*Article*

# Patterns of Transposable Element Distribution Around Chromatin Ligation Points Revealed by Micro-C Data Analysis


[1] DNA Resonance Research Foundation, San Diego, CA, USA,
[2] Institute of Pharmacy and Biotechnology of Lumumba Peoples' Friendship University of Russia, Moscow, Russia
[3] Institute for Bionic Technologies and Engineering, Sechenov First Moscow State Medical University, Moscow, Russia



**Abstract:** Background: Transposable elements constitute a significant portion of eukaryotic genomes, yet their role in chromatin organization remains poorly understood.

Methods: This study computationally investigates the density patterns of transposons around chromatin contact points identified from public Micro-C chromatin conformation data from human cell culture. The density peak patterns of various transposable families and subfamilies were studied within a 100kb window centered on contact points. The analysis was focused on the most abundant transposons, such as Alu and LINE-1.

Results: The computational analysis revealed highly pronounced, non-random density patterns of transposons around the chromatin contact points. The patterns were produced by aligning all ligation points and plotting the average density around them. The patterns were strikingly different between transposable element families and substantially different between the members of the families. The patterns were found to be reproducible across independent studies and biological replicates. Among major families and subfamilies there were no members that didn't have reproducible density patterns around the contact points. Randomly generated coordinates produced less pronounced patterns, which were not correlated between replicates as expected for the negative control. Some families showed enrichment and some - depletion at contact points, while 100Kb window-wide patterns remained correlated between biological replicates. The patterns were asymmetric relative to the chromosomal orientation. Additionally, the patterns were oriented relative to the transposon sequence direction.



**Keywords: transposable elements, chromatin conformation, chromatin folding, chromatin, Micro-C**


## 1. Introduction

Transposable elements (TEs) constitute a significant portion of eukaryotic genomes, often comprising more than half of the genomic content in many species. Once considered "junk DNA", these mobile genetic elements have been increasingly implicated in various aspects of genome function and evolution [1, 2]. Previously, we have emphasized that transposable elements can serve an important positive biological function as anchors and regulators of chromatin folding [3–10]. Recent advancements in chromosome conformation capture techniques, particularly Micro-C and Hi-C, have enabled high-resolution mapping of chromatin interactions [11], revealing the three-dimensional organization of the genome.

The rationale behind our investigation stems from the hypothesis that repetitive elements, due to their sequence homology, might provide sequence-specific anchors for chromatin folding. This aligns with emerging views of spatial-temporal genome regula-

tion, where chromatin organization plays a key role in cellular responses [12]. This hypothesis is based on the idea that homologous sequence structures may form homologous contacts in the nucleoplasm. In other words, two identical double helices would adhere to each other. Since TEs are highly repetitive, we hypothesized that they form adhesive contact points in chromatin folding. To test this, we examined the distribution of TEs around the chromatin contact points identified through micro-C experiments. Since the contact points are measured via ligation and sequencing, the exact positions of ligation points (LPs) can be reconstructed from the Micro-C data. Then, we mapped the density patterns of various families and subfamilies of TEs.

Our study focuses on two of the most abundant TE families in the human genome: Alu elements, short interspersed nuclear elements (SINEs) approximately 300 base pairs in length, and LINE-1 (L1) elements, long interspersed nuclear elements that can span several kilobases. Additional families were also studied.

## 2. Methods

### *2.1. Public chromatin conformation capture datasets*

We analyzed chromatin interaction data from four independent datasets to identify ligation points (LPs). Dataset 1 (DS1) and Dataset 2 (DS2) were produced by Micro-C: SRR12625672 and SRR12625674 (biological replicates from HUDEP cell line, ~43M paired-end reads each, 150nt). https://www.ncbi.nlm.nih.gov/sra/SRR12625672. The Micro-C protocol employs micrococcal nuclease digestion, achieving nucleosome-level (~200bp) resolution.

Datasets 3 and 4 (DS3 and DS4) were produced by Hi-C: SRR27906244 and SRR27906243 from the GM13977 cell line (~75M paired-end reads each). Hi-C libraries were constructed using a standard protocol with restriction enzyme digestion (MboI, MseI, or NlaIII), which typically produces 2-6 kilobase fragments, followed by biotin incorporation and proximity ligation. The Hi-C data were obtained from NCBI Bioproject PRJNA1074296 (https://www.ncbi.nlm.nih.gov/bioproject/1074296) and processed using JuicerTools v1.14.08. Despite the methodological differences between Micro-C and Hi-C, all datasets were analyzed using identical parameters for TE density patterns and correlations.

### *2.2. Identification of ligation points.*

Although contact points in the chromatin conformation capture are called points, they are imprecisely defined and are actually represented by short fragments. Yet, it is possible to define precisely to a single nucleotide which sequences were ligated in the assay. The first step in the analysis was to identify the exact positions of the ligation points on the genome. The method identifies ligation points with single-nucleotide precision by using the alignment of paired-end reads to the genome. In Micro-C, most paired reads overlap, forming a continuous sequence. This sequence corresponds to the real physical fragment (called here ligated_fragment) produced by ligation in chromatin conformation capture. For each paired read, we reconstruct the ligated_fragment sequence by aligning paired reads. Next, we align the ligated_fragment against the genome. Only those paired reads are retained where two parts of the ligated_fragments (called here arms) align to the same chromosome and are positioned at least 1Kb apart. Once the arms are aligned, the exact position of coordinates of the chromosome that were ligated to each other in the assay are identified (called here, LP1 and LP2). The pipeline for the identification of the ligation points was called LigP_finder_v2. The Python code for LigP_finder_v2 can be downloaded from https://github.com/maxrempel/DRRF/tree/main/LigP_finder-main

All datasets were processed using whole-genome alignment in LigP_finder_v2.

When applied to Hi-C datasets, the program produced fewer ligation points since the reads in Hi-C don't overlap, but the number of found ligation points was sufficient to produce TE density pattern plots and correlations. In general, the amount of data from the experiments was in excess since the plots didn't visually deteriorate even when a single small chromosome was used for pattern plotting. The pattern was consistent across chromosomes, although the density of ligation points varied widely, about 10-fold between chromosomes, reflecting variable compaction of chromosomes in given cell lines.

To test whether the program was correct, we ran tests using synthetic data and did a spot-check of the results. Also, the final program was described in English algorithm, and reprogrammed from scratch by another remote programmer. The recreated program produced identical results, demonstrating that both programs are correct.

The Ligation points, LPs, were obtained with hg38, unmasked version: https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz

We used the hg38 transposable element (TE) annotation from UCSC: https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.out.gz

### *2.3. Plotting density patterns*

For each identified LP, we examined the genomic regions extending 50kb in both directions, which we termed "exbors" (from EXtended harBORS). Within these regions, we cataloged the density of transposable elements using bedtools, and function intersect. In that, the rigth and left-oriented elements were analysed separately. We focused on the most abundant TE families and subfamilies based on their genomic copy numbers.

To quantify TE density patterns, we calculated TE density in 1kb bins across the 100kb window centered on each LP. Importantly, we conducted separate analyses for elements on the plus and minus strands to investigate potential strand-specific (same as orientation-specific) patterns. To visualize these density patterns, we smoothed the density curves with the Kernel Density Estimation (KDE) smoothing method.

For negative controls, we performed random control analyses. This involved generating random LP positions across the genome in excess and selecting those random coordinates that were located in non-repeat-masked (unique) regions across the whole genome. Then, real and random coordinates of LPs were compared using density and correlation plots. To control for programming errors, three programmers (DO, AV and MM) reproduced the prlots independently, and the plots came out identical. Correlations between density pattern plots were calculated using the Pearson correlation and visualized as diagonal heatmaps. The Python code for LigP_finder_v2 also contains TE density pattern plotting with KDE smoothing.
https://github.com/maxrempel/DRRF/tree/main/LigP_finder-main

Some plots are shown in the Results section, and many more plots are shown in the Supplementary Materials pdf file.

### *2.4. Analysis of harbor homology*

To measure sequence similarity between harbor pairs, we extracted ±5 Kb sequences around each ligation point (10kb total harbor size). Sequence alignment was performed using minimap2 with the 'sr' preset and minimum 70% identity threshold. To compare within-pair versus between-pair homology, we analyzed sequence similarity between

harbors from the same ligation point pair (REAL) versus harbors from different ligation point pairs (CONTROL). Both unique/low-copy sequences (regions remaining after repeat masking) and total sequence (unmasked) were analyzed. Pairs with overlapping harbors (distance < 10kb) were excluded from the analysis. The Python code for homology analysis is available at https://github.com/maxrempel/DRRF/tree/main/LigP_finder-main.

## 3. Results

For each tested transposable element (TE) family, we produced TE density patterns in 100 Kb windows (called here "harbors") around chromatin ligation points (called here LPs) identified in Micro-C and Hi-C chromatin conformation assays. We examined the density of six major transposable element (TE) families: Alu, L1 (LINE-1), L2 (LINE-2), Medium Reiterated Frequency Repeat (MER), Mammalian-wide interspersed repeats (MIR), and Mammalian Long Terminal repeat (MLT) elements, as well as most frequent subfamilies the most frequent TEs: L1 and Alu. Figure 1 presents the distribution of L1 elements around LPs (ligation points).

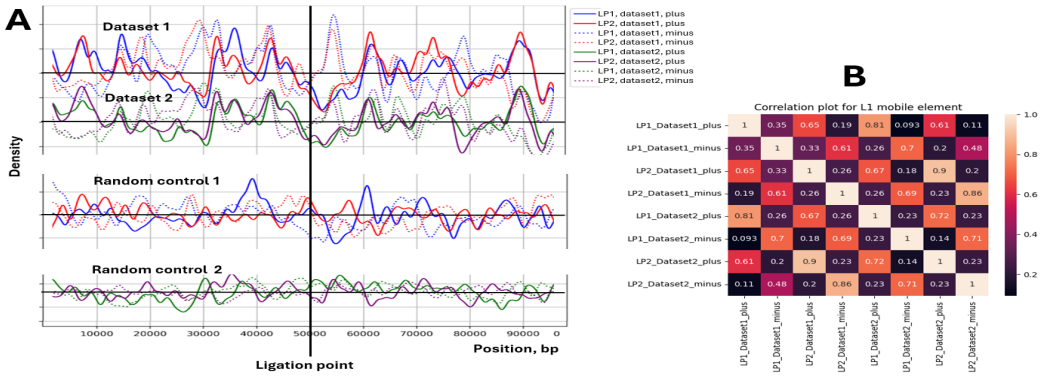



**Figure 1:** Distribution of L1 transposable elements around chromatin ligation point. Panel A. The density of L1 elements (a major subfamily of Long Interspersed Nuclear Elements or LINEs) relative to chromatin ligation points (LPs) in two experimental Micro-C datasets and their corresponding random controls. The x-axis represents the position relative to the LP (vertical black line at 50,000 bp), spanning 100 kb. The y-axis shows the density of L1 elements (count of L1 elements divided by the bin size, bp). Solid lines represent L1 elements in the plus strand (therefore oriented left to right), and dotted lines represent the opposite-oriented L1 elements labeled as the minus strand. Dataset 1 (blue/red) and Dataset 2 (green/purple) show experimental data for LP1 and LP2 regions. LP1 and LP2 ligation points were numbered from left (start) to right (end) of the chromosome. Random controls 1 and 2 correspond to Datasets 1 and 2, respectively, generated using randomized LP positions. The graphs are oriented from left to right on the chromosome. Panel B shows a correlation heatmap comparing L1 element density patterns between different datasets and strands. The color intensity represents the Pearson correlation coefficient, with lighter colors indicating stronger correlations. The heatmap reveals strong correlations between biological replicates (Dataset1 and Dataset2) for the same strand orientation, while plus and minus strands show notably weaker correlations, quantitatively confirming the strand bias observed in panel A.

Figure 1 presents a view of L1 transposable element distribution around chromatin ligation points (LPs) identified through micro-C experiments. L1, or LINE-1, is a major subfamily of Long Interspersed Nuclear Elements (LINEs) comprising about 20% of the human genome. The data reveal patterns that suggest a non-random association between L1 elements and chromatin structure.

Key observations from the figure include:

1. Pattern reproducibility: Panel A shows a similarity in the distribution patterns between Dataset 1 and Dataset 2. This consistency across independent datasets strengthens the biological significance of the observed patterns, which is quantitatively confirmed in Panel B by high correlation coefficients (>0.8) between datasets for corresponding strands.

