



## RESEARCH ARTICLE

## OPEN ACCESS

# Age and Sex Effects on Blood Retrotransposable Element Expression Levels: Findings From the Population-Based Rhineland Study

Valentina Talevi<sup>1</sup> | Hang-Mao Lee<sup>2</sup> | Dan Liu<sup>1</sup> | Marc D. Beyer<sup>3,4</sup> | Paolo Salomoni<sup>2</sup> | Monique M. B. Breteler<sup>1,5</sup> | N. Ahmad Aziz<sup>1,6</sup>

<sup>1</sup>Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany | <sup>2</sup>Fundamental Research, Nuclear Function Group, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany | <sup>3</sup>Immunogenomics and Neurodegeneration, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany | <sup>4</sup>Platform for Single Cell Genomics and Epigenomics (PRECISE) at the German Center for Neurodegenerative Diseases (DZNE) and The University of Bonn and West German Genome Center (WGGC), Bonn, Germany | <sup>5</sup>Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany | <sup>6</sup>Department of Neurology, Faculty of Medicine, University of Bonn, Bonn, Germany

**Correspondence:** N. Ahmad Aziz ([ahmad.aziz@dzne.de](mailto:ahmad.aziz@dzne.de))

**Received:** 4 February 2025 | **Revised:** 7 April 2025 | **Accepted:** 14 April 2025

**Funding:** N. Ahmad Aziz was supported by a European Research Council Starting Grant (Number: 101041677). The Rhineland Study is funded by the German Center for Neurodegenerative Diseases (DZNE). This work was further supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC2151-390873048) and SFB 1454 (Project-ID 432325352); through the Federal Ministry of Education and Research grant (FKZ: 01KX2230; "PreBeDem – Mit Prävention und Behandlung gegen Demenz"), and the Helmholtz Association under the 2023 Innovation Pool. Dan Liu is partly supported by a grant from the Alzheimer's Association (24AARFD-1192360). PS is currently being supported by the following funding bodies: the DZNE and the Ministry of Culture and Science of the State of Northrhine Westphalia (CANCer TARgeting, CANTAR programme as part of Netzwerke 2021) and the German Cancer Aid Program THUNDER. Work in part related to the topic of this manuscript was supported by the ERC (Consolidator Award H3.3Cancer, number 616744), the Wilhelm Sander Stiftung, the Worldwide Cancer Research Fund (WCRF), the DFG under Germany's Excellence Strategy (Grant no. EXC2151-390873048; Excellence Cluster ImmunoSensation2), the Helmholtz-Gemeinschaft Aging and Metabolic Programming (AMPro) Consortium, and along with other funding bodies. PS is recipient of an Honorary Professorship at UCL (2023–2027). PS thanks current and former members of his group for discussions, in particular Dr. Jenny Russ and Dr. Julia Gerber. MDB is supported by the excellence cluster ImmunoSensation2 (EXC 2151, #390873048); the DFG via IRTG2168 (#272482170), SFB1454 (#432325352); the EU-funded project NEUROCOV receiving funding from the RIA HORIZON Research and Innovation under GA No. 101057775; the Else-Kröner-Fresenius Foundation (2018\_A158). The Lifelines initiative has been supported by funding from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University, and the provinces in the north of the Netherlands (Drenthe, Friesland, Groningen).

**Keywords:** aging | biomarkers | heterochromatin | immune response | population-based study | retrotransposable elements | sex-differences

## ABSTRACT

Retrotransposable elements (RTEs) have been implicated in the pathogenesis of several age-associated diseases. Although model systems indicate that age- and sex-dependent loss of heterochromatin increases RTE expression, data from large human studies are lacking. Here we assessed the expression levels of 795 blood RTE subfamilies in 2467 participants of the population-based Rhineland Study. We found that the expression of more than 98% of RTE subfamilies increased with both chronological and biological age. Moreover, the expression of heterochromatin regulators involved in RTE silencing was negatively related to the expression of 690 RTE subfamilies. Finally, we observed sex differences in 42 RTE subfamilies, with higher expression in men. The genes mapped to sex-related RTEs were enriched in immune response-related pathways. Importantly, we validated our key findings in an independent population-based cohort. Our findings indicate that RTEs and their repressors are markers of aging and that their dysregulation is linked to inflammation, especially in men.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Aging Cell* published by Anatomical Society and John Wiley & Sons Ltd.

## 1 | Introduction

The rapid increase in human life expectancy has spurred interest in understanding of the molecular mechanisms underlying aging and late-life health. Several different approaches have been employed to elucidate the molecular and cellular pathways implicated in aging processes (Lopez-Otin et al. 2023), focusing mainly on detecting genetic polymorphisms associated with age-related traits (Slagboom et al. 2018), exploring age-related epigenetic changes (Jones et al. 2015), and defining age-dependent transcriptomic signatures (Peters et al. 2015). In the past few years, thanks to technological advances in sequencing, repetitive segments of DNA previously known as “junk DNA” have been characterized and found to regulate many biological processes (Chuong et al. 2016; Lu et al. 2020). Indeed, by far the largest proportion of repetitive elements in the human genome is composed of transposable elements (TEs), DNA sequences able to move within the genome (Bourque et al. 2018). TEs represent at least 45% of the entire human DNA (Lander et al. 2001), and they are mainly found in the form of retrotransposable elements (RTEs), which use a so-called “copy and paste” mechanism to spread across the genome (Bourque et al. 2018). As a result of their long evolutionary history, RTE sequences have accumulated mutations, rendering them widely different. According to sequence homology, transposition mechanisms and structural features, RTEs are classified hierarchically into super-, sub- and separate distinct families (Bourque et al. 2018). The quantification of RTE expression was initially limited to the subfamily level. However, more recently further characterization of RTEs at loci level has become feasible, enabling localization and measurement of the expression levels of single RTEs (Yang et al. 2019), although validation of these methods in large genomic datasets is lacking.

Despite the fact that many RTEs in humans have lost their retrotransposition ability due to accrued mutations in the course of evolution, RTEs have been proposed to influence cellular function and tissue homeostasis in the host species via additional mechanisms. Specifically, RTEs could modulate the expression of adjacent genes by acting as enhancers and promoters, while RTEs as RNAs or retrotranscribed DNA could engage the nucleic acid-sensing machineries and trigger inflammatory pathways and antiviral mechanisms (Burbage et al. 2023; De Cecco et al. 2019; Gerber et al. 2021; Gorbunova et al. 2021; Merlotti et al. 2023; Volkman and Stetson 2014). Finally, proteins encoded by RTEs have been linked to senescence and propagation of pathogenic aggregates (Liu, Heumuller, et al. 2023; Liu, Liu, et al. 2023). Therefore, to prevent aberrant activation, RTE expression is tightly regulated by several silencing mechanisms, particularly heterochromatin regulation (Marsano and Dimitri 2022; Zhang et al. 2022). An intricate network of factors, including chromatin-modifying enzymes, chromatin-associated proteins, and histone modifications, ensures and maintains a proper heterochromatin configuration (Zhang et al. 2022). Previous studies, mostly conducted in animal models, reported age-related accumulation of RTEs (Chen et al. 2016; De Cecco et al. 2013, 2019; LaRocca et al. 2020; Wood et al. 2016), induced by decreased efficiency of repressive mechanisms during aging (Pal and Tyler 2016; Zhang et al. 2020). Our own work has suggested that derepression of RTEs leads to aging-associated

phenotypes, such as myeloid bias and inflammation (Gerber et al. 2021). Furthermore, a recent study proposed a causal role of RTEs in aging in *C. elegans* (Sturm et al. 2023), while small-scale human studies suggest an association between retrotransposon expression in blood and cellular senescence and inflammation (Tsai et al. 2024). Despite substantial progress in the study of aging and aging-related traits in the context of RTEs, previous research has primarily focused on RTEs at the family level (LaRocca et al. 2020; Tsai et al. 2024). Limited knowledge exists about the expression of specific RTE subfamilies in aging, and specifically about the localization of individual RTE loci, although the biological role of RTEs is also dependent on their genomic location, having the capability to influence the expression of mapped or neighboring genes (“cis effects”) (Elbarbary et al. 2016). Moreover, the amount of RTEs as well as the content of heterochromatin differs not only across age but also between sexes, mainly due to the presence of large repetitive sequences on the Y-chromosome (Nguyen and Bachtrog 2021; Tang et al. 2018). Indeed, aging is a sex-specific process, with sex differences in both lifespan and the occurrence of many diseases (Hägg and Jylhävä 2021), including inflammatory (Klein and Flanagan 2016), cardiovascular (Leening et al. 2014), metabolic (Tramunt et al. 2020) and neurodegenerative diseases (Beam et al. 2018; Moisan et al. 2016). However, scant research is available on potential sex differences in the expression levels of RTEs in the general population.

Thus, to assess whether dysregulation of RTE expression is a marker of human aging, we evaluated the relation between RTE expression levels and chronological age in a large population-based cohort. Given that deregulation of RTE expression could also be a driver of the aging process and accelerate aging traits (De Cecco et al. 2019; LaRocca et al. 2020; Rigal et al. 2022; Sturm et al. 2023), we also assessed the relation between RTE expression levels and biological age employing “GrimAge”, an estimator of epigenetic aging that is associated with both morbidity and mortality risk (Lu et al. 2019). Indeed, we observed an increase in the expression levels of more than 98% of RTE subfamilies with chronological age. Moreover, we found higher expression of specific RTEs in men compared to women. Importantly, we could replicate the associations of both age and sex with RTE expression levels using data from an independent population-based cohort study. Finally, we were able to demonstrate a negative association between the expression levels of RTE subfamilies and those of chromatin-associated methyltransferase and repressors, key factors for heterochromatin maintenance and RTE silencing (Zhang et al. 2022).

## 2 | Results

### 2.1 | Discovery Study Population

A total of 2467 participants were included in our discovery cohort derived from the Rhineland Study (see Section 4 for a detailed description). The mean (standard deviation (SD)) age was 54.95 (14.36) years and 45.1% were men (Table 1). We could quantify the expression levels of 795 RTE subfamilies in peripheral blood transcriptome data from participants of the Rhineland Study, which could further be classified into 20 RTE families and 4 RTE superfamilies. SINE superfamily expression

**TABLE 1** | Demographic characteristics of the participants and stratification by sex.

Variables	Analytical sample Model 1			Analytical sample Model 2		
	Total	Sex stratification		Total	Sex stratification	
		Women	Men		Women	Men
N participants (%)	2467	1354 (54.9)	1113 (45.1)	2321	1278 (55.1)	1043 (44.9)
Age (Years)	54.95 (14.36)	54.90 (11.7)	55.02 (14.61)	54.76 (14.35)	54.68 (14.13)	54.86 (14.62)
Smoking status: Smokers (%)	317 (12.8)	158 (11.7)	159 (14.3)*	292 (12.6)	149 (11.7)	143 (13.7)
Erythrocytes (T/L)	/	/	/	4.68 (0.43)	4.47 (0.34)	4.93 (0.40)***
Leukocytes (G/L)	/	/	/	5.74 (1.50)	5.72 (1.50)	5.77 (1.49)
Basophils (%)	/	/	/	0.80 (0.35)	0.81 (0.36)	0.78 (0.35)**
Eosinophils (%)	/	/	/	3.29 (2.12)	3.12 (2.09)	3.51 (2.13)***
Lymphocytes (%)	/	/	/	32.76 (8.05)	33.46 (8.02)	31.92 (8.00)***
Monocytes (%)	/	/	/	8.37 (1.93)	7.94 (1.75)	8.89 (2.00)***
Neutrophils (%)	/	/	/	54.77 (8.71)	54.67 (8.80)	54.90 (8.61)

Note: The analytical dataset in *Model 1* included all the participants with RNA-Seq expression data available. The analytical dataset in *Model 2* included all the participants with RNA-Seq expression data and cell count information available. The data are expressed as mean (standard deviation) and categorical variables as number (percentage). White blood cell subtypes are expressed as relative measurement (percentage) to leukocytes. The differences between women and men and between the two analyzed datasets were calculated using a *t*-test (continuous) or chi-squared test (categorical). \*\*\**p*-value < 0.01, \*\**p*-value < 0.05, \**p*-value < 0.1. Abbreviations: G/L, Gigacells per Liter; N, number of participants; T/L, Teracells per Liter.

represented almost 52% of the whole RTE superfamily expression levels across 2467 participants, followed by LINE, LTR, and retroposon superfamilies (Figure S1).

