

Brief Report

Not peer-reviewed version

LINE-1 Depletion at Promoters of Neurodevelopmental Disorder Genes: A Genome-Wide Analysis

[Can Sevilimis](#) *

Posted Date: 10 April 2026

doi: 10.20944/preprints202604.0715.v1

Keywords: LINE-1; retrotransposons; neurodevelopmental disorders; promoter regions; purifying selection; pleiotropy; epigenetics; seizure; attention-deficit/hyperactivity disorder; genome-wide analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Brief Report

LINE-1 Depletion at Promoters of Neurodevelopmental Disorder Genes: A Genome-Wide Analysis

Can Sevilmiş

Department of Molecular Biology and Genetics, Bahçeşehir University, 34353 Istanbul, Turkey; can.sevilmis@bahcesehir.edu.tr

Abstract

LINE-1 retrotransposons constitute approximately 17% of the human genome and are capable of influencing gene expression when inserted in proximity to regulatory regions. Neurodevelopmental disorder (NDD) genes require precise spatiotemporal regulation during brain development; however, the relationship between LINE-1 occupancy and the promoter architecture of NDD-associated loci has not been systematically examined. Here, a genome-wide computational analysis was performed comparing LINE-1 element density in promoter regions (± 2 kb from the transcription start site) across five NDD gene sets — including genes annotated to autism spectrum disorder (SFARI Gene database), seizure (HP:0001250), attention-deficit/hyperactivity disorder (HP:0007018), their phenotypic intersection, and syndromic NDD genes — against a curated housekeeping gene set ($n=1,982$) derived from the HRT Atlas. All NDD gene sets exhibited significantly lower LINE-1 promoter occupancy compared to housekeeping genes (Mann-Whitney U test, $p < 0.05$ across all tested NDD subsets). A consistent gradient was observed, with genes annotated to both seizure and ADHD phenotypes showing the lowest LINE-1 occupancy (23.5% vs 31.1% in housekeeping genes; $p = 0.0029$, rank-biserial $r = 0.082$). The observed depletion was further supported by length-matched control analysis ($n=678$ pairs, $p = 0.036$, $r = 0.066$), suggesting that the signal is not fully explained by intronic size differences. Furthermore, no significant differences in GC content ($p = 0.3289$) or CpG observed/expected ratios ($p = 0.9665$) were detected between NDD Tier 1 and housekeeping gene promoters, indicating that the results are not attributable to sequence composition biases. These findings are consistent with stronger selective constraint against LINE-1 insertions at NDD promoters, potentially reflecting the intolerance of these loci to transcriptional dysregulation. The pronounced depletion in genes annotated to both seizure and ADHD phenotypes provides a genomic context for understanding the regulatory vulnerability of pleiotropic neurodevelopmental loci, with implications for interpreting non-coding variation in clinical genomics.

Keywords: LINE-1; retrotransposons; neurodevelopmental disorders; promoter regions; purifying selection; pleiotropy; epigenetics; seizure; attention-deficit/hyperactivity disorder; genome-wide analysis

1. Introduction

Long Interspersed Element-1 (LINE-1, L1) retrotransposons are the most abundant autonomous mobile genetic elements in the human genome, comprising approximately 17% of total genomic sequence[1–4]. Full-length LINE-1 elements span approximately 6 kilobases and encode two open reading frames: ORF1p, an RNA-binding protein, and ORF2p, a bifunctional protein with endonuclease and reverse transcriptase activities that together execute retrotransposition via a target-primed reverse transcription mechanism[3]. Although the vast majority of genomic LINE-1 copies are retrotransposition-incompetent due to 5' truncation or accumulated mutations, an estimated 80–100 loci per haploid genome retain retrotransposition competence, and ongoing LINE-1 activity continues to be a source of inter-individual genomic variation[4].