2. Strand-specific patterns: The plus and minus strands show distinct patterns (Panel A), which is quantitatively supported by the low correlation coefficients (0.2-0.4) between the plus and minus strands in Panel B.
3. Depletion at the ligation point: There is a noticeable dip in L1 density directly at the ligation point (50,000 bp). This depletion indicates the exclusion of L1 elements from immediate chromatin contact points.
4. Random controls: The random control datasets in Panel A show markedly different patterns characterized by less pronounced fluctuations and lack of correlations between datasets and within pairs of ligation points. This contrast underscores the biological significance of the patterns observed in the experimental data.

These findings suggest a sequence-specific role of L1 elements in chromatin folding, with patterns indicating their selective retention in specific chromatin environments. Future studies should focus on understanding the mechanisms driving these distribution patterns.

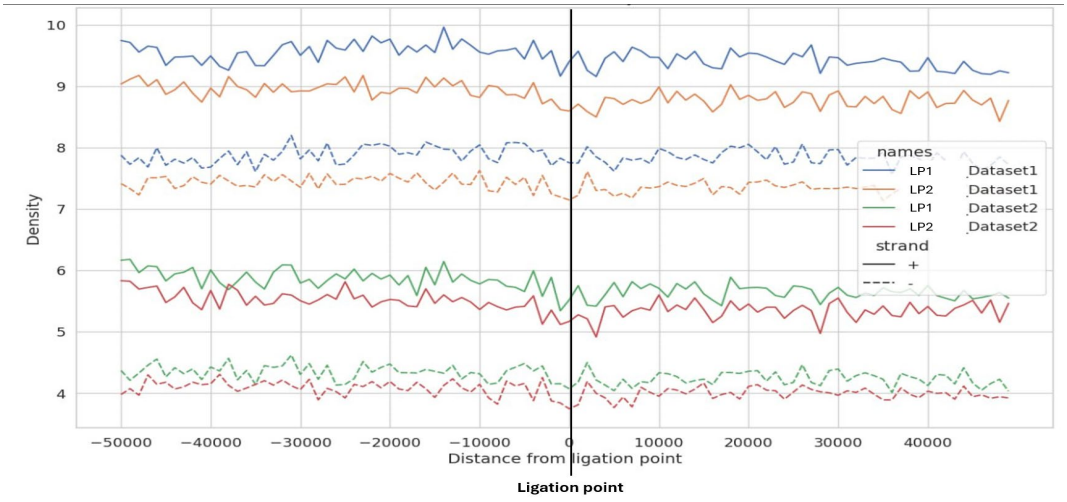



**Figure 2:** Distribution of Alu transposable elements around chromatin ligation points. The density of Alu elements (a subfamily of Short Interspersed Nuclear Elements or SINEs) relative to chromatin ligation points (LPs) in two experimental datasets. The x-axis represents the distance from the LP (at 0), spanning 100 kb (-50,000 to +50,000 bp). The y-axis shows the density of Alu elements. Solid lines represent the plus strand, and dotted lines represent the minus strand. Dataset 1 (blue/orange) and Dataset 2 (green/red) show data for LP1 and LP2 regions. The y-axis shifts were added artificially to prevent overlap of the curves.

Figure 2 presents the distribution of Alu elements around chromatin ligation points (LPs) identified through Micro-C experiments. Alu elements, the most abundant member of Short Interspersed Nuclear Elements (SINEs), comprise approximately 11% of the human genome.

Key observations from Figure 2 reveal consistent patterns in Alu element distribution around chromatin ligation points. The patterns are consistent across datasets and ligation points. As for other TEs, a strand difference was observed, with the plus strand (represented by solid lines) displaying different density patterns compared to the minus strand (dotted lines). This asymmetry is consistent in all tested datasets and was observed in every other tested TE family.

### *3.1. Additional Transposable Elements*

Our analysis of transposable element (TE) distribution around chromatin ligation points (LPs) revealed that each TE family and subfamily has a distinct density pattern around ligation points. We examined the density of six major TE families: Alu, L1, L2, MER, MIR, and MLT elements, as well as several subfamilies within these groups, within

a 100 kb window centered on LPs identified through Micro-C experiments (Fig. 1 for L1, Fig. 2 for Alu; see Supplementary Figures for all tested TE families and subfamilies).

The distribution patterns observed in every examined TE family had very similar trends to the trends observed for L1 elements (Fig. 1), although the density patterns around the ligation points were unique for each tested TE family and subfamily.

Also, the patterns for all the TE families and subfamilies were asymmetric relative to the ligation points, oriented with respect to chromosome direction (with the short arm (p) positioned towards the left, following standard genomic convention). This asymmetry indicates that transposon directionality relative to chromosome orientation participates in chromatin folding. The consistency of these patterns suggests that the evolutionary selection of transposon placement produced long-range chromosome-wide orientations of transposons. Such large-scale genomic patterns may represent fundamental principles of chromosome organization that persist through evolution.

Since every tested TE family and subfamily produced consistent density patterns around the ligation points, this suggests the fundamental role of transposable elements in sequence-specific chromatin folding that transcends the specific characteristics of individual TE types. The observed strand biases and dataset-specific variations appear to be general features of how TEs are distributed relative to chromatin contact sites.

To exclude potential methodological artifacts, we utilized Hi-C data (Datasets 3 and 4, see Methods) to validate our findings. While both Hi-C and Micro-C capture chromatin conformation, they differ fundamentally in their digestion methods: Hi-C uses restriction enzymes producing fragments of several kilobases, while Micro-C employs micrococcal nuclease digestion, achieving nucleosome-level (~200bp) resolution. Despite this resolution difference, Hi-C data reproduced the key patterns of transposable element distribution around contact points, with correlation coefficients between biological replicates reaching 0.7-0.8 for same-strand comparisons and remaining below 0.2 for opposite strands (Supplementary Figures). This cross-method validation confirms the strand-specific organization of transposable elements around chromatin contact points. While absolute correlation values between datasets from different laboratories and cell lines were not expected due to the tissue-specific nature of chromatin organization, the fundamental patterns were consistently reproduced: strong correlations between biological replicates and between members of ligation point pairs and strong strand asymmetry. These patterns were reproduced despite using cell lines from different blood lineages (HUDEP erythroid progenitors (datasets 1 and 2) vs. GM13977 lymphoblastoid cells (datasets 3 and 4)), suggesting that the results reflect a common principle in chromatin organization.

### *3.2. Asymmetric Distribution Suggests Homological Adhesion*

We noticed that the asymmetry of transposable element density patterns relative to chromosome orientation suggests tandem-like organization. However, we found no periodic repetition of these patterns, suggesting an aperiodic repetition of patterns that maintains directional consistency. Coincidentally, this aligns with Erwin Schrödinger's description of the hereditary material as aperiodic crystal [13].

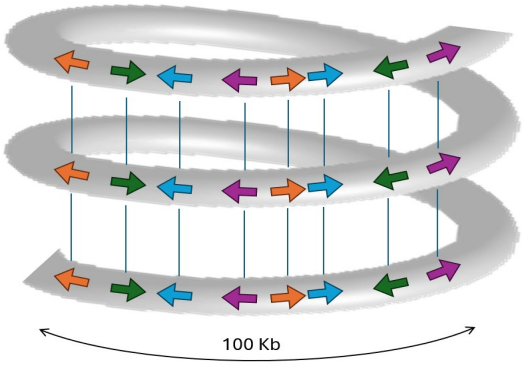


**Figure 3:** Model of homological adhesion in chromatin folding. The figure shows how patterns of identical transposon sequences (arbitrarily placed colored arrows) can serve as contact points through sequence-specific homological adhesion. The 100 Kb window demonstrates the alignment of identical transposons creating interactions between distant chromatin regions.

We propose that this aperiodic but directional repetition of TE patterns functions in chromatin architecture through homological adhesion, where identical sequences, particularly transposable elements, can form contact points through sequence-specific adhesion of identical sequences. Here, we propose for the first time that two parallel double helices of DNA can adhere to each other when their sequences are identical, forming the molecular basis for homological adhesion. The density plots reveal that the patterns are asymmetric and face in one direction on chromosomes. Since they face in one direction on chromosomes, they could serve as specific anchors for homological adhesion and the formation of large-scale helices.

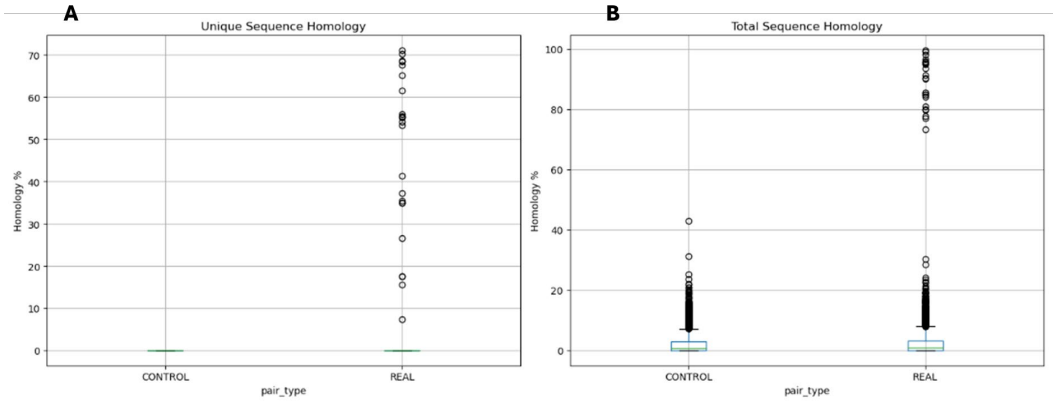



**Figure 4:** Distribution of sequence homology within and between ligation point pairs. Panel A shows homology in unique and low-copy sequences left after masking. Panel B shows total sequence homology from the unmasked sequence. For REAL comparisons, we measured homology between two harbors from the same ligation point pair (within the LP pair). For CONTROL comparisons, we measured homology between harbors from different ligation point pairs (between LP pairs).

We next tested if the sequences that come together in chromatin display homology. As a negative control, we measured homology between unpaired unligated harbors from different harbor pairs. Comparing harbors (ligated fragments) from the same harbor pair versus different pairs revealed a strong enrichment of homology in ligated regions. In unique and low-copy sequences, within-pair homology reached 70%, while between-pair homology remained near zero (mean 0.47% vs 0.02%). The signal persisted in an unmasked sequence, with within-pair homology reaching 100% compared to a maximum of 40% between pairs. This provides quantitative evidence for sequence-specific homological dsDNA-dsDNA adhesion in chromatin contacts.

## 4. Discussion

The patterns of TE distribution around chromatin ligation points observed here provide insight into the potential role of TEs in chromatin folding. We discovered consistent strand asymmetry across TE families, which indicates systematic evolutionary pressures governing transposon organization. While general principles of transposon insertion have been described [14], this chromosome-wide directional bias has not been previously reported.

The density patterns around ligation points were distinct for each TE family. Alu elements showed proportionally scaled patterns with multiple peaks, reflecting their higher copy number, while L1 elements displayed fewer but more pronounced peaks.

### *4.1. Homological adhesion*

The discovery of the density patterns of transposable elements around the chromatin contact sites and of the homology between contacting DNA sequences offers support for homological adhesion. We will nickname it "homadhesion" for brevity. Homadhesion is the adhesion of two DNA duplexes (double helixes) to each other if they have similar sequences.

Several studies provide support for sequence-dependent attraction between DNA duplexes. Yoo et al. [15] demonstrated sequence-dependent attraction between double-stranded DNA molecules using molecular dynamics simulations and single-molecule FRET experiments. They showed that DNA duplexes can attract each other over distances up to 2-3 nm in the presence of polyamines like spermine. Importantly, AT-rich sequences showed stronger attraction than GC-rich sequences, and DNA methylation enhanced these interactions.

Lee et al. [16] provided theoretical support for sequence-dependent attraction between intact DNA duplexes. Their model incorporated electrostatic forces and sequence-dependent DNA shape variations, predicting lower interaction energies for pairs of DNA fragments with parallel homologous sequences compared to those with uncorrelated sequences. This interaction was modeled without strand separation, considering the aqueous environment through parameters that account for electrostatic screening in solution.

Barzel and Kupiec reviewed the evidence for the pairing of similar DNA sequences across different organisms [17]. They noted that in yeast, matching DNA sequences can locate each other and recombine efficiently even when in different genomic locations despite the large amount of genomic DNA present. Their review suggested that similar sequences are paired as part of the genome's basic organization.

