



OPEN Z-DNA formation in promoters conserved between human and mouse are associated with increased transcription reinitiation rates

Nazar Beknazarov^{1,3}, Dmitry Konovalov^{1,3}, Alan Herbert^{1,2}✉ & Maria Poptsova¹✉

A long-standing question concerns the role of Z-DNA in transcription. Here we use a deep learning approach DeepZ that predicts Z-flipons based on DNA sequence, structural properties of nucleotides and omics data. We examined Z-flipons that are conserved between human and mouse genomes after generating whole-genome Z-flipon maps and then validated them by orthogonal approaches based on high resolution chemical mapping of Z-DNA and the transformer algorithm Z-DNABERT. For human and mouse, we revealed similar pattern of transcription factors, chromatin remodelers, and histone marks associated with conserved Z-flipons. We found significant enrichment of Z-flipons in alternative and bidirectional promoters associated with neurogenesis genes. We show that conserved Z-flipons are associated with increased experimentally determined transcription reinitiation rates compared to promoters without Z-flipons, but without affecting elongation or pausing. Our findings support a model where Z-flipons engage Transcription Factor E and impact phenotype by enabling the reset of preinitiation complexes when active, and the suppression of gene expression when engaged by repressive chromatin complexes.

Flipons are important functional genetic elements that regulate various genomic processes by changing conformation^{1–3}. Here we focus on Z-flipons that can form Z-DNA and Z-RNA under physiological conditions. Recent studies highlighted the key-role of Z-RNA in immune response^{4–15}, either switching interferon responses off to limit inflammation or initiating cell death to eliminate virally infected or dysfunctional cells^{15–18}. The flipons involved are subject to natural selection and underscore the importance of flipons in genome evolution.

Other genetic roles for Z-DNA are less well characterized. A role in transcription was suggested early on with evidence provided by Wittig and collaborators in studies of the c-myc gene in permeabilized nuclei¹⁹. A number of experiments show that the negative supercoiling produced by RNA polymerases during gene transcription can induce Z-DNA associated with chromatin remodeling by the SMARCA4 complex^{20,21}. Further evidence for the transient formation of non-B DNA structure in both mouse and human cells is provided by whole genome experimental approaches that involve the use of rapid permanganate/S1 nuclease (KEx) footprinting and kethoxal-assisted, single-stranded DNA sequencing (KAS-seq) performed on the time scale of seconds to minutes^{22,23}. A different approach for inferring Z-DNA function, is based on the detection of Z-DNA using ChIP-seq approaches. Such experiments are currently limited in number: one study in humans used the Z-DNA structure specific Za domain to enrich for Z-DNA²⁴, while a mouse-study used a Z-DNA specific antibody to detect Z-DNA formation in curaxin treated cells¹⁵. These approaches allow the mapping of the chromatin states that are associated with Z-DNA formation, without conditioning on computationally predicted Z-prone sequences.

A general problem in flipon detection lies in the dynamic nature of their formation—they flip-on, execute their function and then flip-off. Any experiment will capture only the subset of all functional flipons that are active at a given moment in time. The problem is especially true for Z-DNA given the rapid speed at which flips occur. The ultimate goal is to compile whole-genome maps of functional Z-flipons, not just maps of Z-prone

¹Laboratory of Bioinformatics, Faculty of Computer Science, National Research University Higher School of Economics, Moscow, Russia. ²InsideOutBio, Charlestown, MA, USA. ³These authors contributed equally: Nazar Beknazarov and Dmitry Konovalov. ✉email: alan.herbert@insideoutbio.com; mpoptsova@hse.ru

sequences. The function roles of Z-flipons can be explored by aggregating information from co-localizing epigenetic marks at the sites of Z-DNA formation.

Earlier we developed the DeepZ approach²⁵ that makes prediction of Z-DNA forming sites by supervised training on a large number of omics datasets, including histone marks (HM), transcription factors (TF), chromatin modification complexes (CC) and RNA polymerase and promoter factor binding sites (RP). We took into account the *in vitro* determined energetic parameters of dinucleotides switching from the B- to Z- conformation, negatively weighing those sequences, such as runs of deoxyadenosine, that are unlikely to form Z-DNA under physiological conditions²⁶. We used experimentally confirmed Z-DNA regions to train and validate DeepZ and then to generate *de novo* predictions of Z-DNA forming sequences, both in human²⁵ and in mouse¹⁵.

Here we expand the DeepZ approach to explore ENCODE generated datasets to identify epigenetic marks that correlate with Z-DNA function. DeepZ scores the importance of each feature in the prediction model it generates, allowing us to identify HM, TF, CC and RP that are associated with Z-DNA formation. We focus on those Z-flipons that are in the conserved regions shared by mouse and human genomes, hypothesizing that significant regulatory features of these regions should also be similar. Findings in one genome could then be validated by replication in the other. We further confirmed the DeepZ predictions using those from Z-DNABERT, an orthogonal transformer algorithm trained on experimental datasets independently derived by the chemical footprinting of Z-DNA in live cells.

Clustering of HM, TF, CC and RP features enabled us to map Z-flipons to known genomic regions. We found that Z-flipons could be associated with both active and repressed chromatin regions, suggesting that they have both a positive and a negative role in regulating transcription. We explored further those Z-flipons located at alternative and bidirectional promoters and noted enrichment of functional Z-flipons in neurogenesis genes. We then focused on the impact that Z-flipons have on experimentally derived measures of transcription. We found that conserved active Z-flipons in living cells are associated with increased reinitiation rates of gene transcription, without effects on pause duration or on the elongation rates of RNA polymerases²⁷.

Results

DeepZ predictions in human and mouse genomes based on the common set of omics features

Here we applied the same DeepZ pipeline as described in²⁵ using for training human²⁴ and mouse genome ChIP-seq data¹⁵. The schematic of the DeepZ approach is presented in Fig. 1A (see Methods, Supplementary Fig. 1). For omics data we took all available HM, TF, DNase accessible sites, CC and RP (see full list in Supplementary Table 1). We selected only experiments available for both genomes. In total there were 544 features used (Fig. 1B, C), including 65 HM, 466 TF and CC, 3 methylation maps, 8 RP, a map of dinucleotide energy transitions from B- to Z-form, and DNase hypersensitivity sites. We generated whole-genome annotations for human and mouse genomes with Z-DNA regions (Supplemental Data 1–2), which comprise 30,083 segments in human and 17,569 in mouse.

The number of potential Z-DNA forming sites in human comprise ~ 3 Mb compared to ~ 2.6 Mb in mouse genome, but due to the different genome sizes, both comprise ~ 0.1% of cumulative genome length in both genomes (Supplementary Table 2). The distributions of the DeepZ predicted Z-DNA over genomic regions for mouse and human are given in Fig. 1D. In both genomes the distribution is qualitatively the same with enrichment in promoters, exons, 5'UTR and 3'UTR.

Since DeepZ distribution over genomic regions is shifted towards enrichment in regulatory regions—promoters, exons, 5'UTR, 3'UTR (Fig. 1D), we verified DeepZ performance metrics on each genomic region separately. The results for promoters are concordant with those found in other regions of the genome (Table 1 and Supplementary Table 3).

To verify how groups of features contribute to model performance, we performed an ablation analysis with gradient boosting used to assess group feature importance (see “Methods” and Supplementary Fig. 2). We confirmed that information only from DNA sequence is not sufficient to predict Z-DNA formation, and that each omics feature group improves model performance as measured by the F1 metric. We further assessed the effects of different thresholds to distinguish between flipons that form Z-DNA under physiological conditions and those that do not. All the results presented are based on the use of threshold 3 for both mouse and human genomes (See Methods and Materials).

Z-flipons are enriched in conserved human-mouse regions

We found significant enrichment of the predicted DeepZ regions in the conserved regions between human and mouse genome (Fig. 1E, F, Supplementary Data 3–4). The enrichment is 2.6-fold ($p < 0.001$, permutation test) in human and 1.6-fold in mouse ($p < 0.001$, permutation test), comprising 20% (human) and 10% (mouse) of all DeepZ predicted Z-DNA regions (Supplementary Table 4). The number of genes with Z-DNA from regions conserved in vertebrate clades is almost three times larger (7188 genes) in human than in mouse (2310 genes) (Supplementary Table 4), likely reflecting the differences in the nature and size of the training sets we used. From both lists, 966 genes are human and mouse orthologs with predicted Z-DNA sites in the body of the gene. GO analysis reveals enrichment of human and mouse orthologs, with conserved Z-flipons in regulation of metabolic process (546 genes, FDR $e-36$), regulation of transcription by RNA polymerase II (220 genes, FDR $e-21$), response to stimulus (495 genes, FDR $e-09$), binding to protein and nucleic acids (758 genes; FDR $e-21$), alternative splicing (614 genes, FDR $e-12$), chromatin organization (77 genes, FDRe-8), MAPK signaling pathway (37 genes; FDR $e-06$), with location in nucleus (590 genes, FDR $e-45$), nuclear lumen (423 genes, FDR $e-37$) (Supplementary Table 5). Full list of the most-enriched pathways and processes are given in Supplementary Table 6.

