)** Distributions of the expression of old (> 44.2 mya) and young (<= 44.2 mya) KRAB-ZNF genes and TEs expression. **(C)** and **(D)** Differentially expressed KRAB-ZNF genes and TEs (|log2FoldChange| > 0.5, $p < 0.05$) in temporal cortex (tcx) and cerebellum (cbe). The expression of KRAB-ZNF genes and TEs in cerebellum used the same log expression scale.

2. Human-specific correlations limited to the healthy human temporal cortex: 2492 negative TE:KRAB-ZNFs not detected in AD

To examine the negative correlations between KRAB-ZNF genes and TEs in the Mayo Dataset, we compared negative TE:KRAB-ZNF pairs between control and AD in the cerebellum and temporal cortex. More negative TE:KRAB-ZNFs were found in the control group in both brain regions (**Figure 6A**). For instance, in the temporal cortex, there were 37781 negative TE:KRAB-ZNF in control and 28121 negative TE:KRAB-ZNF in AD, of which 10039 overlapped between both conditions, indicating that several negative TE:KRAB-ZNF relationships got “lost” in AD progression.

To reduce the number of dataset-specific negative TE:KRAB-ZNF and falsely inferred negative TE:KRAB-ZNF for healthy human brains, we next only considered negative TE:KRAB-ZNF that were also found in the human expression data of the primary and secondary cortices of the Primate Brain Dataset. This list was then further refined by excluding negative TE:KRAB-ZNF found in AD or in any other species (**Figure 6B**). This way, we received 2492 negative TE:KRAB-ZNF that seemed to be specific to healthy human brains (**Figure 6B**) because they were either not significantly correlated (**Figure 6C**) or positively correlated in NHP and AD brains (**Figure 6D**). The most abundant TE families among those TEs with healthy human-specific links to KRAB-ZNFs are ERVs, L1s, and Alu elements (**Figure 6B**). Alu elements, ERVs, hAT elements, L1, and SVAs showed most often also differential expression between species. Interestingly, all Alu elements and SVAs had lower expression in AD relative to control samples, while the other TE families included both up-regulated and down-regulated TEs (**Figure 6E**). Overall, more TEs seemed to be downregulated in AD, which is mirrored by a trend toward more upregulation of KRAB-ZNFs in AD.

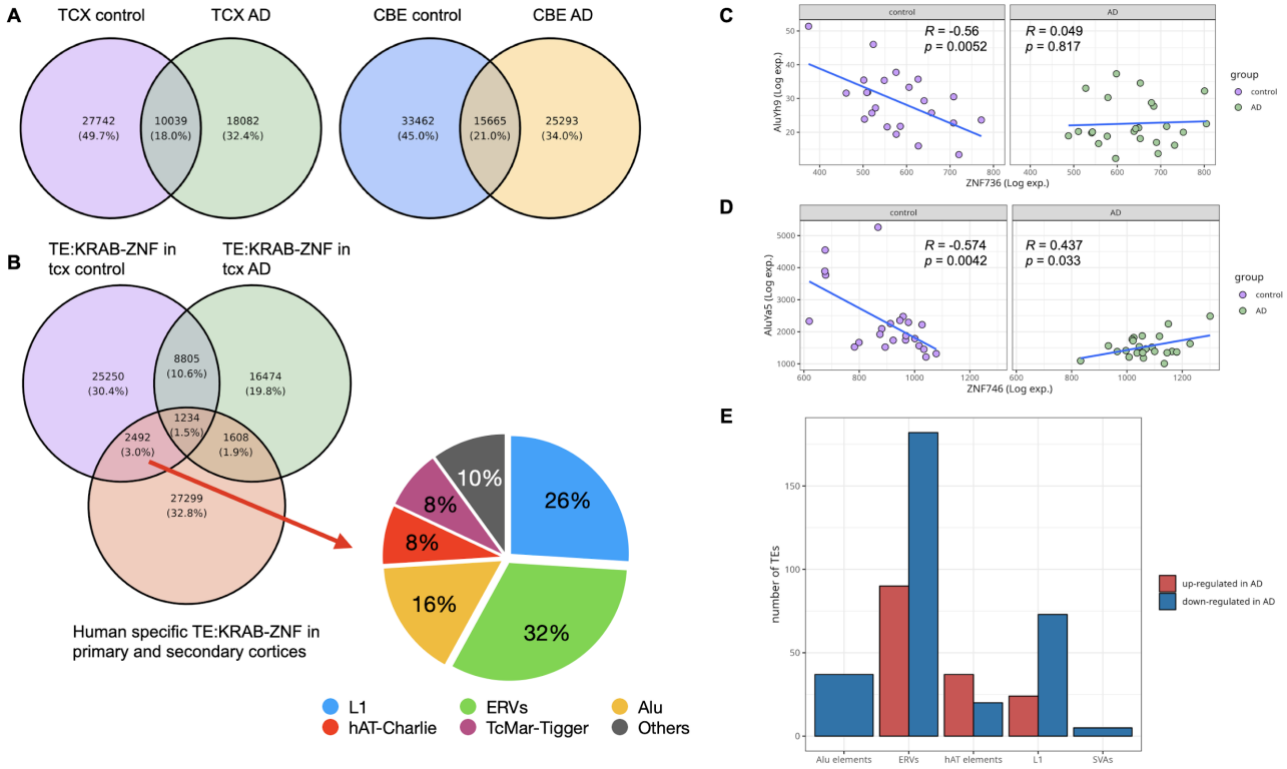
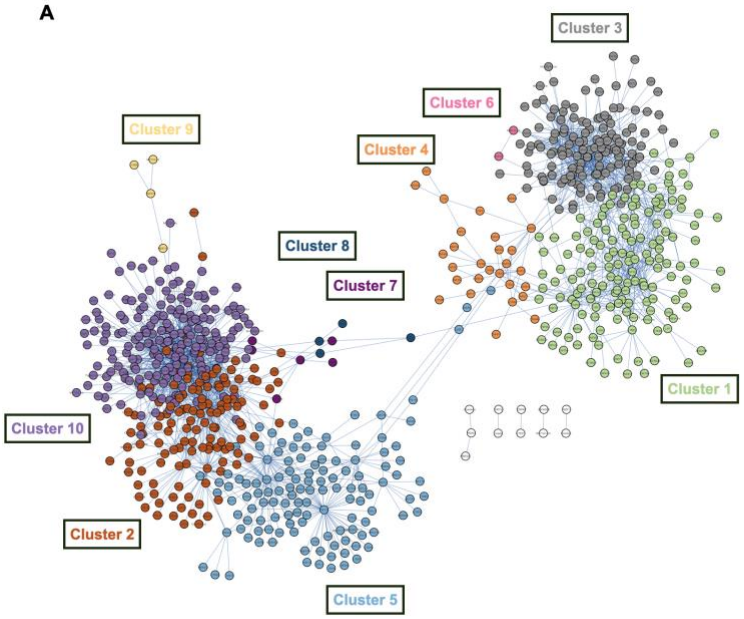


