A role for Flipons and miRNAs in Promoter Specification during Development

Running Title: Turning Torque into Action

Alan Herbert[1*], Fedor Pavlov[2], Dmitrii Konovalov[2], Maria Poptsova[2]

[1]InsideOutBio, 42 8th Street, Charlestown, MA

[2] Laboratory of Bioinformatics, Faculty of Computer Science, National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Moscow, Russia 101000

*Corresponding author

alan.herbert@insideoutbio.com

Abstract

The classical view of gene regulation is based on prokaryotic models and the operon concept with protein-based transcription factors controlling the expression of metabolic pathways essential for bacterial adaptations in response to environmental changes. A new view for establishing cell identity is emerging in eukaryotes where RNA-based pathways provide the framework for the readout of genomic information. Another perspective poses that alternative DNA structures encoded by flipons enable switching of cellular responses from one state to another. Here we provide evidence that these RNA and DNA mechanisms are deeply connected. We present data supporting a model where flipons open up binding sites for microRNAs (miRNAs), leading to the establishment of bivalent promoters early in development whose location structures lineage-specific events. These outcomes are potentially influenced by ovarian and spermatozoan miRNAs, transmissions with evident evolutionary ramifications. The data supports a new perspective on genetic regulation, one in which the genome provides a canvas framed by flipons, sketched with miRNAs and embellished by proteins.

Introduction

The approach we take follows that of Britten and Davidson who first proposed the programming of genomes by *trans*-RNAs that act at locations distant from where they are produced [1]. While there are many classes of trans-RNAs of varying sizes [2], we will focus on microRNAs (miRNAs) and examine the DNA sites to which they bind and their potential role in embryonic development. We posit that flipons, sequences that adopt non-B-DNA conformations under physiological conditions[3], play a role in targeting maternal and paternal miRNAs to particular locations to establish developmentally important bivalent promoters (BiPs), where both active histone 3, lysine 4 trimethylation (H3K4me3) and suppressive histone 3, lysine 27 trimethylation (H3K27me3) are present. The loss of one or other of these modifications at a later stage guides tissue formation and differentiation. We find that different flipon classes and evolutionarily conserved miRNA families seed sequences [4] colocalize with each other in promoter regions, along with the AGO1 and AGO2 argonaute proteins that are miRNAs effectors (AERs). The promoter so characterized are enriched for transcription factor genes that specify cell-fate.

The Links between miRNAs and Gene Regulation

The ability of miRNAs to regulate phenotype was first shown by their effects due to their suppression of transcript translation. By changing the time window in which a protein was produced, such miRNAs alter the size and cellular composition of an organ formed during development [5]. In mice, knockout of conserved miRNAs has profound effects on embryonic development [4]. These miRNAs arise through a number of canonical and non-canonical processing pathways that output AERs [4]. In line with a role for *trans*-RNAs in these early outcomes, *Ago2* knockout is embryonically lethal with a failure to develop extra-embryonic tissue (Table 1). However, survival does not depend on the AGO2 enzymatic slicer activity needed to suppress translation as mice homozygous for a catalytically dead $Ago2^{ADH}$ allele develop normally [6], suggesting that AGO2 is active through a different mechanism. Besides the cytoplasm, miRNAs also localize to the nucleus [2,7]. They modulate gene expression [2,8,9] and potentially a subset of alternative splices [10-12].

Flipons and Nucleosome Phasing

A role for Z-DNA and G4 flipons in gene expression has been proposed almost since the discovery of these non-B-DNA structures [13]. The flip to an alternative conformation requires energy, such as that liberated when polymerases and helicases hydrolyze ATP to separate DNA strands or that energy released instantaneously when nucleosomes are ejected from DNA (Figure 1A and B)[14]. The unwinding of the B-DNA helix due to these events generates negative supercoiling that drives the formation of left-handed DNA [15] and four-stranded quadruplexes (consensus sequence $d(G_3X_{1\text{-}7})_4$ [16-18]. The flip occurs faster than topoisomerases can relax the DNA torque [19]. The enrichment of Z- and G- flipons and the structures they form in promoter regions likely flags these sites for the cellular machinery to find easily, while also keeping transcription start sites (TSS) nucleosome-free [14,20,21].

These and other types of flipon can be detected *in cellulo* by using potassium permanganate ($KMnO_4$) footprinting [22]. In particular, they identify sites of Stress Induced Duplex Destabilization (SIDDs) [23] that potentially bind *trans*-RNA to form R-loops [24,25] and triplexes [26,27]. Both of these alternative structures are incompatible with nucleosome assembly [28] or engagement of B-DNA specific proteins. SIDDs are AT-rich and are referred to here as S-flipons. Besides RNA and DNA, SIDDs can bind single-stranded DNA with high affinity, such as those KH domain proteins that regulate c-MYC gene expression

[29] and alternative RNA splicing [30]. $KMnO_4$ footprinting can also detect H-DNA, which is formed by the fold-back of a single-stranded region of polypurine or polypyrimidine DNA onto an adjacent sequence-matched duplex [22]. H-DNA represents a class of T-flipon that can also form triplexes with either *trans*-RNAs and DNAs [13,14].

The question arises whether the effects of *trans*-RNAs and flipons on gene expression are concerted. Here we present evidence in support of this possibility and for their joint role in early development.

Results

We queried the flipon sequences identified by Kouzine et al. [22] in activated murine B cells by KMnO4 footprinting to assess their match to those seed sequences found in the set of miRNA families highly conserved in placental animals [4] (Figure 2A, Supplementary Data 1) (Z-DNA (n=25,059), G4 (n=20,253), SIDD (n=15,296), H-DNA (n=17,100)). We observed a frequent overlap of miRNA seed sequences with SIDDs (mean length =170 bps), in contrast to other flipon classes (Figure 2A). The most frequent overlap was with miR-203a (~20%) that promotes epidermal differentiation by repressing the proliferative potential of dermal stem cells in mice, starting at embryonic day 13.5 (E13,5) [31]. Other miRNA matches were also common, including the miR-374 (4.56%) that is encoded within the X-inactivation center [32]. The high match frequency observed is consistent with a biological role for such interactions as the typical miRNA has 40-1500 copies per cell [33], a stoichiometry sufficient to cover many of the available 15,296 SIDD sites identified in the Kouzine data assuming that about half the miRNAs are nuclear [7]. There were a few instances of overlap of miRNA seed sequences with other flipons classes (Figure 2A). The Z-prone $d(AC)_n$ sequence implicated in splicing [11,34] has a 4.5% overlap with miR-329/362. Other miRNA seed sequences containing $d(G)_3$ (or its complement $d(C)_3$) overlapped the annotation for G4-quadruplex ($d(G_3X_{1-7})_4$).

We tested the colocalization of experimentally determined SIDDs with other flipon classes by extending the search window to 100 bases either side of the initial SIDD/miRNA seed match (Figure 2B). We found some instances where the same miRNA seed sequence match was present at multiple sites within a region (top panel Figure 2B, the index interaction is shown with blue lines, red lines show additional matches). Allowing matches with other miRNA seeds yielded a potential miRNA binding site within 500 basepairs for 70% of flipons. We also found a high frequency of Z-, G- and T-flipons close to the index SIDD/miRNA seed match (Figure 2B, Supplementary Data 1).

The top hits for a Z-flipon match with a single miRNA seed sequences include miR-377, miR-744, miR-329/362 and miRNA-342 (8.34%, 7.63%, 7.25% and 7.00% overlap respectively), all containing alternating pyrimidine/purine Z-DNA prone sequences. These miRNAs potentially regulate Z-DNA formation by promoting R-loop formation. A similar situation occurs for G4-forming sequences with the top hits containing a run of three purines on one strand or the other, again suggesting a reciprocal relationship between flipon conformations. For experimentally validated H-DNA, the top hits include miR-185, miR-203a and miR-204/211, each with polypurine seeds and overlaps of 15.32%, 6.56% and 6.04% respectively. In most other cases, the match of the miRNA seed sequence to canonical Z-, G- and T-flipon sequences was low, suggesting that these elements are distinct from S-flipons. Indeed, the spectrum of colocalization of miRNA with differs by flipon class (the position of miRNAs named on the x-axis is quite variable) (Figure 2B).

To confirm the interaction of miRNAs with SIDDs, we examined the binding of AERs to DNA using the ENCODE AGO1 and AGO2 ChIP-seq datasets that are available for the human HepG2 liver cell line [35]. Their binding sites are highly enriched in promoters (70.17% and 54.68% overlap within 1 Kb of TSS respectively, Supplementary Figure 1) [35]. For SIDDs located in promoters, we found a 15.05% and 25.47% overlap with AGO1 and AGO2 peaks respectively compared to 1.91% and 4.79% for SIDDs located throughout the genome. The pattern of binding around TSS shows a wide distribution of AGO contacts consistent with multiple different miRNA binding sites (Figure 2C). By comparison, the overlap of AGO1 and AGO2 with the experimentally validated Z-flipons in promoters was 4.23% and 3.76% respectively, consistent with results from the observed overlap of seed and flipon sequences.

The close proximity between the different flipon classes was confirmed genomewide (Figure 3). The enrichment of experimentally determined Z- and G-flipons in promoters is consistent with previous results [14,20,21] (Figure 3A). Multiple exact matches for the same miRNA seed is frequent for SIDDs (41.01% Figure 3B and 67.93% when combined with single exact matches), suggesting that this feature is under selection. The tight association with SIDDs and other classes of flipons increases as the search window is expanded to 500 bp either side of the SIDD(Figure 3B). The pairing of Z-DNA with SIDDs in promoter regions is smaller numerically than with G4 and H-DNA but constitutes a higher percentage of total Z-flipon counts, increasing to 33% within 500 bps(Figure 3C and Supplementary Figure 2, 3). Gene ontogeny shows that the promoters identified in the 500 bp window are highly enriched for TF involved in development, particularly those involved in cell fate commitment (Table 1, Supplementary Data 2)

Discussion

The close proximity of SIDDs to other classes of flipons leads to the general model presented in figure 1 where the exchange of energy between the different conformations is modulated by the power stored within nucleosomes (Figure 1A and B). In Figure 1C, we show that the flip from Z-DNA to B-DNA during nucleosome docking transfers energy to the S-flipon, opening up the SIDD to allow sequence-specific binding of an RNA to form an R-loop. In panel 1D, we show a G-flipon storing energy that can be used by a helicase to place a nucleosome in a position to stabilize a triplex [28].

We found an enrichment of TF involved in cell fate specification in promoters where SIDDs were paired with other flipons (Table 1). To better understand how flipon pairs and miRNAs interact, we turned to early development where the sequence of events can be established by the time at which gene knockout is embryonically lethal. The time of death given in Table 2 is for mice homozygously deficient in key histone methyltransferases. The row ordering informs as to the epigenetic modifications made by each enzyme is necessary for a subsequent developmental step. The first step requires SETDB1 to suppress endogenous retroelements that are actively producing double-stranded RNAs [36,37] and results in the H3K9me3 dependent formation of constitutive heterochromatin [38]. The second critical step is the marking of CpG rich promoters with H3K4me3 by SETD1A [39]. The enzyme is guided by the CXXC1 protein and related family members that bind at CpG steps [40,41]. The lysine methylation likely facilitates the ejection of nucleosomes that potentially powers the flip of DNA to an alternative conformation. The third step is the creation BiPs, when H3K27me3 is written by the enhancer of zeste homolog 2 (EZH2) methyltransferase to sites marked by H3K4me3. EZH2 is part of the polycomb repressive complexes 2 (PRC2) and is essential for normal development [42]. The guidance of PRC2 by small RNAs is well established in worms and yeast but fewer examples are known in mammals [43-45]. The H3K27me3 is thought to protect against DNA methylation that accompanies heterochromatin

formation[39]. The fourth step depends on protein factors that edit BiP marks to produce tissue-specific outcomes and involves PRC1 [46]. The diversity of PRC1 complexes suggest they have a number of divergent roles in tissue development [47]. However, deletion of the PRC1 enzyme EZH1 is not embryonically lethal [48].

