Supplemental Material

Genome-wide profiling reveals functional interplay of DNA sequence composition, transcriptional activity, and nucleosome positioning in driving DNA supercoiling and helix destabilization in C. elegans embryos

Kristina Krassovsky,1 Rajarshi P. Ghosh,1,2 and Barbara J. Meyer1,2

1Department of Molecular and Cell Biology
University of California, Berkeley
Berkeley, California 94720-3204, USA

2Howard Hughes Medical Institute

Corresponding Author:
Barbara J. Meyer
Email: bjmeyer@berkeley.edu
Tel.: +1 510-643-5585

Running Title:
Genome-wide mapping of DNA supercoiling and SS DNA

Key words:
DNA Supercoiling, Single-Stranded DNA, Nucleosome Positioning, X-Chromosome Dosage Compensation, Biotinylated Psoralen Photo-Crosslinking, Potassium Permanganate Reactivity, Caenorhabditis elegans
| Table of Contents |
|-------------------|
| Supplemental Figure S1 ................................................. 2 |
| Supplemental Figure S1 legend ........................................ 3 |
| Supplemental Figure S2 .................................................. 4 |
| Supplemental Figure S2 legend .......................................... 5 |
| Supplemental Figure S3 .................................................. 6 |
| Supplemental Figure S3 legend .......................................... 7 |
| Supplemental Figure S4 .................................................. 8 |
| Supplemental Figure S4 legend .......................................... 9 |
| Supplemental Figure S5 .................................................. 10 |
| Supplemental Figure S5 legend .......................................... 11 |
| Supplemental Figure S6 .................................................. 12 |
| Supplemental Figure S6 legend .......................................... 13 |
| Supplemental Figure S7 .................................................. 14 |
| Supplemental Figure S7 legend .......................................... 15 |
| Supplemental Figure S8 .................................................. 16 |
| Supplemental Figure S8 legend .......................................... 17 |
| Supplemental Figure S9 .................................................. 18 |
| Supplemental Figure S9 legend .......................................... 19 |
| Supplemental Figure S10 ............................................... 20 |
| Supplemental Figure S10 legend ....................................... 21 |
| Supplemental Figure S11 ............................................... 22 |
| Supplemental Figure S11 legend ....................................... 23 |
Run gel with ethidium bromide

Run gel without ethidium bromide

B

Experiment 1
Supercoiled
Relaxed

Experiment 2
Supercoiled
Relaxed

Experiment 3
Supercoiled
Relaxed

C

Fluorescence quantification

Fluorescence intensity

Number of experiment

supercoiled
relaxed
Supplemental Figure S1. Biotinylated psoralen (BP) incorporates preferentially into negatively supercoiled DNA.

(A) Images of supercoiled and relaxed plasmid DNA electrophoresed through agarose gels in the absence and presence of ethidium bromide to verify the supercoiling status of DNA.

(B) Representative dot blot images of supercoiled and relaxed plasmids incubated with BP and probed with streptavidin-conjugated horseradish peroxidase.

(C) Quantification of fluorescence in dot blots exemplified in (B) verifies preferential BP incorporation into supercoiled DNA versus relaxed DNA.
**Supplemental Figure S2.** Open chromatin is not the primary determinant of psoralen binding.

(A) (B) Heatmaps of GC content in the 2 kb region around (A) 13,596 low-to-moderately transcribed coding promoters and (B) 19,231 putative enhancers reveal GC-rich regions with sharp GC/AT boundaries.

(C) Heatmap of supercoils (wild-type embryos, replicate 1) in the coding promoters shown in (A) recapitulate the findings of Figure 2 showing that the boundaries of supercoiling at TSSs are marked by a sharp GC to AT transition.

(D) Heatmap of supercoils in the 2 kb region surrounding putative enhancers shown in (B) reveals little or no supercoiling, correlated with the lack of transcription at these putative enhancer sites. The center of each enhancer was defined by the center of the ATAC-seq peak.