Figure 6. Negative TE:KRAB-ZNF and the "lost" negative TE:KRAB-ZNF in temporal cortex. (A) Overlapped percentage and count of negative TE:KRAB-ZNF of control and AD samples in temporal cortex (tcx) and cerebellum (cbe)(overlaps between two brain regions in **Figure S7**). **(B)** Percentage of overlapping human-specific negative TE:KRAB-ZNF found in control, AD from Mayo Data and human specific negative TE:KRAB-ZNF of the primary and secondary cortices areas from the Primate Brain Data. The subset of 2492 negative TE:KRAB-ZNF is unique to humans and specific to the healthy human brain. Below is the composition of their TE families. **(C)** AluYh9:*ZNF736* was negatively correlated in the control sample but not significantly correlated in the AD sample from the temporal cortex. **(D)** AluYa5:*ZNF746* was negatively correlated in control but positively correlated in AD from the temporal cortex. **(E)** Distribution of differentially expressed TEs from the 2492 healthy human-specific negative TE:KRAB-ZNF subset.

To study the interplay of the KRAB-ZNF genes and TEs of the 2492 healthy human-specific negative TE:KRAB-ZNF, we constructed a bipartite network based on their 239 KRAB-ZNFs and 551 TEs, derived from the 2492 human-specific negative TE:KRAB-ZNF interactions. To analyze the local structures of the network, we calculated their bipartite modularity for partitioning nodes and obtained 10 clusters based on similar modularity (**Figure 7A**). Clusters 2, 5, and 10 have higher numbers of TEs, and clusters 1, 3, and 4 have higher numbers of KRAB-ZNFs (**Figure 7A** and **Figure S8**). All TEs from the Alu subfamilies, except FLAM_A and FLAM_C, belong to cluster 3 (**Figure 7A** and **Figure S8**). The composition of young and old KRAB-ZNF nodes was significantly different between clusters (Fisher's exact pairwise test, p = 1e-06; **Figure 7C**, **Figure S8** and **Table S5**). To test whether young KRAB-ZNFs preferentially interact with young TEs, we calculated the assortativity within the network. We found only weak support for assortativity in each cluster (**Table S6**), indicating no preference for young KRAB-ZNFs repressing young TEs in that network.

We selected 40 hubs from the whole network, which have the top 5% highest degree and strength centrality. These hubs include 39 KRAB-ZNF genes and 1 TE. We also identified "gatekeepers", which have the top 5% highest betweenness centrality. There were 14 TEs and 26 KRAB-ZNFs among these gatekeepers, of which 19 were also classified as hubs in the network (**Figure 7A**). These 19 KRAB-ZNF genes, together with AluYa5, are thus likely important players in the intricate network of negative TE:KRAB-ZNFs in healthy human temporal cortex.



Clusters	Hubs	Gatekeepers
1	**ZNF569**, ZNF627, ZNF658, **ZNF33A**	L1M3c, L1MA5A, X8_LINE, ZKSCAN1, ZNF197, ZNF232, **ZNF33A**, **ZNF569**
2	**ZNF37A**, **ZNF114**, **ZNF91**, ZNF599, **ZNF431**, **ZNF430**, ZNF761	**ZNF114**, ZNF3, **ZNF37A**, **ZNF430**, **ZNF431**, **ZNF91**
3	ZNF454, ZNF324, ZNF460, ZNF169, **ZNF567**, ZFP1, ZNF468, ZNF736, **ZNF655**, **AluYa5**	**AluYa5**, AluYb8, AluYd8, **ZNF567**, **ZNF655**
5	**ZNF573**, **ZNF735**, **ZNF124**, **ZNF730**	FLAM_C, LTR10G, LTR32, MER57B1, **ZNF124**, ZNF493, ZNF550, **ZNF573**, **ZNF730**, **ZNF735**
8	-	L1ME3C, Tigger12, ZNF738
10	**ZNF432**, **ZNF20**, **ZNF205**, ZNF785, ZNF577, ZNF688, **ZNF69**, **ZNF141**, ZNF765, ZNF506, ZNF552, ZNF732, ZNF726, ZNF600	MamRep434, MER94B, **ZNF141**, **ZNF20**, **ZNF205**, **ZNF432**, ZNF492, **ZNF69**