For the adhesion mechanism itself, several forces could be considered: electrostatic (including ionic), hydrophilic-hydrophobic, hydrogen bonding, van der Waals, and other weak forces. The dynamic nature of chromatin condensation and decondensation suggests that DNA duplexes remain nearly intact during these interactions. While hydrated DNA structure may be primary in mediating adhesion, other nucleoplasmic components likely participate, including histones, other proteins, and low molecular weight chemicals. The highly negatively charged DNA duplexes must overcome electrostatic repulsion to adhere. This requires neutralization by positive ions present in the nucleoplasm, including protons ($H^+$), hydronium ions ($H3O^+$), positively charged histones and other proteins, and ions such as $Na^+$, $K^+$, $Mg^{2+}$, and polyamines like spermine and spermidine.

## 5. Conclusions

This discovery of consistent density patterns of transposable elements around chromatin contact points suggests a functional involvement of TEs in sequence-specific chromatin folding and homological adhesion. The observed asymmetry of patterns relative to chromosome orientation and transposon orientation suggests the existence of chromosome-scale sequence organization that was not previously known.

This opens new opportunities for research into the role of transposable elements in sequence-specific chromatin folding and, through that, in genome regulation. The organization of chromatin through networks of interacting domains has been previously described [18], though not through the lens of transposon-mediated interactions we report here. Future work should focus on elucidating the mechanisms underlying the observed density patterns and investigating their potential implications for genome regulation and function. Exploring these patterns across different cell types, developmental stages, and organisms could provide valuable insights into the evolutionary conservation and functional significance of TEs in chromatin folding [19].

The sequence-specific principles of genome organization are known to operate across multiple scales, from local loop formation to chromosome territories. Dixon et al.[20] demonstrated how chromatin organization changes systematically during cellular differentiation. The importance of repetitive elements in nuclear organization was highlighted by Cournac et al.[21], showing correlations between similar repetitive elements and 3D

folding patterns. Studies of chromosome territories by Cremer et al. [22] revealed principles of nuclear organization. Rowley et al. [23] established evolutionarily conserved principles of 3D chromatin organization. The emerging picture of genome architecture was developed by Bonev et al. [24], who mapped dynamic changes in genome organization during development. Our finding of chromosome-orientation-dependent patterns in TE densities around contact points suggests an additional layer of sequence-encoded structural information that may help explain these organizational principles.

As our understanding of genome organization continues to evolve, it is becoming increasingly clear that a comprehensive view of genomic function must include consideration of transposable elements. This work lays the foundation for future investigations of the role of transposable elements in chromatin folding.

**Acknowledgments.** We thank Jekaterina Erenpreisa for the discussion. The work and the DNA Resonance Research Foundation (DRRF) were funded by MM. The publication fee was paid with AV's reviewer's vouchers. AV acknowledges support of his work by the RUDN USAL Program.

**Author contributions:** AV, DO, and MM did genomic computation.

**Conflicts of Interest**: The authors declare no conflicts of interest.

**References**

1. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009;10:691–703.
2. Deininger P. Human Retrotransposable Elements: An End or Just the Beginning? Nature Reviews Genetics. 2021;22:749–71.
3. Polesskaya O, Guschin V, Kondratev N, Garanina I, Nazarenko O, Zyryanova N, et al. On possible role of DNA electrodynamics in chromatin regulation. Prog Biophys Mol Biol. 2017. https://doi.org/10.1016/j.pbiomolbio.2017.12.006.
4. Polesskaya O, Kananykhina E, Roy-Engel AM, Nazarenko O, Kulemzina I, Baranova A, et al. The role of Alu-derived RNAs in Alzheimer's and other neurodegenerative conditions. Med Hypotheses. 2018;115:29–34.
5. V. Savelyev I, V. Zyryanova N, O. Polesskaya O, Myakishev-Rempel M. On the existence of the DNA resonance code and its possible mechanistic connection to the neural code. Neuroquantology. 2019;17.
6. Savelev I, Myakishev-Rempel M. Possible traces of resonance signaling in the genome. Prog Biophys Mol Biol. 2020;151:23–31.
7. Savelev I, Myakishev-Rempel M. Evidence for DNA resonance signaling via longitudinal hydrogen bonds. Prog Biophys Mol Biol. 2020;156:14–9.
8. Myakishev-Rempel M, Savelev IV. How Schrödinger's Mice Weave Consciousness. In: Rhythmic Advantages in Big Data and Machine Learning. Singapore: Springer Nature Singapore; 2022. p. 201–24.
9. Savelev IV, Miller RA, Myakishev-Rempel M. How the Biofield is Created by DNA Resonance. In: Rhythmic Advantages in Big Data and Machine Learning. Singapore: Springer Nature Singapore; 2022. p. 161–99.
10. Vikhorev AV, Savelev IV, Polesskaya OO, Rempel MM, Miller RA, Vetcher AA, et al. The avoidance of purine stretches by cancer mutations. Int J Mol Sci. 2024;25:11050.
11. Krietenstein N, Abraham S, Venev SV. Recent advancements in chromosome conformation capture techniques, particularly Micro-C, have enabled high-resolution mapping of chromatin interactions" in Introduction. Molecular Cell. 2020:554–65.
12. Erenpreisa J, Giuliani A, Yoshikawa K, Falk M, Hildenbrand G, Salmina K, et al. Spatial-temporal genome regulation in stress-response and cell-fate change. Int J Mol Sci. 2023;24:2658.
13. Schrödinger E. What is Life? The Physical Aspect of the Living Cell. Cambridge: Cambridge University Press; 1944.
14. Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet. 2017;18:292–308.
15. Yoo J, Kim H, Aksimentiev A, Ha T. Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation. Nat Commun. 2016;7:11045.
16. Lee DJ, Wynveen A, Albrecht T, Kornyshev AA. Which way up? Recognition of homologous DNA segments in parallel and antiparallel alignment. arXiv [physics.bio-ph]. 2014.
17. Barzel A, Kupiec M. Finding a match: how do homologous sequences get together for recombination? Nat Rev Genet. 2008;9:27–37.
18. Erenpreisa J, Krigerts J, Salmina K, Gerashchenko BI, Freivalds T, Kurg R, et al. Heterochromatin networks: Topology, dynamics, and function (a working hypothesis). Cells. 2021;10:1582.
19. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19:199.

20. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518:331–6.
21. Cournac A, Koszul R, Mozziconacci J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. Nucleic Acids Res. 2016;44:245–55.
22. Cremer T, Cremer M, Hübner B, Strickfaden H, Smeets D, Popken J, et al. The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. FEBS Lett. 2015;589 20 Pt A:2931–43.
23. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily conserved principles predict 3D chromatin organization. Mol Cell. 2017;67:837–52.e7.
24. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D genome rewiring during mouse neural development. Cell. 2017;171:557–72.e24.

# Patterns of Transposable Element Distribution Around Chromatin Ligation Points Revealed by Micro-C Data Analysis

Alexandr V. Vikhorev[1], Michael M. Rempel[1], Oksana O. Polesskaya[1], Ivan V. Savelev[1], Alexandre Vetcher[2], Daniel Oliveira[1], and Max V. Myakishev-Rempel[1]