We posit that interaction of flipons with *trans*-RNA bootstraps the initial phase of embryogenesis by specifying sites destined to become bivalent promoters (Figure 4). The CpG rich regions in layer 1 are enriched for promoter flipons. The adoption of a non-B-DNA conformation is favored by H3K4me3 and ensures that TSS and SIDDs are nucleosome-free and in open chromatin. The *trans*-RNAs that bind SIDDs (layer 2) then specify the location of BiPs by promoting H3K27me3 (layer 3). These tags localize the cellular machinery needed for later steps involving sequence-specific binding by TF and editing of chromatin marks (layer 4). A subset of TF interactions may be further moderated by miRNAs that sequester their binding sites into R-loops or triplexes [49]. Each of the four layers utilizes different sets of information to genetically encode developmental programs. The initial specification of bivalent promoters is by the *trans*-RNAs transmitted to the embryo by each parent. These include those miRNA derived from ovary and spermatozoan specific miRNAs [50,51], transmissions with evident evolutionary ramifications [52].

Future Steps

The results we present are for a subset of conserved miRNAs from a single cell line and likely underestimates the true extent of these interactions that may also involve PIWI [8] and other *trans*-RNA types [2]. The SIDDs involved in specifying outcomes likely vary by tissue. Their detection by $KMnO_4$ footprinting may be incomplete. This technique relies on thymine modification with other unpaired bases that can be detected with kethoxal missed [53]. Both footprinting techniques can be applied to any cell line and also to any early stage of embryonic development to provide more information on how flipons affect the readout of genetic information. The ability to control exposure and duration [22] permits a careful mapping of chronological events to establish cause and effect. The approach can also potentially inform on how miRNAs coordinate nuclear and cytoplasmic gene expression. Other potential applications include the targeting of lineage-specific S-flipons with RNAs that alter binding sites for both enhancer and repressor protein complexes [49].

Methods

Datasets: AGO1 ChIP-seq (GEO:GSE174905), AGO2 ChIP-seq (GEO: GSE136467), $KMnO_4$ mapping of non-B-DNA structures (https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#nonbdna), conserved miRNA seed sequences(Bartel [4])

Code Availability: https://github.com/TheodorrodeohT/article_conserved_miRNA.
The following types of flipons were uplifted from mm9 to mm10 using UCSC liftOver. `bedtools 2.26` software was used to find overlaps between `.bed` files and apply slops of different sizes to the initial flipon coordinates. The input `.bed` files were converted into `fasta` format using `biopython 1.79` python library. Each flipon sequence was tested for embedding miRNA seed sequences via default python string and array operations. The results were saved as a `pandas` dataframe and later converted into `.xlsx` format for better readability. Values for pie charts were calculated using `ChIPpeakAnno` and `ChIPseeker` R libraries. The genomic annotation used as a reference was [GENCODE vM25] (https://www.gencodegenes.org/mouse/release_M25.html) (GRCm38.p6). The pie charts were drawn using `plotly 5.8.0` python library."

For the analysis of AGO1 and AGO2, we used custom R (4.1.3) and python (3.7.13) scripts. The scripts are available in the repository https://github.com/Madund3ad/article-mi-RNA/. The genomic annotation used as a reference was [ENSEMBL GRCh37.87 http://anonymous@ftp.ensembl.org/pub/grch37/release-87/gtf/homo_sapiens/Homo_sapiens.GRCh37.87.gtf.gz (GRCh37.87) Overlaps and distances to TSS were calculated using" GenomicRanges" R library and values for smooth curves were calculated using the "zoo" R library. After that distances to TSS were drawn using "ggplot2" R library. Pie charts' data were calculated using `ChIPpeakAnno` and `ChIPseeker` R libraries and drawn using `plotly 5.8.0` python library.

Analysis:

The ssDNA sequence motifs of non-B DNA from Kouzine paper are used as the peak data set. Since the original version of the mouse genome was mm9, the peaks were uplifted to mm10 using the UCSC liftOver utility (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/), resulting in slightly different N dimensions of each dataset: H-DNA (17,110 → 17,000), G4 (20,260 → 20,253), Z-DNA (25,063 → 25,059), SIDD (15,299 → 15,296). Using Python 3.7 and biopython 1.79 library, the peak coordinates were converted into nucleotide sequences. Then, each sequence was tested for overlap with each miRNA seed region. The final results are compiled into a summary table for comparative analysis of the overlap values. Gene enrichment was performed using the WEB-based GEne SeT AnaLysis Toolkit (http://www.webgestalt.org/).

For the analysis of distance to TSS distribution we used ENCODE annotation (GRCh37.87). Only the distance to the nearest region of interest (AGO1 or AGO2 ChIP peak) is shown on the plot. A rolling average with a window of 8 bases is applied to smooth the curve.

For the analysis of overlaps between ENCODE AGO narrow peaks, experimental Kouzine SIDD peaks and promoters we took 1000bp upstream regions from TSS as a promoter. TSS coordinates were obtained from ENCODE annotation (GRCh37.87). Two regions were considered to be overlapping if they had at least one base pair in common. Significance of the overlap was calculated using Monte-Carlo simulation (n=1000).

Contributions: AH conceived and wrote the paper with input from MP, FP and DK. FP analyzed the interactions between flipons and miRNAs and DK analyzed the locations of AGO1 and AGO2 in K562. MP helped edit the paper and supervised the analysis by FP and DK.

Acknowledgements. We thank Sid Balachandran for reviewing the manuscript

Conflicts of Interest: AH is the founder of InsideOutBio, a company that works in the field of immuno-oncology. All the authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

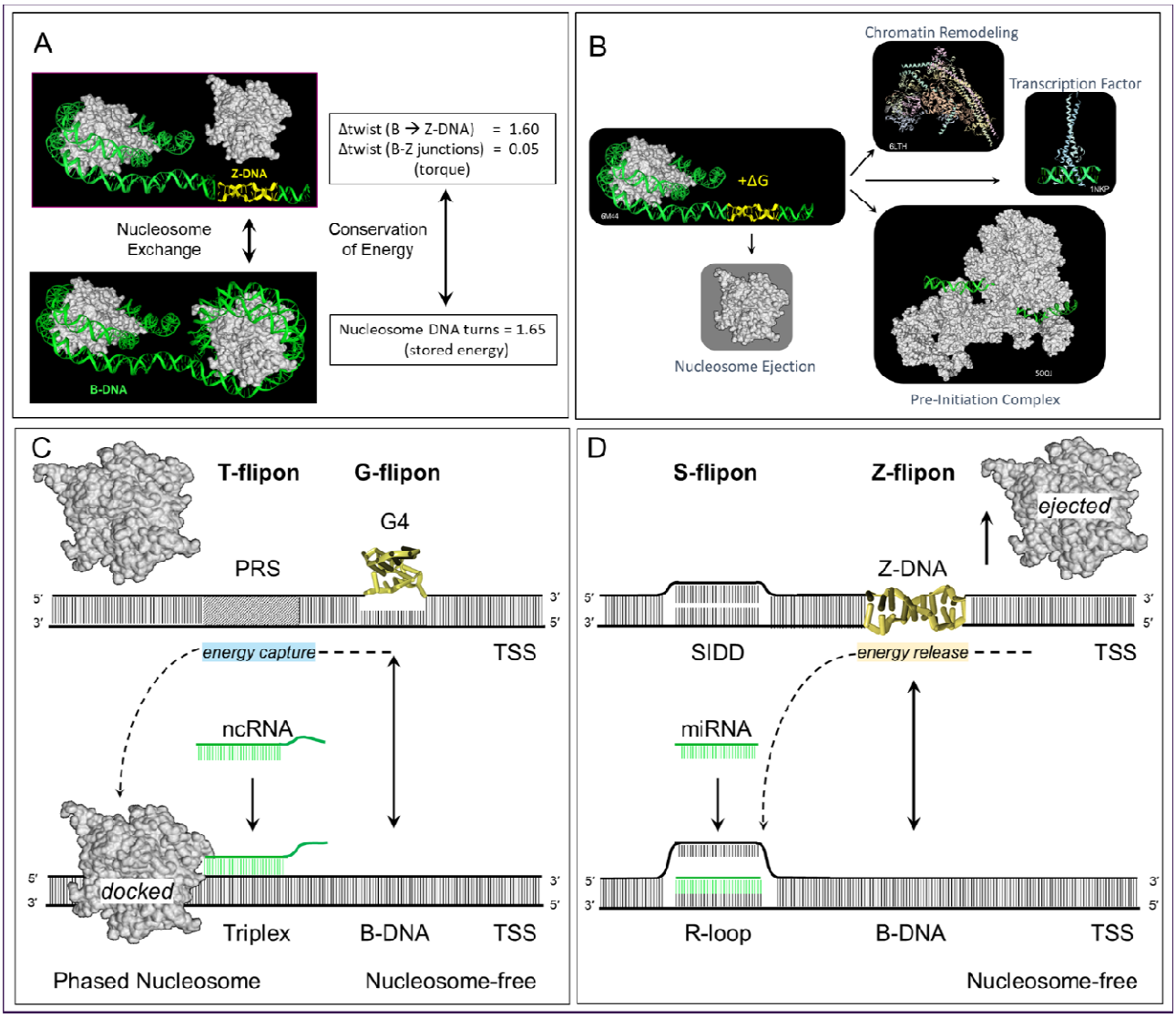
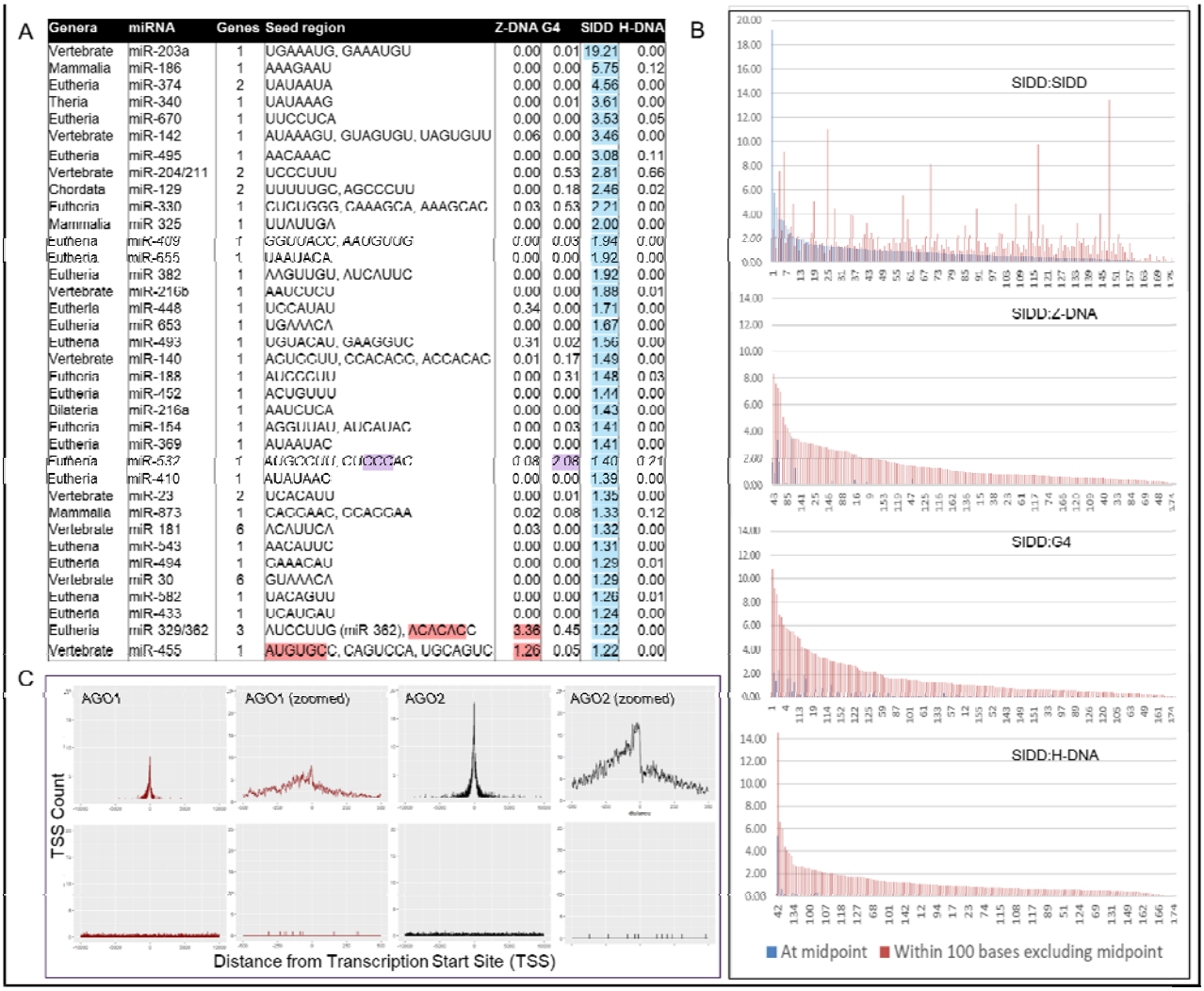