Beyond their role as mobile elements, LINE-1 sequences exert regulatory influence on the transcriptome through multiple mechanisms. The LINE-1 5' untranslated region harbors internal sense promoter activity, the strength of which is substantially modulated by upstream flanking genomic sequences at the insertion site[5]. Additionally, the same region contains an antisense promoter capable of initiating transcription of neighboring sequences and influencing the expression of adjacent protein-coding genes[6]. In human neural progenitor cells, genome-wide loss of CpG methylation leads to selective transcriptional activation of evolutionarily young, hominoid-specific LINE-1 elements, which subsequently act as alternative promoters for hundreds of protein-coding genes enriched for neuronal functions[7]. Notably, several genes regulated through this LINE-1-mediated promoter mechanism have been independently implicated in neurodevelopmental disorders, including autism, cognitive impairment, and epilepsy[7].

LINE-1 elements exhibit elevated activity during neuronal development and are subject to relaxed silencing constraints in neural lineages relative to other somatic tissues. Muotri et al. demonstrated that engineered human LINE-1 elements can retrotranspose in rat hippocampal neural progenitor cells in vitro and produce somatic neuronal mosaicism in transgenic mice in vivo[8]. Coufal et al. subsequently confirmed that human LINE-1 elements retrotranspose in neural progenitors derived from human fetal brain and embryonic stem cells, and detected a statistically significant increase in endogenous LINE-1 copy number in hippocampal tissue relative to non-neural somatic tissues from the same individuals[9]. These findings indicate that LINE-1 retrotransposition occurs preferentially in the neural lineage and may contribute to somatic genomic diversity among post-mitotic neurons[10]. The preferential activity of LINE-1 in neural progenitors is mechanistically linked to developmentally regulated epigenetic remodeling: LINE-1 promoter methylation decreases transiently during the transition from neural stem cells to progenitors, coinciding with a window of permissive retrotransposition activity[7,8].

The proximity of LINE-1 elements to gene regulatory regions raises the question of whether genomic loci under strong functional constraint exhibit systematic differences in LINE-1 occupancy relative to less constrained regions. Genes associated with neurodevelopmental disorders (NDDs) are characterized by unusually strict spatiotemporal regulation during brain development, haploinsufficiency intolerance, and high constraint scores in population genetics databases. If LINE-1 insertion into the promoter regions of NDD-associated genes is deleterious and subject to negative selection, these loci would be expected to show lower LINE-1 occupancy relative to constitutively expressed housekeeping genes, which are under comparatively relaxed regulatory constraints. To the author's knowledge, this prediction has not been systematically examined at genome scale across multiple NDD gene sets.

Here, a genome-wide computational analysis was performed comparing LINE-1 element density in promoter regions (± 2 kb from the transcription start site) across five independently curated NDD gene sets and a housekeeping gene control set derived from the HRT Atlas, using pre-computed RepeatMasker annotations for the GRCh38 assembly and intersection analysis with BEDTools. It was further assessed whether LINE-1 occupancy patterns extend to intronic regions and whether observed differences are confounded by gene length.

2. Materials and Methods

2.1. Data Sources and Gene Set Curation

Five NDD-associated gene sets were curated from two independent, publicly available databases. Autism spectrum disorder (ASD) candidate genes were obtained from the SFARI Gene database (March 2026 release)[11]. Genes were stratified by evidence score: a high-confidence set (Tier 1) comprised genes with SFARI scores 1 or 2 ($n=952$), and an extended set (Tier 2) additionally included score 3 genes ($n=1,173$). Genes annotated as syndromic (syndromic flag = 1) were retained as a separate subgroup ($n=309$ promoters) for subgroup analysis. The Tier 1 set was used as the primary NDD gene set in all main analyses.

Seizure-associated genes (HP:0001250, $n=2,110$) and attention-deficit/hyperactivity disorder-associated genes (HP:0007018, $n=441$) were downloaded from the Human Phenotype Ontology (HPO) database[12]. A phenotypic intersection set was defined as genes annotated to both HPO terms ($n=346$), representing loci implicated in epilepsy–ADHD comorbidity at the annotation level. The phenotypic intersection of seizure and ADHD was specifically selected for analysis because it represents one of the most highly prevalent clinical comorbidities in neurodevelopment, providing a robust model to test the regulatory constraints on pleiotropic loci[13].