## 2.2 | Changes in RTE Expression Levels With Chronological Age

RTE subfamily expression levels increased with chronological age in 97.5% (95% confidence interval (CI): 96.1 to 98.3) of the 795 RTE subfamilies (one proportion Z-test *p*-value < 2.2e-16) (*Model 1*). However, only two subfamilies (i.e., MER54B and LTR43-int) retained statistical significance (false discovery rate (FDR) < 0.05) after multiple comparisons adjustment. The same positive trend was also observed after adjusting for blood cell composition (*Model 2*), indicating higher expression levels in 99.5% (95% CI: 98.7–99.8) of RTE subfamilies with older chronological age (one proportion Z-test *p*-value < 2.2e-16). Similar to *Model 1*, only MER54B was significantly positively associated with chronological age after FDR adjustment (Figure 1A, Table S1, Figure S2A). The per locus analysis, conducted on the 64 MER54B loci, revealed that four of these loci were the main drivers of the association with chronological age. The expression levels of three of these loci increased with age, and they were located in intronic regions or overlapped with splicing sites of the mapped genes. Specifically, two loci were located in the beta-defensin gene cluster on chromosome 8, while the others were positioned within the *NBEA* and *RAB27B* genes (Figure 1B, Table S2, Figure S2B). Further evaluation of the mapped and *cis*-genes revealed that the expression levels of three age-related loci were positively associated with the expression levels of the corresponding genes. Specifically, the expression levels of the two loci on chromosome 8 were positively associated with those of the *cis*-gene *DEFA3*, while the expression levels of the RTE locus

within the *NBEA* gene showed a strong association with those of the gene itself, suggesting co-expression of these two transcripts (Table S3).

The sensitivity analysis, conducted on three additional RTE subfamilies that were most strongly associated with chronological age (i.e., CR1.L3A-Croc, AluYk11, and SVA\_C), showed that two CR1.L3A-Croc loci were positively associated with chronological age (Table S2). Moreover, we observed that one of these loci, located in chromosome 20, was also positively associated with the corresponding mapped gene, *TSHZ2* (Figure S3, Table S3).

Given that the evolutionary youngest human-specific RTE subfamilies (including L1HS and L1PA2) are likely to be functionally most relevant in terms of transposition capacity (Brouha et al. 2003), we also performed a separate locus-specific analysis for these RTEs. This analysis revealed that chronological age was significantly associated with one locus (out of 226) for L1HS and nine loci (out of 602) for L1PA2 (Figure S4, Table S2). Most of these loci were located in intronic regions and were positively associated with the corresponding mapped or *cis*-genes, including *CAMK4*, *ATM*, and *IFI16* (Table S3).

## 2.3 | Changes in RTE Expression Levels With Biological Age

The relation between RTE subfamilies and biological age mirrored the one observed for chronological age. The expression levels of 99.3% (95% CI: 98.5–99.7) of RTE subfamilies increased with higher biological age (one proportion Z-test *p*-value < 2.2e-16) (*Model 1*). Specifically, 650 RTE subfamilies were associated with biological age at FDR < 0.05. A similar trend was also observed after adjustment for blood cell composition, with 98.0% (95% CI: 96.8–98.7) of RTE subfamilies

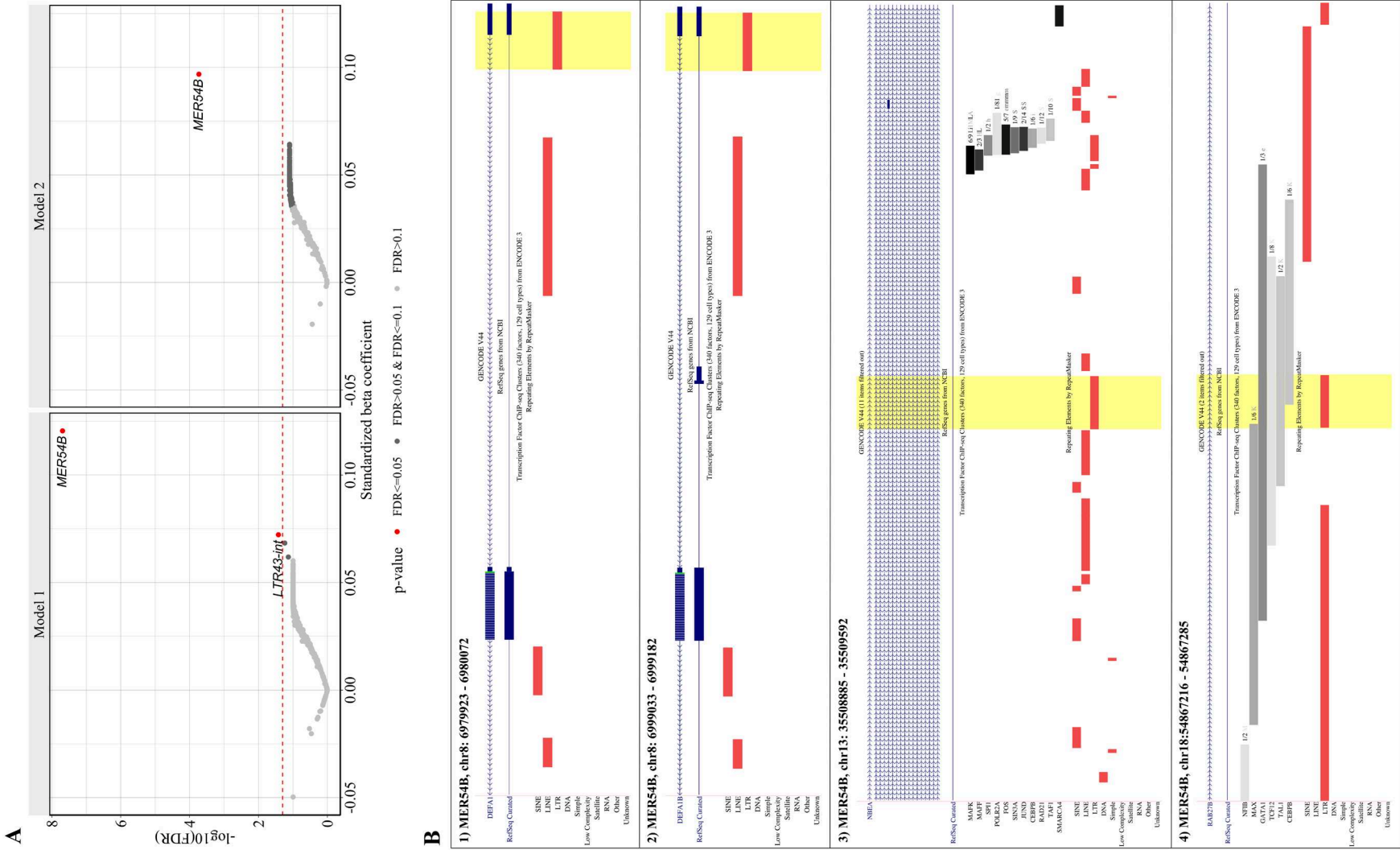
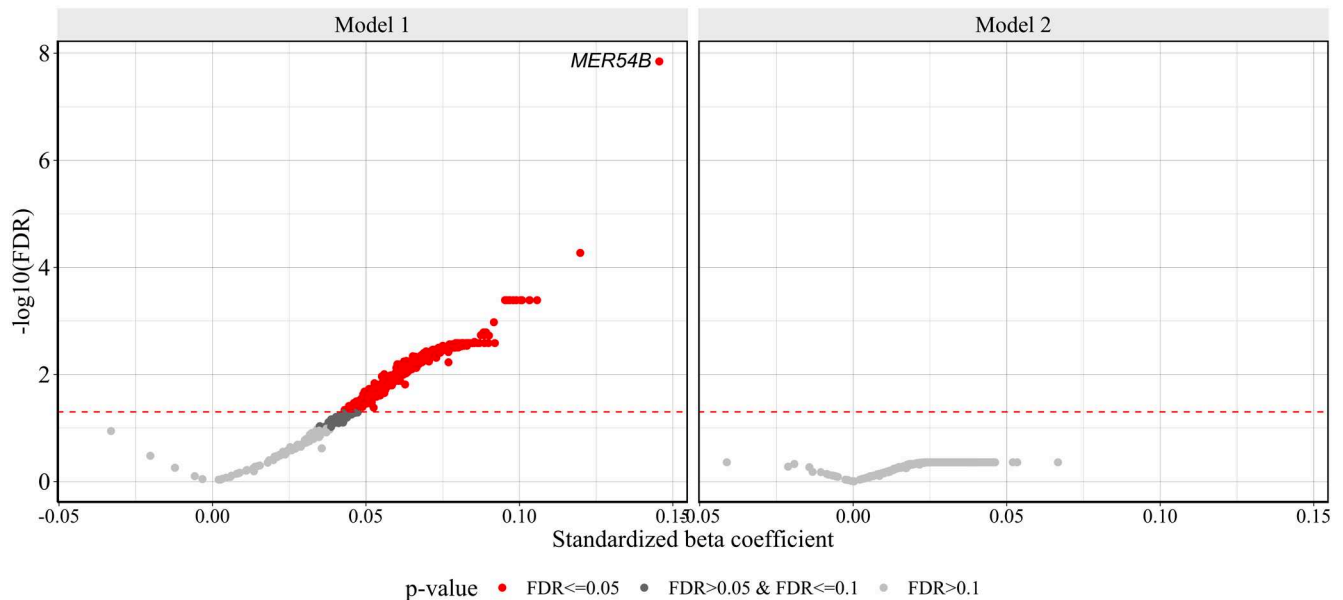


FIGURE 1 | Legend on next page.



**FIGURE 1** | Association of RTE subfamilies and loci with chronological age. (A) Volcano plot showing the change in the standardized expression of RTE subfamilies per one standard deviation in chronological age. Each dot represents an RTE subfamily. The red horizontal line indicates the threshold for statistical significance, set at  $FDR < 0.05$ . *Model 1*: RTE subfamily<sub>i</sub> ~ intercept + age + sex + smoking + batch. *Model 2*: Additionally adjusted for blood cell composition. (B) Visualization of the four MER54B age-associated RTE loci, highlighted by the yellow shading, in the human genome reference GRCh38/hg38 through UCSC Genome Browser. Mapped genes (in blue) and transcription factor binding sites (in gray) are also indicated, if present.