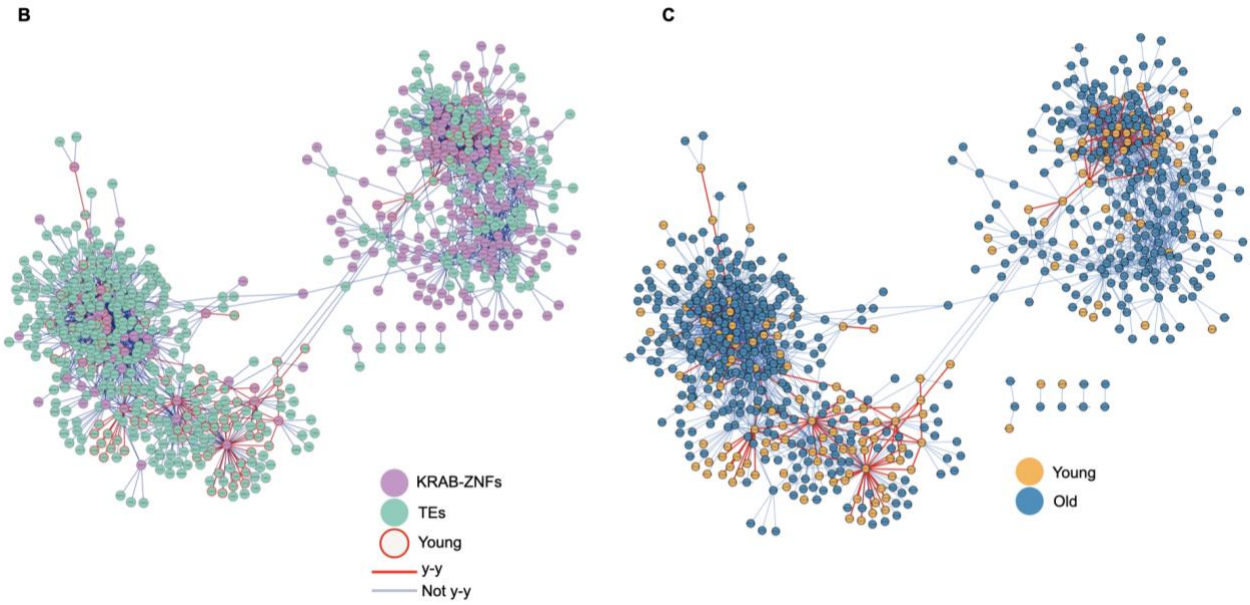


Figure 7. 2492 healthy human-specific negative TE:KRAB-ZNF regulatory network in temporal cortex. Bipartitate network of human-specific negative TE:KRAB-ZNF that were only detected in control but not in AD brains and only in human but not NHP brains **(A)** Clusters were calculated based on bipartite modularity. Hubs and gatekeepers of each cluster are listed in the right-hand table. Nodes that are hubs and gatekeepers are shown in bold font. A larger version of this figure with node names can be found in **Figure S9**. **(B)** KRAB-ZNFs and TEs are colored in purple and green, respectively. Nodes of young KRAB-ZNF and young TE are marked with a red border. Two young nodes are connected with a red link (y-y). **(C)** Young KRAB-ZNFs and TEs are shown in yellow, while old ones are shown in blue.

Discussion

From our systematic examination of KRAB-ZNF genes and TEs within primate brains, we revealed several intriguing patterns and findings. Firstly, high variations in the expression of KRAB-ZNF genes and TEs between species suggests a prominent evolutionary signature in the regulation of KRAB-ZNF genes and TEs in primates. Secondly, younger members of both KRAB-ZNF genes and TEs exhibit significantly lower expression levels compared to older members, indicating that newly emerged genetic elements might still be on the raise to their full expression potential or might be subjected to more stringent regulatory mechanisms, potentially avoiding disrupting impact at the organismal level (**Figure 2C**). Thirdly, we found that humans had more negative TE:KRAB-ZNF pairs, which results in higher connectivity in the human negative TE:KRAB-ZNF network compared to the networks of NHPs (**Figure 4**). This observation is in line with previous findings that humans also showed higher connectivity in the transcription factor networks of the brain compared to NHPs (Berto and Nowick 2018). A substantial portion of young negative TE:KRAB-ZNF seems to be human-specific, pointing to recent additions to the negative TE:KRAB-ZNF network in the human lineage and presumably an increase in the complexity of gene regulation (**Figure 3C**). Furthermore, most of the differentially expressed TEs were found to be correlated with at least one KRAB-ZNF gene and expressed at a significantly lower level compared to non-differentially expressed TEs (**Figure 3E**). The network of negative TE:KRAB-ZNF is considerably changed in AD brains, displaying fewer links and potential loss of control of TE expression. These collective findings suggest an important role of the negative TE:KRAB-ZNF network in shaping the evolution and in maintaining the functionality of the human brain and particularly highlights rapid

rewiring involving young KRAB-ZNFs and young TEs.

