Supplementary Materials

Enrichment of Z-DNA-Forming Sequences Within Super-Enhancers: A Computational and Population-Based Study

Yulia V. Makus [1,2], German A. Ashniev [1,3,4], Alexey V. Orlov [1], Petr I. Nikitin [1], Zoia G. Zaitseva [1], Natalia N. Orlova [1,*]

[1] Prokhorov General Physics Institute of the Russian Academy of Sciences, 38 Vavilov St., 119991 Moscow, Russia;
[2] National Research University Higher School of Economics, st. Myasnickaya, 20, 101000 Moscow, Russia;
[3] Faculty of Biology, Lomonosov Moscow State University, Leninskiye Gory, MSU, 1-12, 119991 Moscow, Russia
[4] Institute for Information Transmission Problems RAS, 127051 Moscow, Russia
* Correspondence: alexey.orlov@kapella.gpi.ru (A.V.O.); natalja191186@gmail.com (N.N.O.)

1. Table S1: Download Links for 100 individuals from the 1000 Genomes Project used in the study

Table S1 contains a list of direct download links to variant call format (VCF) files for 100 individuals selected from the 1000 Genomes Project. These links provide access to raw genotype data. Each row in the table corresponds to an individual and includes a direct FTP link to their respective VCF file. The dataset was used in this study for downstream analyses. Due to the large number of entries, the full table is provided as a separate file (*Table S1.csv*) in the folder of Supplementary Materials.

2. Table S2. Fold-Change in Z-DNA Enrichment Across Enhancer Categories and Background Regions Across Cell Lines

Table S2. Fold-Change in Z-DNA Enrichment Across Enhancer Categories and Background Regions Across Cell Lines

		HCT116	BT_549	HeLa	LNCap	MCF7	MDA-MB-231
TE	Original count	29940	11413	23149	20956	22017	32082
	Intersected count	11315	4902	7195	6122	8521	12733
	Fold-change	***2.646***	***2.328***	***3.217***	***3.423***	***2.584***	***2.520***
SE_E	Original count	15755	3058	9572	2837	8306	11060
	Intersected count	4229	1345	2285	994	2125	2071
	Fold-change	***3.725***	***2.274***	***4.189***	***2.854***	***3.909***	***5.340***
SE	Original count	1191	590	903	608	540	1168
	Intersected count	1092	535	793	536	501	967
	Fold-change	***1.091***	***1.103***	***1.139***	***1.134***	***1.078***	***1.208***
BG	Original count	101634					
	Intersected count	1092					
	Fold-change	***93.071***					

3. Table S3. Disease-Associated Variants Identified Within Z-DNA-Forming Super-Enhancer Regions

Table S3. Disease-Associated Variants Identified Within Z-DNA-Forming Super-Enhancer Regions

Chr.	Position (hg38)	rsID	Nearby Gene	Mutation Type	Potential Pathology	Clinical Significance
10	119 394 275	-	WAC	Long Deletion	Neurodevelopmental	Uncertain
10	119 394 519	-	WAC	Long Indel	Intellectual disability	Likely pathogenic
10	119 394 520	-	WAC/ASH2L	Insertion	Cancer/neurodevelopmental	Uncertain
10	119 394 546	-	WAC	Frameshift	ASD, macrocephaly	Likely pathogenic
10	119 394 789	-	ASH2L	Insertion	Somatic cancers	Likely pathogenic

4. Table S4: Per-individual information on mutation status, reference and alternative alleles, affected motifs, transcription factor binding changes, and intersections with pathogenic variants

Table S2 provides detailed per-individual information on genetic mutations, including their effects on transcription factor (TF) binding and intersections with known pathogenic variants. This dataset is structured to facilitate the analysis of mutation consequences at specific genomic positions across multiple individuals.

The table includes the following columns:

- Individual_id – Unique identifier for each individual.
- Position – Genomic position of the mutation.
- Status – Type of reference or alternate allele (e.g., REF, ALT).
- Value – Full sequence context of the region containing the mutation.
- Reference – Reference allele sequence.
- Mutation – Mutated allele sequence (if applicable).
- Motif – DNA sequence motif affected by the mutation.
- What happened – Description of mutation effects (e.g., gained or lost binding sites).
- TF – Transcription factor affected by the mutation.
- Intersection with pathogenic mutations – List of pathogenic variant IDs and their genomic coordinates.
- Allele frequency – Observed frequency of the mutation in the population.
- Allele frequency in 100 – Frequency of the mutation in a selected subset of 100 individuals.

Due to the large size of the dataset, Table S2 is provided as a separate file (*Table S2.xlsx*) in the folder of Supplementary Materials.