**FIGURE 2** | Associations of RTE subfamilies with biological age. Volcano plot showing the change in the standardized expression of RTE subfamilies per one standard deviation in biological age. Each dot represents a RTE subfamily. The red horizontal line indicates the threshold for statistical significance, set at  $FDR < 0.05$ . *Model 1*: AgeAccelGrim ~ intercept + RTE subfamily<sub>i</sub> + sex + smoking + batch. *Model 2*: Additionally adjusted for blood cell composition.

showing increased expression levels with higher biological age (one proportion Z-test  $p$ -value  $< 2.2e-16$ ). However, after correcting for multiple comparisons to minimize false positive rates, none of the associations between separate RTE subfamilies and biological age remained statistically significant (Figure 2, Table S4).

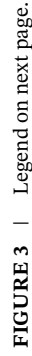
## 2.4 | Sex Differences in RTE Expression Levels

Sex was significantly associated with the expression levels of 55 RTE subfamilies ( $FDR < 0.05$ , *Model 1*). After adjustment for blood cell counts, 42 RTE subfamilies remained significantly associated with sex, with higher expression levels in men compared to women (Figure 3A, Table S5). With the exception of a single RTE subfamily (L1MDB), the entirety of the sex-related RTE subfamilies belonged to the LTR superfamily, mainly comprised of human endogenous retroviruses (HERVs). The per locus level analysis revealed that these associations with sex were driven by 247 RTE loci. Moreover, it indicated that the higher expression of RTE subfamilies in men was mainly due to the location of 168 of the RTE loci on the Y-chromosome, confined to a window of about 18 million base pairs. However, even when considering only autosomes, the majority of sex-related RTE loci still exhibited higher expression levels in men as compared to women. Specifically, 66 out

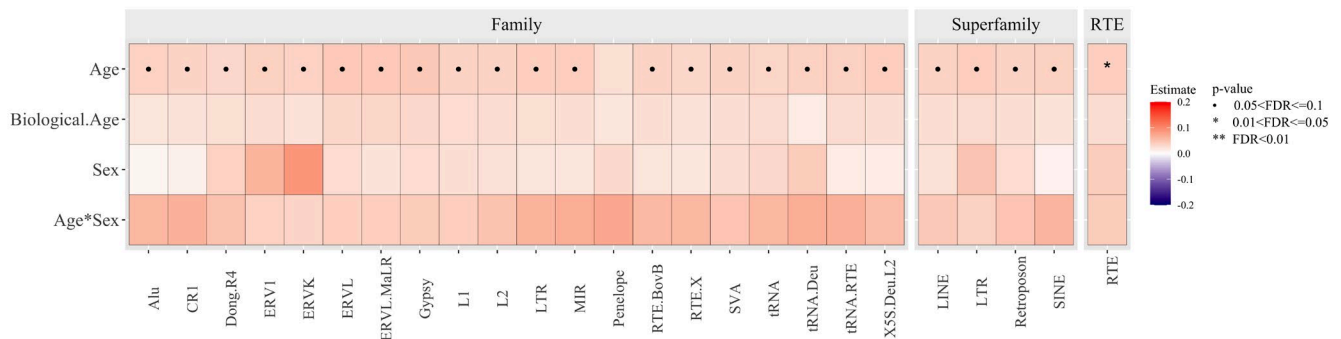
of 79 RTE loci that were located on autosomes showed higher expression levels in men compared to women (Figure 3B,C, Table S6). The genomic annotation of these sex-related loci revealed that they were located mainly in intronic and intergenic regions (Figure S5). By investigating the relation of the RTE loci's expression with the corresponding mapped and/or *cis*-genes, we found that the majority of the associations were in the same direction, with an increase in RTE loci expression levels corresponding to an increase in gene expression levels. This finding suggests that RTE loci, which are located in intronic regions, are co-expressed with the genes they are mapped to, whereas those located in intergenic regions likely function as regulators of the proxy genes (Table S7). Further functional analysis of genes associated with sex-related RTEs revealed significant enrichment for immune responses (e.g., “positive regulation of interleukin-1 beta production” and “cellular response to interferon-beta”) (Figure S6). Sex did not modify the associations of expression levels of RTE subfamilies with chronological age or biological age.

## 2.5 | Sensitivity Analysis

The associations of the RTE scores, computed at family, superfamily, and overall levels with chronological age, biological age, and sex mirrored the ones identified at the subfamily



**FIGURE 3** | Significant association of RTE subfamilies and loci with sex. (A) Box plots showing the different expression levels between men and women for the 42 RTE subfamilies for which we observed a significant sex difference in the RTE expression levels. The RTE subfamilies are grouped based on the family they belong to. (B) Manhattan plot representing RTE loci belonging to the 42 RTE subfamilies associated with sex. The red line indicates the threshold for statistical significance, set at  $FDR < 0.05$ . (C) Ideograms of chromosomes 1, 12, and Y showing the locations of sex-related loci on the respective chromosomes and the mapped genes, when present.



**FIGURE 4** | Association of RTE families, superfamilies, and overall RTE expression levels with age, sex, and biological age. Heatmap representing the strength and directionality of the associations between RTE family, superfamily, and overall RTE expression levels with chronological age, sex, biological age, and age-sex interaction.

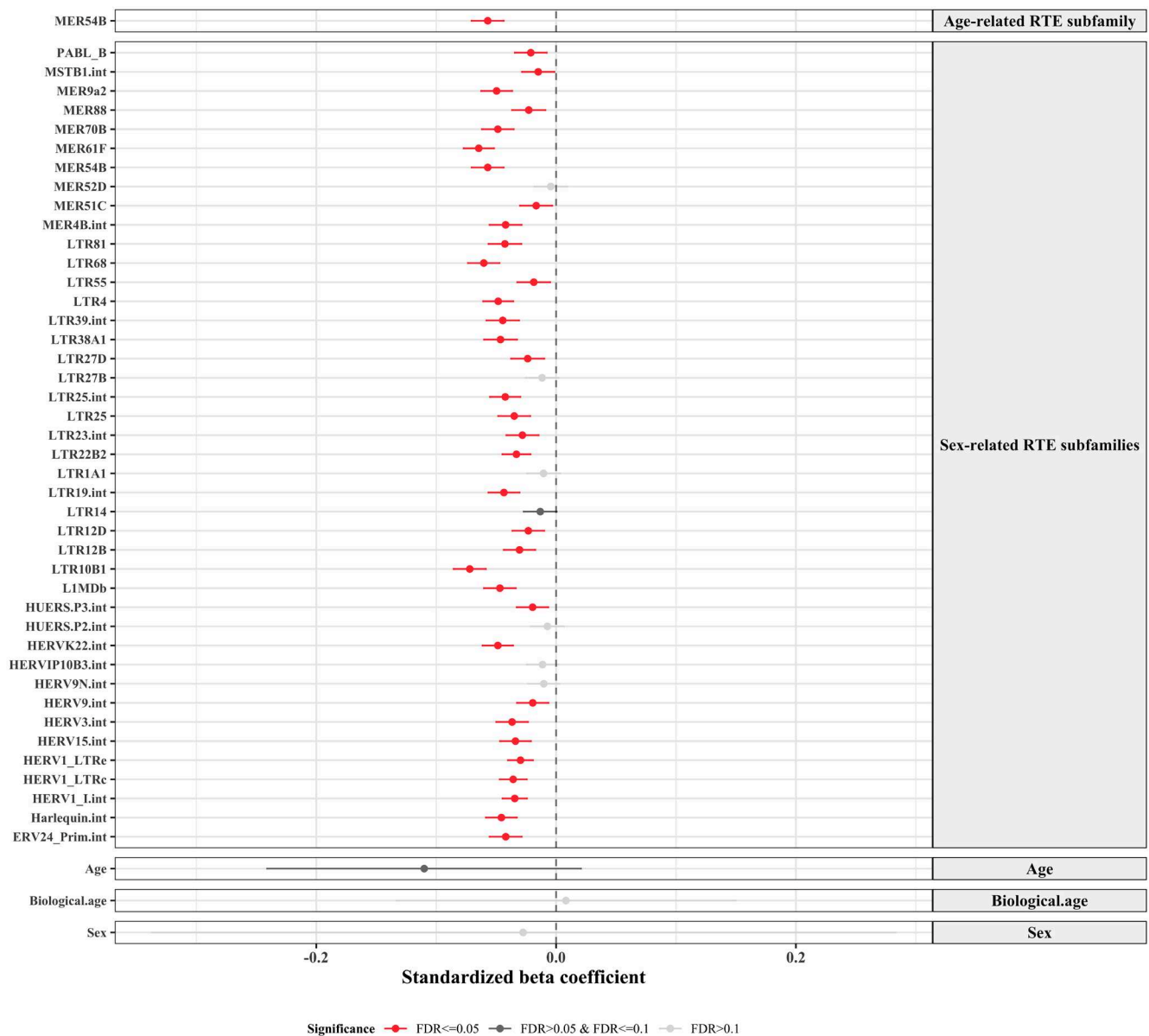
level, confirming the positive association of RTE expression levels with both chronological and biological age, as well as the higher expression levels of RTEs in men compared to women. (Figure 4, Tables S8–S10).

## 2.6 | Association of RTE Repressor Genes With RTE Subfamilies

The analysis of the effect of heterochromatin regulators on RTE subfamilies revealed that the RTE repressor score (calculated by combining the expression levels of *DAXX*, *TRIM28*, *SETDB1b*, and *DNMT1* (Gerber et al. 2021; Haggerty et al. 2021)) is negatively associated with the expression levels of age- and sex-related RTE subfamilies (Figure 5, Table S11). Interestingly, this relation was consistent across most RTE subfamilies. Specifically, a lower RTE repressor score corresponded to higher expression levels of 690 RTE subfamilies at  $FDR < 0.05$  (Model 2) (Figure S7). Moreover, we found that the RTE repressor score tended to be lower in older than younger participants ( $\beta$  estimate:  $-0.11$ ,  $p$ -value =  $0.099$ ). However, no association with biological age and no sex differences in the RTE repressor scores were observed. The absence of an association between RTE repressor score and biological age, even when including only the two DNA methylators (*DNMT1* and *SETDB1*) in the derivation of the score (data not shown), may be attributed to the fact that these two genes alone do not fully capture the complexity of DNA methylation patterns that determine biological age. Moreover, although *DNMT1* and *SETDB1* are important regulators of methylation, they play distinct roles in this process, which may explain the lack of association. *DNMT1* primarily functions as a maintenance methyltransferase, preserving existing DNA methylation patterns rather than establishing new marks (Law and Jacobsen 2010). In contrast, epigenetic age estimators, such as GrimAge, reflect long-term, cumulative methylation changes, which may not be directly influenced by *DNMT1*'s maintenance activity. *SETDB1*, on the other hand, contributes to gene silencing

primarily through histone modification, particularly H3K9me3, which regulates heterochromatin formation (Leung et al. 2011). However, its impact on DNA methylation may not directly affect the CpG sites included in the estimation of GrimAge, as it primarily influences chromatin structure and gene expression rather than directly modifying DNA methylation patterns. Furthermore, *DNMT1* and *SETDB1* are not the only players in this process. For example, other DNA methyltransferases, such as *DNMT3A* and *DNMT3B*, also contribute to the overall methylation patterns. Additionally, DNA methylation is regulated by a broad range of genes, including those encoding for other DNA methyltransferases such as *DNMT3A* and *DNMT3B*, as well as extrinsic factors such as diet and environmental exposures.