Given the generally accepted role of KRAB-ZNF proteins as repressors, our primary focus in this study was on negative correlations between KRAB-ZNF genes and TEs. However, we also observed numerous positive correlations between these genetic players. Keeping in mind, that correlations do not indicate direction and can also be caused by indirect relationships, several plausible explanations exist for these positive interactions. Firstly, some TEs can function as *cis*-acting regulatory elements capable of influencing the expression of nearby genes, which have been mentioned to be tissue-specifically co-opted TEs (Sentmanat and Elgin 2012; Wolf et al. 2020; Coronado-Zamora and González 2023). Additionally, it is conceivable that the expression levels of KRAB-ZNF genes themselves are influenced by their ability to sense the presence and activity of TEs within the genome (Pontis et al. 2019; Senft and Macfarlan 2021). Another perspective is that primate-specific KRAB-ZNFs can bind to gene promoters, regulating gene expression in the primate brain without necessarily targeting TEs (Farmiloe et al. 2020); observed correlations could then stem from indirect regulation. We propose that a comprehensive study of the regulatory networks, including positive correlations and gene promoter binding sites, would contribute to a deeper understanding of not only the interplay between KRAB-ZNFs and TEs but also the complexity of the roles of KRAB-ZNFs in primate evolution.

TEs have also been observed to be dysregulated in certain diseases. Here, we utilized gene expression data from AD patients to examine differences in the regulation of KRAB-ZNF genes and TEs. Notably, we observed that the variances in expression levels were more distinct across various brain regions than between control and AD

samples (**Figure 5A**). This finding points out the challenge of discerning differences between control and AD samples solely through the consideration of differential expression. This is also illustrated by the detection of only six and ten differentially expressed KRAB-ZNFs and TEs, respectively (**Figure 5C**). Upon combing expression data with correlation results and an evolutionary perspective, we determined 2492 human-specific negative TE:KRAB-ZNF that were not detectable in AD brains (**Figure 6B**), suggesting considerable alterations in the intricate interplay between KRAB-ZNFs and TEs. Within the network of the 2492 negative TE:KRAB-ZNF, AluYa5 is the only TE that acts as a hub and gatekeeper. It is from one of the evolutionary youngest Alu subfamilies, which has also been one of the most active Alu subfamilies. Its polymorphic insertions across the genomes contribute noticeably to human genome diversity and divergence from NHPs (Carroll et al. 2001). We identified Alu elements that are down-regulated in AD patients, such as AluYh9 and AluSp, which have been associated with Tau pathology in the human brain (Guo et al. 2018) and AluY and AluSx, which are mediating in-frame deletions on *PSEN1* (Le Guennec et al. 2017). Alu elements have further shown to regulate the expression of *ACE* in neurons from Alu insertions (Wu et al. 2013) and to play a role in A-to-I RNA editing during neurogenesis in the central nervous system and mitochondrial homeostasis (Larsen et al. 2018). The distinct subcluster of Alu elements with their connected KRAB-ZNF genes in the bipartite network (cluster 3 in **Figure 7A**) might contribute to differences in *ACE* expression or A-to-I RNA editing in AD and among primates.

With respect to the arms-race hypothesis (Jacobs et al. 2014), we observed a higher number of negative correlation between young TEs and young KRAB-ZNF genes in humans compared to NHPs. This suggests increased repression from young KRAB-

ZNFs on young TEs in the human brain (**Figure 3C**). However, our findings only weakly support the arms-race hypothesis. Firstly, we noted that young TEs exhibit lower expression levels than old TEs (**Figure 2C** and **Figure 5B**), which might not be expected if they only recently escaped repression. Secondly, some young TEs were also negatively correlated with old KRAB-ZNF genes, leading to weak assortativity regarding age inference. We interpret this finding as indicating that the interplay between the evolution of TE and KRAB-ZNF is more complex than a specific one-to-one relationship.

In response to the need for systematically comparing expression profiles of orthologous genes and TEs across diverse species or conditions, we developed TEKRABber, an R Bioconductor package which equips users to effortlessly adapt its comprehensive pipeline for their analyses. Our pipeline takes care of normalizing expression data across species from transcriptomic data by extending the usage of the scaling-based normalization method (Zhou et al. 2019). To optimize computational efficiency during the calculations of pairwise correlations, we harnessed the capabilities of RCpp (Eddelbuettel and François 2011). Taking these advantages, users can adapt the TEKRABber's pipeline to analyze selected orthologous genes and TEs from their expression data within an acceptable timeframe. Taking one of the examples in this study, calculating pairwise correlations of 284 orthologs with 861 TEs neglecting sample number can be executed in approximately 30 minutes on standard hardware (Intel Core i7 2.6GHz, 16GB memory). Nonetheless, it is crucial to recognize that our present approach is constrained by certain limitations. These constraints primarily stem from variations in gene lengths and positions of the same TE across individual genomes, making it challenging to effectively normalize TE expression levels.

We only used the available reference genomes for our analysis; however, it becomes increasingly clear that substantial individual differences in the presence of TEs exist, which are not covered by a single reference genome. Currently, there are emerging methods designed to detect TE expression, including advancements in sequencing resolution like long-read sequencing (Marx 2023) and tools for de novo TE annotations (Storer et al. 2022; Orozco-Arias et al. 2023). These developments hold promise for achieving a more precise depiction of the variations between samples and the reference, ultimately enhancing our understanding of TE-associated expression patterns. Hence, we acknowledge the significant potential of integrating improved TE orthologous information into our method.

