**SUPPLEMENTAL INFORMATION**

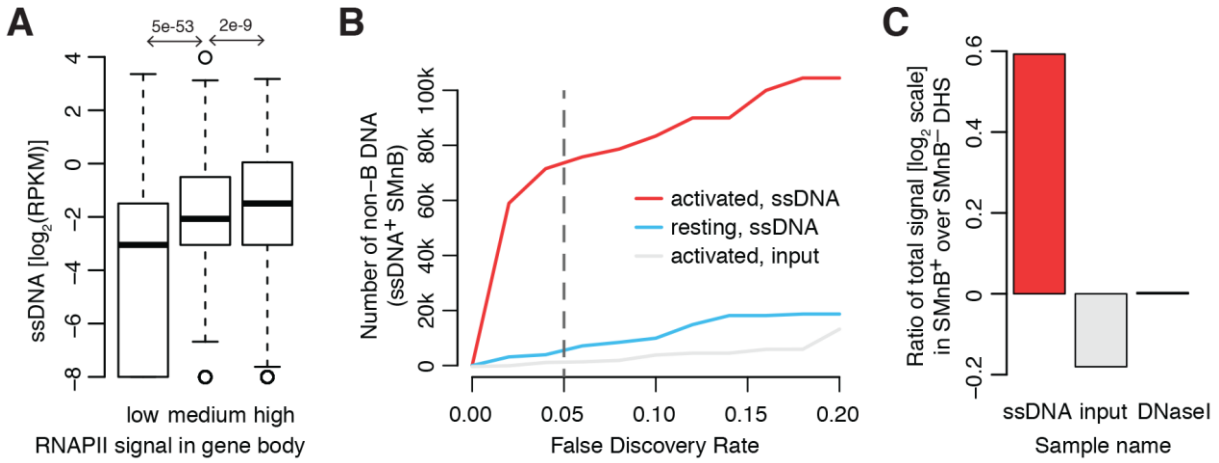


**Figure S1. Related to Figure 2. ssDNA-Seq tags are associated with transcription and with SMnB. A)** Density of ssDNA-Seq tags defined as reads per kilobase per million (RPKM), in the regions starting from –2 Kb to –1 Kb from TSS for low, medium and high transcriptional gene activity classified according to RNAPII occupancy in gene body (measured by ChIP-Seq assay). P-values (one-sided Wilcoxon rank sum test) are shown at the top of the plot; Cohen's d effect size of differences is –0.41 and –0.14, respectively. There are 2682, 1961 and 1907 genes in low, medium and high transcriptional gene activity groups, respectively. **B)** Dependence of the number of identified non-B DNA structures from the stringency of False Discovery Rate (FDR). The inflection from the steep to shallower linear increase in the number of identified structures at unique sites devoid of RNAPII in activated B-cells (red line) at FDR of 0.05 (dashed line) was considered as a sign of saturation and so was chosen as the FDR-threshold for this analysis. The blue line shows the same analysis in resting B-cells. As a control, the number of non-B DNA forming sequences enriched in the sequencing tags of input DNA from activated B-cells is shown. **C)** ssDNA-Seq signal in activated B-cells (red) is enriched in DNase I hypersensitive sites (DHS) containing SMnB as compared with SMnB-lacking DHS of equivalent accessibility (black) in intergenic regions; input DNA signal (gray) at $SMnB^+$ DHS over $SMnB^-$ DHS was shown as a control. P-value of one-sided Wilcoxon rank sum test for the difference in ssDNA-Seq signal between $SMnB^+$ *versus* $SMnB^-$ DHS in activated B-cells is 1.8e-16.