To further explore the role of the RTE repressor score in gene regulation, we assessed its associations with the genes expressed in blood samples of Rhineland Study participants. Interestingly, we found that the RTE repressor score exhibited statistically significant associations with the vast majority of genes (10,397 out of 11,306), with approximately half showing positive associations (5054 out of 10,397) and half showing negative associations (5343) (Table S11). These widespread associations may be explained by the fact that the genes included in the derivation of the RTE repressor score are involved in the regulation of common cellular processes affecting the expression of a large number of other genes, including epigenetic modulation, transcriptional repression, and chromatin remodeling. Focusing on the top five positively associated genes (*POLR2A*, *ARID1A*, *SRRM2*, *MED12*, and *EP400*) and the top five negatively associated genes (*SF3B1*, *SPCS3*, *NEK7*, *ZRANB2*, and *DNAJB14*), which were not already included in the derivation of the RTE repressor score, we further examined their genomic locations in relation to RTEs using the UCSC Genome Browser. Some genes, such as *ARID1A* and *NEK7*, were found to map to a large number of RTE loci across different superfamilies (SINE, LINE, and LTR), while others, such as *SRRM2* and *SPCS3*, mapped to only a few (Figure S8).



**FIGURE 5** | Relation between RTE repressor score and age- and sex-related RTE subfamilies, chronological age, and sex. The forest plot shows the associations of RTE repressor score with the age-related RTE subfamily, MER54B, the 42 sex-related RTE subfamilies, chronological age, and sex. The standardized beta coefficients for RTE subfamilies were obtained through the following linear regression model: RTE subfamily<sub>i</sub> ~ intercept + RTE repressor score + age + sex + smoking + batch + blood cell composition. The standardized beta coefficients for age and sex were obtained through the following linear regression model: RTE repressor score ~ intercept + age + sex + smoking + batch + blood cell composition. The standardized beta coefficients for biological age and sex were obtained through the following linear regression model: RTE repressor score ~ intercept + biological age + sex + smoking + batch + blood cell composition.

## 2.7 | Validation of Findings in an Independent Cohort

For external validation of our results, we used data from a total of 706 participants from the population-based Lifelines Study for whom RTE expression levels, age, sex, smoking status, and blood cell counts were available (see Section 4 for a detailed description). The mean age of the included participants was 47.7 years (SD: 11.1), with 43.5% male. The expression levels of 790 RTE subfamilies could be quantified in the Lifelines participants.

The age analysis revealed that chronological age was significantly associated with eight RTE subfamilies at  $p$ -value

$< 0.05$  (Model 1) (Table S12). The RTE subfamily MER54B, identified as the top hit associated with chronological age in the Rhineland Study, was validated in the Lifelines cohort ( $\beta$  estimate: 0.112,  $p$ -value = 0.005). Additionally, we observed a positive trend between chronological age and the expression levels of RTE subfamilies, consistent with the findings from the Rhineland Study. Specifically, expression levels increased with chronological age in 93.2% (95% CI: 91.2–94.7) of the 790 RTE subfamilies (one proportion Z-test  $p$ -value  $< 2.2e-16$ ). After adjusting for blood cell concentrations (Model 2), the results remained consistent, with chronological age still showing a positive association with the expression levels of eight RTE subfamilies at  $p$ -value  $< 0.05$ , including the MER54B



subfamily (Table 2). Additionally, similar to *Model 1*, 92.3% (95% CI: 90.2–93.9) of the 790 RTE subfamilies exhibited increased expression levels with chronological age (one proportion Z-test  $p$ -value  $< 2.2 \times 10^{-16}$ ), mirroring the findings of the Rhineland Study (Table S12).

The sex analysis in the validation cohort also indicated that sex was significantly associated with the expression levels of 13 RTE subfamilies in *Model 1* and 11 RTE subfamilies in *Model 2* at  $p$ -value  $< 0.05$ . Among these, the sex-related associations observed in the Rhineland Study were replicated for 7 and 6 of these RTE subfamilies in *Model 1* and 2, respectively (Table 2, Table S13).

### 3 | Discussion

To our knowledge, this is the first study reporting a comprehensive overview of genome-wide RTE subfamily expression levels in blood across age and sex in the general population, with further detailed RTE locus-specific characterization. We found that for more than 98% of RTE subfamilies, expression levels increased with both higher chronological and biological age. The chronological age-related findings were replicated in an independent cohort, where also more than 92% of RTE subfamilies exhibited higher expression levels with increasing age. Moreover, we found higher levels of RTE expression in men compared to women, mostly due to the expression of LTR elements on the Y-chromosome, which was also largely replicated in the independent Lifelines cohort. Interestingly, the expression of sex-related RTE loci was primarily associated with the expression of genes involved in immune response pathways. In addition, our results revealed a relation between elevated expression levels of heterochromatin regulator genes and reduced expression levels of RTE subfamilies, indicating that transposable elements are tightly regulated by epigenetic silencing mechanisms.

We observed higher genome-wide expression levels of RTE subfamilies with older chronological age. This association was also apparent when summing up the RTE subfamily expression levels per (super)family, indicating that RTE expression levels increase with age. The increased expression levels of different RTE families with advancing age have also been reported in animal models (De Cecco et al. 2013; Wood et al. 2016) and small-scale human studies (LaRocca et al. 2020). Despite reporting an association between RTE expression and biological age-related events, a recent study did not find a link between chronological age and RTE expression (Tsai et al. 2024), which is likely due to the study's relatively small sample size and use of microarray expression data, precluding comprehensive analyses at the RTE subfamily level. Increased RTE expression with age is thought to be primarily linked to heterochromatin loss (Pal and Tyler 2016), due to the reduction of histone proteins (Feser et al. 2010), decreased DNA methylation (Zampieri et al. 2015), different histone variants and modifications (Dang et al. 2009; Duarte et al. 2014), as well as dysregulation of histone chaperones, such as *DAXX* (Rapkin et al. 2015). Among all the RTE subfamilies, the expression levels of MER54B, classified in the ERVL (LTR) family, showed the strongest association with chronological age. Human endogenous retroviruses (HERVs) are transposable elements originating from the insertion of exogenous retroviruses in the germ line. Although HERVs are

generally retrotranspositionally inactive in humans due to accumulated mutations during evolution (Katzourakis et al. 2005), there is some evidence for the presence of a very recently acquired, apparently intact HERV in a small fraction of the human population (Stoye 2001; Turner et al. 2001). HERVs have been associated with several pathological processes, though the exact underlying mechanisms are still unknown.

It has also been suggested that HERVs might act as transcriptional regulatory elements for neighboring genes (Mao et al. 2021). In support of this hypothesis, our per locus analysis identified two age-related MER54B loci, located within the cluster of defensin genes, whose expression was positively associated with that of the *cis*-gene *DEFA3*. Defensin genes encode for peptides specialized in host innate immune and infection responses (Holly et al. 2017), and they have been reported as potential inflammatory biomarkers of frailty (Cardoso et al. 2018). The co-expression of age-related RTE loci and genes involved in inflammatory pathways further supports the contribution of these elements to the aging process itself (Liu, Liu, et al. 2023). Moreover, we identified another MER54B age-related locus within the intronic region of *RAB27B*, a gene responsible for the production of endogenous exosomes, key regulators of the immune response (Alexander et al. 2017). Indeed, through multimodal experiments, previous studies demonstrated that increased HERVK expression as well as the accumulation of HERVK-Env proteins and HERVK retrovirus-like particles triggers the innate immune response and promotes the intercellular spreading of proteopathic seeds (Liu, Heumüller, et al. 2023; Liu, Liu, et al. 2023). The fourth MER54B age-related locus was located in the intronic region of the *NBEA* gene, which encodes the autism-related protein neurobeachin involved in the regulation of synaptic protein trafficking (Lee et al. 2018). Additionally, several *de novo* *NBEA* genetic variants have been related to early childhood epilepsy (Mulhern et al. 2018). Interestingly, ERVs have also been shown to increase genomic instability and play a critical role in the pathogenesis of autism-spectrum disorders (Lin et al. 2023). However, to what extent similar mechanisms may contribute to age-associated traits, especially cognitive decline, remains to be established. Nonetheless, the strong positive association between the expression levels of this particular RTE locus and *NBEA* indicates that co-localization of RTE loci within or in the vicinity of genes results in their co-transcription (Lanciano and Cristofari 2020).

Interestingly, the expression levels of the youngest human-specific subfamilies, belonging to the long interspersed nuclear element-1 (L1) family, increased with chronological age. Specifically, L1HS, also known as L1PA1, is the youngest human-specific subfamily (Brouha et al. 2003; Lanciano et al. 2024; Lander et al. 2001) and has retained its active retrotransposing ability (Sun et al. 2018). We found that the age-related L1HS locus is located within the *CAMK4* gene. The CaMK4 protein is a multifunctional serine/threonine kinase, involved in various aspects of the immune response. It plays a critical role in regulating the production of cytokines, including interleukin-17, and has been implicated in the pathogenesis of autoimmune diseases such as systemic lupus erythematosus and psoriasis (Xu et al. 2024; Yong et al. 2022). L1PA2 is the evolutionary second youngest human-specific subfamily. The corresponding age-related loci were interspersed across the genome,

**TABLE 2** | Validation of the Rhineland Study findings in the Lifelines cohort.

RTE subfamily	Rhineland Study	Lifelines
	Estimate (FDR)	Estimate ( <i>p</i> -value)
Age analysis		
MER54B	0.096 (1.86e-04) <sup>c</sup>	0.116 (0.003) <sup>c</sup>
MER74C	0.045 (0.08) <sup>a</sup>	0.129 (0.001) <sup>a</sup>
MER70B	0.057 (0.08) <sup>a</sup>	0.084 (0.041) <sup>a</sup>
LTR58	0.041 (0.087) <sup>a</sup>	0.081 (0.049) <sup>a</sup>
LTR26E	0.033 (0.113)	0.094 (0.021) <sup>b</sup>
MER51E	0.022 (0.241)	0.093 (0.023) <sup>b</sup>
AluYb9	0.020 (0.251)	0.084 (0.041) <sup>b</sup>
HERV1_I-int	0.013 (0.373)	0.092 (0.023) <sup>b</sup>
Sex analysis		
LTR27D	0.228 (1.29e-05) <sup>c</sup>	0.368 (5.708e-06) <sup>c</sup>
MER54B	0.313 (5.54e-11) <sup>c</sup>	0.290 (2e-04) <sup>c</sup>
HERVK22.int	0.396 (3.47e-19) <sup>c</sup>	0.209 (0.0098) <sup>c</sup>
LTR22B2	1.100 (2.46e-156) <sup>c</sup>	0.170 (0.0341) <sup>c</sup>
HERV9N.int	0.341 (3.28e-14) <sup>c</sup>	0.178 (0.0298) <sup>c</sup>
MER52D	0.137 (0.0384) <sup>c</sup>	0.193 (0.0173) <sup>c</sup>
LTR25.int	0.447 (9.57e-25)	0.123 (0.1306)
HERV1_LTRc	1.180 (1.42e-186)	0.120 (0.1407)
HUERS.P2.int	0.241 (1.52e-06)	0.113 (0.1685)
LTR23.int	0.207 (6.42e-05)	−0.110 (0.1791)
HERV1_I.int	1.300 (3.76e-254)	0.105 (0.196)
Harlequin.int	0.213 (1.21e-05)	−0.072 (0.3712)
HERV15.int	0.181 (0.000449)	−0.064 (0.4346)
HERV9.int	0.260 (4.44e-08)	0.062 (0.4511)
LTR81	0.192 (0.000493)	−0.060 (0.4646)
HERVIP10B3.int	0.227 (2.7e-06)	−0.057 (0.4837)
MER70B	0.139 (0.0285)	0.057 (0.4844)
LTR25	0.378 (3.51e-16)	−0.052 (0.5238)
LTR12D	0.144 (0.0173)	0.050 (0.5435)
LTR10B1	0.234 (6.19e-06)	−0.047 (0.5668)
MER51C	0.131 (0.0499)	−0.046 (0.5695)
MSTB1.int	0.175 (0.00148)	0.045 (0.5801)
LTR55	0.167 (0.00374)	−0.040 (0.622)
LTR27B	0.135 (0.0489)	0.035 (0.6717)
LTR1A1	0.145 (0.0216)	0.034 (0.6835)
MER88	0.204 (0.000169)	−0.029 (0.7232)
L1MDb	0.181 (0.000809)	0.027 (0.7368)