The housekeeping gene reference set was obtained from the HRT Atlas v1.0[14], which defines housekeeping genes as transcripts stably expressed across 52 human tissue and cell types. The raw set comprised 2,176 unique gene symbols. To ensure independence between gene sets, genes present in the SFARI Tier 2 set were excluded from the housekeeping list prior to analysis, yielding a housekeeping control list of 2,031 gene symbols. Following coordinate mapping against the GENCODE v47 reference and exclusion of out-of-bounds intervals, 1,982 unique promoter regions were successfully retained for downstream intersection analysis.

2.2. Genomic Coordinate Processing

Genomic annotations were based on the GRCh38 (hg38) reference assembly. Protein-coding gene coordinates were extracted from GENCODE v47[15] using the comprehensive annotation GTF file. For each protein-coding gene, promoter regions were defined as ± 2 kb windows centered on the transcription start site (TSS), resolved in a strand-aware manner: for genes on the positive strand, the TSS was defined as the annotated start coordinate; for genes on the negative strand, the TSS was defined as the annotated end coordinate. Windows extending below coordinate zero were excluded. This procedure yielded 20,092 unique protein-coding gene promoter regions.

Intronic coordinates were derived from the same GTF file. Transcript-level exon boundaries were extracted for all protein-coding transcripts ($n=65,049$). For each transcript with two or more exons, intronic intervals were computed as the regions between consecutive exon end and start coordinates, excluding the exon boundaries themselves. This yielded 658,165 unique intronic intervals across all transcripts.

All genomic interval files were formatted as BED files and sorted by chromosome and start coordinate prior to intersection analysis.

2.3. LINE-1 Annotation and Intersection Analysis

LINE-1 element coordinates were obtained from the UCSC RepeatMasker track for the hg38 assembly[16], downloaded as a flat text file (rmsk.txt). Elements were filtered to retain only those classified as LINE/L1 family ($n=1,031,524$), and reformatted as a sorted BED file.

Intersection analyses between gene set promoter regions and LINE-1 elements were performed using BEDTools v2.31.1[17]. For each promoter or intronic interval, the number of overlapping LINE-1 elements was computed using the `bedtools intersect -c` command with the `-sorted` flag. Gene set-specific BED files were generated by filtering the genome-wide promoter and intron coordinate files against each curated gene list.

2.4. Statistical Analysis

All statistical analyses were performed in Python 3.13 using the `scipy.stats` and `pandas` libraries. LINE-1 element counts per promoter were compared between each NDD gene set and the housekeeping control using the two-sided Mann-Whitney U test, a non-parametric test appropriate for count data with non-normal distributions. Effect size was quantified as the rank-biserial correlation coefficient r , calculated as:

$$r = 1 - \frac{2U}{n_1 n_2} \tag{1}$$

where U is the Mann-Whitney U statistic and n_1, n_2 are the sample sizes of the two groups. A positive r indicates that the NDD gene set has lower LINE-1 counts than the housekeeping set. Approximate 95% confidence intervals for r were computed using the normal approximation to the standard error:

$$95\% \text{ CI} \approx r \pm 1.96 \sqrt{\frac{n_1 + n_2 + 1}{3 n_1 n_2}} \tag{2}$$

For the intronic confounding assessment, gene-level summaries were computed by aggregating total intronic length (in kb) and total LINE-1 count per gene. Intronic LINE-1 density was expressed as LINE-1 elements per kilobase. A length-matched control analysis was performed by identifying, for each NDD Tier 1 gene, the housekeeping gene with the most similar total intronic length within a $\pm 50\%$ tolerance window. Each housekeeping gene was matched to at most one NDD gene; matched pairs were excluded from further matching to ensure independent observations. Mann-Whitney U test and rank-biserial r were then computed on the matched pairs. Spearman rank correlation between total intronic length and LINE-1 density was used to quantify the confounding effect.

A p -value threshold of 0.05 was applied for all comparisons. No correction for multiple testing was applied across gene sets, as each comparison was pre-specified and hypothesis-driven.

3. Results and Discussion

3.1. LINE-1 Occupancy Is Reduced in NDD Gene Promoters

Genome-wide intersection of 1,031,524 LINE-1 elements with promoter regions (± 2 kb from TSS) of 20,092 protein-coding genes revealed substantial variability in LINE-1 occupancy across functional gene categories. Among housekeeping genes ($n=1,982$ promoters), 31.1% harbored at least one LINE-1 element, establishing the baseline occupancy rate for constitutively expressed loci.