Conclusion

In summary, our findings emphasize the intricate network of negative interactions between KRAB-ZNF genes and TEs in both evolution and neurodegenerative disease. Thus, to achieve a more comprehensive understanding of KRAB-ZNFs and TEs function, it is not enough to only investigate expression levels, but network analysis as facilitated by TEKRABber needs to be leveraged. Evolutionarily, the human brain exhibits a notably denser negative TE:KRAB-ZNF network compared to non-human primates, particularly in more recently evolved TEs and KRAB-ZNFs. The healthy human brain negative TE:KRAB-ZNF network contains a distinct cluster exclusively comprising Alu elements, which is an evolutionary novelty that is lost in AD brains. These insights emphasize the nuanced dynamics of negative TE:KRAB-ZNF interactions and their relevance in both evolutionary and disease contexts.

Materials and Methods

1. Primate Brain Data

For comparing humans with NHPs, we used published RNA-seq data (GSE127898) including 422 brain samples from biological replicates that consisted of 4 humans, 3 chimpanzees, 3 bonobos, and 3 macaques (**Table 1**). Samples from each individual were from 33 different brain regions, and total RNA was sequenced on the Illumina HiSeq 4000 system with a 150 bp paired-end sequencing protocol. More details about samples and the preparation steps can be found in **Table S1** and (Khrameeva et al. 2020). RNA-seq FASTQ data on Gene Expression Omnibus (Barrett et al. 2012) were retrieved using NCBI SRA Toolkit v3.0.3.

2. Mayo Data

The Mayo RNA-seq study (Allen et al. 2016) was utilized to compare human control and Alzheimer's disease samples. Total RNA was sequenced from both control and AD samples collected from the temporal cortex and cerebellum. The reprocessed FASTQ files, which include only control and AD samples in the temporal cortex and cerebellum, were downloaded via Synapse consortium studies (accession: syn8612203 and syn8612213). The number of samples can be found in **Table 1** and more detailed, including biological sex, age, and Braak staging is in **Table S3.**

Table1. Datasets

Dataset	Categories (Biological replicates)	Total number of samples
Primate Brain Data	Human (4)	132
(GSE127898)	Chimpanzee (3)	96
	Macaque (3)	96
	Bonobo (3)	98
Mayo Data	Control cerebellum (cbe)	23
(syn5550404)	Control-temporal cortex (tcx)	23
	AD cerebellum (cbe)	22
	AD temporal cortex (tcx)	24

3. Transcriptome analysis

Adapters and low-quality reads were removed from the FASTQ files using fastp (0.12.4). The selected FASTQ files from both datasets were then mapped to their respective references downloaded from UCSC Table Browser, including hg38, panTro6, panPan3, and rheMac10 (Karolchik 2004), using STAR (2.7.10b) (Dobin et al. 2013) with the parameters `--outFilterMultimapNmax 100 --withAnchorMultimapNmax 100`. These parameters were chosen to increase the likelihood of capturing TE mapping by allowing for multiple alignments of reads and more loci anchors for mapping. The resulting BAM files were used to quantify counts of TEs and genes using TEtranscripts (2.2.3) (Jin et al. 2015) with the `--sortByPos` parameter to determine the expression levels of genes and TEs. Gene and TE indices were created using UCSC gene annotations and the RepeatMasker track, enabling reads to match with these intervals. If a read was overlapped with both a gene exon and a TE, it would be checked to determine whether it had a unique alignment or multiple locations in the genome. If an annotation existed, the uniquely aligned read

was assigned to the gene. Otherwise, it was assigned to the TE. For reads with ambiguous mappings, they were evenly weighted across TE or gene annotations using the Expectation Maximization (EM) algorithm.

4. Inferences about the evolutionary age of KRAB-ZNFs and TEs

A total of 337 expressed KRAB-ZNF genes were detected by referencing the comprehensive KRAB-ZNF catalog (Huntley et al. 2006). The annotation provided by GenTree (Shao et al. 2019) was utilized to infer the evolutionary age of KRAB-ZNF genes. GenTree employed a synteny-based pipeline to date primate-specific protein-coding genes and presented a branch view to depict the origin of these genes. To complement the age inference of KRAB-ZNF genes, which lacked direct dating annotations in GenTree, we referred to the annotations of primate orthologs in 27 primates, including humans (Jovanovic et al. 2021). By incorporating this complementary age information, we obtained the age inference of all 337 KRAB-ZNF genes for our downstream analysis (**Figure S1**). The evolutionary age inference of TEs was taken from Dfam (Storer et al. 2021) by extracting the species information for each TE subfamily (**Figure S2**).

5. Development of pipeline software, TEKRABber

To compare the expression of orthologous genes and TEs across species, we concatenated steps including normalization, differential expression, and correlation analysis into an R Bioconductor package, TEKRABber. The name was derived from the idea that TEs are bound (“grabbed”) by KRAB-ZNF proteins. TEKRABber adapted the scale-based normalization method (Zhou et al. 2019) to normalize orthologous genes for comparison between two species. In brief, the conserved orthologous gene

lengths from two species were selected from Ensembl data (Cunningham et al. 2022) and combined with the expression data to find an optimal scaling factor that can normalize the data to achieve a minimization of deviation between empirical and nominal type I error in a hypothesis testing framework. The normalization step for TEs followed a similar concept. However, instead of using orthologous genes, we used the subset of TEs including LTR, LINE, SINE, SVAs, and DNA transposons (as defined by RepeatMasker (Smit et al. 2013)), for which homologues can be found in other species, for scaling to normalize the expression of TEs. After normalization, differentially expressed orthologous genes and TEs were analyzed using DESeq2 (Love et al. 2014). The one-to-one correlations between selected orthologous genes and TEs from each species were estimated. In this analysis, we used Pearson's correlations with an adjusted *p-value* using the Benjamini-Hochberg correction (Benjamini and Hochberg 1995) to obtain significant correlations.