## Supplement

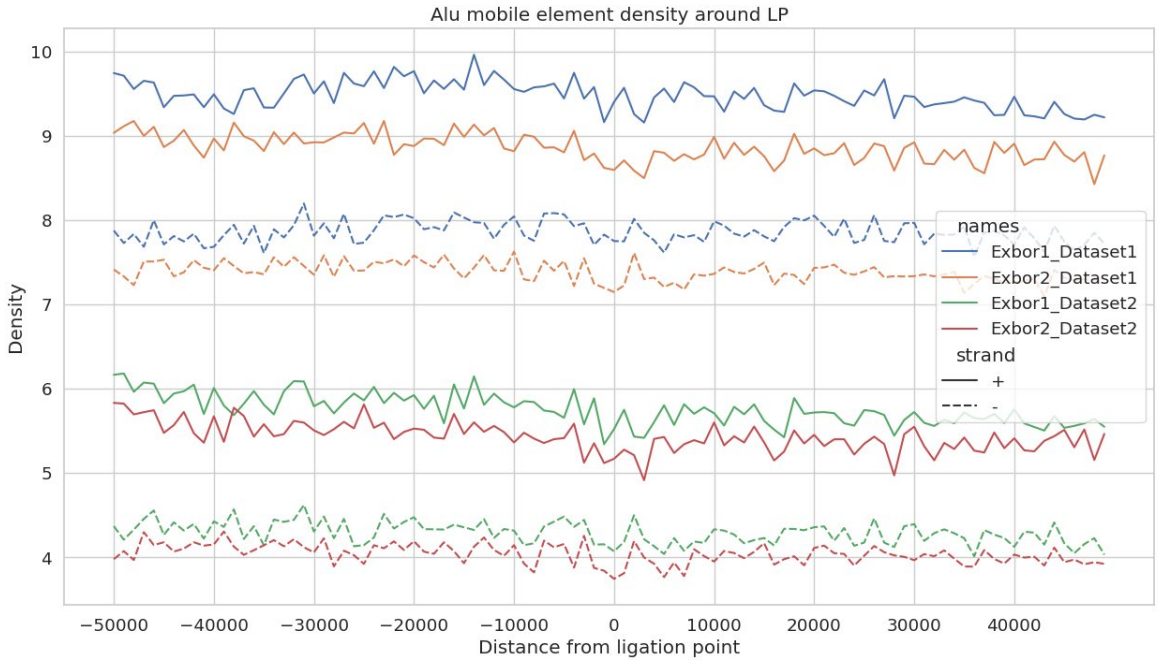



**Fig. S_ALU**

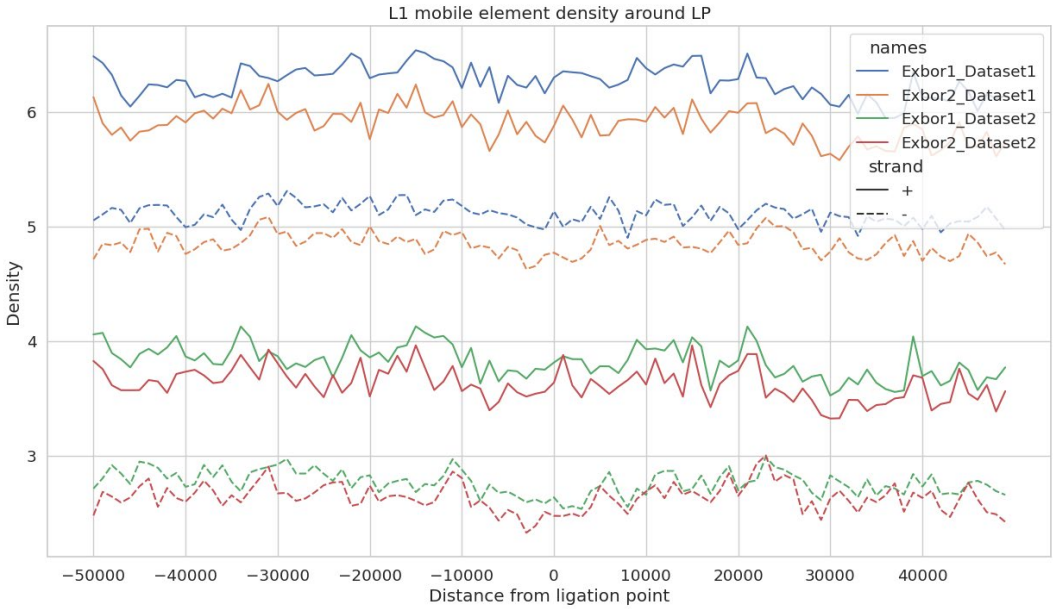



**Fig. S_L1**

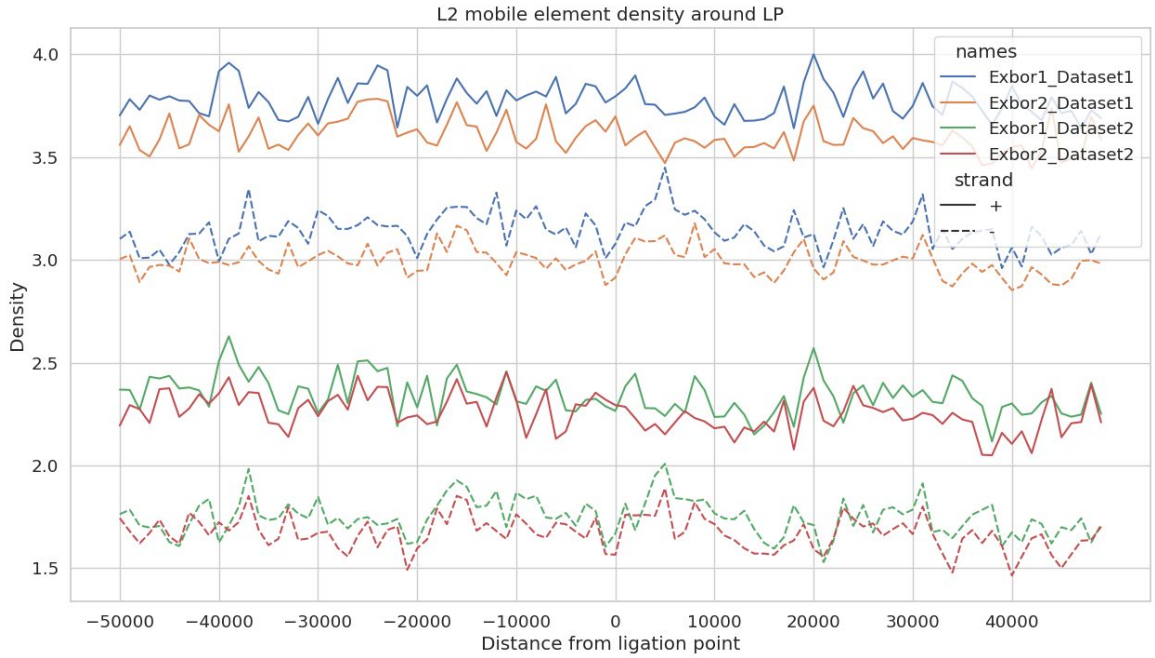



**Fig. S_L2**

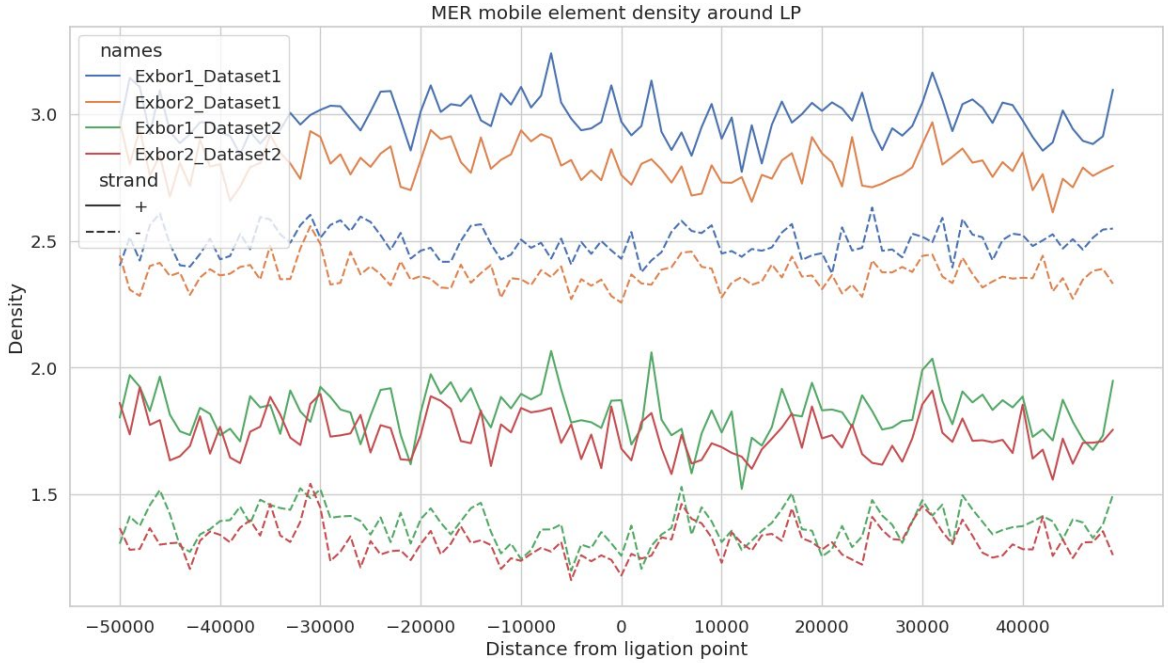

MER mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
strand
+
-
Density
3.0
2.5
2.0
1.5
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_MER**

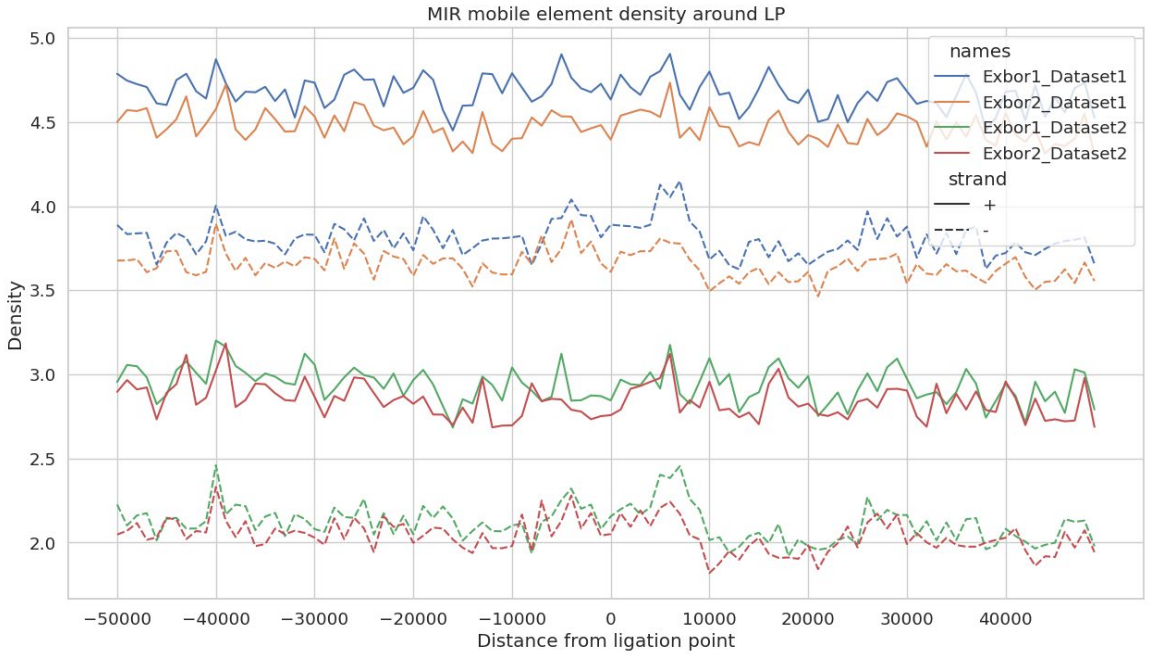

MIR mobile element density around LP
Density
Distance from ligation point
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
strand
+
-
5.0
4.5
4.0
3.5
3.0
2.5
2.0
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000

**Fig. S_MIR**

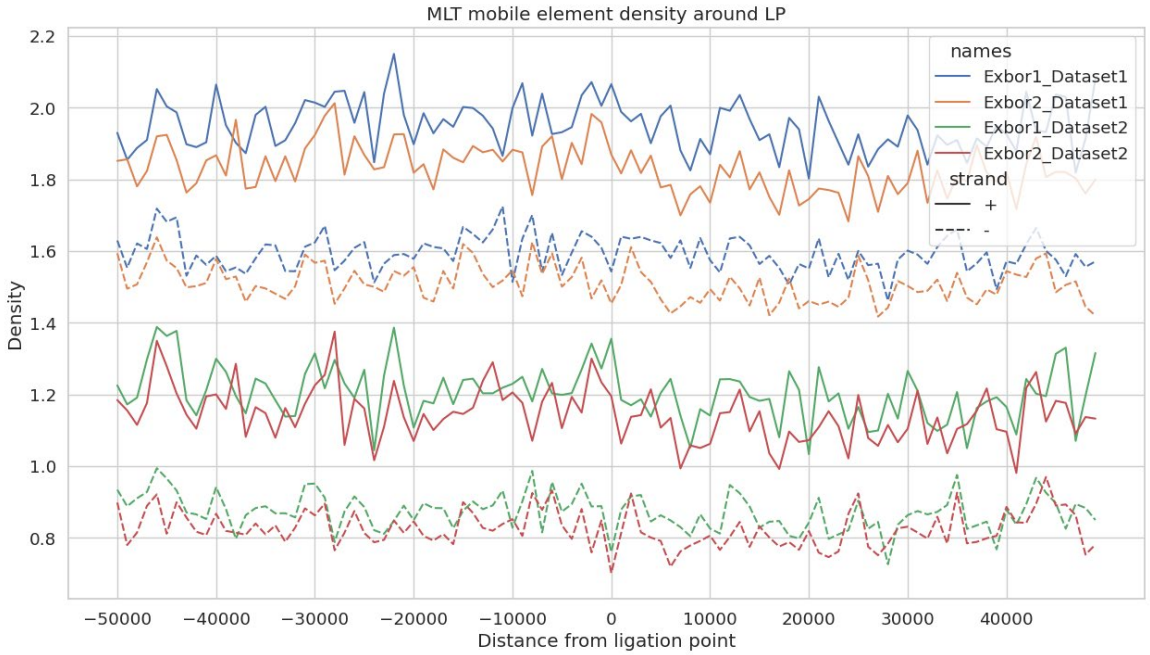

MLT mobile element density around LP
Density
Distance from ligation point
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
strand
+
-
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000

**Fig. S_MLT**

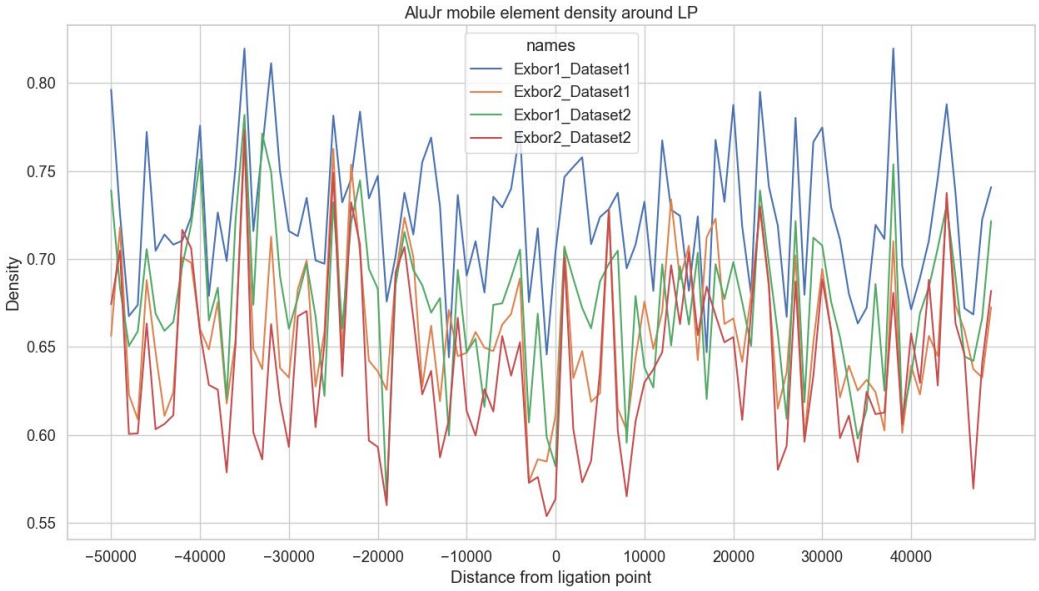

AluJr mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
0.80
0.75
0.70
0.65
0.60
0.55
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_AluJr**