The primary NDD gene set (SFARI Tier 1, $n=945$ promoters) exhibited significantly lower LINE-1 occupancy compared to housekeeping genes (27.2%; Mann-Whitney U test: $p = 0.014$, rank-biserial $r = 0.045$; Table 1). This depletion was consistent across gene sets curated from an independent annotation resource: genes annotated to the seizure phenotype in the Human Phenotype Ontology (HP:0001250, $n=2,072$ promoters) showed 28.6% occupancy ($p = 0.036$, $r = 0.031$), and genes annotated to attention-deficit/hyperactivity disorder (HP:0007018, $n=435$ promoters) showed 24.8% occupancy ($p = 0.005$, $r = 0.069$; Table 1). These results indicate that LINE-1 depletion from promoter regions is a consistent feature of NDD-associated loci, independent of the annotation database used to define the gene sets.

3.2. A Gradient of LINE-1 Depletion Across NDD Subgroups

Subgroup analyses revealed a monotonic gradient of LINE-1 depletion that intensified with increasing phenotypic specificity and comorbidity burden (Figure 1). Syndromic NDD genes ($n=309$ promoters), defined as SFARI genes with confirmed syndromic presentation, showed 25.9% LINE-1 occupancy. This difference from housekeeping genes was statistically significant ($p = 0.0344$, $r = 0.061$), with directionality consistent with the broader depletion pattern.

The most pronounced depletion was observed in the phenotypic intersection set — genes annotated to both HP:0001250 (seizure) and HP:0007018 (ADHD) in the HPO database ($n=340$ promoters). This set exhibited only 23.5% LINE-1 occupancy, representing the lowest value across all gene sets examined and a statistically significant difference from the housekeeping baseline ($p = 0.0029$, $r = 0.082$). It is important to note that this intersection is defined purely at the level of HPO annotation; it does not imply that all genes in this set are causally implicated in both phenotypes, as a single gene may be annotated to multiple phenotypes through distinct mechanisms or alleles.

Across all six groups, LINE-1 occupancy followed a consistent gradient: Housekeeping (31.1%) > HPO Seizure (28.6%) > NDD Tier 1 (27.2%) > Syndromic NDD (25.9%) > HPO ADHD (24.8%) > HPO Seizure $\cap$ ADHD (23.5%). All effect sizes were positive, indicating uniform directionality of depletion (Figure 1B).

3.3. Intronic LINE-1 Enrichment Is Primarily Driven by Gene Length Confounding

To assess whether the promoter depletion pattern extended to other genic regions, LINE-1 intersection was additionally performed on intronic coordinates. At the intron level, NDD Tier 1 genes exhibited higher LINE-1 occupancy than housekeeping genes (29.8% vs 22.7%; $p = 4.3 \times 10^{-121}$, $r = 0.056$). However, this apparent enrichment was driven by a pronounced difference in gene architecture: the median total intronic length of NDD Tier 1 genes was 256.4 kb, compared to 37.2 kb for housekeeping genes — a 6.9-fold difference. A Spearman rank correlation between total intronic length and intronic LINE-1 density confirmed a significant positive association in NDD genes ($\rho = 0.30$, $p = 7.9 \times 10^{-21}$), indicating that longer genes accumulate more LINE-1 elements by virtue of their size.

To directly test whether the intronic enrichment signal was confounded by gene length, a length-matched control analysis was performed. For each NDD Tier 1 gene ($n=933$), a housekeeping gene with total intronic length within $\pm 50\%$ was identified without replacement, yielding 678 matched pairs. After matching, the difference in LINE-1 density remained statistically significant ($p = 0.036$, $r = 0.066$), indicating that the observed depletion is unlikely to be explained solely by differences in total intronic length.

In contrast, the promoter depletion signal was not subject to this confounding, as all promoter windows were defined as fixed-length 4 kb intervals centered on the TSS. Taken together, these results indicate that while the massive intronic LINE-1 enrichment in NDD genes is primarily driven by their extended gene length, a modest underlying depletion in LINE-1 density persists even after length correction, paralleling the promoter-level findings.