Figure 1 Converting torque into action **A.** The stress produced in underwound DNA is relieved by a flipon adopting the left-handed Z-DNA conformation or by wrapping the DNA around a nucleosome. The change in DNA twist when Z-DNA flips to B-DNA is equivalent to the number of turns in the solenoid formed by coiling the DNA around the histone core. The formation of Z-DNA can phase the placement of nucleosomes on DNA so that transcription start sites are accessible [54]. **B.** Release of the energy stored in nucleosome cores can power formation of different chromatin complexes that regulate chromatin state and the readout of information from the genome. **C**. Triplex formation by Purine Rich Sequences (PRS) within a T-flipon can also incorporate miRNAs and long non-coding RNAs [26] G-flipons are more stable than Z-DNA and require resolution by a helicase to generate unwound DNA that then wraps around nucleosomes to capture the free energy produced [20,21].**D.** The energy released by ejection of a nucleosome (indicated by upwards arrow) is captured by Z-DNA formation and accumulates until sufficient to power strand separation at the Stress Induced Sequence Destabilized (SIDD) sites that define S-flipons, allowing base-specific binding of ligands such as RNA, DNA and protein. In both cases,

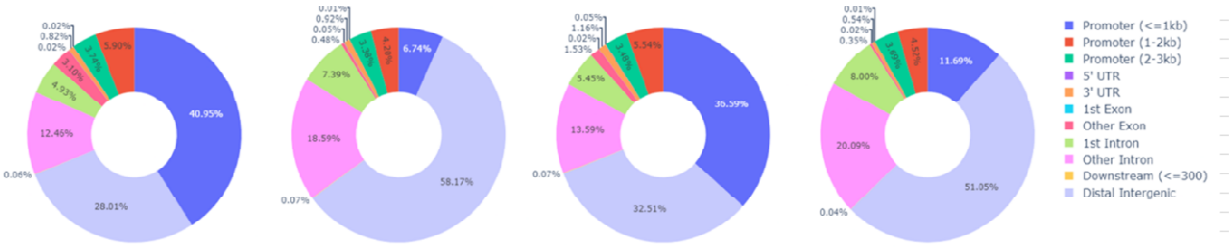
flipons are associated with nucleosome free regions and flag these sites for chromatin and transcriptional factors to bind, reducing the genomic search space.



Genera	miRNA	Genes	Seed region	Z-DNA	G4	SIDD	H-DNA
Vertebrate	miR-203a	1	UGAAAUG, GAAAUGU	0.00	0.01	19.21	0.00
Mammalia	miR-186	1	AAAGAAU	0.00	0.00	5.75	0.12
Eutheria	miR-374	2	UAUAAUA	0.00	0.00	4.56	0.00
Theria	miR-340	1	UAUAAAG	0.00	0.01	3.61	0.00
Eutheria	miR-670	1	UUCCUCA	0.00	0.00	3.53	0.05
Vertebrate	miR-142	1	AUAAAGU, GUAGUGU, UAGUGUU	0.06	0.00	3.46	0.00
Eutheria	miR-495	1	AACAAAC	0.00	0.00	3.08	0.11
Vertebrate	miR-204/211	2	UCCCUUU	0.00	0.53	2.81	0.66
Chordata	miR-129	2	UUUUUGC, AGCCCUU	0.00	0.18	2.46	0.02
Euthoria	miR-330	1	CUCUGGG, CAAAGCA, AAAGCAC	0.03	0.53	2.21	0.00
Mammalia	miR 325	1	UUAUUGA	0.00	0.00	2.00	0.00
Eutheria	miR-409	1	GGUUACC, AAUGUUG	0.00	0.03	1.94	0.00
Eutheria	miR-655	1	UAAUACA	0.00	0.00	1.92	0.00
Eutheria	miR 382	1	AAGUUGU, AUCAUUC	0.00	0.00	1.92	0.00
Vertebrate	miR-216b	1	AAUCUCU	0.00	0.00	1.88	0.01
Eutheria	miR-448	1	UCCAUAU	0.34	0.00	1.71	0.00
Eutheria	miR 653	1	UGAAACA	0.00	0.00	1.67	0.00
Eutheria	miR-493	1	UGUACAU, GAAGGUC	0.31	0.02	1.56	0.00
Vertebrate	miR-140	1	AGUCCUU, CCACACC, ACCACAC	0.01	0.17	1.49	0.00
Euthoria	miR-188	1	AUCCCUU	0.00	0.31	1.48	0.03
Eutheria	miR-452	1	ACUGUUU	0.00	0.00	1.41	0.00
Dilateria	miR-216a	1	AAUCUCA	0.00	0.00	1.43	0.00
Euthoria	miR-154	1	AGGUUAU, AUCAUAC	0.00	0.03	1.41	0.00
Eutheria	miR-369	1	AUAAUAC	0.00	0.00	1.41	0.00
Eutheria	miR-532	1	AUGCCUU, CUCCCAC	0.08	2.08	1.40	0.21
Eutheria	miR-410	1	AUAUAAC	0.00	0.00	1.39	0.00
Vertebrate	miR-23	2	UCACAUU	0.00	0.01	1.35	0.00
Mammalia	miR-873	1	CACCAAC, GCAGGAA	0.02	0.08	1.33	0.12
Vertebrate	miR 181	6	ACAUUCA	0.03	0.00	1.32	0.00
Eutheria	miR-543	1	AACAUUC	0.00	0.00	1.31	0.00
Eutheria	miR-494	1	GAAACAU	0.00	0.00	1.29	0.01
Vertebrate	miR 30	6	GUAAACA	0.00	0.00	1.29	0.00
Eutheria	miR-582	1	UACAGUU	0.00	0.00	1.26	0.01
Eutheria	miR-433	1	UCAUGAU	0.00	0.00	1.24	0.00
Eutheria	miR 329/362	3	AUCCUUG (miR 362), ACACACC	3.36	0.45	1.22	0.00
Vertebrate	miR-455	1	AUGUGCC, CAGUCCA, UGCAGUC	1.26	0.05	1.22	0.00

Figure 2. **A.** Mapping of highly conserved miRNA family seed sequences to SIDDs experimentally mapped in activated B-cells with the percentage of those that embed miRNA seed sequences highlighted in light blue. The SIDD sequences that overlap G- and Z-flipons are colored purple and primrose respectively **B**. Relative position of Z-DNA, G4 and H-DNA to SIDDs. The charts are ordered by frequency, not by miRNA family. The blue vertical bars show the percentage of SIDDs that have a complete overlap with the miRNA seed sequence. The red lines are the percentage of matches for experimentally determined Z-DNA, G4 or H-DNA within 100 bases either side of the SIDD/miRNA match. The additional SIDD matches in this region are for the same miRNA. C. The position of ENCODE AGO1 and AGO2 ChIP-seq peaks in HepG2 liver cells relative to the Transcription Start Site (TSS) is shown in the upper panel along with a zoomed in image for both. For comparison, the overlap with randomly selected sequences is shown in the lower panel. SIDDS located in promoters show a 15.05% and 25.47% overlap with AGO1 and AGO2 peaks respectively compared to 1.91% and 4.79% genomewide. The genomewide features recognized by each protein is shown in supplemental figure 1.

A. Genomic Location of Experimentally Determined Flipons



B. Genomic Overlap of Conserved miRNAs with Flipons singly or paired with SIDD at 100bp, 200bp or 500bp

		Absolute Counts																Percentage %											
		Exact match				Exact match +-100bp				Exact match +-200bp				Exact match +-500bp				Exact match				Exact match +-100bp				Exact match +-200bp				Exact match +-500bp			
	N	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Z-DNA	25,059	21,650	2,616	730	45	2,262	3,936	14,795	4,066	444	869	15,131	8,615	55	26	6,209	18,609	86.47	10.44	2.91	0.18	9.03	15.71	59.04	16.23	1.77	3.47	60.38	34.30	0.22	0.10	25.10	74.58
G4	20,253	15,684	3,736	770	63	1,031	2,372	12,984	3,866	161	359	12,269	7,464	11	14	4,676	15,552	77.44	18.45	3.80	0.31	5.09	11.71	64.11	19.09	0.79	1.77	60.58	36.85	0.05	0.07	23.09	76.79
SIDD	15,296	3,153	4,117	6,273	1,753	135	520	10,354	4,287	10	35	8,762	6,489	1	0	2,977	12,318	20.61	26.92	41.01	11.46	0.88	3.40	67.69	28.03	0.07	0.23	57.28	42.42	0.01	0.00	19.46	80.53
H-DNA	17,100	15,626	1,066	395	13	2,687	3,117	8,549	2,747	670	923	9,800	5,707	43	44	4,715	12,298	91.38	6.23	2.31	0.08	15.71	18.23	49.99	16.06	3.92	5.40	57.31	33.37	0.25	0.26	27.57	71.92
Z-DNA & SIDD (+-100bp)	758	620	95	41	2	58	116	441	143	3	20	472	263	0	0	209	549	81.79	12.53	5.41	0.26	7.65	15.30	58.18	18.87	0.40	2.64	62.27	34.70	0.00	0.00	27.57	72.43
G4 & SIDD (+-100bp)	2,250	1,765	424	59	2	54	197	1,439	560	4	11	1,311	924	0	0	508	1,742	78.44	18.84	2.62	0.09	2.40	8.76	63.96	24.89	0.18	0.49	58.27	41.07	0.00	0.00	22.58	77.42
H-DNA & SIDD (+-100bp)	2,010	1,945	65	5	1	173	372	1,103	368	14	41	1,242	719	0	0	494	1,522	96.77	3.23	0.25	0.05	8.61	18.51	54.88	18.31	0.70	2.04	61.79	35.77	0.00	0.00	24.58	75.72
Z-DNA & SIDD (+-200bp)	1,117	923	140	52	2	97	171	646	203	10	33	692	382	0	0	303	814	82.63	12.53	4.66	0.18	8.68	15.31	57.83	18.17	0.90	2.95	61.95	34.20	0.00	0.00	27.13	72.87
G4 & SIDD (+-200bp)	2,534	1,995	471	65	3	71	238	1,617	608	4	22	1,488	1,020	0	0	581	1,953	78.73	18.59	2.57	0.12	2.80	9.39	63.81	23.99	0.16	0.87	58.72	40.25	0.00	0.00	22.93	77.07
H-DNA & SIDD (+-200bp)	2,249	2,158	79	10	2	217	426	1,200	406	20	62	1,381	786	0	0	569	1,680	95.95	3.51	0.44	0.09	9.65	18.94	53.36	18.05	0.89	2.76	61.41	34.95	0.00	0.00	25.30	74.70
Z-DNA & SIDD (+-500bp)	1,802	1,528	209	63	2	153	289	1,067	293	17	48	1,135	602	0	0	484	1,318	84.79	11.60	3.50	0.11	8.49	16.04	59.21	16.26	0.94	2.66	62.99	33.41	0.00	0.00	26.86	73.14
G4 & SIDD (+-500bp)	3,089	2,418	581	84	6	99	298	1,998	694	6	32	1,833	1,218	0	0	712	2,377	78.28	18.81	2.72	0.19	3.20	9.65	64.68	22.47	0.19	1.04	59.34	39.43	0.00	0.00	23.05	76.95
H-DNA & SIDD (+-500bp)	2,560	2,444	96	17	3	262	483	1,366	449	35	74	1,559	892	0	0	657	1,903	95.47	3.75	0.66	0.12	10.23	18.87	53.36	17.54	1.37	2.89	60.90	34.84	0.00	0.00	25.66	74.34