(Continues)

TABLE 2 | (Continued)

RTE subfamily	Rhineland Study	Lifelines
	Estimate (FDR)	Estimate ( <i>p</i> -value)
LTR14	0.178 (0.00148)	0.025 (0.7615)
LTR68	0.130 (0.0489)	0.024 (0.7717)
LTR4	0.509 (4.87e-32)	−0.022 (0.787)
ERV24_Prim.int	0.252 (2.81e-07)	0.022 (0.788)
LTR39.int	0.267 (7.42e-08)	−0.018 (0.8317)
MER4B.int	0.174 (0.00145)	−0.017 (0.8302)
LTR38A1	0.295 (2.38e-09)	−0.016 (0.8505)
HUERS.P3.int	0.309 (2.73e-11)	0.014 (0.8589)
MER61F	0.353 (3.37e-15)	0.014 (0.8635)
PABL_B	0.178 (0.000809)	0.013 (0.8685)
MER9a2	0.170 (0.00145)	0.012 (0.8782)
LTR19.int	0.403 (3.47e-19)	0.010 (0.9048)
LTR12B	0.279 (2.6e-09)	0.002 (0.9803)
HERV3.int	0.344 (9.67e-14)	0.002 (0.9827)
LTR108c_Mam	0.119 (0.154) <sup>a</sup>	0.17 (0.0302) <sup>a</sup>
HERVFH21.int	−0.0553 (0.992)	−0.218 (0.0071) <sup>b</sup>
L1P4d	−0.0311 (0.992)	−0.229 (0.0059) <sup>b</sup>
MER67D	−0.011 (0.992)	−0.188 (0.0218) <sup>b</sup>
LTR21C	/	−0.78 (2.294e-23) <sup>b</sup>
HERV1_LTRe	1.41 (4.69e-281)	/

Note: The Rhineland Study and Lifelines results are based on *Model 2* (adjusted for smoking, batch and blood cell composition).

<sup>a</sup>Significant findings in the Lifelines cohort at *p*-value < 0.05 and replicated in the Rhineland study at *p*-value < 0.05.

<sup>b</sup>Significant findings in the Lifelines cohort at *p*-value < 0.05.

<sup>c</sup>Significant findings in the Rhineland Study at FDR < 0.05 and replicated in the Lifelines cohort at *p*-value < 0.05.

either located within the intronic regions or adjacent to various genes, including *ATM* and *IFI16*. The *ATM* protein, a serine/threonine protein kinase, plays a key role in regulating the cellular response to DNA damage by activating multiple DNA repair pathways (Blackford and Jackson 2017). Dysfunction of *ATM* can lead to genomic instability and cellular senescence (Lombard et al. 2005). Meanwhile, the *IFI16* protein, an interferon gamma inducible protein, is involved in the immune response by detecting viral DNA and activating the interferon signaling pathway (Jiang et al. 2021; Unterholzner et al. 2010). Although we did not directly quantify RTE transposition activity, the higher expression levels of L1HS and L1PA2 subfamilies with chronological age may reflect their increased transposition activity during aging.

The increase of RTE subfamily expression levels with advancing chronological age was also mirrored in the analysis with biological age as the outcome. We observed that an acceleration of biological age corresponded to higher expression values of almost all RTE subfamilies. This result supports the hypothesis that RTE expression levels could serve as transcriptomic markers of aging (LaRocca et al. 2020). In fact, our findings indicate that not only the expression levels of RTE subfamilies increase with

the actual number of years a person has lived, but also seem to reflect the aging process. This result is in line with the findings reported by LaRocca et al., where the expression of repetitive elements was found to be a good predictor of age (LaRocca et al. 2020).

We not only found differences in the expression of RTE subfamilies across age, but also between sexes. Expression levels of 42 RTE subfamilies were higher in men compared to women. A relatively large number of sex-related RTE subfamilies belonged to the ERV1 family. It has been suggested that HERVs' retroviral DNA/RNA directly activates inflammatory responses (Volkman and Stetson 2014), as also suggested by several enriched biological processes in our pathway analysis (e.g., “positive regulation of interleukin-1 beta production” and “cellular response to interferon-beta”). The higher expression of RTE subfamilies in men than women was mainly due to the presence of 168 expressed loci on the Y-chromosome. This result is consistent with the evidence that the Y-chromosome is a hot spot for RTEs, as reported in animal and human studies (Nguyen and Bachtrog 2021; Tang et al. 2018). However, it is still unknown why the Y-chromosome contains a relatively large number of repeat elements. One plausible explanation could be the presence

of non-recombining genomic regions on the Y-chromosome (Gvozdev et al. 2005).

Besides RTE loci located on the Y-chromosome, we also found several RTE loci located on autosomes, which might also contribute to differences in the biology of men and women. These autosomal sex-related loci were mainly located in the promoter and intronic regions of chromosomes 1 and 12. Notably, on chromosome 12, most of these loci were associated with genes from the killer cell lectin-like receptor family (*KLRD1*, *KLRC1*, *KLRG1*, *KLRC4*, *KLRC4-KLRK1*, *KLRK1*), which play a key role in immune regulation. Although the expression of most RTE loci was higher in men compared to women, the LTR25 locus, located in an intergenic region of chromosome 1, was one of the few loci with a higher expression in women compared to men. The expression levels of this RTE locus were positively associated with, among others, those of *cis*-genes belonging to the cluster Fc Gamma receptor genes (*FCGR2B*, *FCGR2C*, *FCGR3B*, *FCGR3A*). This suggests that the LTR25 locus acts as a regulator of the neighboring genes, controlling the expression of Fc Gamma receptor genes, which is in line with the fact that women have higher antibody responses compared to men (Klein and Flanagan 2016). Interestingly, chromosome 1q23, where the Fc Gamma receptor genes are located, has been identified as a lupus susceptibility locus (Moser et al. 1998). This suggests a potential link between women-related RTEs, immune regulation and autoimmune diseases such as systemic lupus erythematosus. Given that women are generally more susceptible to inflammatory diseases (Klein and Flanagan 2016), the small number of RTE loci exhibiting relatively higher expression levels in females is intriguing, and might reflect the presence of sex-specific RTE silencing mechanisms that remain to be elucidated. A non-mutually exclusive hypothesis would be that the relatively higher levels of LTRs in men could play a physiological role in limiting the immune response in males, potentially contributing to sex-based differences in immune regulation.

Interestingly, for both age- and sex-related RTE subfamilies, we observed a decrease in their expression levels with higher expression levels of heterochromatin regulator genes. This result aligns with the knowledge that the activity of RTEs is tightly regulated to prevent aberrant expression (Almeida et al. 2022). The maintenance of heterochromatin integrity is the main mechanism that ensures the silencing of transposable elements. Several molecules are part of the intricate machinery that is responsible for keeping chromatin highly compact, including DNA and histones (e.g., methyltransferase enzymes (e.g., *DNMTs*, *SETDB1*) and the *KRAB-ZPF-TRIM28* gene complex (Almeida et al. 2022; Geis and Goff 2020; Groh and Schotta 2017)). Previous studies have already shown that the depletion of any of these factors might result in an increase in RTE activity (Rowe et al. 2010; Turelli et al. 2014), mainly attributed to the perturbation of chromatin structure. Our findings reinforce the importance of this phenomenon by demonstrating that within the general population, individuals with relatively higher levels of heterochromatin regulator genes had lower expression levels of RTE subfamilies.

This study has both strengths and limitations. We investigated the effects of age and sex on blood RTE subfamily expression

levels in a large population-based study, including a large number of healthy individuals across a wide age range (30–95 years old), with further validation of the key findings in an independent cohort. In addition, candidate RTE subfamilies were further characterized by comprehensively evaluating all their comprising RTE loci across the genome. Moreover, we explored their potential biological role by relating them to the expression levels of nearby genes. Finally, incorporating heterochromatin regulator genes into the analysis yielded important clues to the potential biological mechanisms underlying the silencing process of RTEs. However, due to the cross-sectional nature of our study, we were unable to assess the temporal directionality of the effects, though our findings are highly consistent with those from previous experimental studies. Additional longitudinal and experimental studies are needed to further explore the biological role of RTEs as modulators of aging and age-related diseases. Moreover, we utilized the SQuIRE pipeline, which employs an expectation-maximization algorithm to assign multimapping reads based on the estimated RTE expression derived from uniquely mapped reads. While this approach allows for the inclusion of multimapping reads, it may also induce artificial positive correlations between transposable elements and nearby genes, as reads are preferentially assigned to highly expressed regions. While the majority of RTE-*cis*-gene associations were positive, we also identified several negative and non-significant associations. For example, in the age analysis, we observed a *negative* association between L1PA2 (RTE) and *DNAJC3-DT* (*cis*-gene), as well as many non-significant associations such as MER54B (RTE) and *RAB27B* (*cis*-gene) and L1PA2 and *UGGT2* (*cis*-gene) (Table S3). Similarly, in the sex analysis, we identified 20 significant *negative* associations (Table S7), as well as 83 non-significant associations (not reported in the table), between expression levels of RTEs and *cis*-genes. Thus, these findings indicate that the observed relationships are largely not driven by the expectation-maximization algorithm. Nevertheless, to address this potential limitation, future studies could leverage long-read sequencing for more accurate RTE expression quantification or perform experimental knockdowns to unequivocally determine whether RTEs and nearby genes are truly transcriptionally co-regulated.

In conclusion, we found that blood RTE expression levels increased with both chronological and biological age and were higher in men compared to women. Additionally, increased expression of RTEs was associated with decreased expression of heterochromatin regulator genes and, in men, with increased expression levels of neighboring genes involved in immune response. Collectively, our findings indicate that RTEs are intimately associated with the aging process and that their dysregulation may be linked to inflammation, especially in men.

## 4 | Methods

### 4.1 | Discovery Study Population

For the discovery phase of our study, we used data from participants of the Rhineland Study, an on-going population-based cohort study in Bonn, Germany. Invitations are sent



to individuals from two geographically defined areas in Bonn who are eligible for inclusion, i.e., aged 30 years and above and having sufficient command of the German language to provide informed consent. The study was approved by the ethics committee of the University of Bonn, Medical Faculty. The Rhineland Study is carried out in accordance with the recommendations of the International Conference on Harmonization (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. We based our study on data from the first 3000 consecutively enrolled participants of the Rhineland Study. The current analysis included 2467 participants for whom RNA sequencing (RNA-Seq) data was available after quality control.