3.4. Promoter GC Content and CpG Density Do Not Explain LINE-1 Depletion

To determine whether LINE-1 depletion could be explained by local nucleotide composition, GC content and CpG observed/expected (O/E) ratios were calculated for all promoter windows. No significant difference in GC content was found between NDD Tier 1 genes (median 51.9%) and housekeeping genes (median 51.2%, $p = 0.3289$). Similarly, CpG O/E ratios did not differ between the two sets (median 0.581 vs 0.573, $p = 0.9665$). These observations suggest that the promoter architecture of NDD genes does not inherently exclude LINE-1 elements due to primary sequence composition alone.

3.5. Interpretation of Main Findings

The principal finding of this brief report is a consistent depletion of LINE-1 occupancy at promoters of neurodevelopmental disorder (NDD) genes relative to housekeeping controls. This pattern is reproduced across independently curated gene sets and reaches its strongest magnitude in genes shared between seizure and ADHD annotations.

This result is consistent with prior genome-scale observations that promoter regions are generally depleted of transposable elements, indicating non-random evolutionary constraint at transcription-initiation regions[18,19]. In this context, the stronger depletion observed for NDD-associated loci is consistent with the unusually high dosage sensitivity and regulatory fragility of many neurodevelopmental genes[20].

Two internal checks support the robustness of this signal. First, intronic differences are strongly influenced by gene length, whereas promoter comparisons are performed on fixed-length windows, reducing structural confounding. Second, promoter GC content and CpG observed/expected ratios do not differ significantly between NDD Tier 1 and housekeeping sets, arguing against simple sequence-composition bias as the primary explanation.

Mechanistically, this interpretation is plausible because promoter-proximal regions in NDD-relevant genes are enriched for convergent transcription-factor binding and active chromatin signatures in neurodevelopmental contexts[21,22], while repetitive-element landscape studies support reduced tolerance of disruptive insertions near core regulatory intervals[23,24]. In addition, genome-organization

data showing enrichment of LINE-dense regions in more repressive nuclear compartments provide a compatible systems-level context for why promoter-proximal LINE-1 insertions could be particularly disruptive at dosage-sensitive loci[25].

Accordingly, the most parsimonious interpretation is that promoter-proximal LINE-1 depletion reflects stronger regulatory constraint at NDD loci. This interpretation remains inferential because the current analysis is reference-genome based and does not directly model insertion age or population-level polymorphism.

3.6. Limitations

Several limitations of this study warrant consideration. First, the analysis is based on binary LINE-1 occupancy (presence or absence per promoter window) and does not account for LINE-1 element age, subfamily composition, or degree of truncation, all of which may influence regulatory impact. Second, all gene sets were defined at the level of database annotation, and individual genes within these sets may vary substantially in their biological roles and phenotypic relationships. In particular, the HPO Seizure $\cap$ ADHD intersection does not represent a curated comorbidity gene set, but rather the overlap of two independent annotation lists. Third, the promoter window size of ± 2 kb was selected as a standard regulatory interval; the depletion pattern may differ with alternative window definitions. Fourth, effect sizes across all comparisons were small (rank-biserial $r = 0.031$ – 0.081), indicating that while the signal is consistent and statistically robust, the biological magnitude of LINE-1 depletion at the individual gene level is modest. Finally, this is a purely computational study based on a single reference genome snapshot; the interpretation that depletion reflects historical negative selection against LINE-1 insertions at NDD promoters would require population-level insertion polymorphism data and evolutionary analyses beyond the scope of this work.