T1: no miRNA overlap; T2:single miRNA; T3: multiple matches to same miRNA seed; T4: matches to different miRNA seeds

C. Genomic location of SIDD paired with Flipons 100bp either side

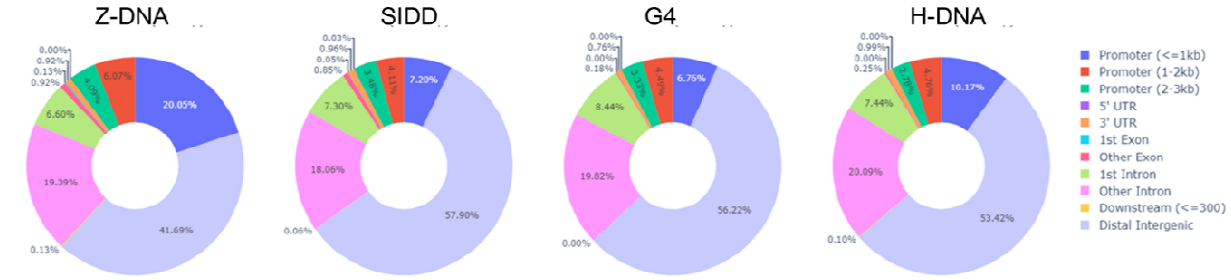


Figure 3 Mapping of Flipon overlaps to miRNAs and with SIDDs **A.** The genomewide distribution of experimentally mapped flipons in murine B-cells [22] **B**. The mapping of miRNA seed sequences to flipons either with an exact match or a match within 100, 200 or 500 basepairs of the initial match site. Grey indicates the number of miRNAs not mapped in the region scanned (T1). T2 represents a single match in the region. Lilac shows the counts for multiple overlaps in a region with the same miRNA seed (T3). Overlaps, such as those highlighted in cream, are frequent for SIDDs. T4 represents the number of overlaps with different miRNA seeds. Overlap percentages are highlighted in green. In addition, the colocalization of a flipon and SIDD within a region is counted along with and the miRNA seed sequence overlap. **C.** The genomic location of flipons within 100 basepairs of the closet SIDD is illustrated with respect to gene features.

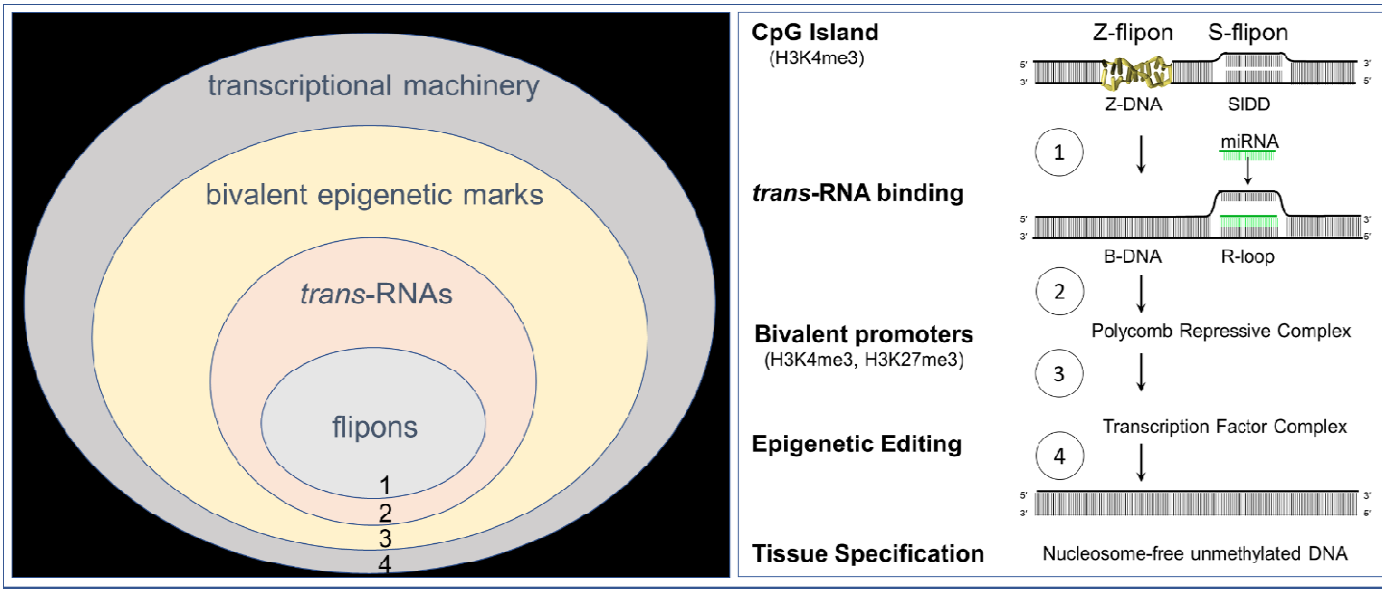


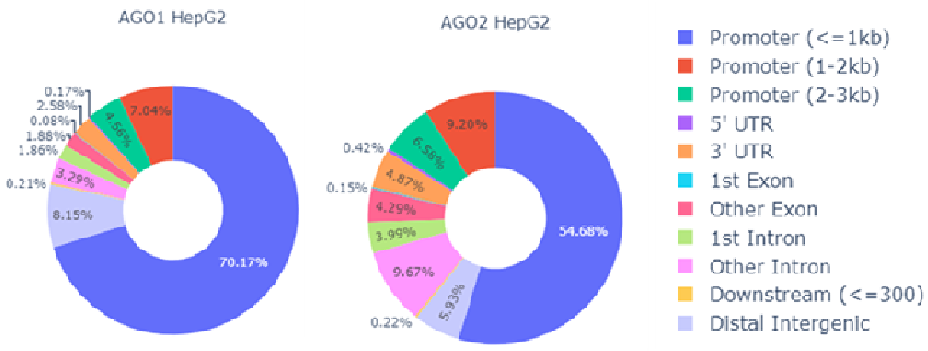
Figure 4. The four numbered layers proposed as necessary to bootstrap development of an embryo. Flipons and *trans*-RNAs have conjoint roles. Flipons, such as those that form Z-DNA and G4 quadruplexes, phase nucleosome to place transcription start sites (TSS) and *trans*-RNA binding sites in regions of open chromatin. Evolutionarily conserved miRNAs represent one class of *trans*-RNAs that recognize SIDDs (Stress-Induced Duplex Destabilization Sites) and form R-loops or triplexes in these AT-rich regions [26,27]. The SIDD sequences are referred to in the text as S-flipons. Early in development, we propose that the engagement of S-flipons by miRNA binding sites in close proximity to TSS establishes bivalent epigenetic tags that underlie the repressive and active states found later during tissue specification. The removal of one or other or both bivalent tags depends on the sequence-preferences of transcription factors and the enzymes present in the complex. Addition of further tags such as K2AK119ub further modify outcomes. Each of the four proposed regulatory layers coordinate development by accessing different types of genetically encoded information.

	Molecular Function					Biological Process (non-redundant)				
Flipon	Geneset	Description	ER	P-value	FDR	Geneset	Description	ER	P-value	FDR
Z-DNA	GO:0043565	sequence-specific DNA binding	4.65	0	0	GO:0045165	cell fate commitment	7.70	1.78E-15	1.44E-12
	GO:0003700	DNA-binding TF activity	4.59	0	0	GO:0001501	skeletal system development	4.48	2.32E-11	9.44E-09
	GO:0001067	regulatory region nucleic acid binding	4.86	0	0	GO:0061564	axon development	4.46	1.44E-10	3.89E-08
	GO:0044212	transcription regulatory region DNA binding	4.87	0	0	GO:0007389	pattern specification process	4.44	3.64E-10	7.38E-08
	GO:0000981	DNA-binding TF, RNA Pol II-specific	5.32	0	0	GO:0048638	regulation. Development and growth	4.79	1.14E-09	1.85E-07
G4	GO:0043565	sequence-specific DNA binding	4.62	0	0	GO:0045165	cell fate commitment	6.86	3.09E-13	2.51E-10
	GO:0003700	DNA-binding transcription factor activity	4.57	0	0	GO:0050808	synapse organization	4.43	4.01E-10	1.63E-07
	GO:0001067	regulatory region nucleic acid binding	4.84	0	0	GO:0007389	pattern specification process	4.30	7.46E-10	2.02E-07
	GO:0044212	transcription regulatory region DNA binding	4.86	0	0	GO:0061564	axon development	4.15	1.55E-09	3.15E-07
	GO:0000976	regulatory region DNA-specific binding	5.00	0	0	GO:0048568	embryonic organ development	4.03	2.88E-09	4.67E-07
H-DNA	GO:0043565	sequence-specific DNA binding	5.20	8.27E-14	8.09E-11	GO:0045165	cell fate commitment	9.19	1.54E-11	1.25E-08
	GO:0000981	DNA-binding TF activity, RNA Pol II-specific	6.31	8.80E-14	8.09E-11	GO:0050808	synapse organization	5.46	3.00E-08	1.22E-05
	GO:0003700	DNA-binding transcription factor activity	5.21	7.07E-13	4.33E-10	GO:0048568	embryonic organ development	4.65	6.65E-07	1.80E-04
	GO:0044212	transcription regulatory region DNA binding	5.37	3.16E-12	1.25E-09	GO:0050890	cognition	5.69	1.23E-06	2.49E-04
	GO:0001067	regulatory region nucleic acid binding	5.35	3.40E-12	1.25E-09	GO:0050803	regulation. synapse structure/ activity	6.10	1.78E-06	2.90E-04

Table 1. Gene Ontogeny Enrichment for Promoters that have SIDDs paired with a Z-, G- or T-flipon 500 basepairs either side. Conserved miRNA seed-sequence location was not used in this analysis. (ER: enrichment ratio; TF: transcription Factor; Pol: Polymerase) (See Supplementary Data for additional details)