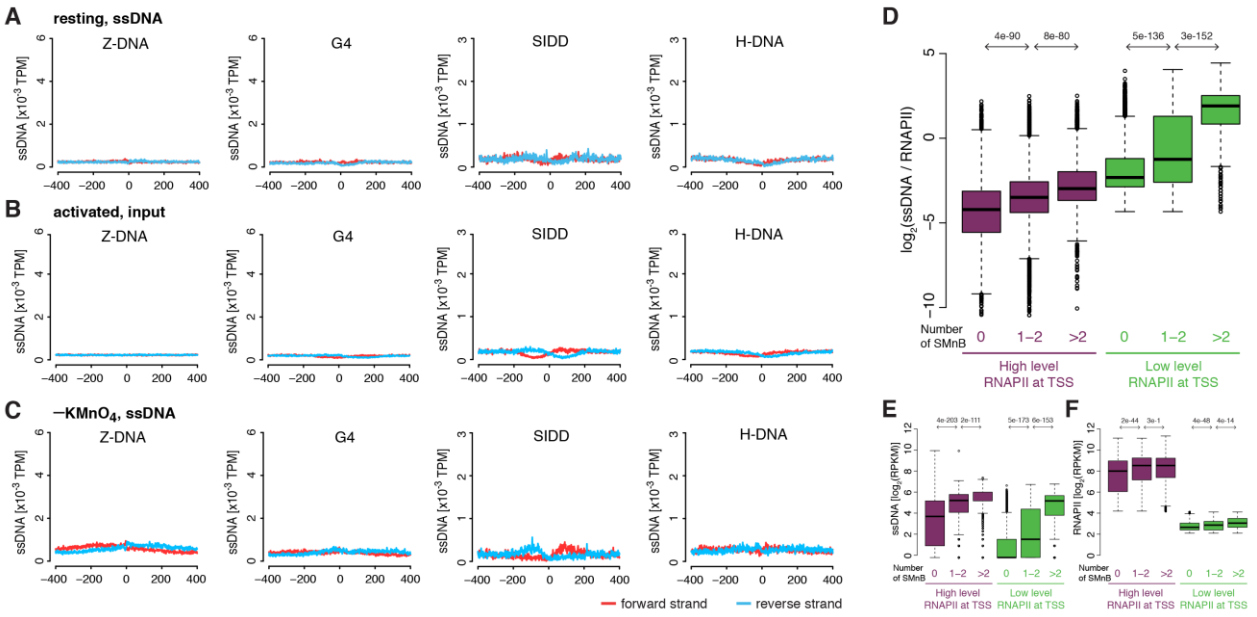


**Figure S2. Related to Figure 3. ssDNA-Seq tags are associated with SMnB in transcriptionally active lymphocytes. A)** Composite profiles of ssDNA-Seq tags for mouse resting B-cells, **(B)** sequencing tags for sonicated DNA from mouse activated B-cells (input), and **(C)** ssDNA-Seq tags for mouse activated B-cells not treated with potassium permanganate ($KMnO_4$) at the SMnB identified in **Figure 2A**. The *y*-axis shows the normalized number of sequence tags from the forward (red) and reverse (blue) strand of DNA at each position shown in the *x*-axis according to the center of the aligned non-B DNA susceptible motifs. **D)** Ratio between ssDNA over RNAPII density around the TSS (–0.5 Kb; +0.5 Kb) classified according to the number of SMnB found near TSS and two levels of RNAPII density (high – violet and low – green). Density of ssDNA-Seq tags **(E)** and density of RNAPII binding measured by ChIP-Seq **(F)** around TSSs classified according to the number of SMnB and density of RNAPII as above. P-values are shown at the top of the plots (one-sided Wilcoxon rank sum test); absolute Cohen's d effect size of all tests is above 0.4 in **(D)** and above 0.5 in **(E)**, and at most 0.3 in **(F)**. In the group of genes with high level of RNAPII at TSS, the number of genes with 0, 1-2 and >2 SMnB near the TSS is 3285, 5540 and 2922, respectively; in the low RNAPII level group the respective numbers are 6048, 3624 and 1044.