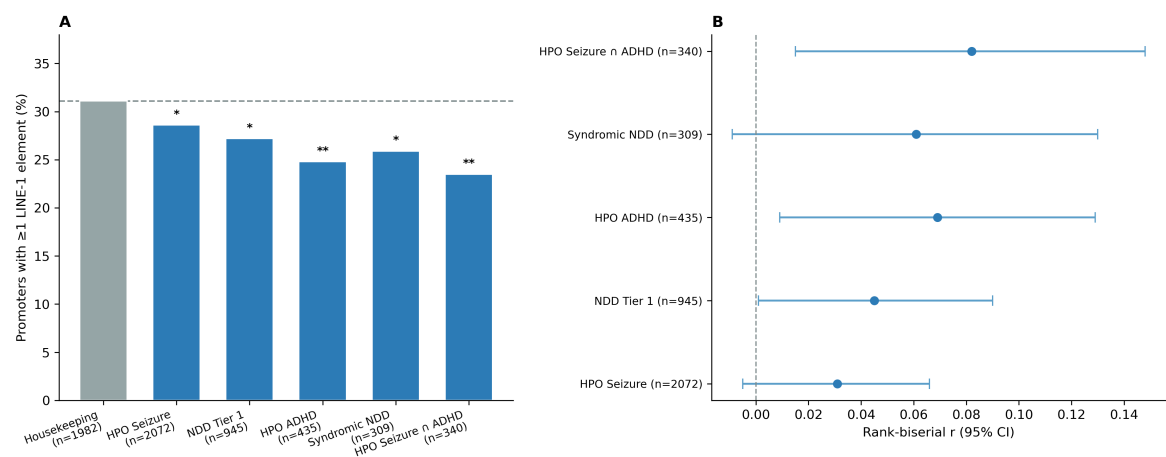
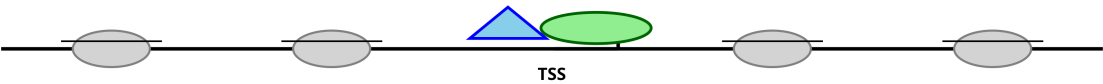


Figure 1. LINE-1 retrotransposon occupancy in promoter regions across neurodevelopmental disorder gene sets. (A) Proportion of promoters (± 2 kb from TSS) containing at least one LINE-1 element across five NDD gene sets and a housekeeping gene control. The dashed line indicates the housekeeping baseline (31.1%). Significance markers above bars denote Mann-Whitney U test results versus the housekeeping gene set (* $p < 0.05$, ** $p < 0.01$, n.s. not significant). (B) Effect sizes (rank-biserial r) with approximate 95% confidence intervals for each NDD gene set compared to the housekeeping gene set. All point estimates are positive and their confidence intervals do not cross zero for statistically significant comparisons, indicating consistent LINE-1 depletion relative to housekeeping genes across all NDD groups. Gene set sample sizes are indicated in parentheses on the x-axis (A) and y-axis labels (B). NDD Tier 1: SFARI Gene database, scores 1–2. Syndromic NDD: SFARI genes with syndromic annotation. HPO Seizure: HP:0001250. HPO ADHD: HP:0007018. HPO Seizure $\cap$ ADHD: genes annotated to both HPO terms.

Panel A: Active Transcription (Housekeeping Gene)



Panel B: Transcriptional Repression (NDD Gene)

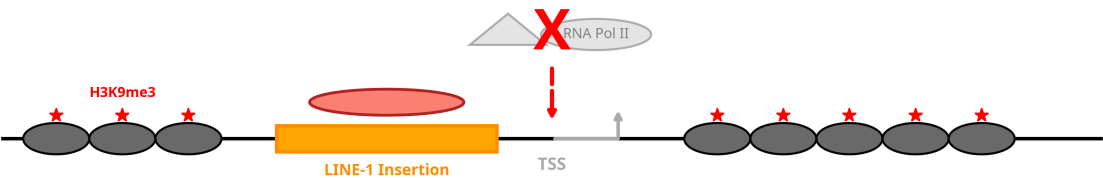


Figure 2. Conceptual model of LINE-1 mediated epigenetic silencing at NDD gene promoters. (Top) House-keeping gene promoters, which are depleted of transposable elements, maintain an open chromatin architecture (euchromatin) that readily permits the binding of transcription factors and RNA Polymerase II at the transcription start site (TSS). (Bottom) De novo insertion of a LINE-1 retrotransposon into the proximal promoter region of a highly constrained neurodevelopmental disorder (NDD) gene is recognized by host silencing machinery, recruiting TRIM28/KAP1. This initiates the local deposition of repressive H3K9me3 histone marks and subsequent chromatin condensation (heterochromatin), physically occluding the TSS and leading to transcriptional repression. Note: This model is proposed based on established epigenetic silencing mechanisms described in existing literature [7,25] and represents a theoretical framework for future experimental validation.