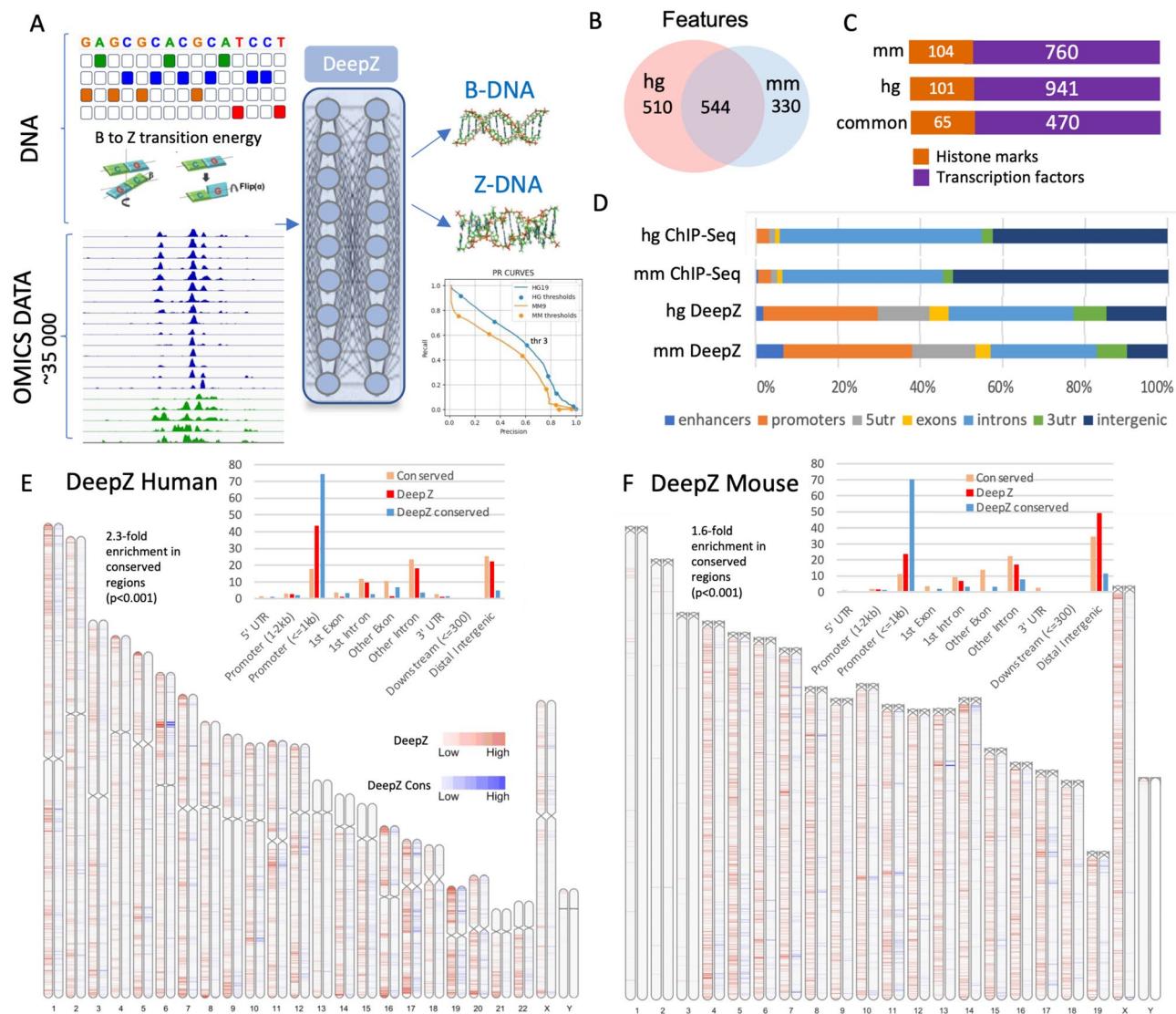


Figure 1. DeepZ predictions in human and mouse genomes based on the common set of omics features. **(A)** General Schema of DeepZ approach, with PR curves showing model performance. **(B)** Number of common and unique features used in DeepZ model. **(C)** Distribution of features over functional groups. **(D)** Distribution of Z-DNA regions over genomic regions. **(E)** Whole-genome distribution of DeepZ predicted Z-flipons in human genome. **(F)** Whole-genome distribution of DeepZ predicted flipons in mouse genome. In **(E,F)** regions conserved between human and mouse are highlighted in blue. Bar-graphs in **(E,F)** present distribution of conserved regions, DeepZ and conserved DeepZ predictions over genomic regions.

Region	PR	RC	F1
hg19			
All regions	0.609	0.517	0.559
Promoter (≤ 1 kB)	0.558	0.529	0.543
Promoter (1–2 kB)	0.655	0.493	0.563
Promoter (2–3 kB)	0.567	0.645	0.603
mm9			
All regions	0.603	0.453	0.517
Promoter (≤ 1 kB)	0.603	0.453	0.518
Promoter (1–2 kB)	0.593	0.467	0.523
Promoter (2–3 kB)	0.544	0.420	0.474

Table 1. DeepZ performance metrics for promoter regions.

Common patterns of transcription factors, regulators and histone marks enriched around Z-flipons conserved in human and mouse

We aimed to find common TF, CC, RP and HM (here and after referred to as an omics feature) that are enriched in regions around conserved Z-flipons, to assess the association of Z-DNA forming regions with active transcription. We performed this analysis initially at the genome-wide level, then in promoter regions, CpG- and non-CpG-promoters and found that the majority of TFs (86%, 404 out of 470 in human and 93%, 439 out of 470 in mouse) and HMs (68%, 46 out of 68 in human and 72%, 49 out of 68 in mouse) showed statistically significant enrichment ($p < 0.001$, permutation test) around conserved Z-regions in promoters (Supplementary Table 7).

We selected the top-20 features that are both most frequently co-localized and significantly enriched ($p < 0.001$, permutation test) with conserved Z-flipons in each category both for human and mouse and the combined feature importance plot is presented in Fig. 2. In line with our previous studies^{15,25,28}, we observed enrichment of Z-flipons in promoters. The result for Z-forming sequences in CpG-promoters is unsurprising as alternating d(CG) d(C₁₋₃G₁₋₃) are prone to Z-DNA formation²⁶. We also observed an enrichment of Z-prone sequences in non-CpG promoters.

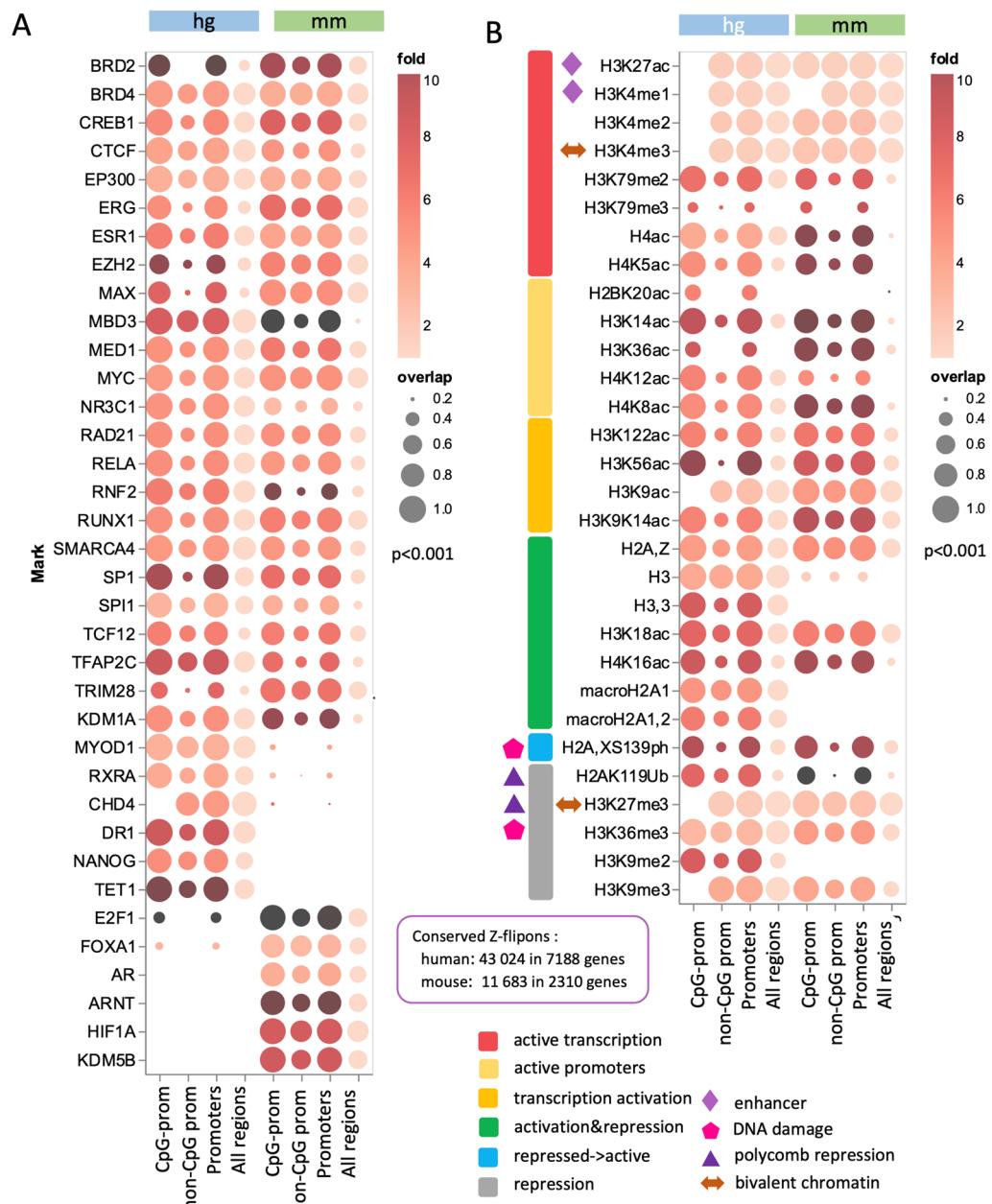


Figure 2. Conserved patterns of transcription factors and histone marks enriched around Z-flipons in human and mouse genome-wide. (A) Enrichment of transcription factors around Z-flipons for the entire genome, promoters, CpG-promoters and non-CpG promoters. (B) Enrichment of histone marks around Z-flipons.

Features that are enriched with conserved Z-flipons in promoter regions, both in human and mouse, are exactly those transcription factors, chromatin remodelers and epigenetic regulators that are enriched in transcriptionally active genes and associated with regions of open chromatin, where DNA conformation is less restrained. For genes presented in Fig. 2 significant Gene Ontology (GO) enrichment was found for the terms “Regulation of transcription by RNA polymerase II” (GO:0006357, FDR 6.08e–25); “Chromatin organization” (GO:0006325, FDR 1.08e–18); “Chromatin remodeling” (GO:0006338, FDR 1.32e–09); “Positive regulation of transcription, DNA-templated” (GO:0045893, FDR 2.99e–23); “Regulation of gene expression, epigenetic” (GO:0040029, FDR 1.40e–05), “Developmental process” (GO:0032502, FDR 1.33e–08), “Cell differentiation” (GO:0030154, FDR 2.71e–05).

Histone marks that are enriched with conserved Z-flipons in promoter regions both in human and mouse include many acetylated marks indicating active promoters and transcription activation (H3K12ac, H3K14ac, H3K18ac, H3K56ac, H3K122ac, H4K5ac, H4K8ac, H4K12ac), H3K4me1/me2/me3 that are indicators of active transcription, and also some specific markers associated with different chromatin states—H2A.Z, H2AXS139ph, and H2AK119Ub.

The same analysis showed that the HM and TF features identified by DeepZ in non-conserved DNA regions are quite variable, highlighting the differences in the evolutionary selection of mouse and human genomes (Supplementary Fig. 3). We also noted that many other HM and TF showed no association with Z-DNA formation, further confirming the specific nature of DeepZ predictions.

Clusters of common conserved Z-flipons between human and mouse reveal functional groups of LINEs, embryonic development, and neurogenesis

To further analyze DeepZ predictions, we applied UMAP clustering of Z-flipons defined as the vectors of omics features. We combined human and mouse data sets and extracted features mapping to experimental and predicted Z-DNA regions, (Fig. 3A). As expected, the UMAP shows that DeepZ predictions incorporate information obtained from many different cell lines and collected under many conditions helped to overcome the limitations of a single experiment by enabling better discrimination of signal from noise. The approaches exploited the differences in the training sets used for human and mouse models. Whereas the human ChIP-seq set was enriched for active Z-flipons that bound Za, the mouse dataset identified inactive flipons that were derepressed by the treatment with curaxin. Together, the datasets helped validate flipons features common to both human and mouse genomes and also improved the mapping of both active and suppressed flipons in each species.

Figure 3B, C shows how Z-flipons are distributed over genomic regions. The clusters with conserved, overlapping human and mouse flipons are highlighted in Fig. 3B. Interestingly, the UMAP generates clusters that correspond to differences between genomic regions that were previously annotated in other ways, even though this prior information was not included in the training data. For example, SINE and LINE repeats form a separate cluster in the map. These regions were strongly enriched for Z-DNA formation when cells were treated with the curaxin CBL0137 that derepresses these regions by altering chromatin structure, enabling their transcription (light blue in Fig. 3A)¹⁵. We did not find enrichment of the ALU family of human SINE in our analysis, even though they contain Z-DNA forming sequences¹. Most likely they were absent from our training set as these sequences are routinely cleaned out by the pipeline used for processing experimental data.