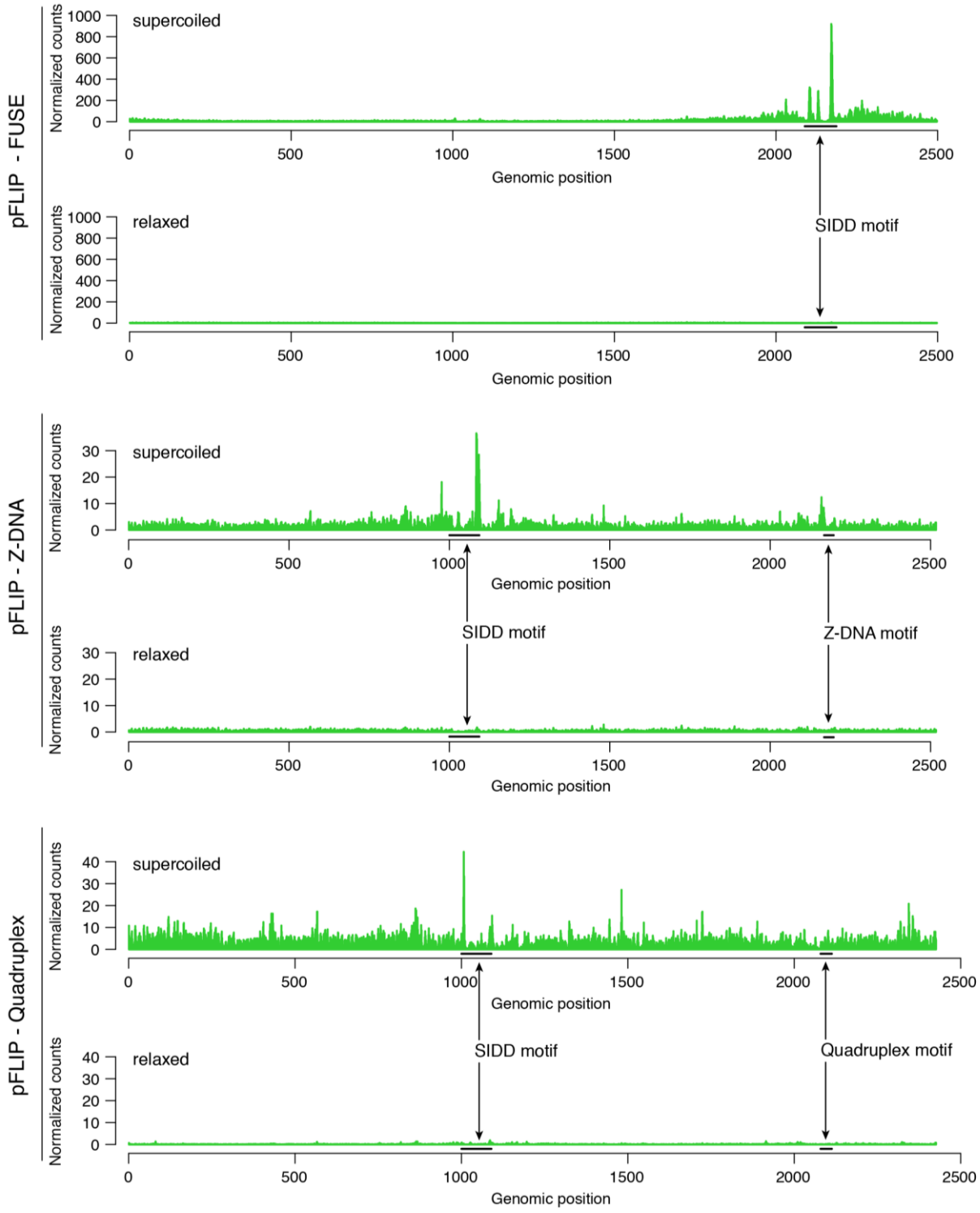


**Figure S3. Related to Figure 4. The relationship between peak morphology, supercoiling and SMnB, *in vitro.*** ssDNA-Seq tag profiles of relaxed and supercoiled plasmids (as in **Figure 4**) containing different prominent non-B DNA forming sequences (SMnB) indicated by black bars. These SMnB are associated with high ssDNA read counts in the supercoiled pFLIP-FUSE (SIDD, p-value 0.0009) and the pFLIP-ZDNA (Z-DNA and SIDD, p-value 0.003) plasmids.