5. Figure S1: Heat maps of Z-DNA distribution across enhancer elements in five cancer cell lines (HeLa, MDA-MB-231, MCF7, BT-549, LNCap)

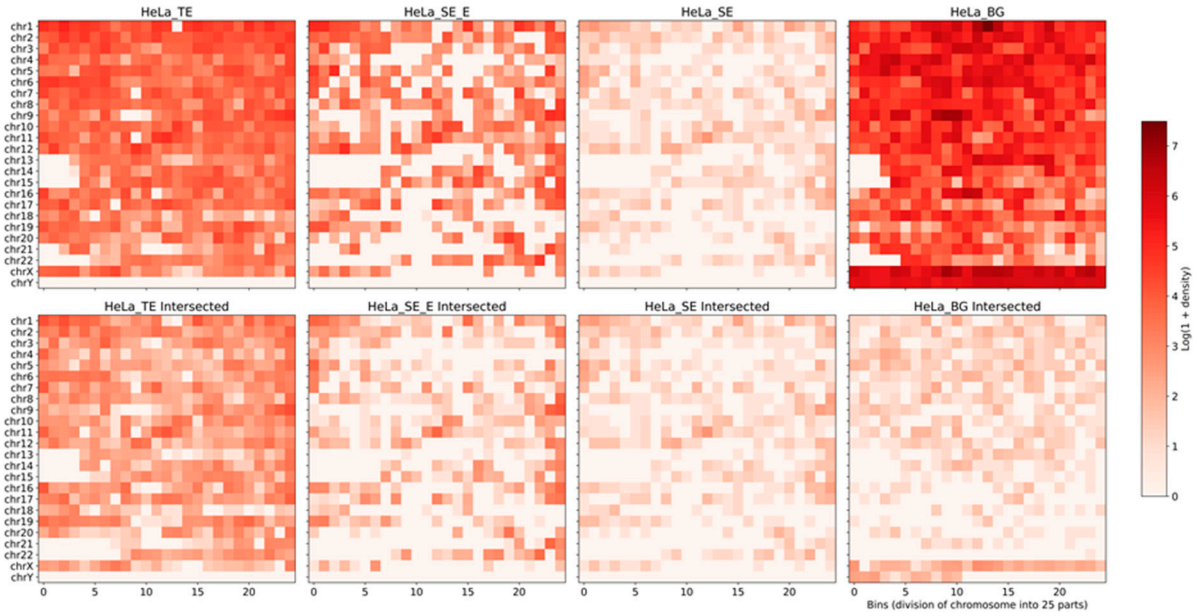

Figure S1a. Heat Maps of Z-DNA Distribution Across Enhancer Elements in the Hela Cell Line: Comparison of Original (Top Row) and Z-DNA-Intersected (Bottom Row) Datasets Across Typical Enhancers, Super-Enhancer Enhancers, Super-Enhancers, and Background Regions (BG) with Chromosomal Binning (25 Bins per Chromosome, hg38 Assembly).