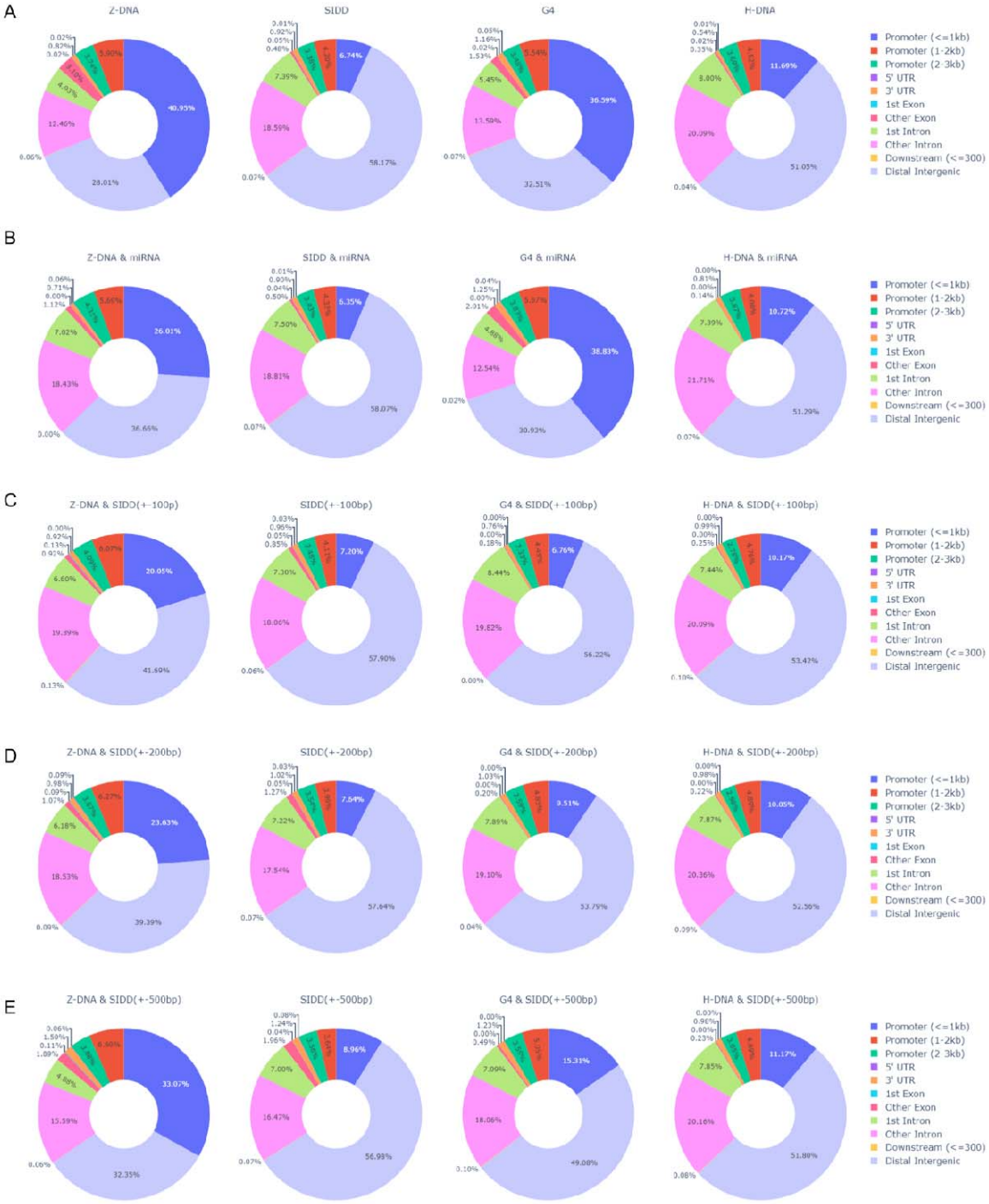
Methyltransferase	Targeting	Modification	Null (Lethality)	Chromatin Feature	Outcome
SETDB1	siRNA	H3K9me3	E3.5-E5.5 [55]	Constitutive Chromatin	Suppression of retroelements [36,37]
SETD1	CXXC1	H3K4me3	E6.5–E7.5 [40]	CpG Rich Regions Tag	Prevention of DNA methylation
PRC2(EZH2)	miRNA	H3K27me3	E7.5-E10.5 [56]	Bivalent Promoters	Fate Specification
EHMT1, EHMT2	Transcription Factors	H3K9me3	E9.5–E12.5 [57]	Facultative Chromatin	Tissue Differentiation
PRC1(EZH1)	CBX proteins	H2AK119ub	Normal [48]	Active genes [58]	Cell-specific programs

Table 2: RNA-Directed DNA Transcription. Histone writers described in the text, listing, modification, lethality in homozygous null embryos and the chromatin feature they are associated with. (me: methyl; ub: ubiquitin; CBX: chromobox proteins; CXXC: Cysteines separated by two other amino acids; PRC: Polycomb Repressive Complex; SET: **S**u(var)3-9, **E**nhancer-of-Zeste and **T**rithorax; EZH: **E**nhancer-of-**Z**este; EHMT: Euchromatic histone-lysine N-methyltransferase; siRNA: small interfering RNA)

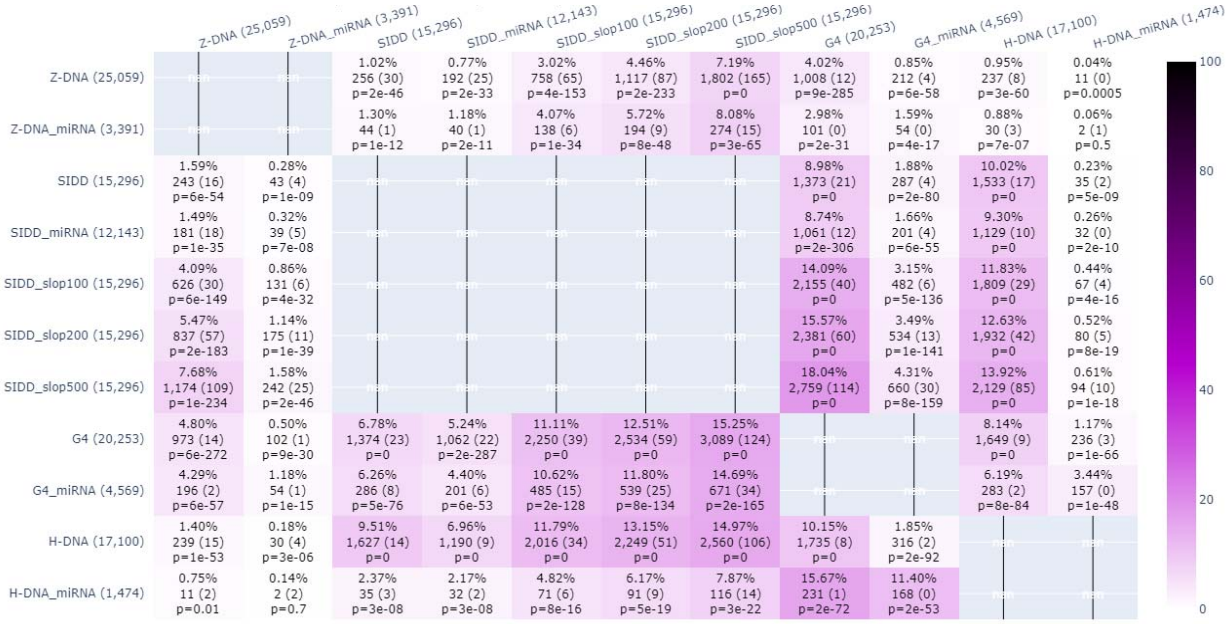


Supplementary Figure 1

The genomic location of ChIP-seq peaks for AGO1 and AGO2 in human HepG2 liver cells annotated by feature.



Supplementary Figure 2 The location of flipons within genomic features A. Mouse genome **B.** Location of flipons matching a conserved miRNA seed **C**. Flipons within 100 basepairs of a SIDD **D.** Flipons within 200 basepairs of a SIDD **E.** Flipons within 500 basepairs of a SIDD.



Supplementary Figure 3

Enumeration of the overlap at different distances between the different features examined in the KMnO4 footprinting dataset [22].