## 4.2 | Measurements

### 4.2.1 | Whole Blood RNA Isolation and Sequencing

We extracted total RNA from peripheral blood. Blood was collected in PAXgene Blood RNA tubes (PreAnalytix/Qiagen) and processed according to manufacturer's guidelines. PAXgene tubes were thawed and incubated at room temperature to increase RNA yields. Total RNA was isolated according to manufacturer's instructions using PAXgene Blood miRNA Kit and following the automated purification protocol (PreAnalytix/Qiagen). RNA integrity and quantity was evaluated using the tapestation RNA assay on a Tapestation 4200 instrument (both from Agilent). We used 750 ng total RNA to generate NGS libraries using the TruSeq stranded total RNA kit (Illumina) following manufacturer's instructions with Ribo-Zero Globin reduction. We checked library size distribution via Tapestation using D1000 on a Tapestation 4200 instrument (Agilent) and quantified the libraries via Qubit HS dsDNA assay (Invitrogen). We clustered the libraries at 250 pM final clustering concentration on a NovaSeq 6000 instrument using S2 v1 chemistry (Illumina) in XP mode and sequenced paired-end 2\*50 cycles before demultiplexing using bcl2fastq v2.20. Quality control of the sequencing was evaluated through FastQC (v0.11.9).

### 4.2.2 | Retrotransposable Element Subfamily and Loci Profiling

To quantify the expression of retrotransposable elements (RTEs) in RNA-Seq data, we used the SQUIRE pipeline (Software for Quantifying Interspersed Repeat Elements) (Yang et al. 2019). SQUIRE quantifies the expression of retrotransposable elements at both the subfamily and locus level. The reads were aligned to the human reference genome hg38 from UCSC, and RTE annotation was based on RepeatMasker (Smit et al. 2013–2015). The quantification step uses an iterative method to quantify unique and multi-mapping reads. SQUIRE outputs both read counts and FPKM (fragments per kilobase transcript per million reads) for RTE subfamilies as well as for RTE loci. RTE subfamilies that were expressed in at least 5% of the participants, with an average of mapped reads equal to or greater than 15, were considered for further analyses in order to improve sensitivity since low expressed

RTEs might reflect noise in the quantification phase. We log-transformed FPKM counts to conform the data to a normal distribution.

### 4.2.3 | Gene Expression Profiling

To quantify the expression of genes in RNA-Seq data, the sequencing reads were first aligned to the human reference genome GRCh38.p13 provided by Ensembl using STAR v2.7.1. The count matrix was generated with STAR-quant-Mode GeneCounts using the human gene annotation version GRCh38.101. Genes with an overall mean expression greater than 15 reads and expressed in at least 5% of the participants were used for the following analyses. Finally, in order to normalize for library size and to log-transform the raw data, we applied the varianceStabilizingTransformation function from DESeq2 v1.30.1 R package.

### 4.2.4 | Estimation of RTE Family and Superfamily Expression Scores and Overall RTE Expression

RTE family and superfamily expression levels were estimated by first standardizing the expression levels of RTE subfamilies and then summing the Z-scores of the RTE subfamilies belonging to the same family/superfamily. Moreover, we obtained the overall RTE expression levels by summing the Z-scores of the 795 RTE subfamilies.

### 4.2.5 | DNA Methylation Quantification and Biological Age Estimation

Genomic DNA was extracted from buffy coat fractions of anti-coagulated blood samples using Chemagic DNA buffy coat kit (PerkinElmer, Germany) with Chemagic Magnetic Separation Module 1 and Chemagic Prime 8 Automated Workstation, and was subsequently bisulfite converted using the EZ-96 DNA Methylation-LightningMagPrep from Zymo according to the manufacturer's instructions. DNA methylation levels were measured on Illumina iScan using Illumina's Human MethylationEPIC BeadChip. Sample-level and probe-level quality control was performed using the "minfi" package in R (version 3.5.0). GrimAge was calculated based on the algorithm developed by Lu et al., using 1030 CpG sites. Biological age was defined as the residual (in years) that results from regressing GrimAge on chronological age (Lu et al. 2019).

### 4.2.6 | Demographic and Biochemical Variables

Data on age, sex, and smoking status were based on self-reports. Smoking status was defined as a dichotomous variable (e.g., current smoker and not-current smoker). Missing values for smoking status were imputed using cotinine metabolite levels, measured through the Metabolon HD4 platform (St Helen et al. 2012). Differential blood cell counts (e.g., erythrocytes, leukocytes, basophils, eosinophils, lymphocytes, monocytes, neutrophils) were performed at the Central Laboratory of the University Hospital in Bonn using EDTA-whole blood samples

on a hematological analyzer Sysmex XN9000. For the analysis of biological age, defined as described below, blood cell type proportions, estimated from methylation data, were used as covariates (Houseman et al. 2012).

### 4.3 | Statistical Analysis

#### 4.3.1 | Association of Chronological Age and Sex With RTE Subfamily Expression Levels

We performed multivariable regression analyses to assess the association of chronological age and sex with the expression levels of each RTE subfamily. We also tested possible age and sex interaction effects on RTE subfamily expression levels. As covariates, we included smoking status and batch (*Model 1*). To account for blood cell composition, we further adjusted the model for white and red blood cell counts, and the relative fractions of basophils, eosinophils, lymphocytes, monocytes, and neutrophils (*Model 2*). All the continuous variables were standardized in order to enable comparisons among variables with different units. To account for multiple comparisons, the results from the regression models were corrected using the false discovery rate (FDR) method. The level for statistical significance was set at  $FDR < 0.05$ .

#### 4.3.2 | RTE Subfamily Expression Levels and Biological Age

Multivariable regression was also used to test the association of expression levels of RTE subfamilies with biological age. First, we regressed out the batch effect of methylation from biological age. Next, we used the biological age residuals as the outcome to investigate its relation with RTE subfamilies. As previously, *Model 1* was adjusted for sex, smoking status, and batch of RNA-Seq data. To further account for changes in blood composition with age and/or sex, we additionally adjusted for white and red blood cell counts and the relative fractions of CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, natural killer cells, B cells, monocytes, and granulocytes estimated from methylation data (*Model 2*).

#### 4.3.3 | Association of RTE Locus Expression Levels With Chronological Age and Sex

To identify which RTE loci were driving the effects of age and/or sex, for RTE subfamilies that were significantly associated with age or sex (at  $FDR < 0.05$ ), we additionally assessed the associations of age and sex with RTE expression at the level of each single locus. To this end, we employed a zero-inflated Gaussian mixed model in order to account for excessive zeros in the expression data of RTE loci. Specifically, the expression at each RTE locus was used as the outcome, and age and sex were included as independent variables. All models were adjusted for smoking status, white and red blood cell counts, as well as relative fractions of different white blood cell subtypes, including batch as a random effect. The level for statistical significance was set at  $FDR < 0.05$ . As a further sensitivity analysis, we also performed single locus analyses for those

RTE subfamilies that exhibited the largest effect sizes for the association with chronological age, selecting one from each superfamily (i.e., short interspersed nuclear element (SINE), long interspersed nuclear element (LINE), long terminal repeat (LTR) and retroposon).

#### 4.3.4 | Relation of RTE Locus Expression Levels With Mapped/CIS Genes

Using multivariable regression, we also investigated the association of candidate RTE loci (independent variables) with mapped and *cis*-genes (outcome variables), while adjusting for age, sex, batch, smoking status, white and red blood cell counts, and the relative fractions of white blood cell subtypes. *Cis*-genes were defined as those genes lying within a 100 Kb window (upstream or downstream) of a candidate RTE locus, and a *p*-value of  $< 0.05$  was considered statistically significant for this targeted approach. Pathway enrichment analysis was conducted on genes that were significantly associated with the corresponding RTE locus. The pathway enrichment analysis was performed using the “clusterProfiler” R package (v. 3.18.1), querying Gene Ontology: Biological Processes (GO:BP). Redundant terms collected from Gene Ontology were grouped using the *rrvgo* R Bioconductor package (v. 1.2.0) with default parameters.

#### 4.3.5 | Sensitivity Analysis

To investigate whether the changes in RTE expression levels across different subfamilies were additive with regard to chronological age, biological age, and sex, we performed additional regression analyses using *Model 1* and *Model 2* as described previously, but now using the summary scores of RTE expression as outcomes (i.e., RTE family Z-score, RTE superfamily Z-score and overall RTE Z-score as defined above).

#### 4.3.6 | Exploring the Associations of “RTE Repressor” Genes With RTE Subfamilies

To evaluate the association of age- and sex-dependent RTE subfamilies with genes involved in the maintenance of heterochromatin structure, we selected four key methyltransferase and co-repressor genes whose expression levels were available in the Rhineland Study transcriptomics dataset (i.e., *DAXX*, *TRIM28*, *SETDB1b*, and *DNMT1*) (Gerber et al. 2021). Specifically, we created a summary score for these four “RTE repressor” genes by regressing out batch effects from the expression levels of each gene, followed by summing up the Z-standardized residuals of the four genes. A multivariable linear regression analysis was then performed to explore the association between the RTE repressor score (main independent variable) and each age- and sex-related RTE subfamily, as well as all genes expressed in blood samples of Rhineland Study participants (dependent variable). The analysis was adjusted for age, sex, and smoking status (*Model 1*), and additionally for white and red blood cell counts, and the relative fractions of basophils, eosinophils, lymphocytes, monocytes, and neutrophils (*Model 2*). The significance level was set at  $FDR < 0.05$ . All the statistical analyses were performed in R 4.0.3.

### 4.3.7 | External Validation

We used data from a subset of the Lifelines Study, an independent population-based cohort study, to validate our key results regarding the associations of both chronological age and sex with the expression levels of different RTE subfamilies. The Lifelines Study is a large population-based cohort study in the north of the Netherlands, which collects deep pheno- and genotypic data (Sijtsma et al. 2022). To ensure comparability with the participants of the Rhineland Study, we only included participants aged 30 years and above. Of these, 706 had both RNA-seq and the required covariate data available and were included in our validation study.

In Lifelines, RNA was isolated from whole blood, collected in PAXgene tubes using the PAXgene Blood miRNA Kit (Qiagen, California, USA). Illumina's HiSeq2000 instrument was used for sequencing with paired-end 2×50 bp reads. Detailed information about the RNA isolation and sequencing process is described in the LifeLines DEEP manuscript (Tigchelaar et al. 2015). To obtain the RTE subfamily counts in the Lifelines dataset, we applied the SQuIRE pipeline to the RNA-Seq data, using the same normalization and filtering steps as described for the Rhineland Study data. Smoking status was collected through questionnaires, while the concentrations of blood cells (i.e., leukocytes, mononuclear cells, and neutrophilic granulocytes) were assessed in plasma extracted from fresh blood samples at the laboratory center of the University Medical Center Groningen.

We assessed the association of chronological age and sex with the expression levels of RTE subfamilies through multivariable regression analyses, adjusting for smoking status (*Model 1*), and additionally for leukocytes, neutrophils, and monocyte concentrations (*Model 2*). All the continuous variables were standardized. For validation of the associations that survived FDR correction in the discovery cohort, the level of statistical significance was set at  $p$ -value < 0.05 in the replication cohort.

### Author Contributions

V.T.: conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft; H.-M.L.: conceptualization, methodology, writing – original draft; D.L.: conceptualization, methodology, data curation, writing – original draft; M.D.B.: data acquisition, methodology, writing – review and editing; P.S.: conceptualization, methodology, validation, supervision, writing – original draft; M.M.B.B.: conceptualization, methodology, data acquisition, validation, writing – original draft, supervision, project administration, funding acquisition; N.A.A.: conceptualization, methodology, data acquisition, validation, writing – original draft, supervision, project administration.