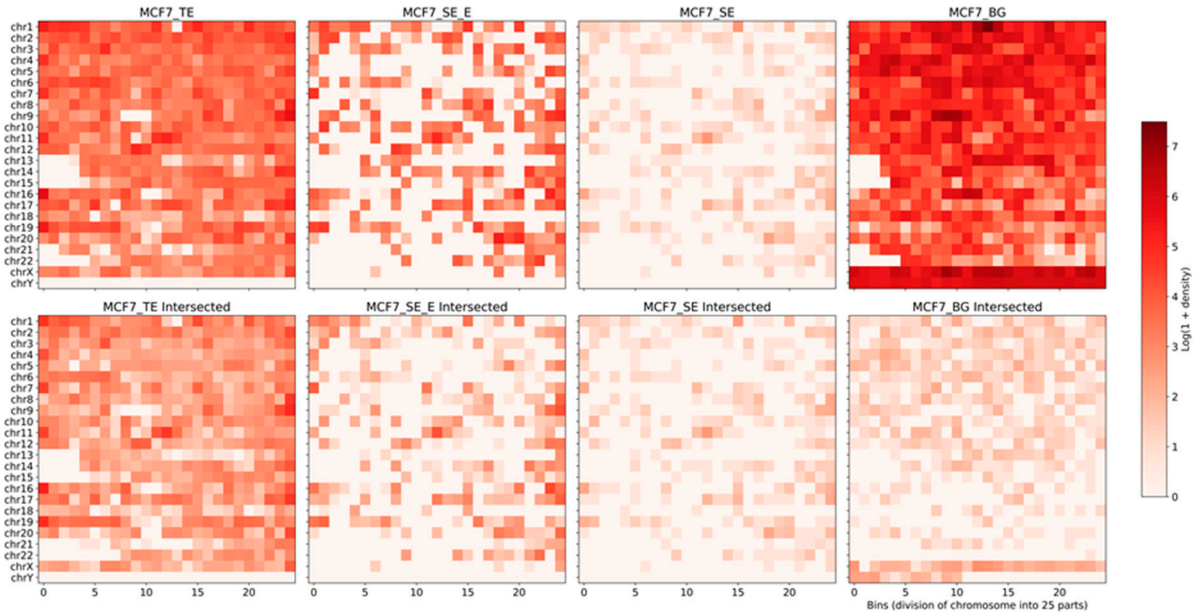

Figure S1b. Heat Maps of Z-DNA Distribution Across Enhancer Elements in the MCF7 Cell Line: Comparison of Original (Top Row) and Z-DNA-Intersected (Bottom Row) Datasets Across Typical Enhancers, Super-Enhancer Enhancers, Super-Enhancers, and Background Regions (BG) with Chromosomal Binning (25 Bins per Chromosome, hg38 Assembly).

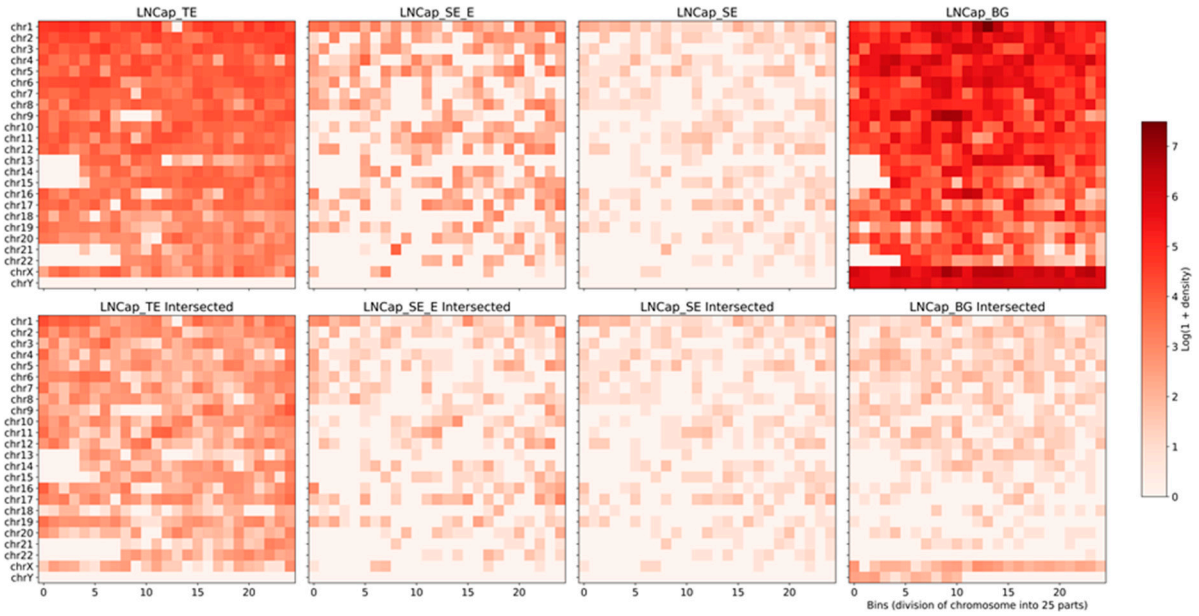

Figure S1c. Heat Maps of Z-DNA Distribution Across Enhancer Elements in the LNCap Cell Line: Comparison of Original (Top Row) and Z-DNA-Intersected (Bottom Row) Datasets Across Typical Enhancers, Super-Enhancer Enhancers, Super-Enhancers, and Background Regions (BG) with Chromosomal Binning (25 Bins per Chromosome, hg38 Assembly).