6. Differentially expressed genes and TEs and their correlations

In this study, all differentially expressed genes and TEs were quantified using DESeq2 (Love et al. 2014) utilizing an absolute log2FoldChange threshold greater than 1.5 (adjusted $p < 0.05$). However, the Mayo Data adopted an absolute log2FoldChange threshold greater than 0.5 due to a lower number of detected differentially expressed genes and TEs. Heatmaps and upset plots were visualized using ComplexHeatmap (Gu et al. 2016; Gu 2022). For notation, we employ negative TE:KRAB-ZNF to signify a negative correlation between a TE and a KRAB-ZNF gene. For instance, AluYa5:*ZNF658* indicates a detected negative correlation between AluYa5 and *ZNF658*.

7. Constructing negative TE:KRAB-ZNF regulatory network

Nodes and edges were first selected from the negative TE: KRAB-ZNF results and processed into a dataframe object. The network structure was created using RCy3 (v2.16) to import into Cytoscape (v3.10.1) for visualization and editing. To visualize the regulatory networks, we applied the yFiles Organic Layout for an undirect graph by assigning nodes as objects that had repulsive or attractive forces between them (https://www.yworks.com).

8. Network analysis

Previously constructed negative TE:KRAB-ZNF networks were analyzed as bipartite networks, consisting of two classes of nodes, KRAB-ZNF genes and TEs, with connections existing only between the two classes and no edges between elements from the same class. Node properties were first checked using a bipartite module in networkx (v2.8.4) (Hagberg et al. 2008). These values include bipartite degree centrality, bipartite strength centrality, and bipartite betweenness centrality for each KRAB-ZNF and TE node. The top 5% scores were selected as hubs (bipartite degree and bipartite strength centrality) and gatekeepers (bipartite betweenness centrality). For cluster detection, leading eigenvector community methods (Csárdi et al. 2023) were used to specify unipartite community structures, and then bipartite modularity (Barber 2007) was calculated with CONDOR (v1.1.1) (Platig et al. 2016). To conduct enrichment analysis among clusters, a Fisher exact test using 1 million simulations with a p-value < 0.05 was used.

Software and Code Availability

TEKRABber: https://bioconductor.org/packages/release/bioc/html/TEKRABber.html

Source code: https://github.com/ferygood/primateBrain_TEKRABZNF

Acknowledgement

1. We extend our sincere thanks to the data provider and all members of the research team involved in The MayoRNAseq study. We also appreciate the invaluable data resources provided (https://doi.org/10.7303/syn5550404), which have supported the progression of this research.
2. Our sincere thanks go to Vladimir M. Jovanovic, Rebecca Saager, Jeong-Eun Lee, Fatemeh Zebardast, and Melanie Sarfert for their contributions through critical reading and feedback.

Author Contributions

Conceptualization, KN and YC; Methodology, YC and AM; Visualization, YC and AM; Writing, KN, YC and AM; Data Curation, YC; Software Development, YC; Formal Analysis, YC and AM; Supervision, KN; Project Administration, KN; Funding Acquisition, KN. All authors edited and reviewed the manuscript.

Funding

This work supported by funding from the Deutsche Forschungsgemeinschaft awarded to KN (NO 920/8-1) and by the SPP “Die Genomischen Grundlagen Evolutionärer Innovationen (GEvol)” (NO 920/10-1).

References

Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, Burgess JD, Chai H-S, Crook J, Eddy JA, et al. 2016. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data* 3:160089.

Barber MJ. 2007. Modularity and community detection in bipartite networks. *Phys. Rev. E* 76:066102.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. 2012. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* 41:D991–D995.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57:289–300.

Berto S, Nowick K. 2018. Species-Specific Changes in a Primate Transcription Factor Network Provide Insights into the Molecular Evolution of the Primate Prefrontal Cortex.O'Connell M, editor. *Genome Biology and Evolution* 10:2023–2036.

Bodea GO, McKelvey EGZ, Faulkner GJ. 2018. Retrotransposon-induced mosaicism in the neural genome. *Open Biol.* 8:180074.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol* 19:199.

Carroll ML, Roy-Engel AM, Nguyen SV, Salem A-H, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, et al. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *Journal of Molecular Biology* 311:17–40.

Colonna Romano N, Fanti L. 2022. Transposable Elements: Major Players in Shaping Genomic and Evolutionary Patterns. *Cells* 11:1048.

Coronado-Zamora M, González J. 2023. Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across *Drosophila* natural strains. *Genome Res.*:genome;gr.277565.122v1.

Csárdi G, Nepusz T, Müller K, Horvát S, Traag V, Zanini F, Noom D. 2023. igraph for R: R interface of the igraph library for graph theory and network analysis. Available from: https://zenodo.org/record/7682609

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. 2022. Ensembl 2022. *Nucleic Acids Research* 50:D988–D995.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Development* 144:2719–2729.