The UMAP clusters observed are associated with different feature vectors (Fig. 4, Supplementary Table 8). Feature importance analysis of these vectors allows their assignment to processes involved in embryonic development and morphogenesis, RNA Polymerase II and RNA Polymerase III dependent transcription, heterochromatin formation associated with LINEs, chromatin binding complexes, negative regulation of transcription and cellular component organization. The LINE cluster in mouse genome due to CBL0137 is enriched for MECP2 (methyl-CpG binding protein 2), a reader of DNA methylation that is a feature of the heterochromatin present in intergenic regions.

GO-analysis of orthologs with Z-flipons from the common cluster (Supplementary Table 9), revealed that genes with conserved Z-flipons are enriched in the development and differentiation processes. Interestingly, Z-flipons with a strict purine-pyrimidine alternation were clustered together with morphogenesis related genes that are evolutionarily old. This alignment suggests the early evolutionary selection of Z-flipons composed of GT- or GC-repeats. Notably, the GO enrichment we observe mirrors that found for Z-flipons bound for conserved microRNAs that produce phenotypic variation during development²⁹.

Many of the enriched genes are involved in neurogenesis, and in particular, are related to synapse organization and function (Supplementary Table 9). Examples of flipons in the human and mouse Wnt Family Member 5A (WNT5A) gene, which is involved in the presynapse assembly pathway, are given in Supplementary Fig. 4. This gene harbors many Z-flipons that are detected by three methods (DeepZ, Z-DNABERT, chemical footprinting), and which are located in 5'UTR, near splice-sites, in exons, at alternative promoters. Such analyses independently confirmed that Z-flipons have conserved functional roles in both genomes and are marked by shared omic features.

Functional Z-flipons at promoters is consistent with a role in transcription

We found significant enrichment of conserved Z-flipons in alternative and bidirectional promoters. We then explored in more detail cases where DeepZ predicted Z-DNA regions were detected in both human and mouse orthologs by two other methods KEx and the Z-DNABERT. The results we show in these figures are representative of findings genome-wide. We supply tracks to enable other researchers to evaluate genome regions that they study (Supplemental Data 1–4).

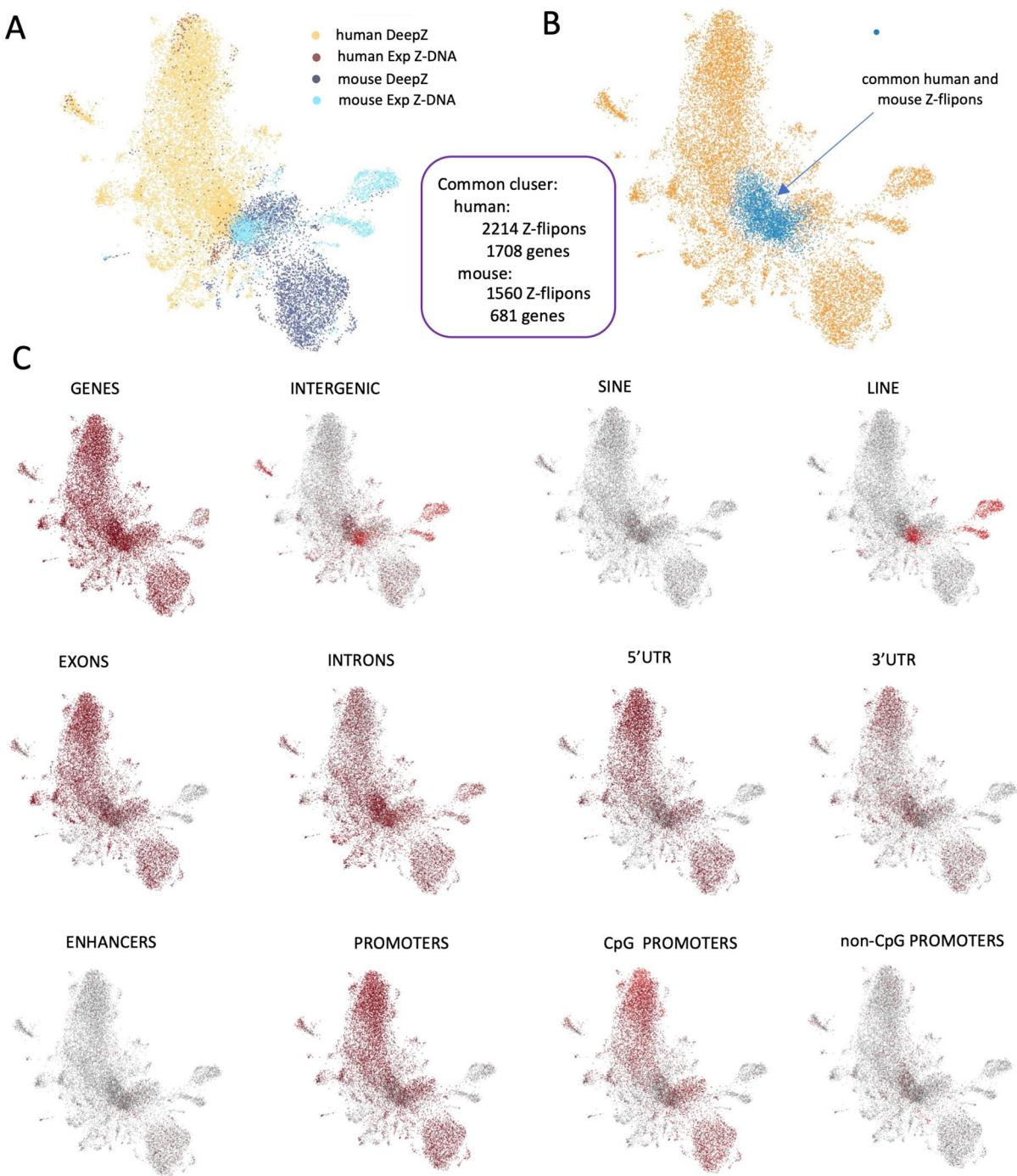


Figure 3. UMAP clustering of Z-flipons based on vectors of common omics features. (A) UMAP clustering of experimental and DeepZ predicted Z-flipons in human and mouse genomes. (B) Cluster of common human and mouse genome. (C) Distribution of Z-flipons over genomic regions, regulatory elements, and transposons.

Z-flipons at alternative promoters

Many genes have alternative promoters, but the mechanism of activation and how they control gene expression is not clear. Our analysis revealed that, depending on the size of the upstream region, the Z-flipons predicted by DeepZ show a 30-fold enrichment within 10-bp and sevenfold within 1 kb from TSS in alternative promoters ($p < 0.001$, permutation test) (Supplementary Table 10). Here we highlight different cases with alternative promoters that are very close to transcription start sites or located at a distance from the main promoter. The alternative PLEKHA7 promoters overlap with Z-flipons detected by three different methods (DeepZ, KEx and ZDNABERT) (Fig. 5A,B). The alternative promoters in both genomes have many noticeable columns of omics signals (Fig. 5C,D). Because DeepZ was trained on a broad ChIP-seq data, the width of DeepZ prediction is comparable to ChIP-seq peak widths and does not have the higher resolution possible with KEx and Z-DNABERT.

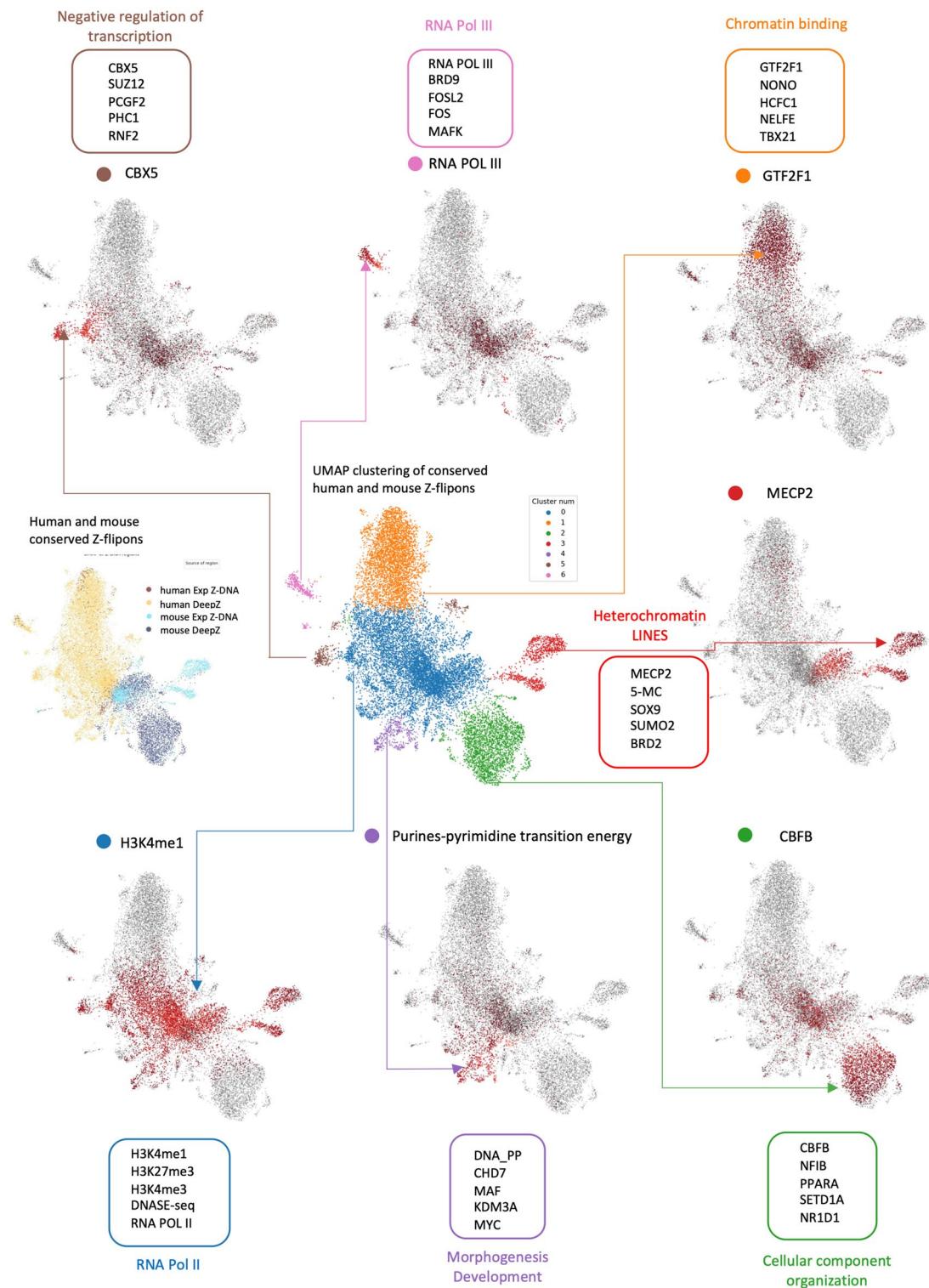


Figure 4. UMAP clusters of human and mouse Z-flipons with one top marker feature highlighted for each cluster. The first 5 marker features are given in the boxes.