Table 1. Summary of LINE-1 promoter occupancy analysis across gene sets. All comparisons were performed against the housekeeping gene set ($n=1,982$) using the two-sided Mann-Whitney U test on the full distribution of LINE-1 counts per promoter. The LINE-1 (%) column represents the proportion of promoters containing at least one element for descriptive purposes. Effect size is reported as the rank-biserial correlation coefficient (r); positive values indicate lower LINE-1 counts in the NDD gene set relative to housekeeping genes. Approximate 95% confidence intervals for r are provided in parentheses. n.s. = not significant ($p \geq 0.05$).

Gene Set	<i>n</i>	LINE-1 (%)	<i>p</i> -value	<i>r</i> (95% CI)
Housekeeping	1,982	31.1	—	—
HPO Seizure	2,072	28.6	0.0360	0.031 (−0.00, 0.07)
NDD Tier 1 (SFARI score 1–2)	945	27.2	0.0138	0.045 (0.00, 0.09)
HPO ADHD	435	24.8	0.0051	0.069 (0.01, 0.13)
Syndromic NDD	309	25.9	0.0344	0.061 (−0.01, 0.13)
HPO Seizure ∩ ADHD	340	23.5	0.0029	0.082 (0.02, 0.15)

4. Conclusions

This analysis shows that the promoter regions of neurodevelopmental disorder-associated genes are consistently depleted of LINE-1 retrotransposons relative to housekeeping gene promoters. The depletion is statistically significant across multiple independently curated NDD gene sets and follows

a gradient that intensifies with phenotypic specificity, with the strongest signal observed in genes annotated to both seizure and ADHD phenotypes in the Human Phenotype Ontology. Intronic LINE-1 differences between NDD and housekeeping genes were heavily confounded by gene length, though a residual depletion was still detectable after length-matched control analysis.

The robust depletion at fixed-length promoter windows, alongside the residual intronic depletion after length correction, is consistent with broad functional constraint acting against LINE-1 insertions across NDD gene loci. What this analysis shows is a consistent pattern: across five independently curated gene sets, NDD promoters carry fewer LINE-1 elements than housekeeping promoters, and the magnitude of depletion tracks with phenotypic specificity. Future work combining this approach with population-level insertion polymorphism data (e.g., leveraging structural variant cohorts like gnomAD SV/MEI) could explicitly test whether this depletion represents an ongoing evolutionary signal or a neutral consequence of local sequence composition.

Author Contributions: Can Sevilmiş: Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable. This study was conducted exclusively using publicly available, anonymized genomic and phenotypic databases. No human or animal subjects were involved in this research.

Informed Consent Statement: Not applicable.

Data Availability Statement: All underlying datasets used in this study are publicly available. Gene set annotations were obtained from the SFARI Gene database (March 2026 release; <https://gene.sfari.org>) and the Human Phenotype Ontology database (<https://hpo.jax.org>). Genomic coordinates were sourced from GENCODE v47 (<https://www.genecodegenes.org>) and the UCSC RepeatMasker track for hg38 (<https://hgdownload.soe.ucsc.edu>). The housekeeping gene reference set was obtained from the HRT Atlas v1.0 (<https://housekeeping.unicamp.br>). All Python and Bash scripts used for data processing, BEDTools intersection, and statistical analysis are freely available in a dedicated GitHub repository at https://github.com/Bilmem2/LINE1_NDD_Promoters.

Use of Artificial Intelligence: During the preparation of this manuscript, the author used Claude (Anthropic) to assist with writing and troubleshooting bioinformatic scripts, as well as for grammatical and stylistic refinement of the text. The author thoroughly reviewed and edited the content following the use of these tools and takes full responsibility for the final content of the publication.

Acknowledgments: The author thanks the developers of SFARI Gene, Human Phenotype Ontology, GENCODE, UCSC Genome Browser, HRT Atlas, and BEDTools for providing open-access resources and software essential to this work.

Conflicts of Interest: The author declares no competing interests.