### Acknowledgements

We would like to thank all participants of the Rhineland Study and the study personnel involved in the extensive data collection and all members of PRECISE for performing RNA-sequencing. We would like to also thank the Lifelines Cohort study and its contributing research centers for providing the data, as well as all the study participants. Open Access funding enabled and organized by Projekt DEAL.

### Ethics Statement

The study was approved by the ethics committee of the University of Bonn, Medical Faculty. The Rhineland Study is carried out in accordance with the recommendations of the International Conference on Harmonization (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The Rhineland Study's dataset is not publicly available because of data protection regulations. Access to data can be provided to scientists in accordance with the Rhineland Study's Data Use and Access Policy. Requests for further information or to access the Rhineland Study's dataset should be directed to rs-duac@dzne.de. Lifelines data may be obtained from a third party and are not publicly available. Researchers can apply to use the Lifelines data used in this study. More information about how to request Lifelines data and the conditions of use can be found on the Lifelines study website (<https://www.lifelines-biobank.com/researchers/working-with-us>).

### References

- Alexander, M., A. G. Ramstead, K. M. Bauer, et al. 2017. "Rab27-Dependent Exosome Production Inhibits Chronic Inflammation and Enables Acute Responses to Inflammatory Stimuli." *Journal of Immunology* 199, no. 10: 3559–3570. <https://doi.org/10.4049/jimmunol.1700904>.
- Almeida, M. V., G. Vernaz, A. L. K. Putman, and E. A. Miska. 2022. "Taming Transposable Elements in Vertebrates: From Epigenetic Silencing to Domestication." *Trends in Genetics* 38, no. 6: 529–553. <https://doi.org/10.1016/j.tig.2022.02.009>.
- Beam, C. R., C. Kaneshiro, J. Y. Jang, C. A. Reynolds, N. L. Pedersen, and M. Gatz. 2018. "Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease." *Journal of Alzheimer's Disease* 64, no. 4: 1077–1083. <https://doi.org/10.3233/jad-180141>.
- Blackford, A. N., and S. P. Jackson. 2017. "ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response." *Molecular Cell* 66, no. 6: 801–817. <https://doi.org/10.1016/j.molcel.2017.05.015>.
- Bourque, G., K. H. Burns, M. Gehring, et al. 2018. "Ten Things You Should Know About Transposable Elements." *Genome Biology* 19, no. 1: 199. <https://doi.org/10.1186/s13059-018-1577-z>.
- Brouha, B., J. Schustak, R. M. Badge, et al. 2003. "Hot L1s Account for the Bulk of Retrotransposition in the Human Population." *Proceedings of the National Academy of Sciences of the United States of America* 100, no. 9: 5280–5285. <https://doi.org/10.1073/pnas.0831042100>.
- Burbage, M., A. Rocanin-Arjo, B. Baudon, et al. 2023. "Epigenetically Controlled Tumor Antigens Derived From Splice Junctions Between Exons and Transposable Elements." *Science Immunology* 8, no. 80: eabm6360. <https://doi.org/10.1126/sciimmunol.abm6360>.
- Cardoso, A. L., A. Fernandes, J. A. Aguilar-Pimentel, et al. 2018. "Towards Frailty Biomarkers: Candidates From Genes and Pathways Regulated in Aging and Age-Related Diseases." *Ageing Research Reviews* 47: 214–277. <https://doi.org/10.1016/j.arr.2018.07.004>.
- Chen, H., X. Zheng, D. Xiao, and Y. Zheng. 2016. "Age-Associated de-Repression of Retrotransposons in the Drosophila Fat Body, Its Potential Cause and Consequence." *Aging Cell* 15, no. 3: 542–552. <https://doi.org/10.1111/ace.12465>.
- Chuong, E. B., N. C. Elde, and C. Feschotte. 2016. "Regulatory Evolution of Innate Immunity Through Co-Option of Endogenous Retroviruses."



- Science 351, no. 6277: 1083–1087. <https://doi.org/10.1126/science.aad5497>.
- Dang, W., K. K. Steffen, R. Perry, et al. 2009. "Histone H4 Lysine 16 Acetylation Regulates Cellular Lifespan." *Nature* 459, no. 7248: 802–807. <https://doi.org/10.1038/nature08085>.
- De Cecco, M., S. W. Criscione, A. L. Peterson, N. Neretti, J. M. Sedivy, and J. A. Kreiling. 2013. "Transposable Elements Become Active and Mobile in the Genomes of Aging Mammalian Somatic Tissues." *Aging (Albany NY)* 5, no. 12: 867–883. <https://doi.org/10.18632/aging.100621>.
- De Cecco, M., T. Ito, A. P. Petrashen, et al. 2019. "L1 Drives IFN in Senescent Cells and Promotes Age-Associated Inflammation." *Nature* 566, no. 7742: 73–78. <https://doi.org/10.1038/s41586-018-0784-9>.
- Duarte, L. F., A. R. Young, Z. Wang, et al. 2014. "Histone H3.3 and Its Proteolytically Processed Form Drive a Cellular Senescence Programme." *Nature Communications* 5: 5210. <https://doi.org/10.1038/ncomms6210>.
- Elbarbary, R. A., B. A. Lucas, and L. E. Maquat. 2016. "Retrotransposons as Regulators of Gene Expression." *Science* 351, no. 6274: aac7247. <https://doi.org/10.1126/science.aac7247>.
- Feser, J., D. Truong, C. Das, et al. 2010. "Elevated Histone Expression Promotes Life Span Extension." *Molecular Cell* 39, no. 5: 724–735. <https://doi.org/10.1016/j.molcel.2010.08.015>.
- Geis, F. K., and S. P. Goff. 2020. "Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview." *Viruses* 12, no. 8: 884. <https://doi.org/10.3390/v12080884>.
- Gerber, J. P., J. Russ, V. Chandrasekar, et al. 2021. "Aberrant Chromatin Landscape Following Loss of the H3.3 Chaperone Daxx in Haematopoietic Precursors Leads to Pu.1-Mediated Neutrophilia and Inflammation." *Nature Cell Biology* 23, no. 12: 1224–1239. <https://doi.org/10.1038/s41556-021-00774-y>.
- Gorbunova, V., A. Seluanov, P. Mita, et al. 2021. "The Role of Retrotransposable Elements in Ageing and Age-Associated Diseases." *Nature* 596, no. 7870: 43–53. <https://doi.org/10.1038/s41586-021-03542-y>.
- Groh, S., and G. Schotta. 2017. "Silencing of Endogenous Retroviruses by Heterochromatin." *Cellular and Molecular Life Sciences* 74, no. 11: 2055–2065. <https://doi.org/10.1007/s00018-017-2454-8>.
- Gvozdev, V. A., G. L. Kogan, and L. A. Usakin. 2005. "The Y Chromosome as a Target for Acquired and Amplified Genetic Material in Evolution." *BioEssays* 27, no. 12: 1256–1262. <https://doi.org/10.1002/bies.20321>.
- Hägg, S., and J. Jylhävä. 2021. "Sex Differences in Biological Aging With a Focus on Human Studies." *eLife* 10: e63425. <https://doi.org/10.7554/eLife.63425>.
- Haggerty, C., H. Kretzmer, C. Riemenschneider, et al. 2021. "Dnmt1 has de Novo Activity Targeted to Transposable Elements." *Nature Structural & Molecular Biology* 28, no. 7: 594–603. <https://doi.org/10.1038/s41594-021-00603-8>.
- Holly, M. K., K. Diaz, and J. G. Smith. 2017. "Defensins in Viral Infection and Pathogenesis." *Annual Review of Virology* 4, no. 1: 369–391. <https://doi.org/10.1146/annurev-virology-101416-041734>.
- Houseman, E. A., W. P. Accomando, D. C. Koestler, et al. 2012. "DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution." *BMC Bioinformatics* 13: 86. <https://doi.org/10.1186/1471-2105-13-86>.
- Jiang, Z., F. Wei, Y. Zhang, et al. 2021. "IFI16 Directly Senses Viral RNA and Enhances RIG-I Transcription and Activation to Restrict Influenza Virus Infection." *Nature Microbiology* 6, no. 7: 932–945. <https://doi.org/10.1038/s41564-021-00907-x>.
- Jones, M. J., S. J. Goodman, and M. S. Kobor. 2015. "DNA Methylation and Healthy Human Aging." *Aging Cell* 14, no. 6: 924–932. <https://doi.org/10.1111/ace.12349>.
- Katzourakis, A., A. Rambaut, and O. G. Pybus. 2005. "The Evolutionary Dynamics of Endogenous Retroviruses." *Trends in Microbiology* 13, no. 10: 463–468. <https://doi.org/10.1016/j.tim.2005.08.004>.
- Klein, S. L., and K. L. Flanagan. 2016. "Sex Differences in Immune Responses." *Nature Reviews. Immunology* 16, no. 10: 626–638. <https://doi.org/10.1038/nri.2016.90>.
- Lanciano, S., and G. Cristofari. 2020. "Measuring and Interpreting Transposable Element Expression." *Nature Reviews. Genetics* 21, no. 12: 721–736. <https://doi.org/10.1038/s41576-020-0251-y>.
- Lanciano, S., C. Philippe, A. Sarkar, et al. 2024. "Locus-Level L1 DNA Methylation Profiling Reveals the Epigenetic and Transcriptional Interplay Between L1s and Their Integration Sites." *Cell Genomics* 4, no. 2: 100498. <https://doi.org/10.1016/j.xgen.2024.100498>.
- Lander, E. S., L. M. Linton, B. Birren, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409, no. 6822: 860–921. <https://doi.org/10.1038/35057062>.
- LaRocca, T. J., A. N. Cavalier, and D. Wahl. 2020. "Repetitive Elements as a Transcriptomic Marker of Aging: Evidence in Multiple Datasets and Models." *Aging Cell* 19, no. 7: e13167. <https://doi.org/10.1111/ace.13167>.
- Law, J. A., and S. E. Jacobsen. 2010. "Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals." *Nature Reviews. Genetics* 11, no. 3: 204–220. <https://doi.org/10.1038/nrg2719>.
- Lee, B., E. Bang, W. S. Yang, et al. 2018. "The Possible Role of Neurobeachin in Extinction of Contextual Fear Memory." *Scientific Reports* 8, no. 1: 13752. <https://doi.org/10.1038/s41598-018-30589-1>.
- Leening, M. J., B. S. Ferket, E. W. Steyerberg, et al. 2014. "Sex Differences in Lifetime Risk and First Manifestation of Cardiovascular Disease: Prospective Population Based Cohort Study." *BMJ* 349: g5992. <https://doi.org/10.1136/bmj.g5992>.
- Leung, D. C., K. B. Dong, I. A. Maksakova, et al. 2011. "Lysine Methyltransferase G9a Is Required for de Novo DNA Methylation and the Establishment, but Not the Maintenance, of Proviral Silencing." *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 14: 5718–5723. <https://doi.org/10.1073/pnas.1014660108>.
- Lin, C. W., J. Ellegood, K. Tamada, et al. 2023. "An Old Model With New Insights: Endogenous Retroviruses Drive the Evolvement Toward ASD Susceptibility and Hijack Transcription Machinery During Development." *Molecular Psychiatry* 28, no. 5: 1932–1945. <https://doi.org/10.1038/s41380-023-01999-z>.
- Liu, S., S. E. Heumuller, A. Hossinger, et al. 2023. "Reactivated Endogenous Retroviruses Promote Protein Aggregate Spreading." *Nature Communications* 14, no. 1: 5034. <https://doi.org/10.1038/s41467-023-40632-z>.
- Liu, X., Z. Liu, Z. Wu, et al. 2023. "Resurrection of Endogenous Retroviruses During Aging Reinforces Senescence." *Cell* 186, no. 2: 287–304. <https://doi.org/10.1016/j.cell.2022.12.017>.
- Lombard, D. B., K. F. Chua, R. Mostoslavsky, S. Franco, M. Gostissa, and F. W. Alt. 2005. "DNA Repair, Genome Stability, and Aging." *Cell* 120, no. 4: 497–512. <https://doi.org/10.1016/j.cell.2005.01.028>.
- Lopez-Otin, C., M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer. 2023. "Hallmarks of Aging: An Expanding Universe." *Cell* 186, no. 2: 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>.
- Lu, A. T., A. Quach, J. G. Wilson, et al. 2019. "DNA Methylation GrimAge Strongly Predicts Lifespan and Healthspan." *Aging (Albany NY)* 11, no. 2: 303–327. <https://doi.org/10.18632/aging.101684>.
- Lu, J. Y., W. Shao, L. Chang, et al. 2020. "Genomic Repeats Categorize Genes With Distinct Functions for Orchestrated Regulation." *Cell Reports* 30, no. 10: 3296–3311. <https://doi.org/10.1016/j.celrep.2020.02.048>.