The right column of omics features at the main promoter (right column in Fig. 5C,D) has Z-flipons confirmed by three methods. DeepZ predicted two Z-flipons around this promoter based on two peaks of omics features corresponding to tandem promoters. There are species differences. For example, the Human Z-flipon present in a distant alternative promoter (left column at Figs. 5C,D) is composed of CA-repeats (GT-repeats at the opposite strand), while in mouse the equivalent Z-flipon lacks this sequence. The splicing graph (Fig. 5A,B) also shows that there are alternative promoters nearby start sites for this gene that are not conserved between human and mouse.

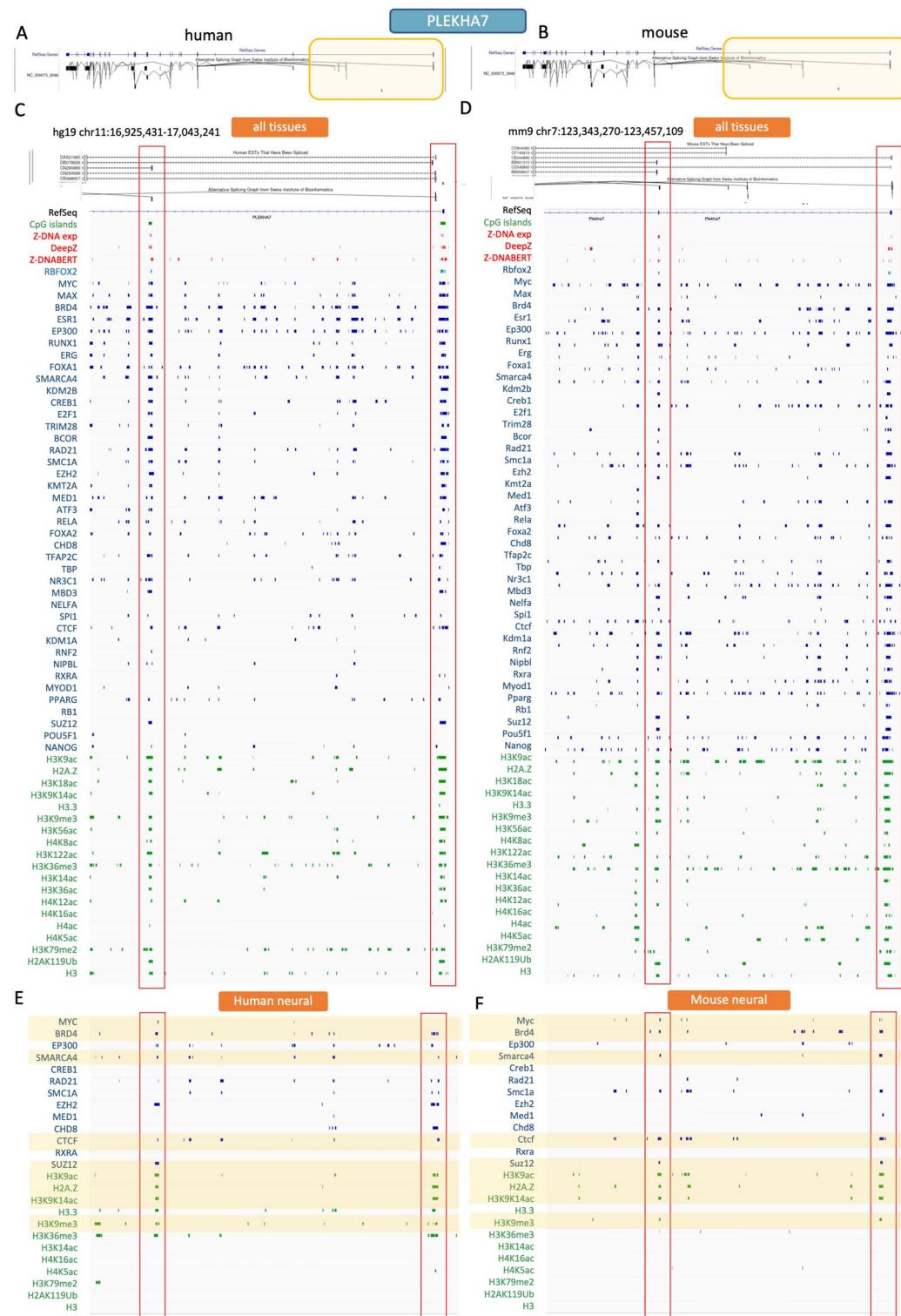


Figure 5. Z-flipons at alternative promoters of PLEKHA7 orthologs in human and mouse genome. (A,B) Alternative splicing graph for PLEKHA7 in (A) human and (B) mouse genomes. (C,D) Region of PLEKHA7 in human (C) and mouse (D) genomes with Z-flipon signals detected by three methods—DeepZ, KEx, Z-DNABERT, and signals from omics features enriched in Z-flipons both in human and mouse genomes. Omics features are aggregated signals from all tissues as they were used in DeepZ model. (E,F) Selected omics features for human and mouse genome for neural tissue type. Features that are common in human and mouse are highlighted in yellow.

Since DeepZ were trained on aggregated experimental signals from different tissues, the regions mapped to the genome do not capture tissue-specific or species-specific differences. We find many examples where feature distributions differ between mouse and human. An example from neural tissue is presented in Fig. 5E,F. We can see that despite the differences, many proteins and histone marks are common to both species. These include MYC, BRD4, SMARCA4, CTCF and SUZ12—all regulators involved in chromatin organization. Presence of both H3K9ac and H3K9me3 indicates that this promoter is being turned on and turned off in neural tissue, and this is conserved between human and mouse.

We also can observe that the complexity of marks for tissue specific gene expression is greater than suggested by viewing our aggregated maps, reflecting differences in genomic structure at a particular locus in each lineage and can differ by species (Supplementary Fig. 4). For the TMEM51-AS1 transcript, we see that the pattern for blood tissue type differs from other tissues, with a bidirectional promoter present in mouse but not in humans. Such a variation is not apparent from the “All Tissues” track (Supplementary Fig. 4G). Similar results were obtained for data generated from liver tissue (Supplementary Fig. 5). The patterns observed in blood tissues are complex with binding of different TF occurring at one or two of the two, three, or even four separated peaks. The variable coding of bidirectional and alternative promoters allows gene expression to change in a tissue- and context-specific manner through their effects on Z-flipon conformation.

Mapping of marks around transcription start sites at bidirectional promoters

We made anchor plots for human bidirectional promoters to further explore the connection between Z-flipons, histone marks and transcription start sites (Fig. 6). Bidirectional promoters are of interest as the energy available to flip sequences to Z-DNA is highest when RNA polymerases transcribe each DNA strand in an opposite direction. As a result, negative supercoiling that is generated 5' of each polymerase accumulates in the region between the two TSS. The prediction then is that active marks of transcription should be highest in these regions, and that these marks should be enhanced in those promoters containing Z-DNA forming elements compared to the non-Z-DNA set. We see such an outcome in the anchor plots centered on bidirectional promoters. The upper part of each panel shows promoters with predicted Z-DNA elements and the lower part is for non-Z-DNA promoters. The difference in the vertical length of each set is because Z-DNA containing promoters are less frequent than non-Z-DNA promoters. The relative proportion of bidirectional promoters that are enriched for a particular mark is higher for promoters with Z-flipons than for those promoters without, with permutation p-values < 0.001 (as given in Fig. 2). For example, H2A.Z marks are present at most Z-DNA containing promoters, but a much smaller fraction of non-Z-DNA promoters. Overall, the analysis revealed 130-fold enrichment within 500-bp and 85-fold within 1 kb from TSS of DeepZ-predicted Z-flipons in bidirectional promoters ($p < 0.001$, permutation test).

When examining particular examples of bidirectional promoters, we observed cases where Z-flipons were detected with three methods (KEx, DeepZ, Z-DNABERT) (Supplementary Table 10). The Z-DNA forming segments align best with active HM at each promoter. Marks for TF and CC are enriched either side of the bidirectional promoter region, as seen for the bidirectional promoter of TMEM51-AS1 (Supplementary Fig. 5) and of BRCA1-NBR2 (Supplementary Fig. 6).

Association of Z-flipons with transcription reinitiation rate

How then do Z-flipons affect gene expression? Early models proposing sequence- and Z-DNA- specific TF currently lack experimental support³⁰. To explore this question further, we explored the association of Z-flipons with parameters that measure transcription kinetics, using the data generously provided by the Cramer laboratory³¹. We tested the initialization rate per cell, the elongation rate per cell, and polymerase pause duration, exploring all human regions with DeepZ and Z-DNABERT predictions (Fig. 7 and Supplementary Table 11). We note that the initiation frequency is actually a measure of transcription reinitiation rather than initiation, which depends on a different set of pioneering transcription factors to activate gene expression³².

Our analysis of these regions revealed significant differences in the distributions of reinitiation frequencies for promoters with conserved DeepZ predicted Z-flipons, as compared to promoters without Z-flipons ($p = 5.62e-5$, Kolmogorov-Smirnov test, Fig. 8A). The reinitiation frequency trends higher for conserved Z-flipon promoters. We also see a similar difference in reinitiation for conserved non-CpG promoters ($p = 1.76e-2$, Kolmogorov-Smirnov test, Fig. 8B) and for conserved CpG-promoters with Z-flipons ($p = 1.2e-4$, Kolmogorov-Smirnov test, Fig. 8C). There was no difference in pause duration or elongation rate associated with Z-flipons.

We verified the results with Z-DNABERT predictions that are based on an independent algorithm and an independent training set ($p = 2.35e-2$, Kolmogorov-Smirnov test, Fig. 8D). We also confirmed the results in a set with overlapping DeepZ and Z-DNABERT predicted Z-flipons that we compared to promoters lacking Z-flipons ($p = 1.52e-3$, Kolmogorov-Smirnov test, Fig. 8E) and to any promoter excluded from the overlapping DeepZ and Z-DNABERT set ($p = 7.70e-4$, Kolmogorov-Smirnov test, Fig. 8F). We extended the analysis by using statistical tests with different underlying assumptions, applying the t-test, median-test and variance-test to all groups in Fig. 8. With all these approaches we confirmed the significance of the relationship between conserved DeepZ predicted Z-flipons and higher reinitiation rate (Supplementary Fig. 7).