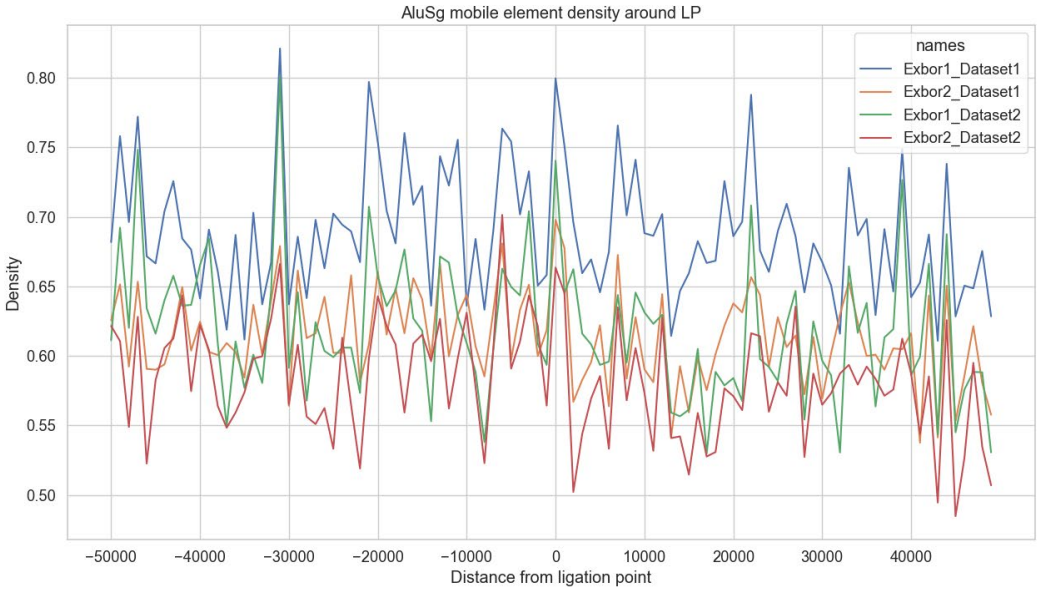

AluSg mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
0.80
0.75
0.70
0.65
0.60
0.55
0.50
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. AluSg**

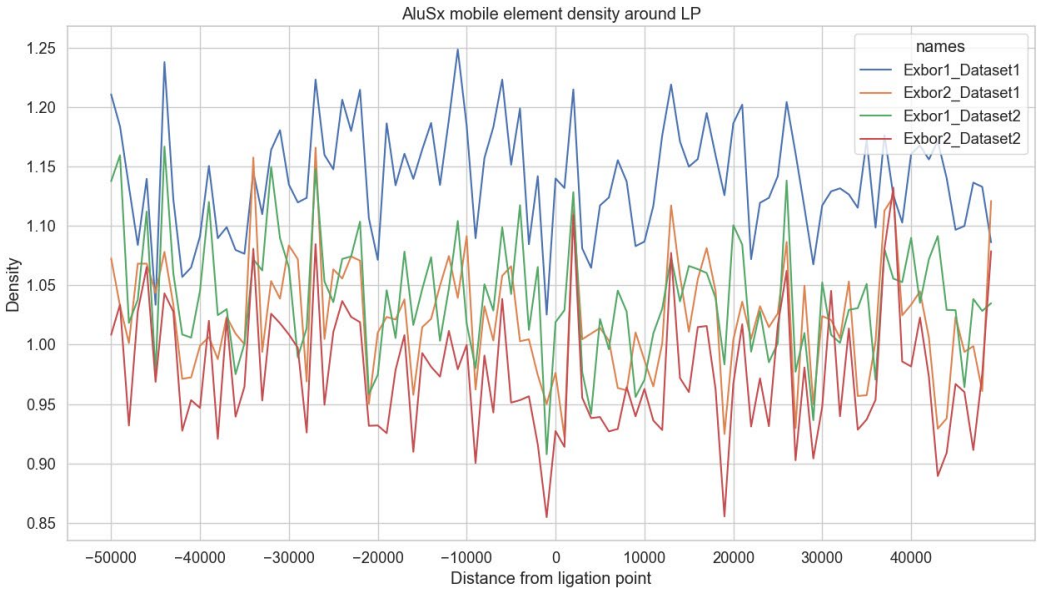

AluSx mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
1.25
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_AluSx**

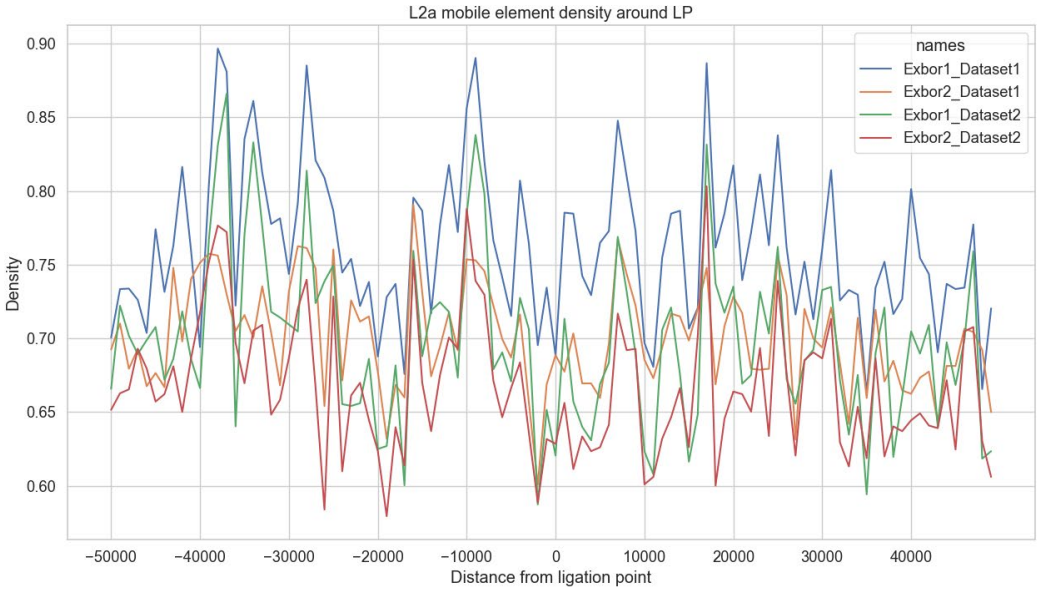

L2a mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
0.90
0.85
0.80
0.75
0.70
0.65
0.60
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_L2a**

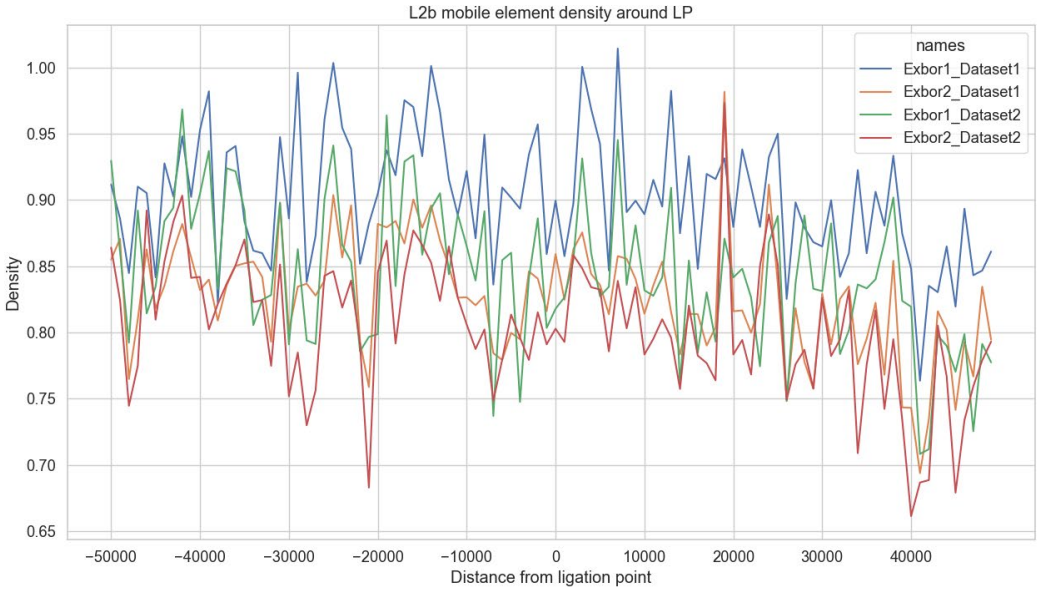

L2b mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_L2b**

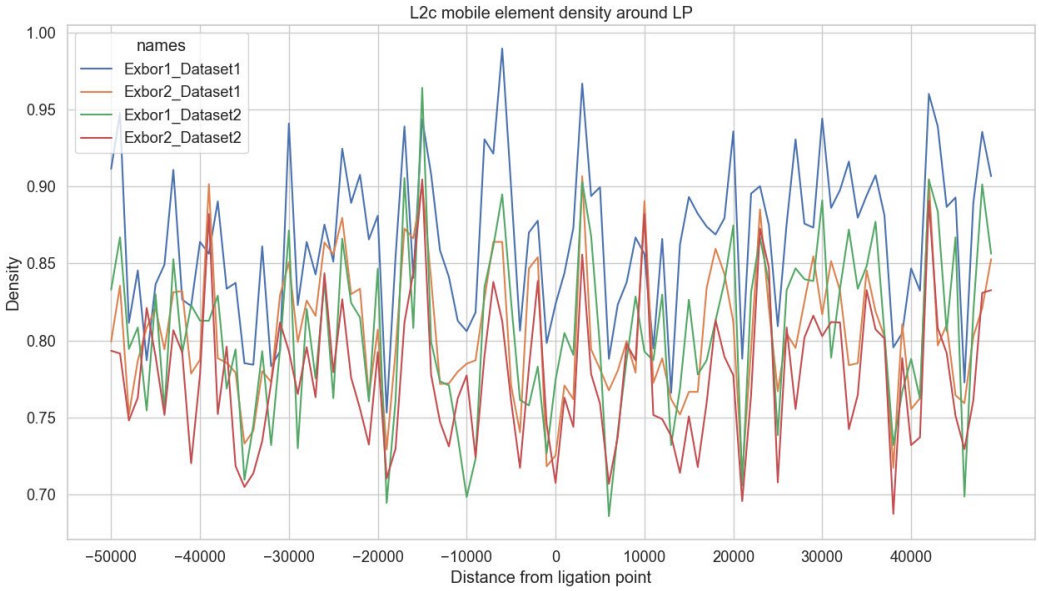

L2c mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
1.00
0.95
0.90
0.85
0.80
0.75
0.70
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_L2c**

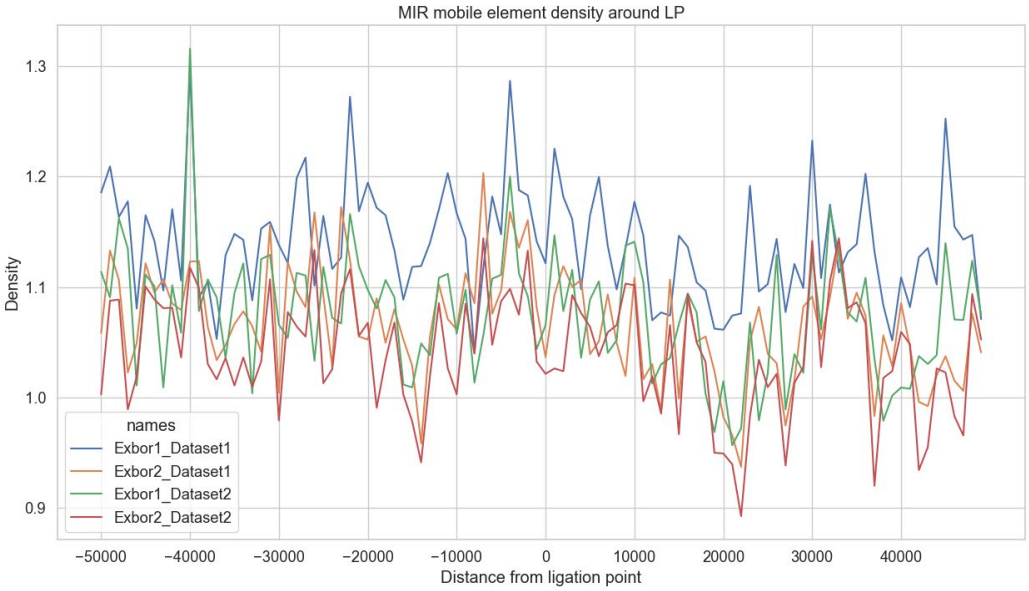



**Fig. S_MIR subfamily**

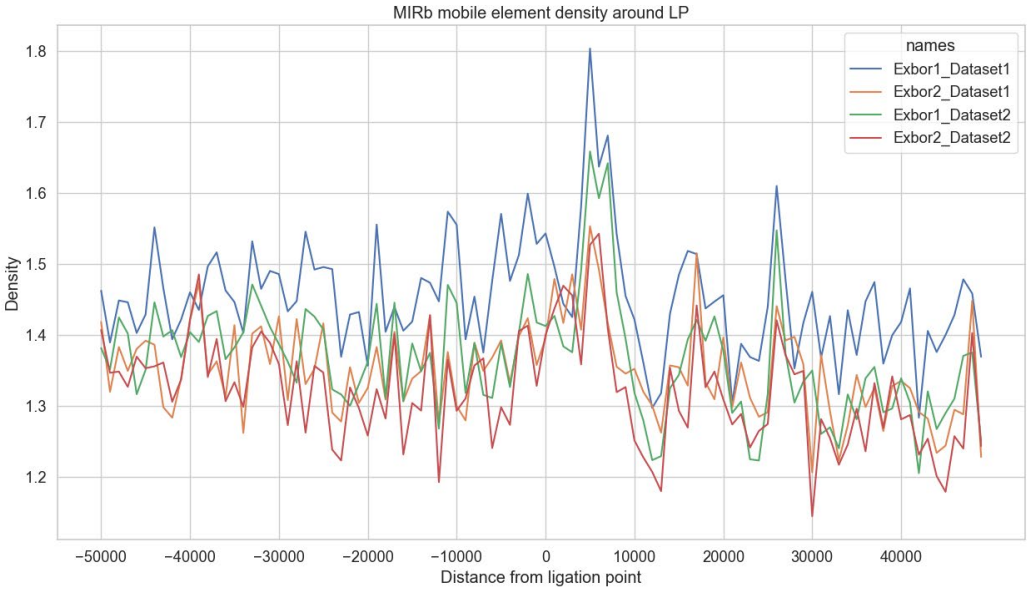

MIRb mobile element density around LP
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
Density
1.8
1.7
1.6
1.5
1.4
1.3
1.2
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_MIRb**