- Mao, J., Q. Zhang, and Y. S. Cong. 2021. "Human Endogenous Retroviruses in Development and Disease." *Computational and Structural Biotechnology Journal* 19: 5978–5986. <https://doi.org/10.1016/j.csbj.2021.10.037>.
- Marsano, R. M., and P. Dimitri. 2022. "Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements." *Cells* 11, no. 5: 761. <https://doi.org/10.3390/cells11050761>.
- Merlotti, A., B. Sadacca, Y. A. Arribas, et al. 2023. "Noncanonical Splicing Junctions Between Exons and Transposable Elements Represent a Source of Immunogenic Recurrent Neo-Antigens in Patients With Lung Cancer." *Science Immunology* 8, no. 80: eabm6359. <https://doi.org/10.1126/sciimmunol.abm6359>.
- Moisan, F., S. Kab, F. Mohamed, et al. 2016. "Parkinson Disease Male-To-Female Ratios Increase With Age: French Nationwide Study and Meta-Analysis." *Journal of Neurology, Neurosurgery, and Psychiatry* 87, no. 9: 952–957. <https://doi.org/10.1136/jnnp-2015-312283>.
- Moser, K. L., B. R. Neas, J. E. Salmon, et al. 1998. "Genome Scan of Human Systemic Lupus Erythematosus: Evidence for Linkage on Chromosome 1q in African-American Pedigrees." *Proceedings of the National Academy of Sciences of the United States of America* 95, no. 25: 14869–14874. <https://doi.org/10.1073/pnas.95.25.14869>.
- Mulhern, M. S., C. Stumpel, N. Stong, et al. 2018. "NBEA: Developmental Disease Gene With Early Generalized Epilepsy Phenotypes." *Annals of Neurology* 84, no. 5: 788–795. <https://doi.org/10.1002/ana.25350>.
- Nguyen, A. H., and D. Bachtrog. 2021. "Toxic Y Chromosome: Increased Repeat Expression and Age-Associated Heterochromatin Loss in Male *Drosophila* With a Young Y Chromosome." *PLoS Genetics* 17, no. 4: e1009438. <https://doi.org/10.1371/journal.pgen.1009438>.
- Pal, S., and J. K. Tyler. 2016. "Epigenetics and Aging." *Science Advances* 2, no. 7: e1600584. <https://doi.org/10.1126/sciadv.1600584>.
- Peters, M. J., R. Joehanes, L. C. Pilling, et al. 2015. "The Transcriptional Landscape of Age in Human Peripheral Blood." *Nature Communications* 6: 8570. <https://doi.org/10.1038/ncomms9570>.
- Rapkin, L. M., K. Ahmed, S. Dulev, et al. 2015. "The Histone Chaperone DAXX Maintains the Structural Organization of Heterochromatin Domains." *Epigenetics & Chromatin* 8: 44. <https://doi.org/10.1186/s13072-015-0036-2>.
- Rigal, J., A. Martin Anduaga, E. Bitman, E. Rivellese, S. Kadener, and M. T. Marr. 2022. "Artificially Stimulating Retrotransposon Activity Increases Mortality and Accelerates a Subset of Aging Phenotypes in *Drosophila*." *eLife* 11: e80169. <https://doi.org/10.7554/eLife.80169>.
- Rowe, H. M., J. Jakobsson, D. Mesnard, et al. 2010. "KAP1 Controls Endogenous Retroviruses in Embryonic Stem Cells." *Nature* 463, no. 7278: 237–240. <https://doi.org/10.1038/nature08674>.
- Sijsma, A., J. Rienks, P. van der Harst, G. Navis, J. G. M. Rosmalen, and A. Dotinga. 2022. "Cohort Profile Update: Lifelines, a Three-Generation Cohort Study and Biobank." *International Journal of Epidemiology* 51, no. 5: e295–e302. <https://doi.org/10.1093/ije/dyab257>.
- Slagboom, E., N. van den Berg, and J. Deelen. 2018. "Phenome and Genome Based Studies Into Human Ageing and Longevity: An Overview." *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1864, no. 9 Pt A: 2742–2751. <https://doi.org/10.1016/j.bbadis.2017.09.017>.
- Smit, A., R. Hubley, and P. Green. 2013–2015. "RepeatMasker Open-4.0." <http://www.repeatmasker.org>.
- St Helen, G., M. Novalen, D. F. Heitjan, et al. 2012. "Reproducibility of the Nicotine Metabolite Ratio in Cigarette Smokers." *Cancer Epidemiology, Biomarkers & Prevention* 21, no. 7: 1105–1114. <https://doi.org/10.1158/1055-9965.Epi-12-0236>.
- Stoye, J. P. 2001. "Endogenous Retroviruses: Still Active After All These Years?" *Current Biology* 11, no. 22: R914–R916. [https://doi.org/10.1016/s0960-9822\(01\)00553-x](https://doi.org/10.1016/s0960-9822(01)00553-x).
- Sturm, A., E. Saskoi, B. Hotzi, et al. 2023. "Downregulation of Transposable Elements Extends Lifespan in *Caenorhabditis elegans*." *Nature Communications* 14, no. 1: 5278. <https://doi.org/10.1038/s41467-023-40957-9>.
- Sun, X., X. Wang, Z. Tang, et al. 2018. "Transcription Factor Profiling Reveals Molecular Choreography and Key Regulators of Human Retrotransposon Expression." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 24: E5526–E5535. <https://doi.org/10.1073/pnas.1722565115>.
- Tang, W., S. Mun, A. Joshi, K. Han, and P. Liang. 2018. "Mobile Elements Contribute to the Uniqueness of Human Genome With 15,000 Human-Specific Insertions and 14 Mbp Sequence Increase." *DNA Research* 25, no. 5: 521–533. <https://doi.org/10.1093/dnares/dsy022>.
- Tigchelaar, E. F., A. Zhernakova, J. A. Dekens, et al. 2015. "Cohort Profile: LifeLines DEEP, a Prospective, General Population Cohort Study in the Northern Netherlands: Study Design and Baseline Characteristics." *BMJ Open* 5, no. 8: e006772. <https://doi.org/10.1136/bmjopen-2014-006772>.
- Tramunt, B., S. Smati, N. Grandgeorge, et al. 2020. "Sex Differences in Metabolic Regulation and Diabetes Susceptibility." *Diabetologia* 63, no. 3: 453–461. <https://doi.org/10.1007/s00125-019-05040-3>.
- Tsai, Y. T., N. Seymen, I. R. Thompson, et al. 2024. "Expression of Most Retrotransposons in Human Blood Correlates With Biological Aging." *eLife* 13: RP96575. <https://doi.org/10.7554/eLife.96575>.
- Turelli, P., N. Castro-Diaz, F. Marzetta, et al. 2014. "Interplay of TRIM28 and DNA Methylation in Controlling Human Endogenous Retroelements." *Genome Research* 24, no. 8: 1260–1270. <https://doi.org/10.1101/gr.172833.114>.
- Turner, G., M. Barbulescu, M. Su, M. I. Jensen-Seaman, K. K. Kidd, and J. Lenz. 2001. "Insertional Polymorphisms of Full-Length Endogenous Retroviruses in Humans." *Current Biology* 11, no. 19: 1531–1535. [https://doi.org/10.1016/s0960-9822\(01\)00455-9](https://doi.org/10.1016/s0960-9822(01)00455-9).
- Unterholzner, L., S. E. Keating, M. Baran, et al. 2010. "IFI16 Is an Innate Immune Sensor for Intracellular DNA." *Nature Immunology* 11, no. 11: 997–1004. <https://doi.org/10.1038/ni.1932>.
- Volkman, H. E., and D. B. Stetson. 2014. "The Enemy Within: Endogenous Retroelements and Autoimmune Disease." *Nature Immunology* 15, no. 5: 415–422. <https://doi.org/10.1038/ni.2872>.
- Wood, J. G., B. C. Jones, N. Jiang, et al. 2016. "Chromatin-Modifying Genetic Interventions Suppress Age-Associated Transposable Element Activation and Extend Life Span in *Drosophila*." *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 40: 11277–11282. <https://doi.org/10.1073/pnas.1604621113>.
- Xu, H., L. Yong, X. Gao, et al. 2024. "CaMK4: Structure, Physiological Functions, and Therapeutic Potential." *Biochemical Pharmacology* 224: 116204. <https://doi.org/10.1016/j.bcp.2024.116204>.
- Yang, W. R., D. Ardeljan, C. N. Pacyna, L. M. Payer, and K. H. Burns. 2019. "SQUIRE Reveals Locus-Specific Regulation of Interspersed Repeat Expression." *Nucleic Acids Research* 47, no. 5: e27. <https://doi.org/10.1093/nar/gky1301>.
- Yong, L., Y. Yu, B. Li, et al. 2022. "Calcium/Calmodulin-Dependent Protein Kinase IV Promotes Imiquimod-Induced Psoriatic Inflammation via Macrophages and Keratinocytes in Mice." *Nature Communications* 13, no. 1: 4255. <https://doi.org/10.1038/s41467-022-31935-8>.
- Zampieri, M., F. Ciccarone, R. Calabrese, C. Franceschi, A. Burkle, and P. Caiafa. 2015. "Reconfiguration of DNA Methylation in Aging." *Mechanisms of Ageing and Development* 151: 60–70. <https://doi.org/10.1016/j.mad.2015.02.002>.
- Zhang, Q., J. Pan, Y. Cong, and J. Mao. 2022. "Transcriptional Regulation of Endogenous Retroviruses and Their Misregulation in

Human Diseases.” *International Journal of Molecular Sciences* 23, no. 17: 10112. <https://doi.org/10.3390/ijms231710112>.

Zhang, W., J. Qu, G. H. Liu, and J. C. I. Belmonte. 2020. “The Ageing Epigenome and Its Rejuvenation.” *Nature Reviews. Molecular Cell Biology* 21, no. 3: 137–150. <https://doi.org/10.1038/s41580-019-0204-5>.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.