Eddelbuettel D, François R. 2011. **Rcpp** : Seamless *R* and *C*++ Integration. *J. Stat. Soft.* [Internet] 40. Available from: http://www.jstatsoft.org/v40/i08/

Evering TH, Marston JL, Gan L, Nixon DF. 2023. Transposable elements and Alzheimer's disease pathogenesis. *Trends in Neurosciences* 46:170–172.

Farmiloe G, Lodewijk GA, Robben SF, Van Bree EJ, Jacobs FMJ. 2020. Widespread correlation of KRAB zinc finger protein binding with brain-developmental gene expression patterns. *Phil. Trans. R. Soc. B* 375:20190333.

Fehlbaum-Beurdeley P, Sol O, Désiré L, Touchon J, Dantoine T, Vercelletto M, Gabelle A, Jarrige A-C, Haddad R, Lemarié JC, et al. 2012. Validation of

AclarusDx™, a Blood-Based Transcriptomic Signature for the Diagnosis of Alzheimer's Disease. *JAD* 32:169–181.

Francis M, Cheng H, Ma P, Grider A. 2019. Genomic Characterization of the Zinc Transcriptional Regulatory Element Reveals Potential Functional Roles of ZNF658. *Biol Trace Elem Res* 192:83–90.

Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G, Dénervaud N, Bucher P, Trono D. 2010. KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading.Madhani HD, editor. *PLoS Genet* 6:e1000869.

Gu Z. 2022. Complex heatmap visualization. *iMeta* 1:e43.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847–2849.

Guo C, Jeong H-H, Hsieh Y-C, Klein H-U, Bennett DA, De Jager PL, Liu Z, Shulman JM. 2018. Tau Activates Transposable Elements in Alzheimer's Disease. *Cell Reports* 23:2874–2880.

Hagberg AA, Schult DA, Swart PJ. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th Python in Science Conference. Pasadena, CA USA. p. 11–15.

Hancks DC, Kazazian HH. 2010. SVA retrotransposons: Evolution and genetic instability. *Seminars in Cancer Biology* 20:234–245.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* 16:669–677.

Imbeault M, Helleboid P-Y, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543:550–554.

Iouranova A, Grun D, Rossy T, Duc J, Coudray A, Imbeault M, de Tribolet-Hardy J, Turelli P, Persat A, Trono D. 2022. KRAB zinc finger protein ZNF676 controls the transcriptional influence of LTR12-related endogenous retrovirus sequences. *Mobile DNA* 13:4.

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

Jin Y, Tam OH, Paniagua E, Hammell M. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31:3593–3599.

Jovanovic VM, Sarfert M, Reyna-Blanco CS, Indrischek H, Valdivia DI, Shelest E, Nowick K. 2021. Positive Selection in Gene Regulatory Factors Suggests Adaptive Pleiotropic Changes During Human Evolution. *Front. Genet.* 12:662239.

Karolchik D. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research* 32:493D – 496.

Khrameeva E, Kurochkin I, Han D, Guijarro P, Kanton S, Santel M, Qian Z, Rong S, Mazin P, Sabirov M, et al. 2020. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* 30:776–789.

Kim S, Cho C-S, Han K, Lee J. 2016. Structural Variation of *Alu* Element and Human Disease. *Genomics Inform* 14:70.

Larsen PA, Hunnicutt KE, Larsen RJ, Yoder AD, Saunders AM. 2018. Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Res* 26:93–111.

Le Guennec K, Veugelen S, Quenez O, Szaruga M, Rousseau S, Nicolas G, Wallon D, Fluchere F, Frébourg T, De Strooper B, et al. 2017. Deletion of exons 9 and 10 of the Presenilin 1 gene in a patient with Early-onset Alzheimer Disease generates longer amyloid seeds. *Neurobiology of Disease* 104:97–103.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.

Marx V. 2023. Method of the year: long-read sequencing. *Nat Methods* 20:6–11.

Nowick K, Carneiro M, Faria R. 2013. A prominent role of KRAB-ZNF transcription factors in mammalian speciation? *Trends in Genetics* 29:130–139.

Nowick K, Gernat T, Almaas E, Stubbs L. 2009. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci. U.S.A.* 106:22358–22363.

Ogo OA, Tyson J, Cockell SJ, Howard A, Valentine RA, Ford D. 2015. The Zinc Finger Protein ZNF658 Regulates the Transcription of Genes Involved in Zinc Homeostasis and Affects Ribosome Biogenesis through the Zinc Transcriptional Regulatory Element. *Molecular and Cellular Biology* 35:977–987.

Orozco-Arias S, Sierra P, Durbin R, González J. 2023. MCHelper automatically curates transposable element libraries across species. Genomics Available from: http://biorxiv.org/lookup/doi/10.1101/2023.10.17.562682

Patel S, Howard D, Chowdhury N, Derieux C, Wellslager B, Yilmaz Ö, French L. 2021. Characterization of Human Genes Modulated by Porphyromonas gingivalis Highlights the Ribosome, Hypothalamus, and Cholinergic Neurons. *Front. Immunol.* 12:646259.

Platig J, Castaldi PJ, DeMeo D, Quackenbush J. 2016. Bipartite Community Structure of eQTLs.Markowetz F, editor. *PLoS Comput Biol* 12:e1005033.

Playfoot CJ, Duc J, Sheppard S, Dind S, Coudray A, Planet E, Trono D. 2021. Transposable elements and their KZFP controllers are drivers of transcriptional innovation in the developing human brain. *Genome Res.* 31:1531–1545.

Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* 24:724-735.e5.