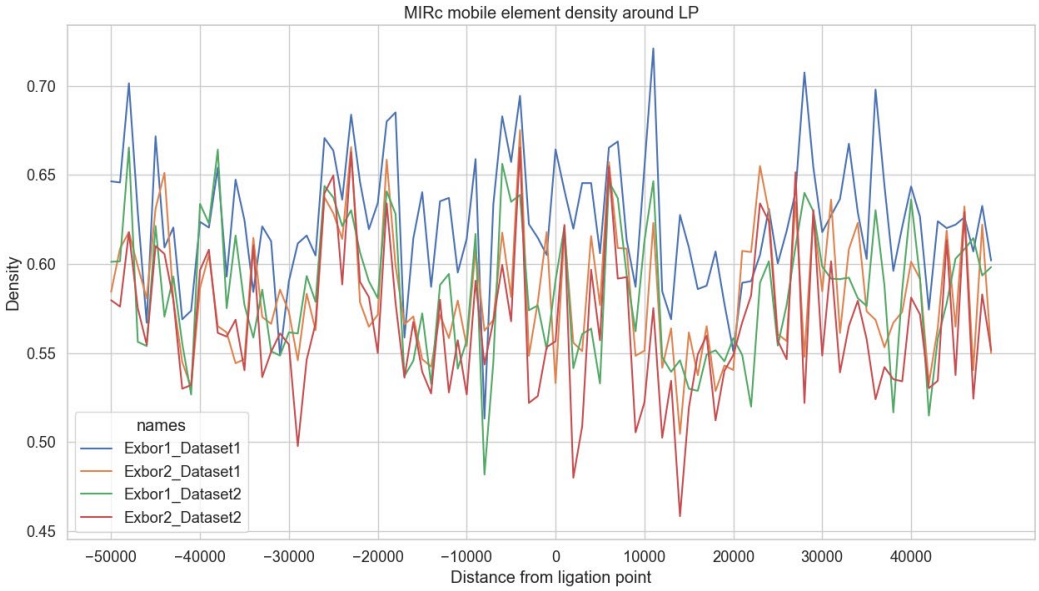

MIRc mobile element density around LP
Density
0.70
0.65
0.60
0.55
0.50
0.45
names
Exbor1_Dataset1
Exbor2_Dataset1
Exbor1_Dataset2
Exbor2_Dataset2
−50000
−40000
−30000
−20000
−10000
0
10000
20000
30000
40000
Distance from ligation point

**Fig. S_MIRc**

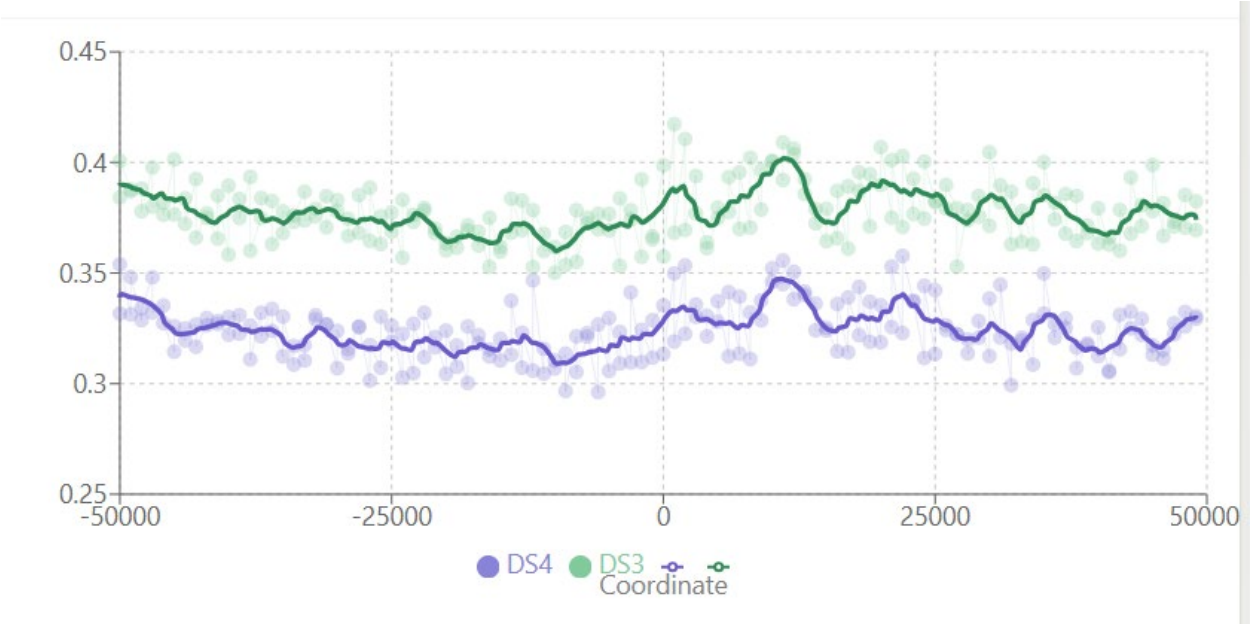



**Fig. S_DS3-DS4 Transposon L2b DS3 is correlated with DS4 (strands averaged, strand info ignored)**

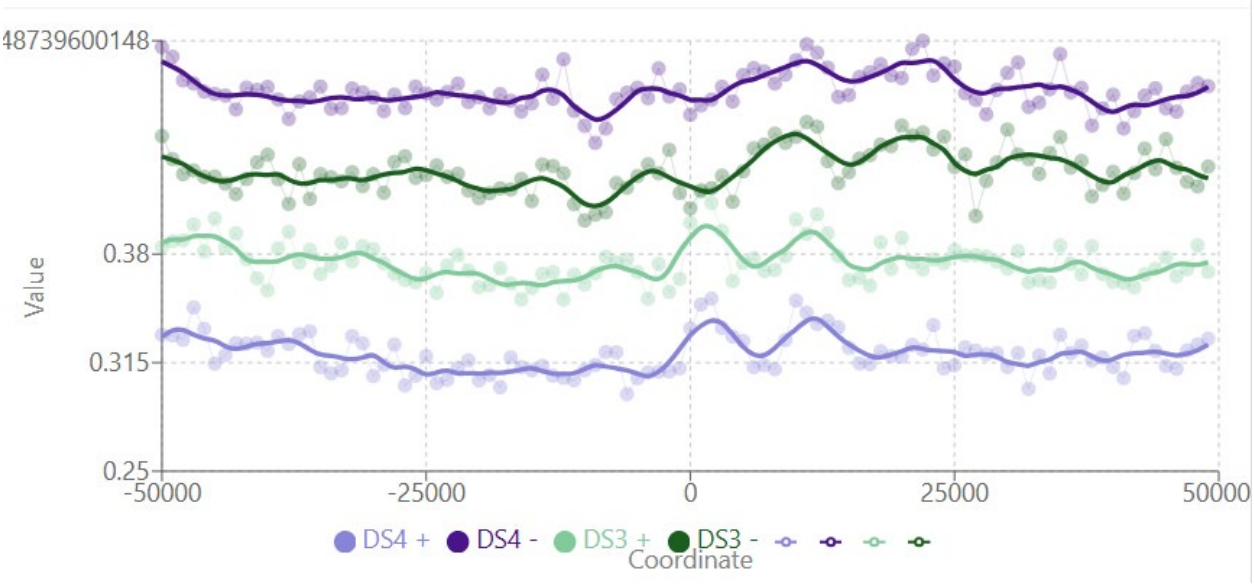
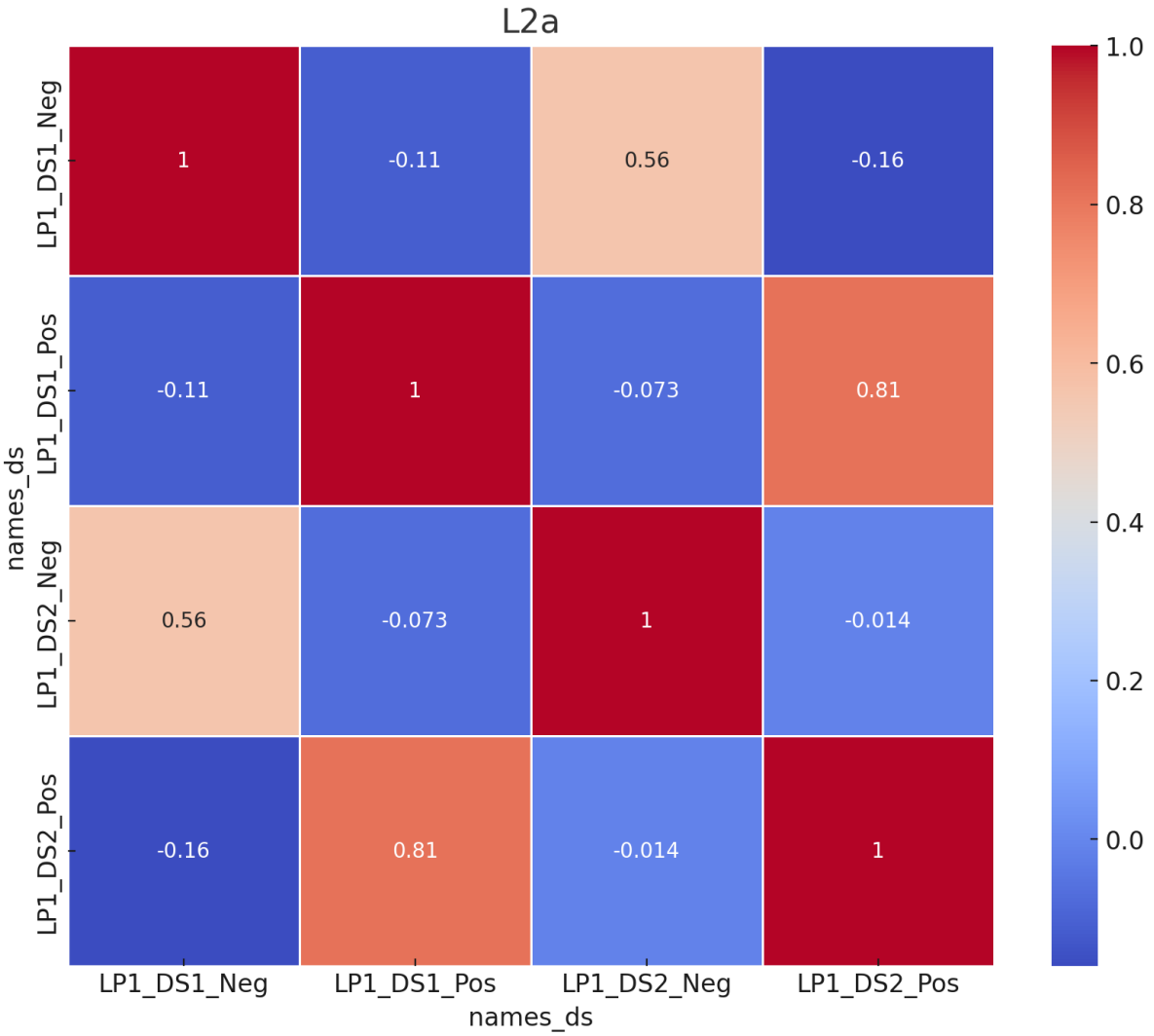