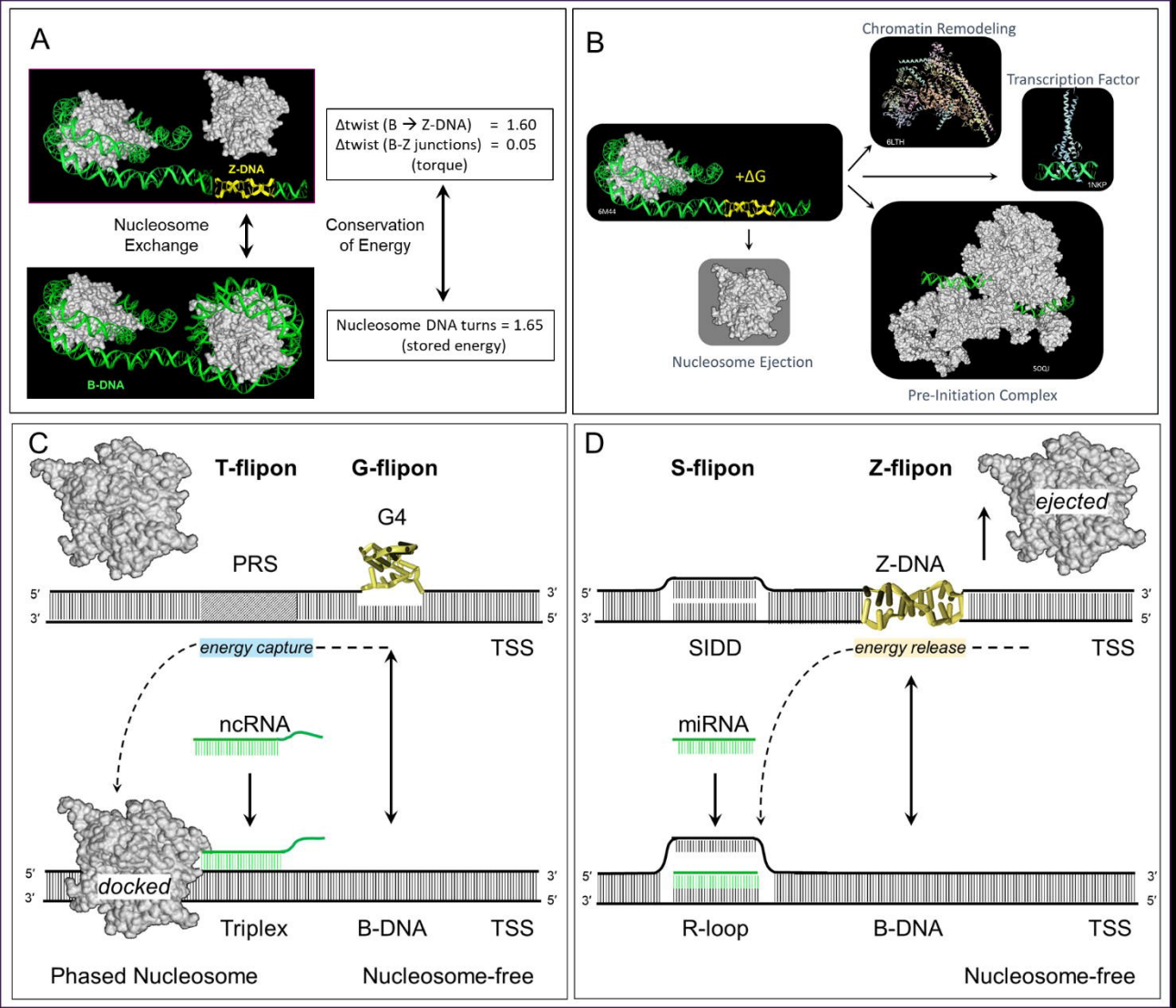
References

1. Britten, R.J.; Davidson, E.H. Gene Regulation for Higher Cells: A Theory. *Science* **1969**, *165*, 349-357, doi:10.1126/science.165.3891.349.
2. Shi, J.; Zhou, T.; Chen, Q. Exploring the expanding universe of small RNAs. *Nat Cell Biol* **2022**, doi:10.1038/s41556-022-00880-5.
3. Herbert, A. A Genetic Instruction Code Based on DNA Conformation. *Trends Genet* **2019**, *35*, 887–890, doi:10.1016/j.tig.2019.09.007.
4. Bartel, D.P. Metazoan MicroRNAs. *Cell* **2018**, *173*, 20-51, doi:10.1016/j.cell.2018.03.006.
5. Horvitz, H.R. Worms, Life, and Death (Nobel Lecture). *ChemBioChem* **2003**, *4*, 697-711, doi:10.1002/cbic.200300614.
6. Cheloufi, S.; Dos Santos, C.O.; Chong, M.M.; Hannon, G.J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **2010**, *465*, 584-589, doi:10.1038/nature09092.
7. Liu, H.; Lei, C.; He, Q.; Pan, Z.; Xiao, D.; Tao, Y. Nuclear functions of mammalian MicroRNAs in gene regulation, immunity and cancer. *Mol Cancer* **2018**, *17*, 64, doi:10.1186/s12943-018-0765-5.
8. Ozata, D.M.; Gainetdinov, I.; Zoch, A.; O'Carroll, D.; Zamore, P.D. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **2019**, *20*, 89-108, doi:10.1038/s41576-018-0073-3.
9. Natarelli, L.; Weber, C. A Non-Canonical Link between Non-Coding RNAs and Cardiovascular Diseases. *Biomedicines* **2022**, *10*, doi:10.3390/biomedicines10020445.
10. Allo, M.; Agirre, E.; Bessonov, S.; Bertucci, P.; Gomez Acuna, L.; Buggiano, V.; Bellora, N.; Singh, B.; Petrillo, E.; Blaustein, M.; et al. Argonaute-1 binds transcriptional enhancers and controls constitutive and alternative splicing in human cells. *Proc Natl Acad Sci U S A* **2014**, *111*, 15622-15629, doi:10.1073/pnas.1416858111.
11. Agirre, E.; Bellora, N.; Allo, M.; Pages, A.; Bertucci, P.; Kornblihtt, A.R.; Eyras, E. A chromatin code for alternative splicing involving a putative association between CTCF and HP1alpha proteins. *BMC Biol* **2015**, *13*, 31, doi:10.1186/s12915-015-0141-5.
12. Chu, Y.; Yokota, S.; Liu, J.; Kilikevicius, A.; Johnson, K.C.; Corey, D.R. Argonaute binding within human nuclear RNA and its impact on alternative splicing. *Rna* **2021**, *27*, 991-1003, doi:10.1261/rna.078707.121.
13. Herbert, A. ALU non-B-DNA conformations, flipons, binary codes and evolution. *Royal Society Open Science* **2020**, *7*, 200222, doi:10.1098/rsos.200222.
14. Herbert, A. The Simple Biology of Flipons and Condensates Enhances the Evolution of Complexity. *Molecules* **2021**, *26*, doi:10.3390/molecules26164881.
15. Peck, L.J.; Wang, J.C. Energetics of B-to-Z transition in DNA. *Proc Natl Acad Sci U S A* **1983**, *80*, 6206-6210.
16. Kouzine, F.; Wojtowicz, D.; Yamane, A.; Resch, W.; Kieffer-Kwon, K.R.; Bandle, R.; Nelson, S.; Nakahashi, H.; Awasthi, P.; Feigenbaum, L.; et al. Global regulation of promoter melting in naive lymphocytes. *Cell* **2013**, *153*, 988-999, doi:10.1016/j.cell.2013.04.033.
17. Naughton, C.; Avlonitis, N.; Corless, S.; Prendergast, J.G.; Mati, I.K.; Eijk, P.P.; Cockroft, S.L.; Bradley, M.; Ylstra, B.; Gilbert, N. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* **2013**, *20*, 387-395, doi:10.1038/nsmb.2509.

18. Teves, S.S.; Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nat Struct Mol Biol* **2013**, *21*, 88-94, doi:10.1038/nsmb.2723.
19. Baranello, L.; Wojtowicz, D.; Cui, K.; Devaiah, B.N.; Chung, H.J.; Chan-Salis, K.Y.; Guha, R.; Wilson, K.; Zhang, X.; Zhang, H.; et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **2016**, *165*, 357-371, doi:10.1016/j.cell.2016.02.036.
20. Shin, S.I.; Ham, S.; Park, J.; Seo, S.H.; Lim, C.H.; Jeon, H.; Huh, J.; Roh, T.Y. Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res* **2016**, doi:10.1093/dnares/dsw031.
21. Marsico, G.; Chambers, V.S.; Sahakyan, A.B.; McCauley, P.; Boutell, J.M.; Antonio, M.D.; Balasubramanian, S. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res* **2019**, *47*, 3862-3874, doi:10.1093/nar/gkz179.
22. Kouzine, F.; Wojtowicz, D.; Baranello, L.; Yamane, A.; Nelson, S.; Resch, W.; Kieffer-Kwon, K.R.; Benham, C.J.; Casellas, R.; Przytycka, T.M.; et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst* **2017**, *4*, 344-356, doi:10.1016/j.cels.2017.01.013.
23. Wang, H.; Noordewier, M.; Benham, C.J. Stress-Induced DNA Duplex Destabilization (SIDD) in the E. coli Genome: SIDD Sites Are Closely Associated With Promoters. *Genome Res* **2004**, *14*, 1575-1584, doi:10.1101/gr.2080004.
24. Nadel, J.; Athanasiadou, R.; Lemetre, C.; Wijetunga, N.A.; P, O.B.; Sato, H.; Zhang, Z.; Jeddeloh, J.; Montagna, C.; Golden, A.; et al. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* **2015**, *8*, 46, doi:10.1186/s13072-015-0040-6.
25. Chedin, F.; Benham, C.J. Emerging roles for R-loop structures in the management of topological stress. *Journal of Biological Chemistry* **2020**, *295*, 4684-4695, doi:10.1074/jbc.REV119.006364.
26. Paugh, S.W.; Coss, D.R.; Bao, J.; Laudermilk, L.T.; Grace, C.R.; Ferreira, A.M.; Waddell, M.B.; Ridout, G.; Naeve, D.; Leuze, M.; et al. MicroRNAs Form Triplexes with Double Stranded DNA at Sequence-Specific Binding Sites; a Eukaryotic Mechanism via which microRNAs Could Directly Alter Gene Expression. *PLoS Comput Biol* **2016**, *12*, e1004744, doi:10.1371/journal.pcbi.1004744.
27. Senturk Cetin, N.; Kuo, C.C.; Ribarska, T.; Li, R.; Costa, I.G.; Grummt, I. Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res* **2019**, *47*, 2306-2321, doi:10.1093/nar/gky1305.
28. Maldonado, R.; Schwartz, U.; Silberhorn, E.; Langst, G. Nucleosomes Stabilize ssRNA-dsDNA Triple Helices in Human Cells. *Mol Cell* **2019**, *73*, 1243-1254 e1246, doi:10.1016/j.molcel.2019.01.007.
29. Levens, D. Cellular MYCro economics: Balancing MYC function with MYC expression. *Cold Spring Harb Perspect Med* **2013**, *3*, doi:10.1101/cshperspect.a014233.
30. Valverde, R.; Edwards, L.; Regan, L. Structure and function of KH domains. *Febs J* **2008**, *275*, 2712-2726, doi:10.1111/j.1742-4658.2008.06411.x.
31. Yi, R.; Poy, M.N.; Stoffel, M.; Fuchs, E. A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* **2008**, *452*, 225-229, doi:10.1038/nature06642.
32. Bian, H.; Zhou, Y.; Zhou, D.; Zhang, Y.; Shang, D.; Qi, J. The latest progress on miR-374 and its functional implications in physiological and pathological processes. *J Cell Mol Med* **2019**, *23*, 3063-3076, doi:10.1111/jcmm.14219.
33. Bissels, U.; Wild, S.; Tomiuk, S.; Holste, A.; Hafner, M.; Tuschl, T.; Bosio, A. Absolute quantification of microRNAs by using a universal reference. *Rna* **2009**, *15*, 2375-2384, doi:10.1261/rna.1754109.

34. Herbert, A. Simple Repeats as Building Blocks for Genetic Computers. *Trends Genet* **2020**, doi:10.1016/j.tig.2020.06.012.
35. Van Nostrand, E.L.; Freese, P.; Pratt, G.A.; Wang, X.; Wei, X.; Xiao, R.; Blue, S.M.; Chen, J.Y.; Cody, N.A.L.; Dominguez, D.; et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **2020**, *583*, 711-719, doi:10.1038/s41586-020-2077-3.
36. Watanabe, T.; Totoki, Y.; Toyoda, A.; Kaneda, M.; Kuramochi-Miyagawa, S.; Obata, Y.; Chiba, H.; Kohara, Y.; Kono, T.; Nakano, T.; et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **2008**, *453*, 539-543, doi:10.1038/nature06908.
37. Tam, O.H.; Aravin, A.A.; Stein, P.; Girard, A.; Murchison, E.P.; Cheloufi, S.; Hodges, E.; Anger, M.; Sachidanandam, R.; Schultz, R.M.; et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **2008**, *453*, 534-538, doi:10.1038/nature06904.
38. Kang, Y.K. Surveillance of Retroelement Expression and Nucleic-Acid Immunity by Histone Methyltransferase SETDB1. *Bioessays* **2018**, *40*, e1800058, doi:10.1002/bies.201800058.
39. Kumar, D.; Cinghu, S.; Oldfield, A.J.; Yang, P.; Jothi, R. Decoding the function of bivalent chromatin in development and cancer. *Genome Res* **2021**, doi:10.1101/gr.275736.121.
40. Bledau, A.S.; Schmidt, K.; Neumann, K.; Hill, U.; Ciotta, G.; Gupta, A.; Torres, D.C.; Fu, J.; Kranz, A.; Stewart, A.F.; et al. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* **2014**, *141*, 1022-1035, doi:10.1242/dev.098152.
41. Xu, C.; Bian, C.; Lam, R.; Dong, A.; Min, J. The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat Commun* **2011**, *2*, 227, doi:10.1038/ncomms1237.
42. Guo, Y.; Zhao, S.; Wang, G.G. Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 'Readout', and Phase Separation-Based Compaction. *Trends Genet* **2021**, *37*, 547-565, doi:10.1016/j.tig.2020.12.006.
43. Kim, D.H.; Villeneuve, L.M.; Morris, K.V.; Rossi, J.J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **2006**, *13*, 793-797, doi:10.1038/nsmb1142.
44. Zardo, G.; Ciolfi, A.; Vian, L.; Billi, M.; Racanicchi, S.; Grignani, F.; Nervi, C. Transcriptional targeting by microRNA-polycomb complexes: a novel route in cell fate determination. *Cell Cycle* **2012**, *11*, 3543-3549, doi:10.4161/cc.21468.
45. Kalantari, R.; Chiang, C.M.; Corey, D.R. Regulation of mammalian transcription and splicing by Nuclear RNAi. *Nucleic Acids Res* **2016**, *44*, 524-537, doi:10.1093/nar/gkv1305.
46. Jeon, A.J.; Tucker-Kellogg, G. Bivalent genes that undergo transcriptional switching identify networks of key regulators of embryonic stem cell differentiation. *BMC Genomics* **2020**, *21*, 614, doi:10.1186/s12864-020-07009-8.
47. Ai, S.; Yu, X.; Li, Y.; Peng, Y.; Li, C.; Yue, Y.; Tao, G.; Li, C.; Pu, W.T.; He, A. Divergent Requirements for EZH1 in Heart Development Versus Regeneration. *Circ Res* **2017**, *121*, 106-112, doi:10.1161/CIRCRESAHA.117.311212.
48. Ezhkova, E.; Lien, W.H.; Stokes, N.; Pasolli, H.A.; Silva, J.M.; Fuchs, E. EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev* **2011**, *25*, 485-498, doi:10.1101/gad.2019811.
49. Chetta, M.; Di Pietro, L.; Bukvic, N.; Lattanzi, W. Rising Roles of Small Noncoding RNAs in Cotranscriptional Regulation: In Silico Study of miRNA and piRNA Regulatory Network in Humans. *Genes (Basel)* **2020**, *11*, doi:10.3390/genes11050482.
50. Yang, Q.; Lin, J.; Liu, M.; Li, R.; Tian, B.; Zhang, X.; Xu, B.; Liu, M.; Zhang, X.; Li, Y.; et al. Highly sensitive sequencing reveals dynamic modifications and activities of small RNAs in mouse oocytes and early embryos. *Sci Adv* **2016**, *2*, e1501482, doi:10.1126/sciadv.1501482.