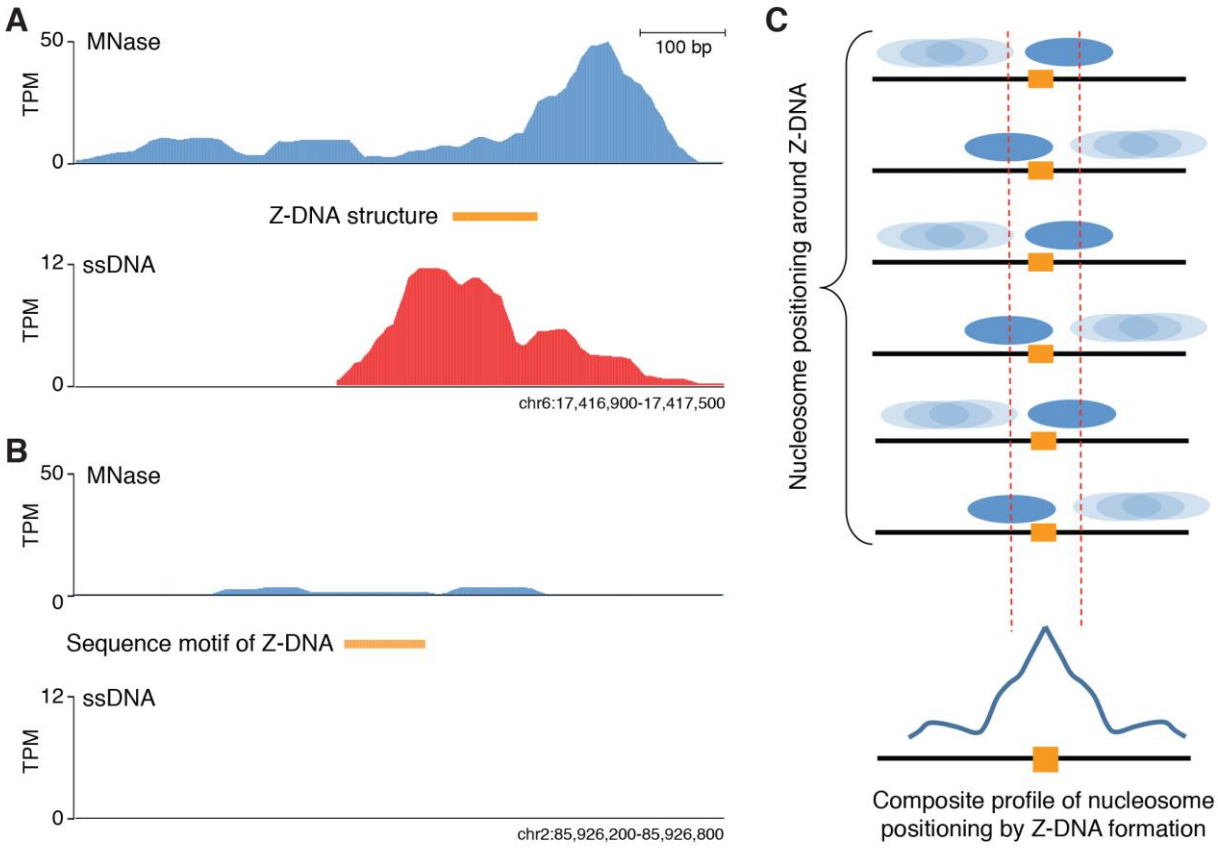


**Figure S4. Related to Figure 5. Z-DNA formation results in strong nucleosome positioning near the Z-B DNA junction. A)** The density of nucleosome occupancy in mouse activated B-cells near a sequence that forms Z-DNA structure is displayed in the UCSC genome browser. A segment of the MET proto-oncogene is shown with the genomic coordinates indicated. **B)** The density of nucleosome occupancy near an almost identical sequence as in **(A)**, with potential to form Z-DNA, but lacking the ssDNA-Seq signal, is shown in the UCSC genome browser. A segment of the intergenic region is shown with the genomic coordinates indicated. Lacking of prominent Mnase-Seq signal indicates that this DNA sequence though protected from Mnase, is not able to position nucleosomes. **C)** Schematic explanation of nucleosome positioning observed in the composite profile near Z-DNA structures as shown in the **Figure 5**.