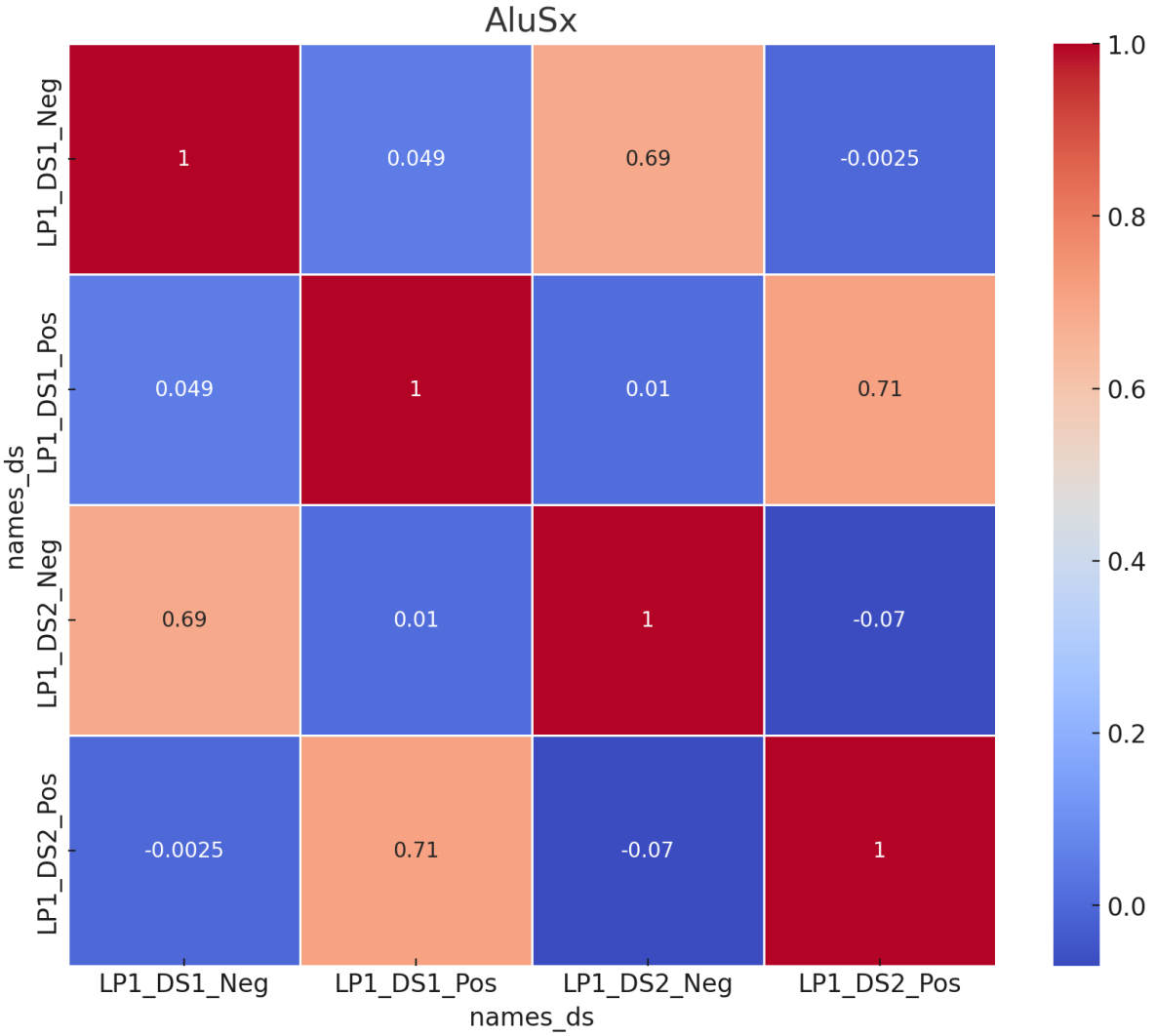

**Fig. S_DS3-DS4 Transposon L2b DS3 is correlated with DS4 (strads plotted separately)**

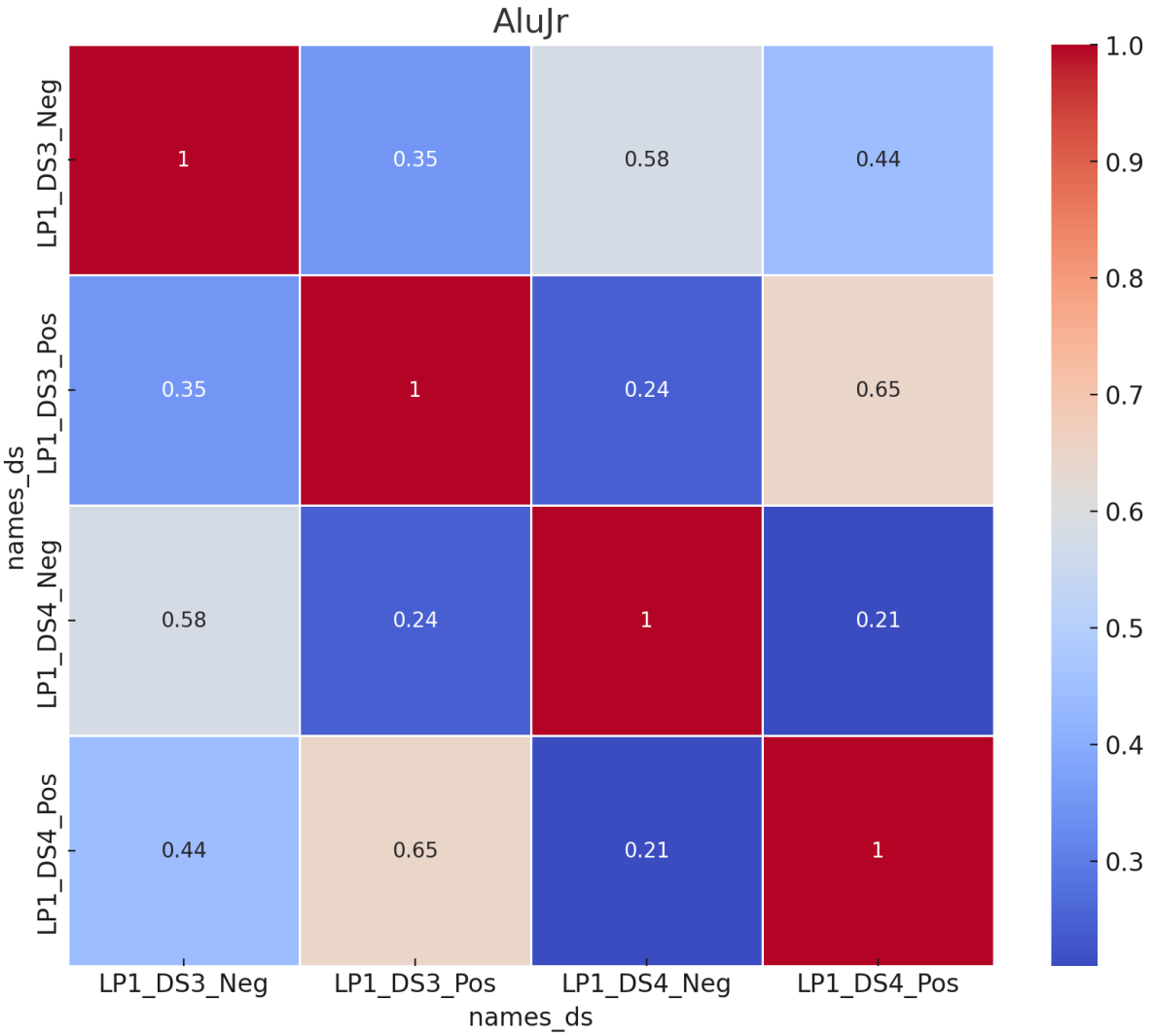
## Correlation

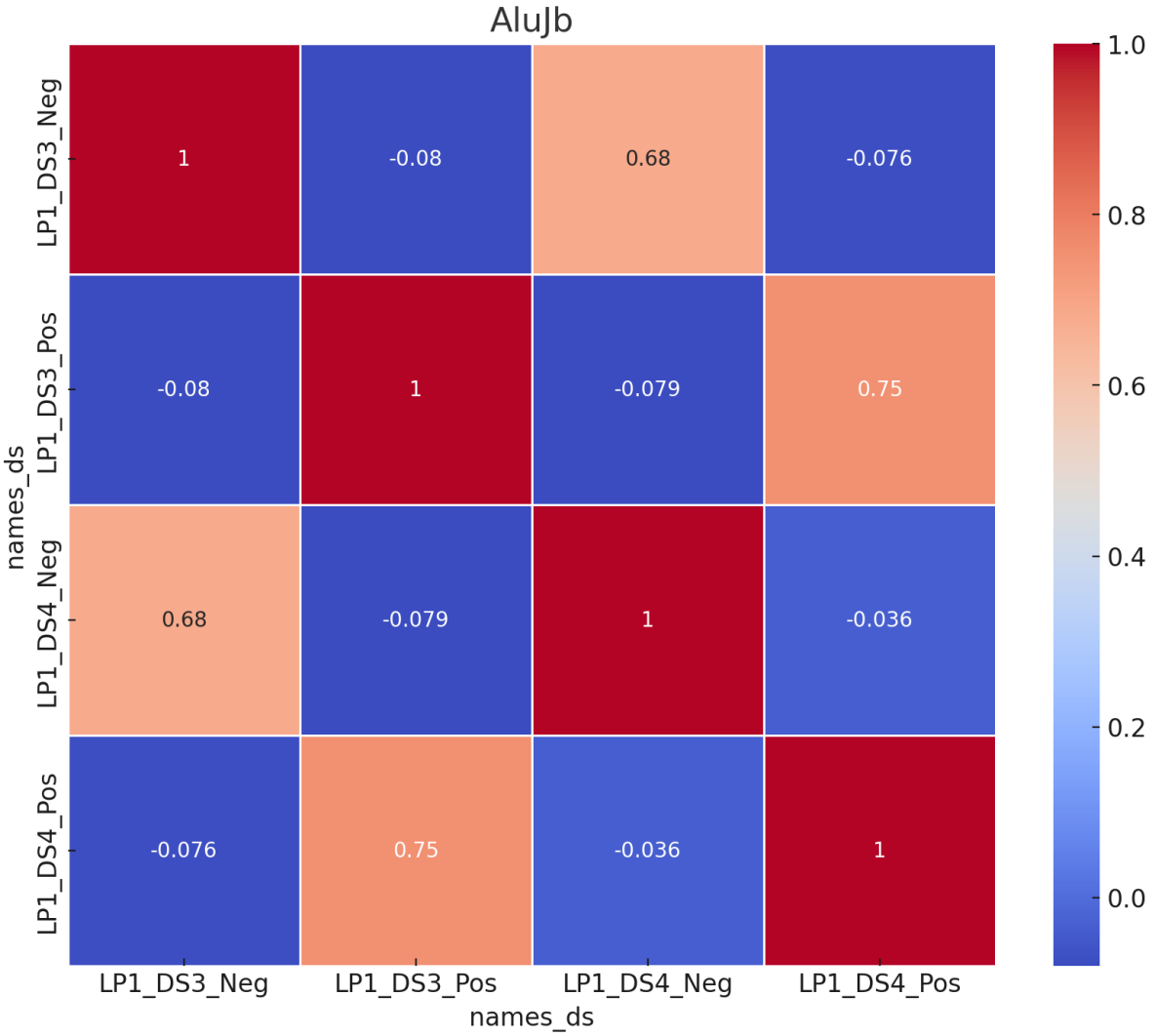
The following graphs are correlations for density curves between Datasets D1, D2, D3, D4. The transposon subfamily is marked on the top of each graph.