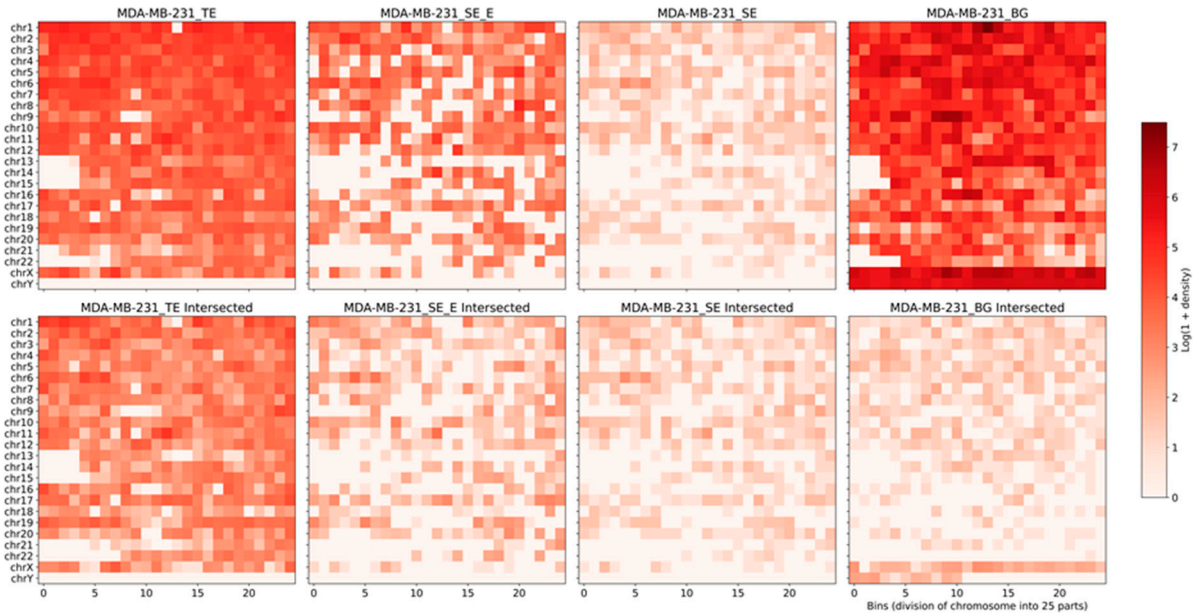

Figure S1d. Heat Maps of Z-DNA Distribution Across Enhancer Elements in the MDA-MB-321 Cell Line: Comparison of Original (Top Row) and Z-DNA-Intersected (Bottom Row) Datasets Across Typical Enhancers, Super-Enhancer Enhancers, Super-Enhancers, and Background Regions (BG) with Chromosomal Binning (25 Bins per Chromosome, hg38 Assembly).

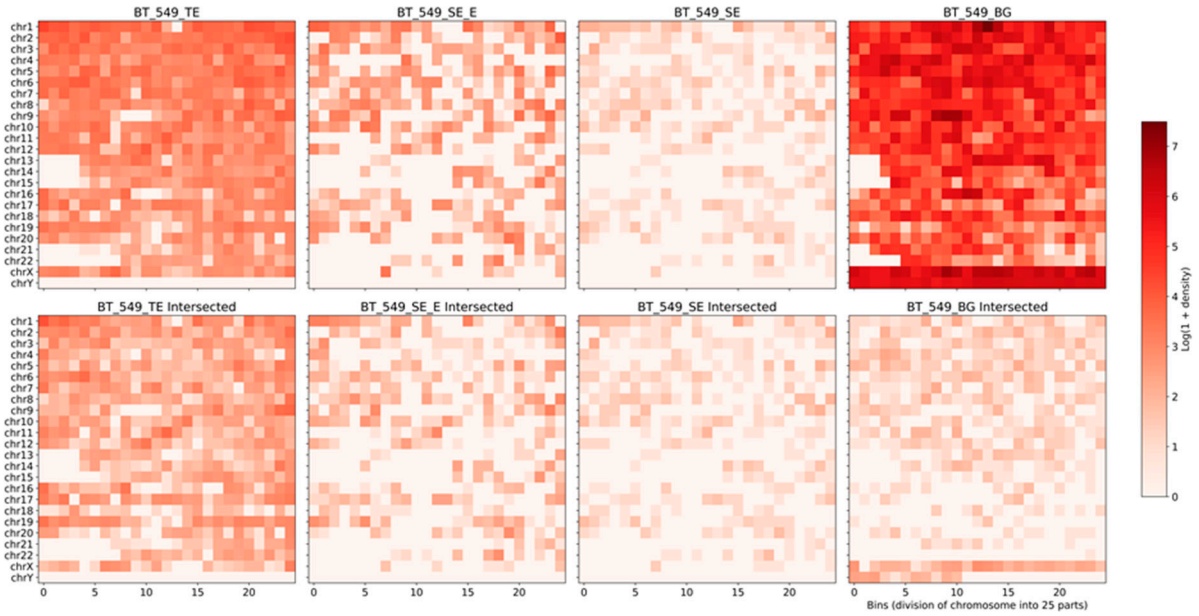

Figure S1e. Heat Maps of Z-DNA Distribution Across Enhancer Elements in the DT_549 Cell Line: Comparison of Original (Top Row) and Z-DNA-Intersected (Bottom Row) Datasets Across Typical Enhancers, Super-Enhancer Enhancers, Super-Enhancers, and Background Regions (BG) with Chromosomal Binning (25 Bins per Chromosome, hg38 Assembly).