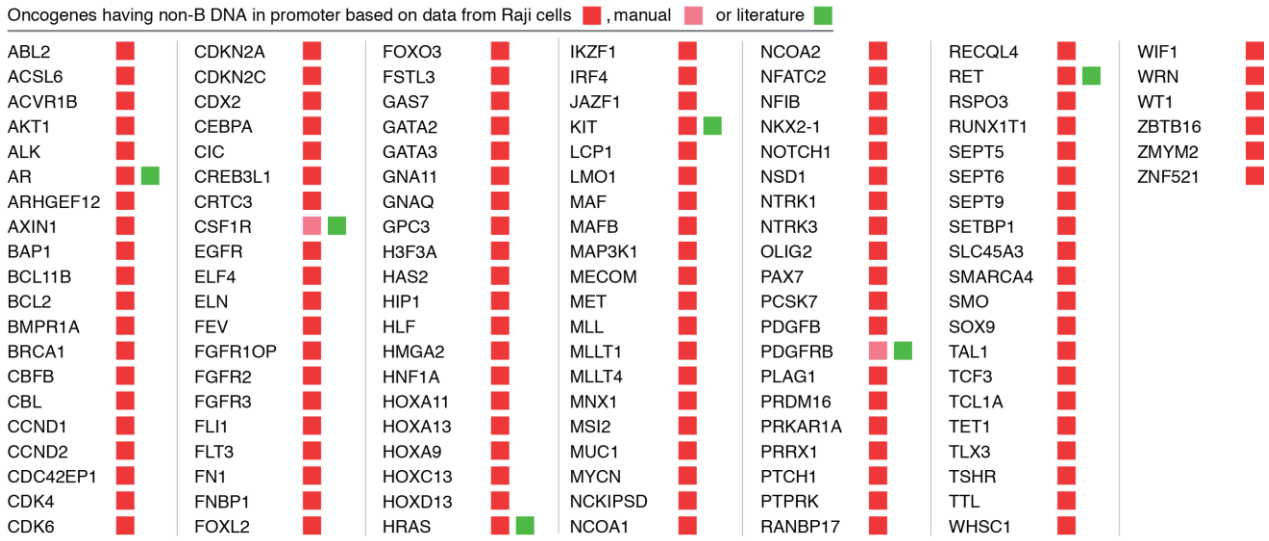


**Figure S5. Related to Figure 6. List of cancer-related genes with non-B DNA structures at the promoter area in Raji cells.** The green squares mark genes in which the importance of non-B DNA for gene function has been proposed in the literature. The pink squares mark individually inspected genes in which the ssDNA signal in the vicinity of the promoter is indicative of non-B DNA formation.