We also examined DeepZ predictions by plotting Z-DNABERT scores for each promoter against the reinitiation rate, dividing the graph into four quadrants (Supplementary Fig. 8). In quadrant 1, promoters have low reinitiation rate, while in quadrant 4 they have low Z-DNABERT-scores. In quadrant 3, both measures are low, while in quadrant 2 both Z-DNABERT scores and reinitiation rates are high. We calculated the ratio of counts in quadrant 2 relative to quadrant 4 for both DeepZ and non-DeepZ promoters. Consistent with our previous analyses, the ratio was significantly higher for DeepZ promoters (Supplementary Fig. 8, $p < 0.00001$). The Z-DNABERT scores in quadrant 2 were within a relatively restricted band, with a slight upwards trend as reinitiation rate moves higher. The lack of correlation between these measures suggests that Z-DNA formation

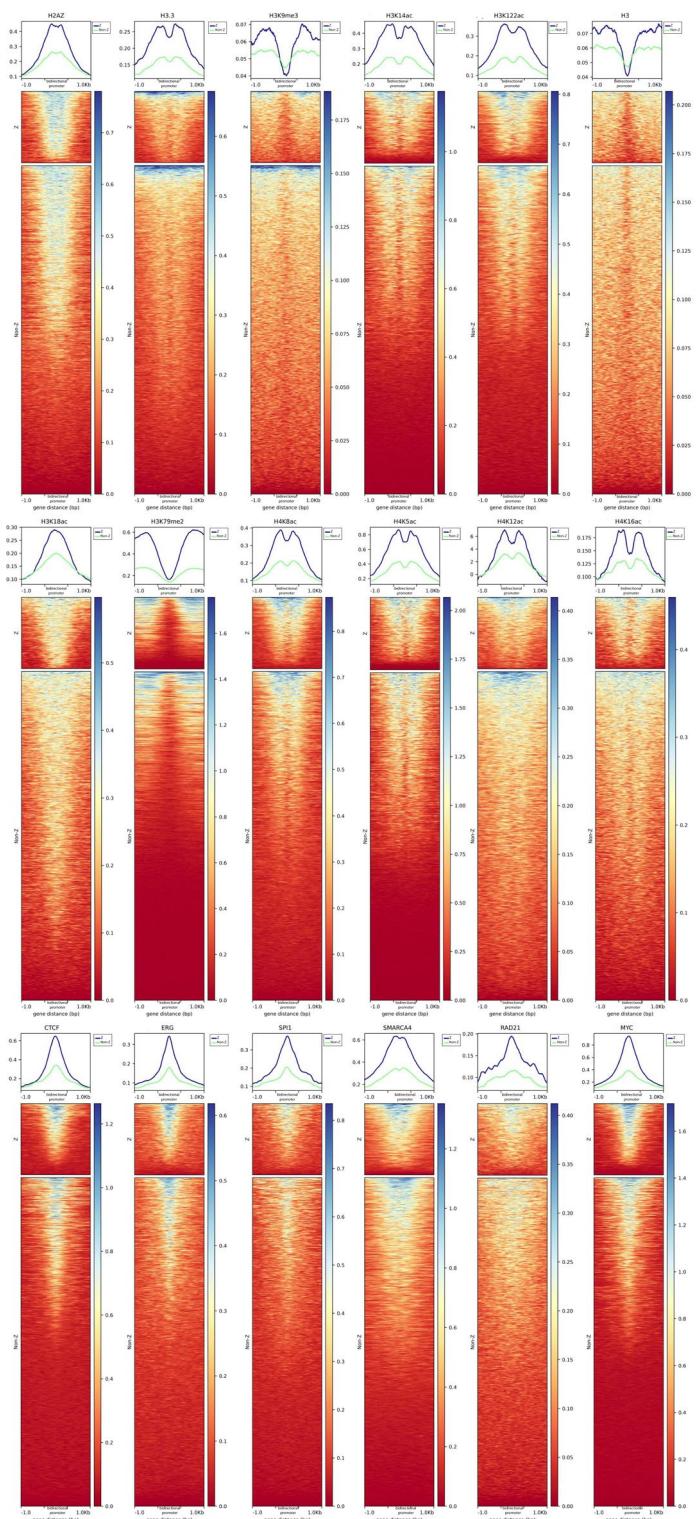


Figure 6. Anchor maps of features colocalization around transcription start sites of bidirectional promoters with and without Z-flipons.

in itself does not solely determine the overall transcription rate. The results are consistent with a mechanism where Z-DNA formation is under selection to optimize reinitiation by resetting the promoter for reuse, with other factors determining pause release and reformation of pre-initiation complexes.

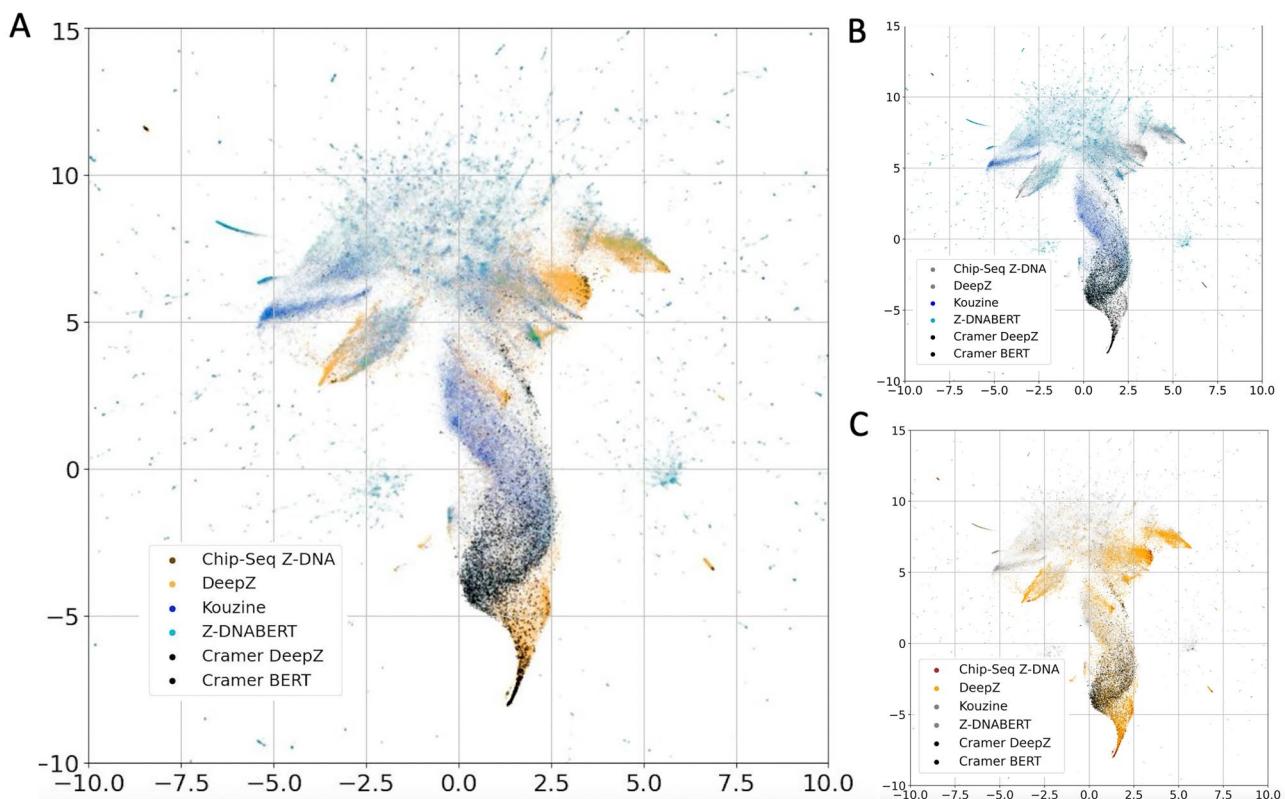


Figure 7. (A) UMAP clusters of human and mouse Z-flipons predicted by different experimental and in silico methods. (B) Deep-Z is highlighted over Z-DNABERT to show the colocalization of predictions. (C) Z-DNABERT is highlighted over Deep-Z to show the colocalization of predictions. The black dots represent the genes from regions conserved between mouse and human genomes that have transcription data provided by the Cramer laboratory and that are analyzed in Fig. 8.

The plot of non-DeepZ promoters also reveals that there are functional Z-flipons, according to their Z-DNABERT scores, that are not classified as such by DeepZ. This outcome could reflect the use of threshold 3 for DeepZ to optimize the F1 metric. While we reduced false positive calls with this choice, we also have increased the false negative rate. The result also could arise from the various limitations in the training sets available for calibrating DeepZ in which some sets of tissue-specific promoters were not well represented in the data used to fully specify the model, or a set of promoters that were not active at the time the ChIP-seq experiments were performed.

The proposed mechanism of action for how Z-DNA modulates transcription reinitiation is presented in Fig. 8G. Here Z-DNA captures the negative supercoiling generated 5' to an elongating polymerase³³. The accumulated energy can then be used to either turn a promoter on or off by powering the assembly of the complexes required. Either the negatively supercoiled DNA can be absorbed by chromatin remodelers as they wrap DNA around a nucleosome to suppress transcription, or by mediator proteins that promote the reassembly of preinitiation complexes³⁴. These outcomes can be further tuned by topoisomerases that regulate local supercoiling³⁵ and potentially by small RNAs that alter promoter conformation³⁶. Our data shows enrichment in promoters of topoisomerase I (odds ratio of 2.25 in human and 5.9 in mouse, $p < 0.01$) and topoisomerase 2B (odds ratio of 9.2 in human and 8.18 in mouse, $p < 0.0001$) relative to promoters without Z-flipons (Supplementary Table 7).

To further examine the role of Z-DNA in reinitiation, we checked for GTF2E2 localization to Z-flipons. The GTF2E2 data was not available at the time DeepZ was trained, so it provides information independent of that used to calibrate the model. The GTF2E2 protein product Transcription Factor E (TFE), subunit B, is of interest given previous yeast experimentation showing the essential role of TFE in transcriptional reinitiation³⁷, and given that transcription-induced Z-DNA formation in yeast has been shown to occur in a promoter-specific fashion³⁸. We observed odds ratio of 23 for human (p -value $< 2.2e-16$, Fisher's exact test, Fig. 8H) for GTF2E2 in DeepZ and Z-DNABERT promoters, a result consistent with involvement of Z-DNA in reinitiation of transcription. The finding is also supported by evolutionary analyses of the relationship between TFE and the Zα domain³⁹, one that awaits experimental validation.