References

1. Lander, E.S.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921.
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
3. Ardeljan, D.; Taylor, M.S.; Ting, D.T.; Burns, K.H. The human LINE-1 retrotransposon: an emerging biomarker of neoplasia. *Clinical Chemistry* **2017**, *63*, 816–822.
4. Beck, C.R.; Collier, P.; Macfarlane, C.; Malig, M.; Kidd, J.M.; Eichler, E.E.; Badge, R.M.; Moran, J.V. LINE-1 retrotransposition activity in human genomes. *Cell* **2010**, *141*, 1159–1170.
5. Lavie, L.; Maldener, E.; Brouha, B.; Meese, E.U.; Mayer, J. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Research* **2004**, *14*, 2253–2260.
6. Honda, T.; Nishikawa, Y.; Nishimura, K.; Teng, D.; Takemoto, K.; Ueda, K. Effects of activation of the LINE-1 antisense promoter on the growth of cultured cells. *Scientific Reports* **2020**, *10*, 22136.

7. Jönsson, M.E.; et al. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nature Communications* **2019**, *10*, 3182.
8. Muotri, A.R.; Chu, V.T.; Marchetto, M.C.; Deng, W.; Moran, J.V.; Gage, F.H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **2005**, *435*, 903–910.
9. Coufal, N.G.; et al. L1 retrotransposition in human neural progenitor cells. *Nature* **2009**, *460*, 1127–1131.
10. Singer, T.; McConnell, M.J.; Marchetto, M.C.; Coufal, N.G.; Gage, F.H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences* **2010**, *33*, 345–354.
11. Abrahams, B.S.; et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders. *Molecular Autism* **2013**, *4*, 36.
12. Köhler, S.; et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Research* **2021**, *49*, D1207–D1217.
13. Fan, H.C.; Chiang, K.L.; Chang, K.H.; Chen, C.M.; Tsai, J.D. Epilepsy and attention deficit hyperactivity disorder: connection, chance, and challenges. *International Journal of Molecular Sciences* **2023**, *24*, 5270.
14. Hounkpe, B.W.; Chenou, F.; de Lima, F.; De Paula, E.V. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Research* **2021**, *49*, D947–D955.
15. Frankish, A.; et al. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Research* **2023**, *51*, D942–D949.
16. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0, 2013. <http://www.repeatmasker.org>.
17. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842.
18. Simonti, C.N.; Pavličev, M.; Capra, J.A. Transposable Element Exaptation into Regulatory Regions Is Rare, Influenced by Evolutionary Age, and Subject to Pleiotropic Constraints. *Molecular Biology and Evolution* **2017**, *34*, 2856–2869.
19. You, J.S.; Pierce, S.; Liang, G.; Jones, P.A. Roles of transposable elements and DNA methylation in the formation of CpG islands and CpG-depleted regulatory elements. *Proceedings of the National Academy of Sciences* **2025**, *122*, e2502963122.
20. Zug, R. Developmental disorders caused by haploinsufficiency of transcriptional regulators: a perspective based on cell fate determination. *Biology Open* **2022**, *11*, bio058896.
21. Darbandi, S.F.; An, J.Y.; Lim, K.; Page, N.F.; Liang, L.; Young, D.M.; Ypsilanti, A.R.; State, M.W.; Nord, A.S.; Sanders, S.J.; et al. Five autism-associated transcriptional regulators target shared loci proximal to brain-expressed genes. *Cell Reports* **2024**, *43*, 114329.
22. Koesterich, J.; An, J.Y.; Inoue, F.; Sohota, A.; Ahituv, N.; Sanders, S.J.; Kreimer, A. Characterization of De Novo Promoter Variants in Autism Spectrum Disorder with Massively Parallel Reporter Assays. *International Journal of Molecular Sciences* **2023**, *24*, 3509.
23. Liang, K.C.; Tseng, J.T.; Tsai, S.J.; Sun, H.S. Characterization and distribution of repetitive elements in association with genes in the human genome. *Computational Biology and Chemistry* **2015**, *57*, 29–38.
24. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* **2016**, *18*, 71–86.
25. Lu, J.Y.; Shao, W.; Chang, L.; Yin, Y.; Li, T.; Zhang, H.; Hong, Y.; Percharde, M.; Guo, L.; Wu, Z.; et al. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports* **2020**, *30*, 3296–3311.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.