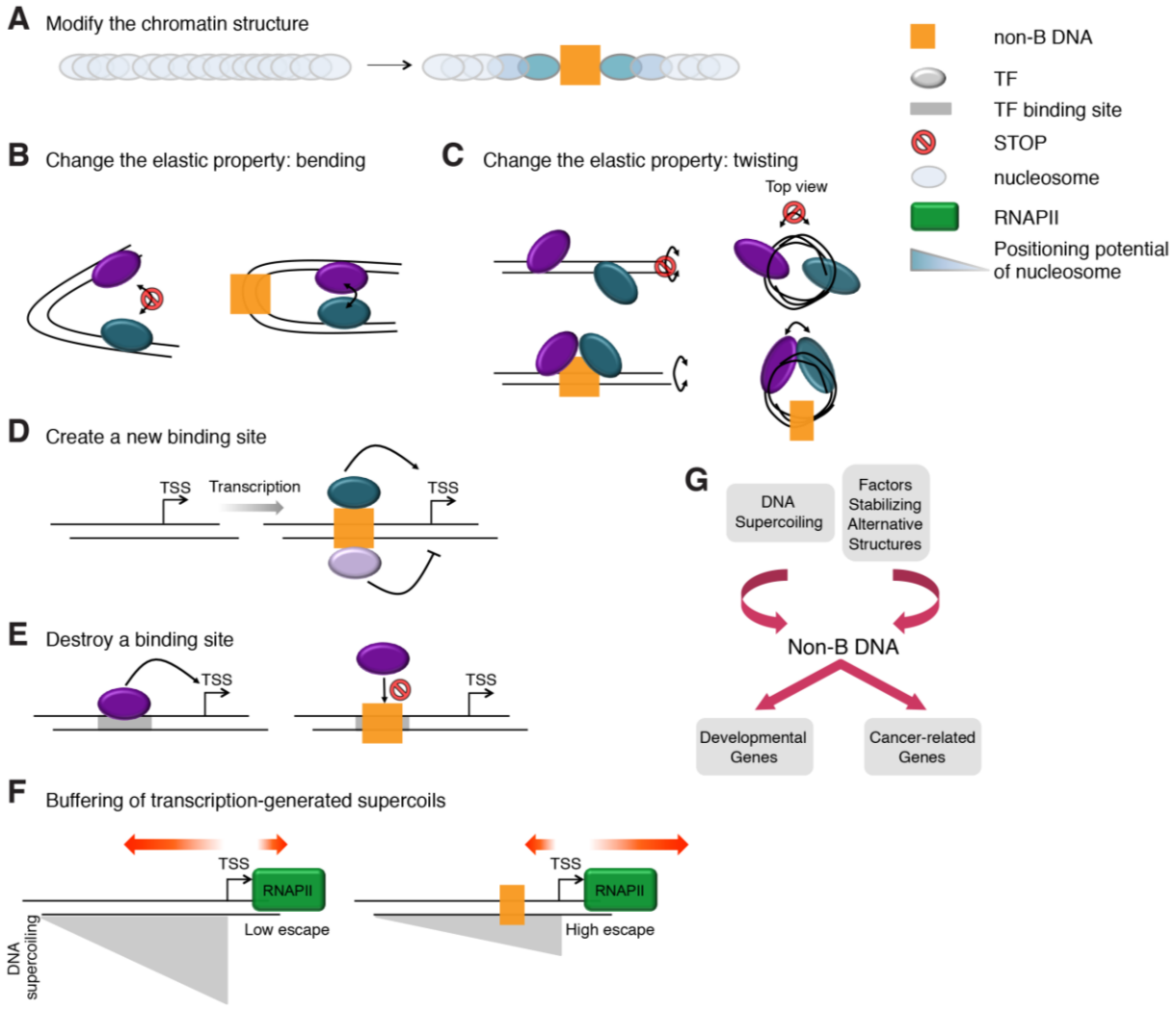


**Figure S6. Related to Figure 6. Potential regulatory mechanisms mediated by non-B DNA formation. A)** Non-B DNA structures reposition flanking nucleosomes. Non-B DNA structures provide flexibility for bending **(B)** and for twisting **(C)** to juxtapose separated regulatory elements. **D)** Non-B DNA structure formation results in the recruitment of DNA conformation-sensitive proteins with regulatory potential. **E)** The formation of non-B DNA structures disrupts B DNA binding sites by sterically blocking the binding of a transcription factor. **F)** A high level of transcription-generated supercoiling might alter the progression of RNAPII through rate-controlling steps of transcription. By absorbing supercoils dynamically, non-B DNA formation is positioned to buffer torsional stress and so it helps to coordinate transitions throughout the transcription cycle via mechanical coupling. **G)** The regulatory role of non-B DNA structures promoted by DNA supercoiling or stabilized by molecular partners is supported by the correlation of unusual DNA conformations with particular transcriptome characteristics, such as the manifestation of unusual DNAs structures near the TSSs of developmental and cancer-related genes.

**(A)** resting mouse B-cell

| Type of SMnB | SMnB not overlapping with RNAPII | | | SMnB overlapping with RNAPII | | |
|---|---|---|---|---|---|---|
| | Analyzed SMnB | ssDNA+ SMnB | Percent | Analyzed SMnB | ssDNA+ SMnB | Percent |
| Z-DNA | 248912 | 1023 | 0.4% | 42739 | 220 | 0.5% |
| G4 | 265630 | 2780 | 1.0% | 32495 | 194 | 0.6% |
| SIDD | 417815 | 878 | 0.2% | 29559 | 85 | 0.3% |
| H-DNA | 322131 | 1054 | 0.3% | 19300 | 79 | 0.4% |

**(B)** human Raji cells

| Non-B DNA type | SMnB not overlapping with RNAPII | | | SMnB overlapping with RNAPII | | |
|---|---|---|---|---|---|---|
| | Analyzed SMnB | ssDNA+ SMnB | Percent | Analyzed SMnB | ssDNA+ SMnB | Percent |
| Z-DNA | 194898 | 39023 | 20.0% | 31668 | 4488 | 14.2% |
| G4 | 180327 | 47628 | 26.4% | 32102 | 5454 | 17.0% |
| SIDD | 357389 | 11754 | 3.3% | 35978 | 154 | 0.4% |
| H-DNA | 144643 | 8008 | 5.5% | 13727 | 348 | 2.5% |