Discussion

Z-flipons and increased transcription reinitiation rate

Here we found evidence that conserved Z-flipons are associated with increased transcription reinitiation rates, regardless of whether or not promoters contained CpG islands. This finding is in line with a mechanism of transcriptional regulation based on chromatin conformation, as proposed in³² where Z-flipons dynamically

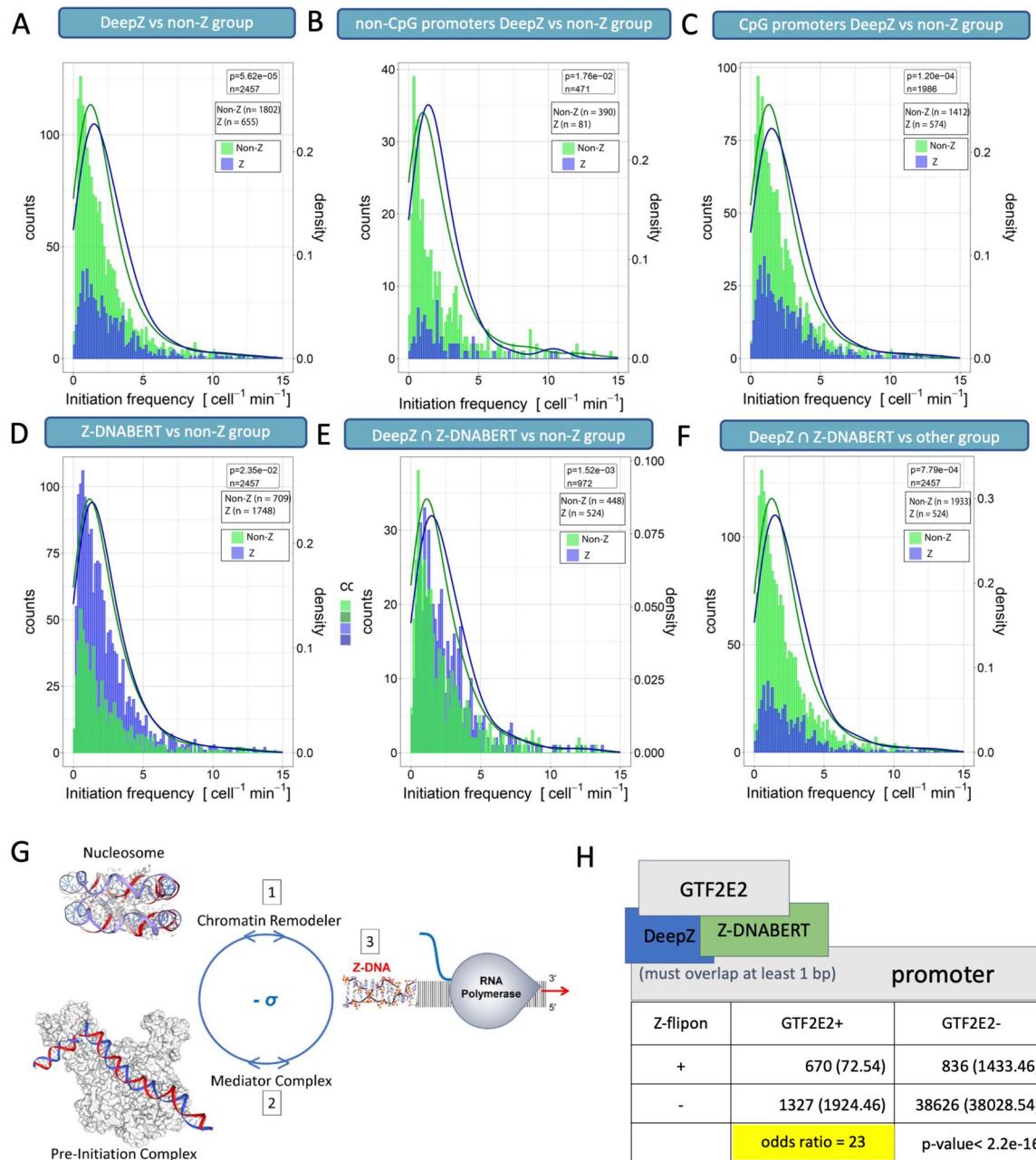


Figure 8. (A,B) Transcription initiation rate for different types of promoters with or without Z-flipons. (A) Transcription initiation rate for promoters with conserved DeepZ-predicted Z-flipons is higher than the rate for promoters without Z-flipons. (B) Transcription initiation rate for promoters with conserved Z-flipons, but lacking CpG islands, is higher than rate for promoters that lack Z-flipons and CpG islands. (C) Transcription initiation rate for CpG-promoters with Z-flipons is higher than for CpG promoters without Z-flipons. (D) Transcription initiation rate for promoters with conserved Z-DNABERT-predicted Z-flipons is higher than the rate for promoters without predicted Z-flipons. (E) Transcription initiation rate for promoters with Z-flipons identified by both DEEP-Z and Z-DNABERT is higher than for all other groups. (F) Transcription initiation rate for promoters containing Z-flipons identified by both DEEP-Z and Z-DNABERT is higher compared to those lacking any predicted Z-DNA forming elements. (G) Schematic representation of mechanism of action of Z-flipons in resetting the transcription initiation complex. The negative supercoiling generated by the RNA polymerase can be used to either wrap DNA around histones or for the assembly of preinitiation complexes. The outcome depends on the action of chromatin remodelers and mediator proteins.

actuate the reset of promoters after each round of transcription. In this scenario, the negative supercoiling captured by Z-DNA offsets the positive supercoiling induced by the preinitiation complex (PIC) used to open up the transcription bubble for docking of the RNA polymerase. The loss of positive supercoiling destabilizes the

PIC, allowing either mediator proteins to reform the PIC or chromatin remodelers to bookmark the promoter for later use.

Other transcription kinetic parameters, such as polymerase pause duration and transcription elongation rates, were not impacted by the presence of Z-flipons. In studies by the Cramer lab, other factors were found to influence these parameters. For example, a role for CDK9 in the regulation of pause duration was demonstrated using inhibitors of this enzyme, while elongation rates were found to be gene specific, likely reflecting the differences in their exon and intron compositions²⁷.

Chromatin organization around conserved Z-flipons

Conserved Z-flipons are highly enriched with proteins engaged in chromatin organization. They are associated with chromatin remodelers that rewrite the histone code by changing histone marks such as acetylation, methylation and ubiquitination. This constant remodeling of histones absorbs and releases negative supercoiling and is also moderated by topoisomerases. Highly enriched chromatin remodelers such as SMARCA4 alter Z-flipon conformation by recruiting topoisomerase 1 and further help by correcting folding DNA into higher order complexes⁴⁰.

We show that both positive and negative transcriptional outcomes are enriched in promoters with Z-flipons. Z-flipons in active promoters are characterized by an active profile (Figs. 3 and 4). The repressive effects are evidenced by Z-flipons in LINE and SINE that are associated with negative chromatin marks. These modifications are established early in development, during a stage where widespread transcription occurs throughout the genome following the removal of parental epigenetic marks. Controlling the ability of a Z-flipons to form Z-DNA is a quite general mechanism for rapidly switching any gene promoter “on” or “off”.

Clustering of conserved Z-flipons connects their role in transcriptional regulation with chromatin organization

The findings here add to previous studies of the role Z-flipons play in cellular biology. It has been elegantly shown that Z-DNA and Z-RNA trigger necroptosis by binding with ZBP1, and that ADAR1, which also has Z-DNA binding domain, acts as a negative regulator of innate immune response^{1,16–18}. Here we show that Z-flipons affect transcription reinitiation rates and demonstrate the association with chromatin remodeling. We extend our findings by clustering features in UMAP and detect colocalization of Z-flipons with features characteristic of active Pol II and Pol III transcription. The enrichment of Z-flipons in conserved promoters is statistically significant for promoters, even those without CpG islands (Supplementary Table 7). The findings support a specific role of Z-flipons in transcriptional regulation

Z-flipons composed of simple purine-pyrimidine alternating repeats formed a separate cluster that also included TF involved in developmental processes, morphogenesis^{29,41} and in neurogenesis. The finding suggests that evolutionary selection for Z-flipons was favored in genes that respond rapidly to changing contexts. The scheme also allows systems to learn by varying the reset by Z-flipons of promoter bound complexes⁴². The level of gene expression is then set by the stability of the PIC: high stability of the complex leads to lower rates of initiation and rapid termination of gene expression once the inducing stimulus is lost. In contrast, lower stability of the PIC enables a rapid reset of promoters resulting in higher levels of gene expression. Here variant histones like H2A.Zac play a part. They are enriched around conserved Z-flipons both in human and mouse, and are known to have an important role in neuronal activity and differentiation^{43,44} and in trophoblast differentiation of human embryonic stem cells⁴⁵. Interestingly, the H2A.Z histone variant has a lysine/alanine dipeptide repeat that can dock to Z-DNA⁴⁶. Of further note is our previous mapping of conserved miRNA seed matches to promoter flipons in developmental genes, providing evidence that regulation of flipon shape by small RNAs provides an additional way to regulate gene expression. This mechanism allows flipon conformations to be set by small RNAs that bind in a sequence-specific manner, the engagement of structure-specific proteins then boot-straps embryogenic development by setting the chromatin states of promoters and enhancers²⁹. The canvas so formed is then further elaborated by the protein transcription factors to drive tissue differentiation.

Consistent with the use of Z-flipons to promote a variety of epigenetic modifications is the clustering of genomic regions based on DeepZ predictions (Figs. 3 and 4). The clusters of functional Z-flipons are associated with a variety of epigenetic marks that are involved with both the positive regulation of transcription (markers H3K4me1, RNA Pol II) and its negative modulation (repressive chromatin complexes) (Fig. 4). In these processes, two different mechanisms of action are involved: positive regulation is related to initiating and resetting transcription complexes, while negative regulation is effectuated via attracting proteins promoting heterochromatin formation with proteins like CBX5 and the set of polycomb repression complexes composed of SUZ12, RNF, PCGF2 and RNF2. What both positive and negative outcomes potentially have in common is the use of Z-DNA formation to power the reset of chromatin structure from one state to the other (Fig. 8G). Overall, the findings support a role for Z-flipons in regulating transcription during development, especially at times when cell cycles in the epiblast are as short as 5.5 h and when transitions from one cell state to another require rapid changes in gene expression⁴⁷.

Z-flipons enrichment at alternative and bidirectional promoters

As it was shown by us and others based on sequence analysis alone^{19–21,25,28,29,48–57}, Z-flipons are enriched in promoters. Here we find Z-flipon enrichment in conserved human-mouse sequences is even higher. We also find that there is often a common pattern of omics features that colocalize with the set of conserved Z-flipons within bidirectional promoters.

Around 52% of human RefSeq genes are subject to regulation by alternative promoters⁵⁸. They can be located in the first 10–50 bp of the main promoter or at a distance of more than 1 Mb. The distant, upstream,

alternative promoters are marked with the same pattern of omics features as those downstream. However, when there are many downstream alternative promoters close to each other, promoter choice may depend on other mechanisms. In such situations, transcription of a 5' promoter can disrupt the use of a 3' promoter by transcriptional interference. In these situations, Z-flipons (and other types of flipons as well) can play a role in enabling transcription from the 3' promoter by either preventing transcriptional read through from the 5' promoter, or by switching transcription from one promoter to the other, as previously described³².

The role for Z-DNA contrasts with the models for regulation of TATAA box promoters by sequence-specific transcription factors that allow for fine tuning of expression in a tissue-dependent manner. The two schemes may in fact complement each other: transcription factors play an important role in the initial pioneering round of gene expression that involves remodeling of the chromatin laid down earlier in development. Subsequently, the Z-DNA produced by the transcribing RNA polymerase then enables regulation of the later rounds of transcription reinitiation by the mediator complex (Fig. 8G). The transcription rate then is influenced by the energetic cost of the promoter reset. Factors such as the strength of PIC binding to the promoter and the duration of polymerase pausing would affect this outcome. The Z-DNABERT scores for the flipons involved in this process appear bounded (Supplementary Fig. 8), indicating that the sequence variability of Z-flipon actuators is limited, and subject to natural selection³².