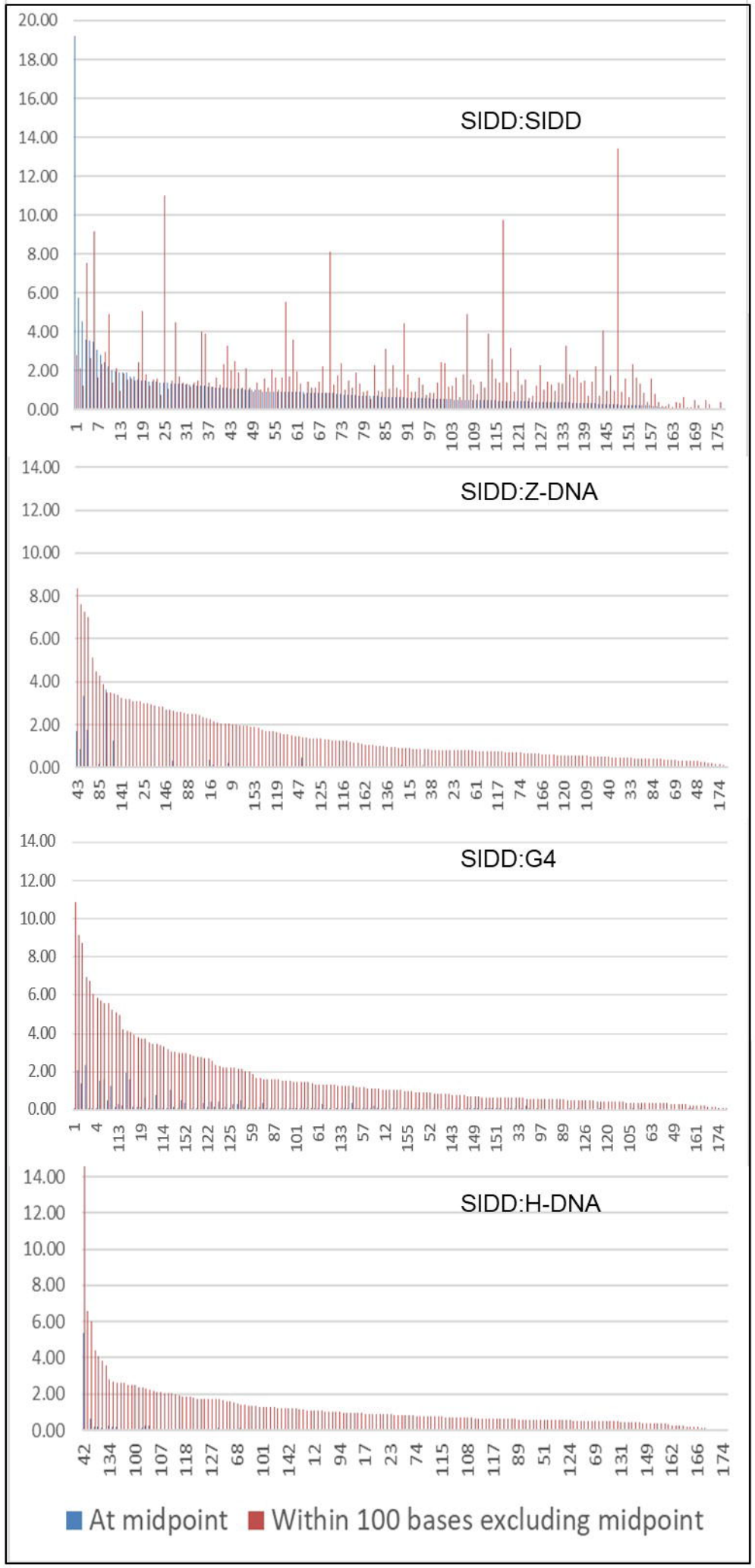
51. Gapp, K.; Bohacek, J. Epigenetic germline inheritance in mammals: looking to the past to understand the future. *Genes Brain Behav* **2018**, *17*, e12407, doi:10.1111/gbb.12407.
52. Herbert, A. The four Rs of RNA-directed evolution. *Nat Genet* **2004**, *36*, 19-25, doi:10.1038/ng1275.
53. Wu, T.; Lyu, R.; You, Q.; He, C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat Methods* **2020**, *17*, 515-523, doi:10.1038/s41592-020-0797-9.
54. 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 U S A* **2007**, *104*, 2229-2234, doi:10.1073/pnas.0611447104.
55. Dodge, J.E.; Kang, Y.K.; Beppu, H.; Lei, H.; Li, E. Histone H3-K9 methyltransferase ESET is essential for early development. *Mol Cell Biol* **2004**, *24*, 2478-2486, doi:10.1128/MCB.24.6.2478-2486.2004.
56. O'Carroll, D.; Erhardt, S.; Pagani, M.; Barton, S.C.; Surani, M.A.; Jenuwein, T. The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **2001**, *21*, 4330-4336, doi:10.1128/MCB.21.13.4330-4336.2001.
57. Tachibana, M.; Sugimoto, K.; Nozaki, M.; Ueda, J.; Ohta, T.; Ohki, M.; Fukuda, M.; Takeda, N.; Niida, H.; Kato, H.; et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **2002**, *16*, 1779-1791, doi:10.1101/gad.989402.
58. Mousavi, K.; Zare, H.; Wang, A.H.; Sartorelli, V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol Cell* **2012**, *45*, 255-262, doi:10.1016/j.molcel.2011.11.019.


A
Z-DNA
B-DNA
Nucleosome Exchange
Δtwist (B → Z-DNA) = 1.60
Δtwist (B-Z junctions) = 0.05
(torque)
Conservation of Energy
Nucleosome DNA turns = 1.65
(stored energy)
B
Chromatin Remodeling
Transcription Factor
6LTH
1NKP
+ΔG
6M44
Nucleosome Ejection
5OQJ
Pre-Initiation Complex
C
T-flipon
G-flipon
G4
PRS
5′
3′
3′
5′
energy capture
TSS
ncRNA
docked
Triplex
B-DNA
TSS
Phased Nucleosome
Nucleosome-free
D
S-flipon
Z-flipon
ejected
Z-DNA
SIDD
energy release
TSS
miRNA
R-loop
B-DNA
TSS
Nucleosome-free