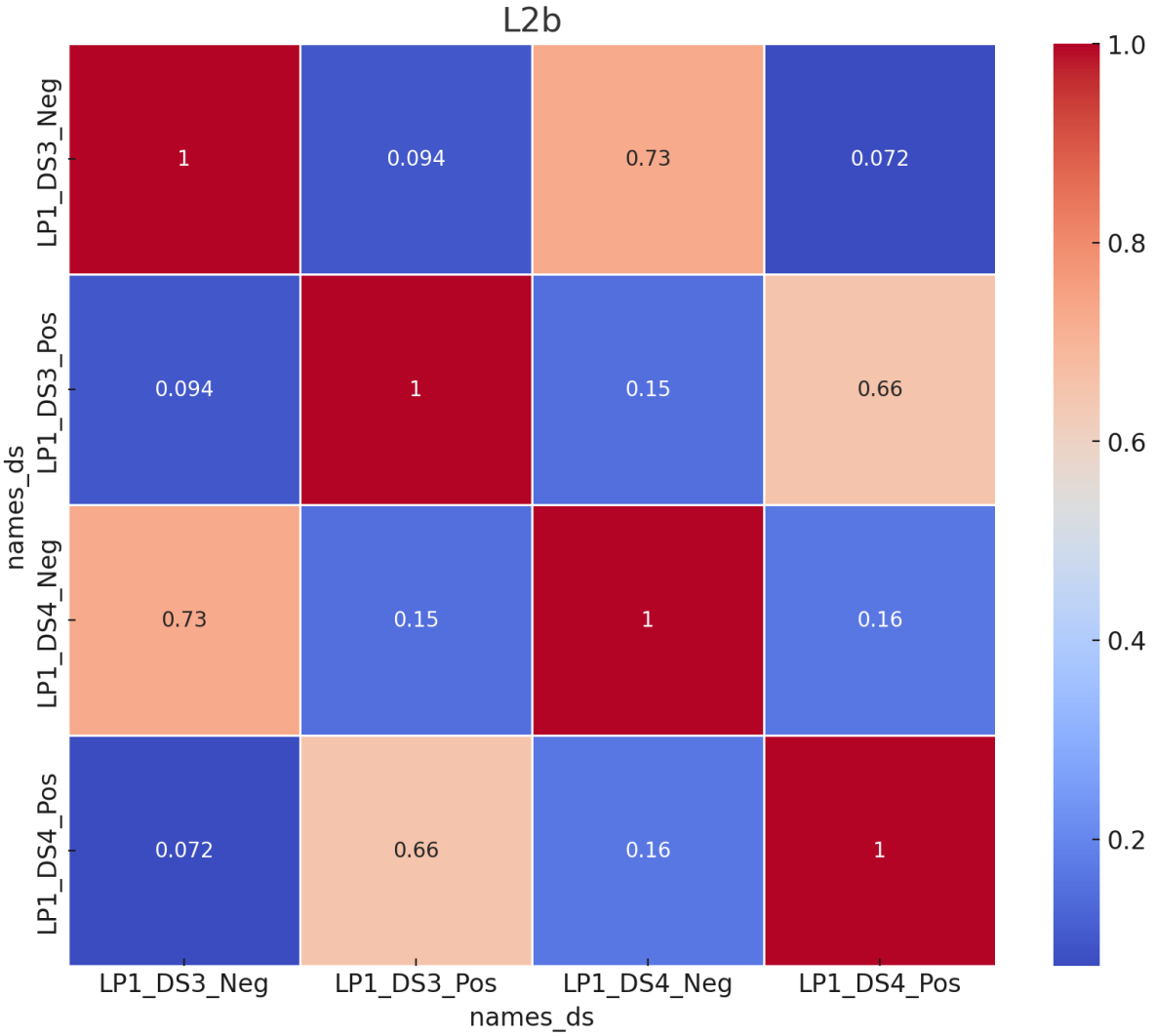
Correlation Analysis of Transposable Element Distributions These correlation plots demonstrate the reproducibility of transposable element density patterns around chromatin contact points across different datasets and methods. Each plot shows pairwise correlations between biological replicates, with lighter colors indicating stronger correlations. Plus (+) and minus (-) strands are analyzed separately to reveal strand-specific patterns. Correlation values between corresponding strands of biological replicates typically reach 0.7-0.8, while correlations between opposite strands remain below 0.2, quantitatively confirming strand specificity. Results are shown for Alu (X), L1 (Y), L2a-c (Z) and their subfamilies. The consistent patterns across Hi-C (DS3, DS4) and Micro-C (DS1, DS2) datasets validate the biological authenticity of these organizational features.

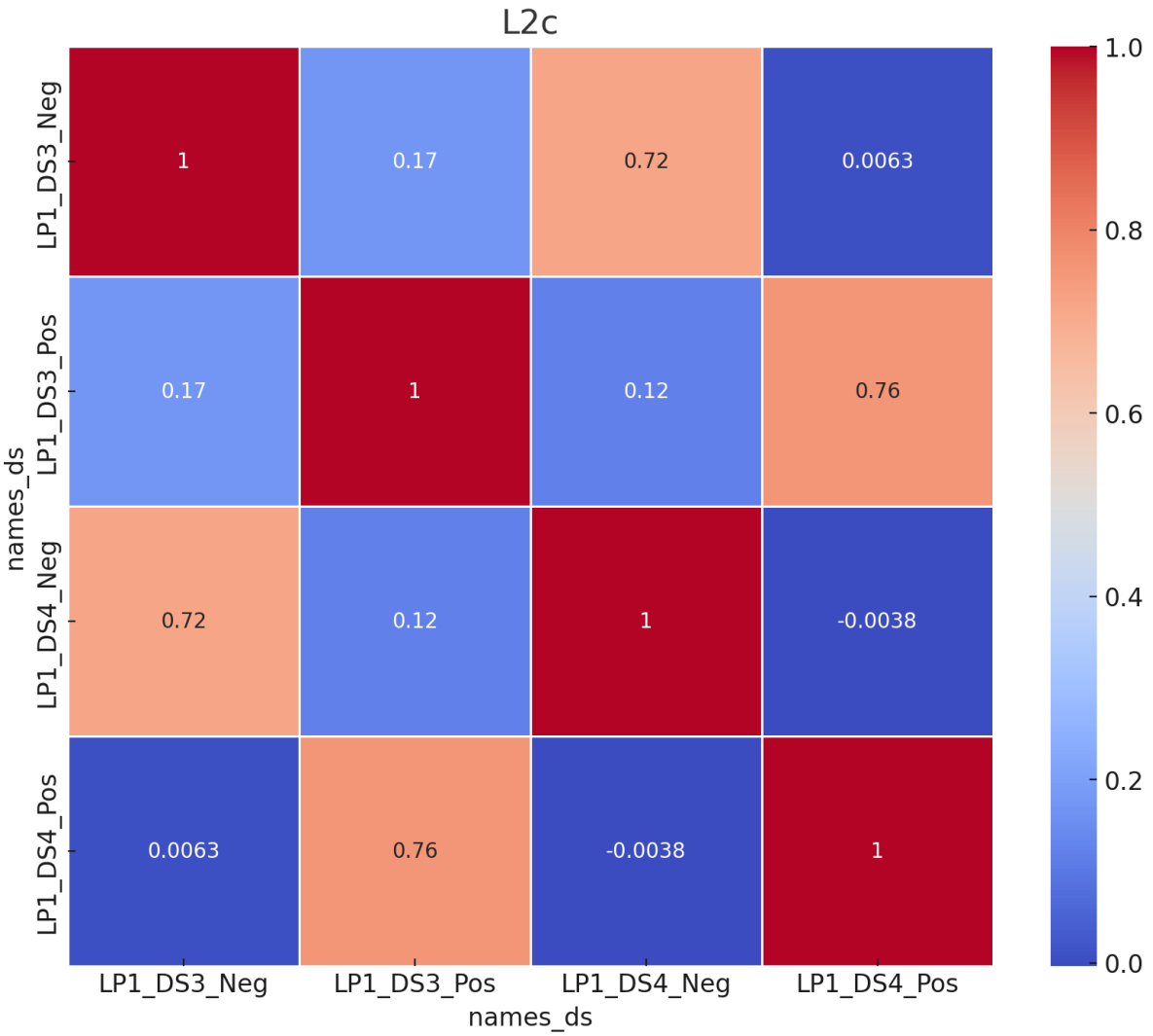



AluSx
LP1_DS1_Neg
LP1_DS1_Pos
LP1_DS2_Neg
LP1_DS2_Pos
names_ds
1
0.049
0.69
-0.0025
0.049
1
0.01
0.71
0.69
0.01
1
-0.07
-0.0025
0.71
-0.07
1
1.0
0.8
0.6
0.4
0.2
0.0


AluJr
LP1_DS3_Neg
LP1_DS3_Pos
LP1_DS4_Neg
LP1_DS4_Pos
names_ds
1
0.35
0.58
0.44
0.35
1
0.24
0.65
0.58
0.24
1
0.21
0.44
0.65
0.21
1
LP1_DS3_Neg
LP1_DS3_Pos
LP1_DS4_Neg
LP1_DS4_Pos
names_ds
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3


AluJb
LP1_DS3_Neg
LP1_DS3_Pos
LP1_DS4_Neg
LP1_DS4_Pos
names_ds
1
-0.08
0.68
-0.076
-0.08
1
-0.079
0.75
0.68
-0.079
1
-0.036
-0.076
0.75
-0.036
1
1.0
0.8
0.6
0.4
0.2
0.0


L2b
LP1_DS3_Neg
LP1_DS3_Pos
LP1_DS4_Neg
LP1_DS4_Pos
names_ds
1
0.094
0.73
0.072
0.094
1
0.15
0.66
0.73
0.15
1
0.16
0.072
0.66
0.16
1
1.0
0.8
0.6
0.4
0.2


L2c
LP1_DS3_Neg
LP1_DS3_Pos
LP1_DS4_Neg
LP1_DS4_Pos
names_ds
1
0.17
0.72
0.0063
0.17
1
0.12
0.76
0.72
0.12
1
-0.0038
0.0063
0.76
-0.0038
1
1.0
0.8
0.6
0.4
0.2
0.0

## Random Controls

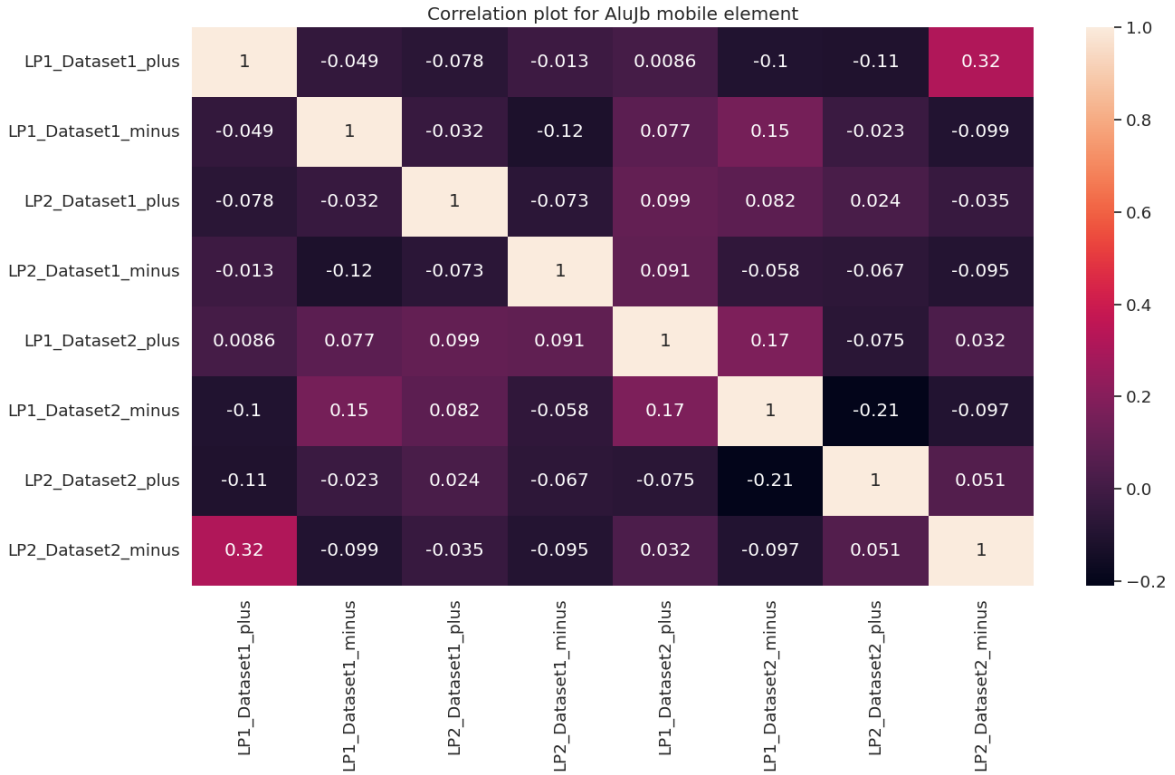



**Fig.S. Random controls**