Ravel-Godreuil C, Znaidi R, Bonnifet T, Joshi RL, Fuchs J. 2021. Transposable elements as new players in neurodegenerative diseases. *FEBS Lett* 595:2733–2755.

Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. *Molecular Ecology* 28:1537–1549.

Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* 22:691–711.

Sentmanat MF, Elgin SCR. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* 109:14104–14109.

Shao Y, Chen C, Shen H, He BZ, Yu D, Jiang S, Zhao S, Gao Z, Zhu Z, Chen X, et al. 2019. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res* 29:682–696.

Shulman JM, Chibnik LB, Aubin C, Schneider JA, Bennett DA, De Jager PL. 2010. Intermediate Phenotypes Identify Divergent Pathways to Alzheimer's Disease.Domschke K, editor. *PLoS ONE* 5:e11244.

Smit, AFA, Hubley, R & Green, P. 2013. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org

Storer J, Hubley R, Rosen J, Smit A. 2022. Methodologies for the De novo Discovery

of Transposable Element Families. *Genes* 13:709.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* 12:2.

Suarez NA, Macia A, Muotri AR. 2018. LINE-1 retrotransposons in healthy and diseased human brain: L1 Activity in the Human Brain. *Devel Neurobio* 78:434–455.

Taşkesen M, Collin GB, Evsikov AV, Güzel A, Özgül RK, Marshall JD, Naggert JK. 2012. Novel Alu retrotransposon insertion leading to Alström syndrome. *Hum Genet* 131:407–413.

Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* 21:1800–1812.

Wolf G, De Iaco A, Sun M-A, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan TS. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife* 9:e56337.

Wu S-J, Hsieh T-J, Kuo M-C, Tsai M-L, Tsai K-L, Chen C-H, Yang Y-H. 2013. Functional regulation of Alu element of human angiotensin-converting enzyme gene in neuron cells. *Neurobiology of Aging* 34:1921.e1-1921.e7.

Yang H, Wang L, Zheng Y, Hu G, Ma H, Shen L. 2022. Knockdown of zinc finger protein 267 suppresses diffuse large B-cell lymphoma progression, metastasis, and cancer stem cell properties. *Bioengineered* 13:1686–1701.

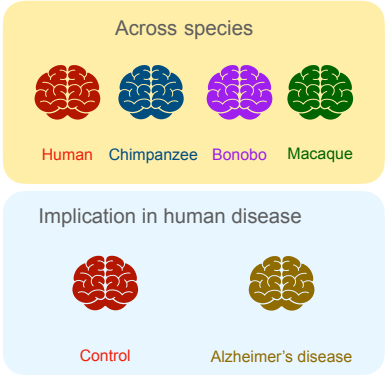
Zhang YE, Landback P, Vibranovski MD, Long M. 2011. Accelerated Recruitment of New Brain Development Genes into the Human Genome.Wolfe KH, editor. *PLoS Biol* 9:e1001179.

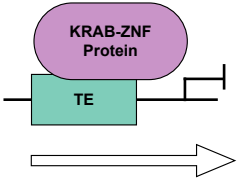
Zhou Y, Zhu J, Tong T, Wang J, Lin B, Zhang J. 2019. A statistical normalization method and differential expression analysis for RNA-seq data between different

species. *BMC Bioinformatics* 20:163.

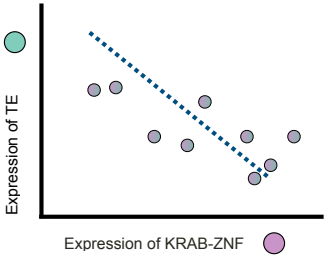
Workflow

Analyzing **KRAB zinc finger genes (KRAB-ZNF)** and **Transposable Elements (TEs)** from two perspectives:

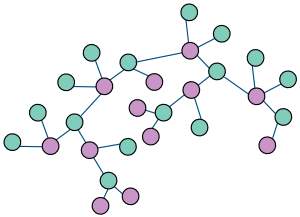




One-to-one negative correlation between TE and KRAB-ZNF denoted as **negative TE:KRAB-ZNF**



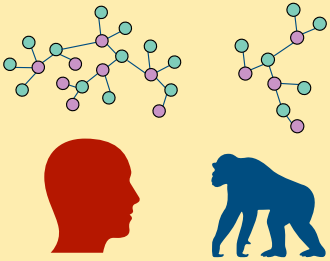
Construct **negative TE:KRAB-ZNF** regulatory network



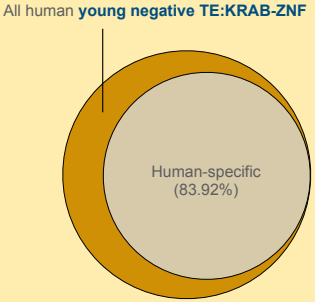
Findings

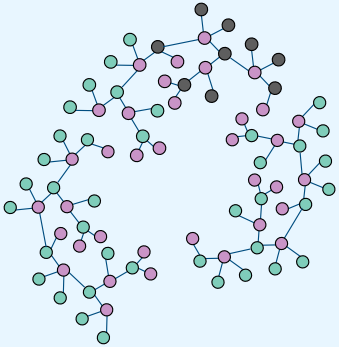
Across species

Humans have more connected **negative TE:KRAB-ZNF** than non human primates.



High percentage of evolutionary **young negative TE:KRAB-ZNF** is human-specific.



Implication in human disease

A subcluster enriched for evolutionary young Alu elements is lost in the network derived from AD patients.