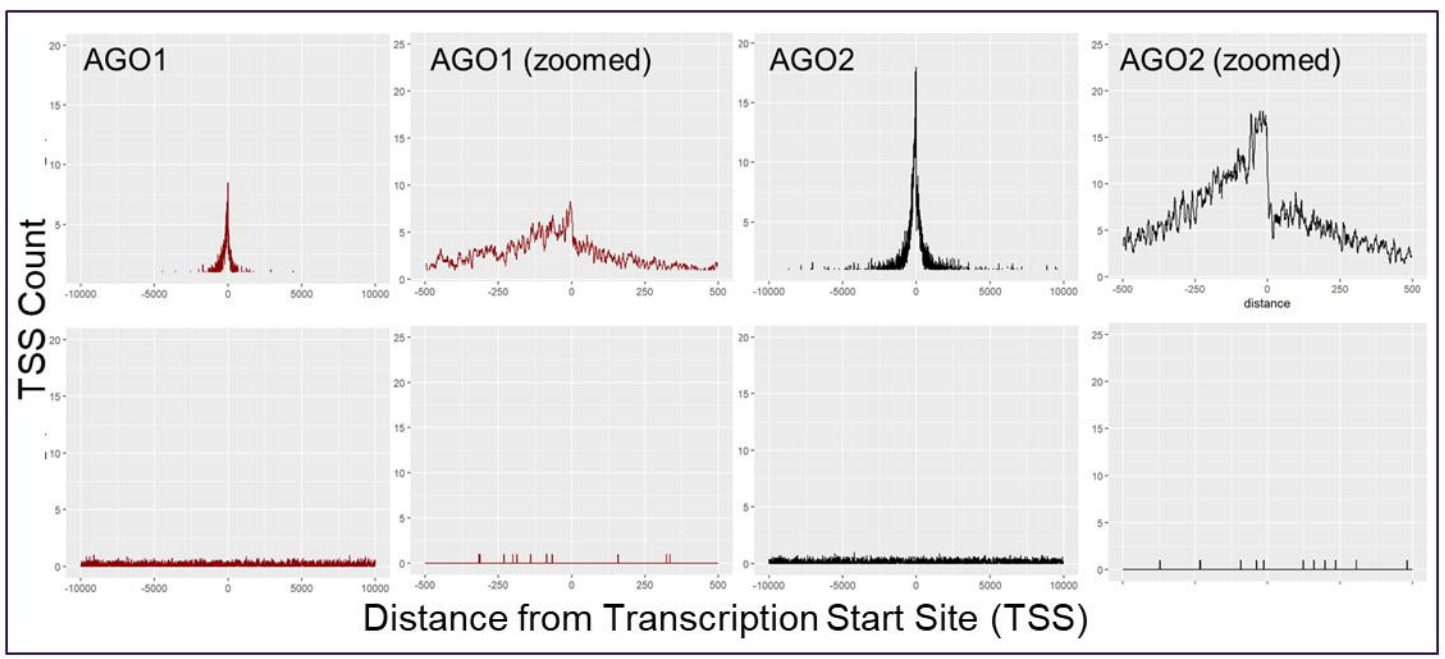
A

Genera	miRNA	Genes	Seed region	Z-DNA	G4	SIDD	H-DNA
Vertebrate	miR-203a	1	UGAAAUG, GAAAUGU	0.00	0.01	19.21	0.00
Mammalia	miR-186	1	AAAGAAU	0.00	0.00	5.75	0.12
Eutheria	miR-374	2	UAUAAUA	0.00	0.00	4.56	0.00
Theria	miR-340	1	UAUAAAG	0.00	0.01	3.61	0.00
Eutheria	miR-670	1	UUCCUCA	0.00	0.00	3.53	0.05
Vertebrate	miR-142	1	AUAAAGU, GUAGUGU, UAGUGUU	0.06	0.00	3.46	0.00
Eutheria	miR-495	1	AACAAAC	0.00	0.00	3.08	0.11
Vertebrate	miR-204/211	2	UCCCUUU	0.00	0.53	2.81	0.66
Chordata	miR-129	2	UUUUUGC, AGCCCUU	0.00	0.18	2.46	0.02
Eutheria	miR-330	1	CUCUGGG, CAAAGCA, AAAGCAC	0.03	0.53	2.21	0.00
Mammalia	miR-325	1	UUAUUGA	0.00	0.00	2.00	0.00
Eutheria	miR-409	1	GGUUACC, AAUGUUG	0.00	0.03	1.94	0.00
Eutheria	miR-655	1	UAAUACA	0.00	0.00	1.92	0.00
Eutheria	miR-382	1	AAGUUGU, AUCAUUC	0.00	0.00	1.92	0.00
Vertebrate	miR-216b	1	AAUCUCU	0.00	0.00	1.88	0.01
Eutheria	miR-448	1	UGCAUAU	0.34	0.00	1.71	0.00
Eutheria	miR-653	1	UGAAACA	0.00	0.00	1.67	0.00
Eutheria	miR-493	1	UGUACAU, GAAGGUC	0.31	0.02	1.56	0.00
Vertebrate	miR-140	1	AGUGGUU, CCACAGG, ACCACAG	0.01	0.17	1.49	0.00
Eutheria	miR-188	1	AUCCCUU	0.00	0.31	1.48	0.03
Eutheria	miR-452	1	ACUGUUU	0.00	0.00	1.44	0.00
Bilateria	miR-216a	1	AAUCUCA	0.00	0.00	1.43	0.00
Eutheria	miR-154	1	AGGUUAU, AUCAUAC	0.00	0.03	1.41	0.00
Eutheria	miR-369	1	AUAAUAC	0.00	0.00	1.41	0.00
Eutheria	miR-532	1	AUGCCUU, CUCCCAC	0.08	2.08	1.40	0.21
Eutheria	miR-410	1	AUAUAAC	0.00	0.00	1.39	0.00
Vertebrate	miR-23	2	UCACAUU	0.00	0.01	1.35	0.00
Mammalia	miR-873	1	CAGGAAC, GCAGGAA	0.02	0.08	1.33	0.12
Vertebrate	miR-181	6	ACAUUCA	0.03	0.00	1.32	0.00
Eutheria	miR-543	1	AACAUUC	0.00	0.00	1.31	0.00
Eutheria	miR-494	1	GAAACAU	0.00	0.00	1.29	0.01
Vertebrate	miR-30	6	GUAAACA	0.00	0.00	1.29	0.00
Eutheria	miR-582	1	UACAGUU	0.00	0.00	1.26	0.01
Eutheria	miR-433	1	UCAUGAU	0.00	0.00	1.24	0.00
Eutheria	miR-329/362	3	AUCCUUG (miR-362), ACACACC	3.36	0.45	1.22	0.00
Vertebrate	miR-455	1	AUGUGCC, CAGUCCA, UGCAGUC	1.26	0.05	1.22	0.00

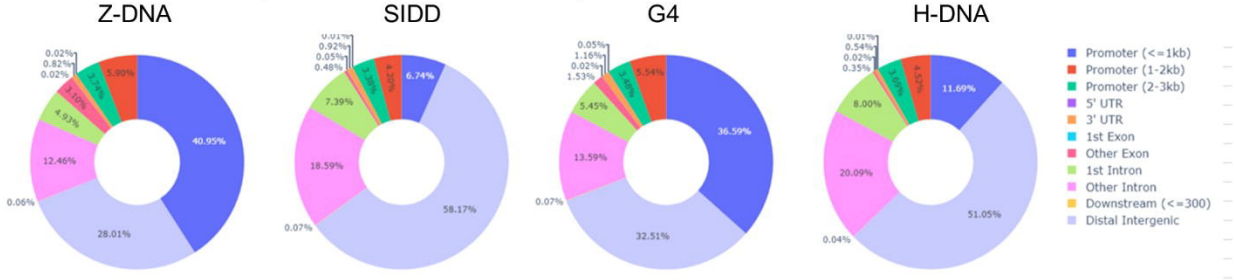
B



C



A. Genomic Location of Experimentally Determined Flipons

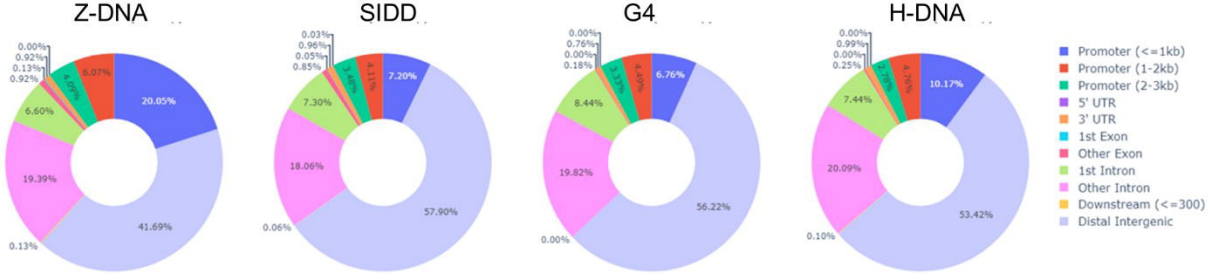


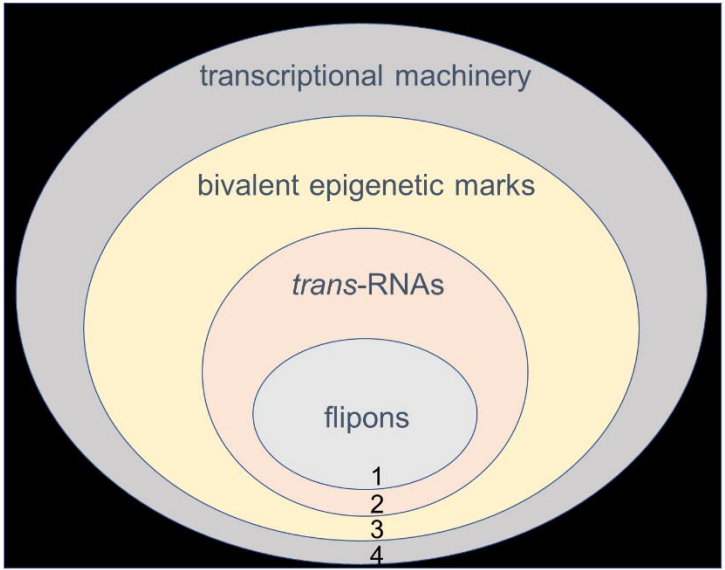
B. Genomic Overlap of Conserved miRNAs with Flipons singly or paired with SIDD at 100bp, 200bp or 500bp

		Absolute Counts															Percentage %											
		Exact match				Exact match +-100bp				Exact match +-200bp				Exact match +-500bp				Exact match				Exact match +-100bp				Exact match +-200bp				Exact match +-500bp			
	N	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Z-DNA	25,059	21,668	2,616	730	45	2,262	3,936	14,795	4,066	444	869	15,131	8,615	55	26	6,289	18,689	86.47	10.44	2.91	0.18	9.03	15.71	59.04	16.23	1.77	3.47	60.38	34.38	0.22	0.10	25.10	74.58
G4	20,253	15,684	3,736	770	63	1,031	2,372	12,984	3,866	161	359	12,269	7,464	11	14	4,676	15,552	77.44	18.45	3.80	0.31	5.09	11.71	64.11	19.09	0.79	1.77	60.58	36.85	0.05	0.07	23.09	76.79
SIDD	15,296	3,153	4,117	6,273	1,753	135	520	10,354	4,287	10	35	8,762	6,489	1	0	2,977	12,318	20.61	26.92	41.01	11.46	0.88	3.40	67.69	28.03	0.07	0.23	57.28	42.42	0.01	0.00	19.46	80.53
H-DNA	17,100	15,626	1,066	395	13	2,687	3,117	8,549	2,747	670	923	9,800	5,707	43	44	4,715	12,298	91.38	6.23	2.31	0.08	15.71	18.23	49.99	16.06	3.92	5.40	57.31	33.37	0.25	0.26	27.57	71.92
Z-DNA & SIDD (+-100bp)	758	620	95	41	2	58	116	441	143	3	20	472	263	0	0	209	549	81.79	12.53	5.41	0.26	7.65	15.30	58.18	18.87	0.40	2.64	62.27	34.70	0.00	0.00	27.57	72.43
G4 & SIDD (+-100bp)	2,250	1,765	424	59	2	54	197	1,439	560	4	11	1,311	924	0	0	508	1,742	78.44	18.84	2.62	0.09	2.40	8.76	63.96	24.89	0.18	0.49	58.27	41.07	0.00	0.00	22.58	77.42
H-DNA & SIDD (+-100bp)	2,010	1,945	65	5	1	173	372	1,103	368	14	41	1,242	719	0	0	494	1,522	96.77	3.23	0.25	0.05	8.61	18.51	54.88	18.31	0.70	2.04	61.79	35.77	0.00	0.00	24.58	75.72
Z-DNA & SIDD (+-200bp)	1,117	923	140	52	2	97	171	646	203	10	33	692	382	0	0	303	814	82.63	12.53	4.66	0.18	8.68	15.31	57.83	18.17	0.90	2.95	61.95	34.20	0.00	0.00	27.13	72.87
G4 & SIDD (+-200bp)	2,534	1,995	471	65	3	71	238	1,617	608	4	22	1,488	1,020	0	0	581	1,953	78.73	18.59	2.57	0.12	2.80	9.39	63.81	23.99	0.16	0.87	58.72	40.25	0.00	0.00	22.93	77.07
H-DNA & SIDD (+-200bp)	2,249	2,158	79	10	2	217	426	1,200	406	20	62	1,381	786	0	0	569	1,680	95.95	3.51	0.44	0.09	9.65	18.94	53.36	18.05	0.89	2.76	61.41	34.95	0.00	0.00	25.30	74.70
Z-DNA & SIDD (+-500bp)	1,802	1,528	209	63	2	153	289	1,067	293	17	48	1,135	602	0	0	484	1,318	84.79	11.60	3.50	0.11	8.49	16.04	59.21	16.26	0.94	2.66	62.99	33.41	0.00	0.00	26.86	73.14
G4 & SIDD (+-500bp)	3,089	2,418	581	84	6	99	298	1,998	694	6	32	1,833	1,218	0	0	712	2,377	78.28	18.81	2.72	0.19	3.20	9.65	64.68	22.47	0.19	1.04	59.34	39.43	0.00	0.00	23.05	76.95
H-DNA & SIDD (+-500bp)	2,560	2,444	96	17	3	262	483	1,366	449	35	74	1,559	892	0	0	657	1,903	95.47	3.75	0.66	0.12	10.23	18.87	53.36	17.54	1.37	2.89	60.90	34.84	0.00	0.00	25.66	74.34

T1: no miRNA overlap; T2:single miRNA; T3: multiple matches to same miRNA seed; T4: matches to different miRNA seeds

C. Genomic location of SIDD paired with Flipons 100bp either side





CpG Island
(H3K4me3)

Z-flipon S-flipon

5′ 3′ 3′ 5′

Z-DNA SIDD

1

miRNA

***trans*-RNA binding**

5′ 3′ 3′ 5′

B-DNA R-loop

2

Bivalent promoters
(H3K4me3, H3K27me3)

Polycomb Repressive Complex

3

Epigenetic Editing

Transcription Factor Complex

4

5′ 3′ 3′ 5′

Tissue Specification

Nucleosome-free unmethylated DNA