Limitations of experimental whole-genome approaches to detect Z-DNA

Whole-genome experimental methods to detect Z-DNA are few and each has its own limitations. ChIP-seq is challenging with stringent controls, like those used by ENCODE, necessary to counter the inherent noisiness of the approach. Expression of the Z-DNA specific Zα domain in cells as a capture reagent also runs the risk of artifactually inducing Z-DNA formation, either *in vivo* or during *in vitro* processing steps. The small dataset generated by this approach reflects the steps taken to minimize false positive results. Other approaches based on Z-DNA antibodies are also performed on cell lysates, not intact cells. As the antibodies diffuse in, other proteins that restrain negative supercoiling diffuse out. The width of peaks obtained in these approaches is broad and exceeds the length of Z-DNA regions by more than 5 times. CUT&Tag approaches that are now replacing ChIP-seq experiments have not yet been performed for Z-DNA according to our knowledge.

KEx is a nucleotide modification method that is performed in intact cells with a 70 s exposure to the potassium permanganate reagent. It detects single-stranded regions, and then Z-DNA is identified by additional computational methods. Similar to KEx in principle is KAS-seq, which uses N3-kethoxal to chemically modify guanine in regions of single-stranded DNA. Both KEx and KAS-seq can be combined with computation algorithms to interrogate the conformation of flipons at nucleotide resolution. These chemical footprinting approaches provide a snapshot of cell state that informs on the results obtained using other techniques. Further, just because sequences can flip to Z-DNA *in vitro* does not mean that they will do so *in vivo*. Together, experimental and computational methods enable insights that are not supplied when either is used alone. Multiple orthogonal approaches like the ones we have used here are necessary to confidently conclude that a Z-flipon is functional within a cell.

Directions of future research

DeepZ is trained on aggregated omics signals from all tissues. With this approach we find signals that are common for all tissues. In the presented study we refined the aggregated signals by selecting only Z-flipons from conserved human-mouse regions. The DeepZ model captured common conserved Z-flipons together with their common conserved omics features. Since alternative promoter usage and splicing are tissue specific, it is of interest to train DeepZ only on tissue-specific omics features and define sets of tissue-specific Z-flipons. That task is the subject of future research.

Also, the data sets used for training were collected from bulk cell cultures. Gene regulation can differ by cell cycle, requiring single cell sequencing to examine how Z-flipons function during the different phases. Also, similar studies performed during embryonic development will allow the exploration of imprinted genes, whose expression depends on whether a chromosome is inherited from the mother or father. Similar to tissue specificity, it is of great importance to compare different developmental stages, and train DeepZ on omics data obtained at various time points from both embryonic and extraembryonic tissues. A recently described method to characterize imprinted genes may also generate relevant data. The study involved single-cell sequencing of cell lines obtained from parent:child trios⁵⁹. This approach can easily be extended by potassium permanganate chemical footprinting of live cells from each trio²². Collectively, the newly produced data will help improve our understanding of the role Z-DNA plays in imprinting.

Conclusion

Here we present evidence for the regulation of transcription reinitiation rates by conserved Z-flipons in both CpG- and non-CpG-promoters. In this model, Z-flipons actuate a reset of promoters. Z-flipons play a dynamic role by capturing and releasing the energy necessary to prepare promoters for another round of transcriptional elongation³². We revealed common patterns of transcription factors, regulators, and epigenetic marks defining conserved Z-flipons at bidirectional and alternative promoters in actively transcribed genes. We also revealed the existence of orthologous omics features associated with Z-flipons that are conserved during the evolution of both human and mouse. The strict purine-pyrimidine repeats that favor Z-DNA formation are enriched in promoters, supporting the view that they remain under selection. These Z-flipons are significantly overrepresented in genes with Gene Ontology annotations for developmental processes and morphogenesis. In other less conserved regions of the genome, including those populated by LINE and SINE retroelements, DeepZ is able to identify differences that map to their functional roles without explicit training of the algorithm to perform that

task. Flipons thereby act as genetic elements that allow the switch of promoters “on” of “off”. They dynamically change shape to modulate gene expression. By doing so, flipons alter the phenotypic pliability of cells, tissues, individuals and species.

Methods

Deep Z approach

We applied here the DeepZ approach as described in detail in²⁵. As input data DeepZ accepts matrices of size $L \times W$, where L equals to 5000 bp and W represents the number of features from omics data. Input features consist of two major groups: the first group incorporates information on DNA sequence, it includes one-hot-encoded DNA sequence of length L , Boolean features representing different simple repeats, and energy transitions to switch from B- to Z-conformation for each dinucleotide as it is implemented in Z-Hunt^{26,60}. The second group of features comprise information from omics data and includes histone marks (HM), DNase I hypersensitive sites (DNase-Seq), transcription factors (TF) and RNA-polymerase (RNAP). The omics data is taken from chip-atlas.org⁶¹. Since different sets of experiments are available for the human and mouse genomes, we took the intersection of features that are present in both genomes (Fig. 1C). The distribution of feature groups can be found in Fig. 1E. Each feature is linearly scaled to the interval [0, 1]. Full list of features is given in Supplementary Table 1. As a training set we took ChIP-seq data from Zhang et al.¹⁵ for mouse genome and ChIP-seq data from Shin et al.²⁴ for human genome.

For experimental Z-DNA ChIP-seq data, both mouse and human genomes were divided into equal segments of the length L . Segments from blacklisted regions⁶² were excluded. Similarly, segments with more than a half of the nucleotides undefined in the genomic assembly were also excluded. All the segments were marked based on the presence of Z-DNA sites as 1 and 0. All Z-DNA regions were assigned to a positive class. The negative class was composed from non-Z-DNA segments 20-fold in size. The resulting segments were stratified into five folds where 4 folds were included in a training set, and the remaining fold formed a test set. We took the best RNN-based model architecture as in²⁵. All training parameters could be found in source code.

We trained the deep neural network described in²⁵. ROC-, PR-, and F-curves of the DeepZ model performances are presented in Supplementary Fig. 1. PR-curve where the precision is low as it is applied to the whole genome without annotation of Z-DNA forming sequences. To overcome this issue, we validated DeepZ by overlap with Z-DNABERT predictions that are based on experimental data is generalized using a transformer language model algorithm. The two approaches are independent of each other, using different algorithms trained on orthogonal datasets. We present the metrics for six different thresholds that are indicated in Fig. 1A and defined by the score quantile ($1-10^{-i}$, $i = 1, 2, 3, 4, 5, 6$). For this study we present results for threshold 3 as it has the optimum value for precision and recall as determined using the F1 metric.

Whole-genome annotation with Z-DNA regions

The trained DeepZ model was used to generate Z-DNA whole-genome annotation. The model predicted probability to form Z-DNA for every segment from all five folds. We generated the annotation with seven thresholds where i -th threshold equals to the quantile $1-10^{-i}$.

Ablation analysis

A total of 60 separate experiments were conducted to evaluate the efficacy of different feature groups. Each experiment focused on a specific subset of features, such as DNA-based features, dinucleotide structural features, RNA polymerase and transcription factor binding sites, histone marks, open chromatin, and DNA methylation. Ratios of the training and testing sets were consistent with those used in for the large DeepZ model. For each feature group, five different Gradient boosting models⁶³ were trained across five different fold-groups, resulting in a total of $5 \times 6 \times 5 = 60$ models. Due to the extensive volume of data, the number of training and testing intervals had to be significantly reduced—more than tenfold—to accommodate RAM limitations. Consequently, the F1 scores presented are not reliable for the absolute performance comparisons; we used them only for relative comparisons within this study.

Conserved Z-DNA regions

Conserved Z-flipons were identified by overlapping DeepZ predicted Z-flipons with vertebrate conservation tracks from UCSC genome browser: Vertebrate Multiz Alignment & Conservation (100 Species) for human and Vertebrate Multiz Alignment & Conservation (60 Species) for mouse.

Enrichment analysis for omics features

Statistical significance of Z-flipon enrichment/depletion with transcription factors or histone marks was calculated using permutation test ($n = 1000$) where the same number of Z-DNA regions was randomly simulated 1000 times and each time the overlap of Z-flipons with a feature was calculated. Since conserved regions tend to be in regulatory regions as are Z-DNA sequences, it was necessary to evaluate the statistical significance of the association of an epigenetic feature with Z-DNA forming sequences, using the null hypothesis that the distributions in regulatory regions are independent of each other. Promoter regions were defined as regions of 1000 bp upstream and 200 bp downstream of TSS. TSS locations were based on Gencode v44⁶⁴ annotation. UCSC CpG islands⁶⁵ (CpGI) track was used for CpG-/non-CpG-promoter classification. Promoter was considered CpG if it had an overlap with the CpGI (CpG Island) annotation.

Transcription kinetics parameters analysis

For each transcript from³¹ the CpG/non-CpG label was assigned in a way described above. P-values were calculated for 6 different DeepZ thresholds (all predictions and conservative) using KS-test. After that they were adjusted for multiple testing (FDR). Reinitiation rate plots were generated via custom R scripts.

UMAP clustering of Z-flipons based on omics vectors

In our study, we aimed to visualize all predicted Z-DNA regions using a UMAP two-dimensional embedding. For each interval, we extracted the omics representation and then computed the intensity (number of non-zero base pairs in every interval) of every feature within these intervals. Subsequently, we normalized the mean of each feature across each organism. The vectors derived from these calculations were utilized as feature vectors for the subsequent UMAP analysis. For the UMAP projection, we employed the Manhattan metric and selected 55 nearest neighbors to ensure a connected representation using a seed value of 42. All other parameters were set to their default values.

For segmenting the Z-DNA regions represented in the UMAP, we applied the Leiden clustering method based on signal intensities. A resolution parameter of 0.25 was chosen for this purpose. The same approach was used to identify the central cluster in the UMAP, but with a higher resolution setting of 1.

To identify distinguishing features of a cluster, we compared the mean and variance of every feature both within and outside each cluster, utilizing Z-statistics. A high Z-statistic indicated that a feature was representative of that particular cluster.

Data availability

All codes and data used for deep learning models for mouse and human genome as well as generated whole-genome DeepZ annotations are available at https://github.com/Nazar1997/DeepZ_MM.