**Table S1. Related to Figure 6 and to Table 1. The extent of non-B DNA formation in mouse and human genomes. A)** Total number of predicted SMnB and the number of identified non-B DNA structures (ssDNA$^+$ SMnB motifs) in activated and resting B-cells for SMnB that do not overlap and do overlap with RNAPII binding sites. **B)** Total number of predicted SMnB and the number of identified non-B DNA structures (ssDNA$^+$ SMnB motifs) in Raji cells, a human Burkitt lymphoma cell line, for SMnB that do not overlap and do overlap with RNAPII binding sites.

| Type of data | | Number of ssDNA islands | Percentage |
|---|---|---|---|
| Total | | 252,828 | 100% |
| Non-B DNA susceptible motifs (SMnB) | Z-DNA | 46,513 | 18% |
| | Quadruplex | 38,237 | 15% |
| | SIDD | 73,349 | 29% |
| | H-DNA | 51,332 | 20% |
| | All | 144,819 | 57% |
| Other types of data | SINE | 145,014 | 57% |
| | Annotated genes | 102,116 | 40% |
| | Observed RNAPII binding | 48,330 | 19% |
| | All | 185,830 | 74% |
| **All SMnB + other types** | | **213,048** | **84%** |
| **NOT** other types of data | | 66,998 | 26% |
| All SMnB & **NOT** other types of data | | 27,218 | 41% |

**Table S2. Related to Table 1. Overlapping of ssDNA sites with other genomic features.**

Total number ssDNA-Seq islands identified by SICER v1.03 (window 100 bp, gap 100 bp, E-value 10,000) in mouse activated B-cells and the number and the percentage of islands that do overlap (+/– 100bp) with different types of SMnB (Z-DNA, quadruplex, SIDD, and H-DNA), SINE, annotated genes, and experimentally identified RNAPII binding sites (identified by SICER from RNAPII ChIP-Seq data). The cumulative numbers for all types of non-B DNA motifs, all other types of data and all types of data are also shown. The last two rows of the table show the number of islands that do not overlap with other types of data and how many of them do overlap with SMnB.

| Gene Ontology | GO Term | P-value |
|---|---|---|
| anatomical structure development | GO:0048856 | 9.60E-77 |
| cell differentiation | GO:0030154 | 4.09E-40 |
| cell-cell signaling | GO:0007267 | 1.39E-26 |
| cell motility | GO:0048870 | 1.85E-23 |
| anatomical structure formation involved in morphogenesis | GO:0048646 | 3.66E-22 |
| locomotion | GO:0040011 | 1.12E-21 |
| cell adhesion | GO:0007155 | 3.04E-16 |
| transmembrane transport | GO:0055085 | 4.04E-15 |
| extracellular matrix organization | GO:0030198 | 9.64E-10 |
| cell morphogenesis | GO:0000902 | 5.46E-08 |
| signal transduction | GO:0007165 | 7.64E-08 |

**Table S3. Related to Figure 6. The results of Gene Ontology (GO) enrichment analysis.**

The analysis was performed using GOrilla tool and generic GO slim filtering. All genes bearing non-B DNA structures in their promoter regions were queried with background consisting of all protein-coding genes. All generic GO slim terms with uncorrected p-value below 1.0e-7 are shown.

| Set of quadruplexes | G4 not overlapping with RNAPII | | | G4 overlapping with RNAPII | | |
|---|---|---|---|---|---|---|
| | Analyzed G4 | ssDNA+ G4 | Percent | Analyzed G4 | ssDNA+ G4 | Percent |
| All predicted G4 (our dataset) | 180327 | 47628 | 26.4% | 32102 | 5454 | 17.0% |
| Subset of predicted G4 verified by G4-Seq | 99222 | 28879 | 29.1% | 19055 | 3177 | 16.7% |

**Table S4. Related to Table 1. Comparison of identified quadruplexes in Raji cells (*in vivo*) and quadruplexes detected *in vitro* by G4-Seq.** Total number of predicted canonical quadruplex motifs and the number of identified quadruplex structures in Raji cells for motifs that do not overlap and do overlap with RNAPII binding sites (top rows). The same for canonical quadruplex motifs that do overlap with regions detected by G4-Seq (bottom rows). Combining the computational SMnB prediction with *in vitro* verification for structure formation slightly increases percentage of identified quadruplex structures *in vivo*.