Received: 5 March 2024; Accepted: 23 July 2024

Published online: 01 August 2024

References

- Herbert, A. Z-DNA and Z-RNA in human disease. *Commun. Biol.* **2**, 1–10 (2019).
- Herbert, A. A genetic instruction code based on DNA conformation. *Trends Genet.* **35**, 887–890 (2019).
- Herbert, A. ALU non-B-DNA conformations, flipons, binary codes and evolution. *R Soc. Open Sci.* **7**, 200222. <https://doi.org/10.1098/rsos.200222> (2020).
- Kaiser, W. J., Upton, J. W. & Mocarski, E. S. Receptor-interacting protein homotypic interaction motif-dependent control of NF-κB activation via the DNA-dependent activator of IFN regulatory factors. *J. Immunol.* **181**, 6427–6434 (2008).
- Lin, J. *et al.* RIPK1 counteracts ZBP1-mediated necroptosis to inhibit inflammation. *Nature* **540**, 124–128 (2016).
- Maelfait, J. *et al.* Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *EMBO J.* **36**, 2529–2543 (2017).
- Nakahama, T. *et al.* Mutations in the adenosine deaminase ADAR1 that prevent endogenous Z-RNA binding induce Aicardi-Goutières-syndrome-like encephalopathy. *Immunity* **54**, 1976–1988.e1977 (2021).
- Newton, K. *et al.* RIPK1 inhibits ZBP1-driven necroptosis during development. *Nature* **540**, 129–133 (2016).
- Pham, T. H., Kwon, K. M., Kim, Y.-E., Kim, K. K. & Ahn, J.-H. DNA sensing-independent inhibition of herpes simplex virus 1 replication by DAI/ZBP1. *J. Virol.* **87**, 3076–3086 (2013).
- Rebsamen, M. *et al.* DAI/ZBP1 recruits RIP1 and RIP3 through RIP homotypic interaction motifs to activate NF-κB. *EMBO Rep.* **10**, 916–922 (2009).
- Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
- Mo, S. *et al.* NeurIPS 2021 AI for Science Workshop.
- Upton, J. W., Kaiser, W. J. & Mocarski, E. S. DAI/ZBP1/DLM-1 complexes with RIP3 to mediate virus-induced programmed necrosis that is targeted by murine cytomegalovirus vIRA. *Cell Host Microbe* **11**, 290–297 (2012).
- Zhang, T. *et al.* Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell* **180**, 1115–1129.e1113 (2020).
- Zhang, T. *et al.* ADAR1 masks the cancer immunotherapeutic promise of ZBP1-driven necroptosis. *Nature* **2**, 1–9 (2022).
- Herbert, A. ADAR and immune silencing in cancer. *Trends Cancer* **5**, 272–282. <https://doi.org/10.1016/j.trecan.2019.03.004> (2019).
- Balachandran, S. & Mocarski, E. S. Viral Z-RNA triggers ZBP1-dependent cell death. *Curr. Opin. Virol.* **51**, 134–140. <https://doi.org/10.1016/j.coviro.2021.10.004> (2021).
- Zhang, T. *et al.* Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell* **180**, 1115–1129.e1113. <https://doi.org/10.1016/j.cell.2020.02.050> (2020).
- Wittig, B., Wolf, S., Dorbic, T., Vahrson, W. & Rich, A. Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.* **11**, 4653–4663. <https://doi.org/10.1002/j.1460-2075.1992.tb05567.x> (1992).
- Maruyama, A., Mimura, J., Harada, N. & Itoh, K. Nrf2 activation is associated with Z-DNA formation in the human HO-1 promoter. *Nucleic Acids Res.* **41**, 5223–5234. <https://doi.org/10.1093/nar/gkt243> (2013).
- Zhang, J. *et al.* BRG1 interacts with Nrf2 to selectively mediate HO-1 induction in response to oxidative stress. *Mol. Cell Biol.* **26**, 7942–7952. <https://doi.org/10.1128/MCB.00700-06> (2006).
- Kouzine, F. *et al.* Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356.e347. <https://doi.org/10.1016/j.cels.2017.01.013> (2017).
- Wu, T., Lyu, R., You, Q. & He, C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat. Methods* **17**, 515–523. <https://doi.org/10.1038/s41592-020-0797-9> (2020).
- Shin, S.-I. *et al.* Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res.* **23**, 477–486 (2016).
- Beknazarov, N., Jin, S. & Poptsova, M. Deep learning approach for predicting functional Z-DNA regions using omics data. *Sci. Rep.* **10**, 19134. <https://doi.org/10.1038/s41598-020-76203-1> (2020).
- Ho, P. S., Ellison, M. J., Quigley, G. J. & Rich, A. A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *EMBO J.* **5**, 2737–2744 (1986).
- Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* <https://doi.org/10.7554/elife.29736> (2017).

28. Umerenkov, D. *et al.* Z-flipon variants reveal the many roles of Z-DNA and Z-RNA in health and disease. *Life Sci Alliance*. <https://doi.org/10.26508/lسا.202301962> (2023).
29. Herbert, A., Pavlov, F., Konovalov, D. & Poptsova, M. Conserved microRNAs and flipons shape gene expression during development by altering promoter conformations. *Int. J. Mol. Sci.* **24**, 4884 (2023).
30. Rich, A., Nordheim, A. & Wang, A. H. The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53**, 791–846. <https://doi.org/10.1146/annurev.bi.53.070184.004043> (1984).
31. Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**, e29736 (2017).
32. Herbert, A. Flipons and small RNAs accentuate the asymmetries of pervasive transcription by the reset and sequence-specific microcoding of promoter conformation. *J. Biol. Chem.* **299**, 105140. <https://doi.org/10.1016/j.jbc.2023.105140> (2023).
33. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027. <https://doi.org/10.1073/pnas.84.20.7024> (1987).
34. Richter, W. F., Nayak, S., Iwasa, J. & Taatjes, D. J. The mediator complex as a master regulator of transcription by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **23**, 732–749. <https://doi.org/10.1038/s41580-022-00498-3> (2022).
35. Wang, J. C. DNA topoisomerases. *Annu. Rev. Biochem.* **65**, 635–692 (1996).
36. Herbert, A. Flipons and small RNAs accentuate the asymmetries of pervasive transcription by the reset and sequence-specific microcoding of promoter conformation. *J. Biol. Chem.* **299** (2023).
37. Blombach, F., Smollett, K. L., Grohmann, D. & Werner, F. Molecular mechanisms of transcription initiation—Structure, function, and evolution of TFE/TFIIE-like factors and open complex formation. *J. Mol. Biol.* **428**, 2592–2606 (2016).
38. Wong, B., Chen, S., Kwon, J. A. & Rich, A. Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **104**, 2229–2234. <https://doi.org/10.1073/pnas.0611447104> (2007).
39. Herbert, A. The ancient Z-DNA and Z-RNA specific Za fold has evolved modern roles in immunity and transcription through the natural selection of Flipons. *R. Soc. Open Sci.* **11**, 240080 (2024).
40. Husain, A. *et al.* Chromatin remodeler SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability. *Nat. Commun.* **7**, 10549. <https://doi.org/10.1038/ncomms10549> (2016).
41. Herbert, A. Simple repeats as building blocks for genetic computers. *Trends Genet.* **36**, 739–750. <https://doi.org/10.1016/j.tig.2020.06.012> (2020).
42. Marshall, P. R. *et al.* Dynamic regulation of Z-DNA in the mouse prefrontal cortex by the RNA-editing enzyme Adar1 is required for fear extinction. *Nat. Neurosci.* **23**, 718–729. <https://doi.org/10.1038/s41593-020-0627-5> (2020).
43. Bosch, A. & Suau, P. Changes in core histone variant composition in differentiating neurons: The roles of differential turnover and synthesis rates. *Eur. J. Cell Biol.* **68**, 220–225 (1995).
44. Dunn, C. J. *et al.* Histone hypervariants H2A.Z.1 and H2A.Z.2 play independent and context-specific roles in neuronal activity-induced transcription of Arc/Arg3.1 and other immediate early genes. *eNeuro* <https://doi.org/10.1523/ENEURO.0040-17.2017> (2017).
45. Kafer, G. R. *et al.* Sequential peripheral enrichment of H2A.Zac and H3K9me2 during trophoblast differentiation in human embryonic stem cells. *J. Cell Sci.* <https://doi.org/10.1242/jcs.245282> (2020).
46. Herbert, A. Nucleosomes and flipons exchange energy to alter chromatin conformation, the readout of genomic information, and cell fate. *Bioessays* **44**, e2200166. <https://doi.org/10.1002/bies.202200166> (2022).
47. Snow, M. Gastrulation in the mouse: Growth and regionalization of the epiblast. *Development* **42**, 293–303 (1977).
48. Georgakopoulos-Soares, I. *et al.* High-throughput characterization of the role of non-B DNA motifs on promoter function. *Cell Genomics* <https://doi.org/10.1016/j.xgen.2022.100111> (2022).
49. Zavarykina, T., Atkarskaya, M. & Zhizhina, G. The structural and functional properties of Z-DNA. *Biophysics* **64**, 671–682 (2019).
50. Kang, H. J. *et al.* Novel interaction of the Z-DNA binding domain of human ADAR1 with the oncogenic c-Myc promoter G-quadruplex. *J. Mol. Biol.* **426**, 2594–2604. <https://doi.org/10.1016/j.jmb.2014.05.001> (2014).
51. Mulholland, N., Xu, Y., Sugiyama, H. & Zhao, K. SWI/SNF-mediated chromatin remodeling induces Z-DNA formation on a nucleosome. *Cell Biosci.* **2**, 3. <https://doi.org/10.1186/2045-3701-2-3> (2012).
52. Li, H. *et al.* Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Res.* **37**, 2737–2746. <https://doi.org/10.1093/nar/gkp124> (2009).
53. Wang, G., Carbalajal, S., Vlijg, J., DiGiovanni, J. & Vasquez, K. M. DNA structure-induced genomic instability in vivo. *J. Natl. Cancer Inst.* **100**, 1815–1817. <https://doi.org/10.1093/jnci/djn385> (2008).
54. Kwon, J. A., Lowenhaupt, K., Wong, B. & Rich, A. A role for Z-DNA in transcriptional activation of the human c-myc oncogene by the vaccinia virus E3L protein. *Cancer Res.* **68**, 65–65 (2008).
55. Liu, H., Mulholland, N., Fu, H. & Zhao, K. Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling. *Mol. Cell Biol.* **26**, 2550–2559. <https://doi.org/10.1128/MCB.26.7.2550-2559.2006> (2006).
56. Champ, P. C., Maurice, S., Vargason, J. M., Camp, T. & Ho, P. S. Distributions of Z-DNA and nuclear factor I in human chromosome 22: A model for coupled transcriptional regulation. *Nucleic Acids Res.* **32**, 6501–6510. <https://doi.org/10.1093/nar/gkh988> (2004).
57. Liu, R. *et al.* Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309–318. [https://doi.org/10.1016/s0092-8674\(01\)00446-9](https://doi.org/10.1016/s0092-8674(01)00446-9) (2001).
58. Kimura, K. *et al.* Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.* **16**, 55–65 (2006).
59. Santoni, F. A. *et al.* Detection of imprinted genes by single-cell allele-specific gene expression. *Am. J. Hum. Genet.* **100**, 444–453 (2017).
60. Schroth, G. P., Chou, P. J. & Ho, P. S. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.* **267**, 11846–11855 (1992).
61. Zou, Z., Ohta, T., Miura, F. & Oki, S. ChIP-Atlas 2021 update: A data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and bisulfite-seq data. *Nucleic Acids Res.* **50**, W175–182. <https://doi.org/10.1093/nar/gkac199> (2022).
62. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354. <https://doi.org/10.1038/s41598-019-45839-z> (2019).
63. Dorogush, A. V., Ershov, V. & Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363* (2018).
64. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773. <https://doi.org/10.1093/nar/gky955> (2019).
65. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282. [https://doi.org/10.1016/0022-2836\(87\)90689-9](https://doi.org/10.1016/0022-2836(87)90689-9) (1987).

Acknowledgements

We thank Patrick Cramer for supplying the data for analysis. We thank Sid Balachandran for helpful discussion.

Author contributions

M.P. and A.H. conceived the study. N.B. and D.K. performed the data analysis. M.P. and A.H. wrote the manuscript. All authors analysed and discussed the results. All authors reviewed the manuscript.

Funding

This study was implemented in the framework of the Basic Research Program at HSE University, RF.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68439-y>.

Correspondence and requests for materials should be addressed to A.H. or M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024