The heatmaps for GC content were made as in Figure 2A, but with the bin for promoters being 20 bp × 136 bp, and the bin for enhancers being 20 bp × 192 bp.

(E) GRO-seq counts at the 13,596 coding promoters (blue) and 19,231 putative enhancers (green) are compared with GRO-seq counts at the TSSs of 4221 highly transcribed genes (red).

(F) Average profiles of ATAC-seq signal in embryos are plotted for the 2 kb region surrounding coding promoters (13,596), highly accessible putative enhancers (9677), and TSSs of highly transcribed genes (4221) at 50 bp resolution.

(G) Average profiles of BP-seq in embryos is plotted for a 2 kb region around coding promoters (13,596), highly accessible putative enhancers (9677), and TSSs of highly transcribed genes (4221) at 50 bp resolution.
Supercoils around random GC-rich sequences

ATAC-seq around random GC-rich sequences

Supercoils around random AT-rich sequences

Figure S3
**Supplemental Figure S3.** Random GC-rich sequences lack negative supercoiling and have weaker GC/AT boundaries than GC-rich sequences of coding promoters, putative enhancers and highly active TSSs.

(A) Heatmap of GC content around 5,862 random GC-rich (GC % >0.45) sequences. Sequences were sorted by their GC content, aligned on top of each other, and binned at 20 bp × 59 bp. To select DNA for the heatmap, 400 bp DNA sequences with an average GC content greater than 45% were identified, 1 kb was added to each side, and sequences were aligned at the center of the original 400 bp region. Each bin was colored according to its GC percentage. These random GC-rich regions have weaker GC/AT boundaries than promoters. The dashed line highlights the sequences with strong GC/AT boundaries.

(B) Heatmap of supercoils for wild-type embryos (replicate 1) at random GC-rich sequences does not reveal negative supercoiling.

(C) Heatmap of ATAC-seq profile around random GC-rich sequences. The dashed line highlights the ATAC-seq signal at sequences with strong GC/AT boundaries.

(D) Heatmap of supercoils for wild-type embryos (replicate 1) around random AT-rich sequences shows greatly attenuated BP-seq signal.
**Figure S4**

**A**

| Transcription (GRO-seq deciles) |
|----------------------------------|
| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |

**GRO-seq signal in starved L1s**

- Paused TSSs are TSSs with Pause index > 5

**B**

**GRO-seq signal in starved L1s**

**C**

**Average supercoils at paused and non-paused TSSs in starved L1**

**D**

**Average supercoils at paused and non-paused genes**

**E**

**GRO-seq at paused and non-paused genes**
**Supplemental Figure S4.** TSSs in starved L1 larvae have attenuated BP-seq signal and exhibit similar supercoiling profiles with and without RNA Pol II pausing.

(A) Formula for determining the pause index of a TSS. More paused genes occur in starved L1s than other developmental stages (Baugh and Sternberg 2006; Kruesi et al. 2013).

(B) Box and whisker plots of GRO-seq values for 4221 genes in starved L1 larvae after partitioning into deciles based on their average GRO-seq signal in the gene body. The lowest decile (decile 1) includes genes with the lowest 10% of GRO-seq values. A trend line through the medians of box plots for each decile shows a graded increase in GRO-seq values up to decile 6, followed by a much sharper ascent from deciles 7 to 10, similar to the trend seen in embryos (Figure 2D).

(C) Box and whisker plots of BP-seq signal in 100 bp intervals around TSSs correspond to the GRO-seq deciles shown in (B). BP-seq signal is generally not correlated with transcription level in starved L1 larvae.

(D) BP-seq profile at 50 bp resolution in the 2 kb region around TSSs with paused (344 TSSs) and non-paused (3876 TSSs) transcription.

(E) GRO-seq signal in bodies of genes having paused and non-paused transcription in starved L1s.
A

\[ \text{TMP} = \frac{\text{TMP nuclear DNA}}{\text{TMP genomic DNA}} \]

B Average supercoiling at active human TSS

C GC content around active human TSSs

D Average supercoiling at fly TSSs

E GC content around fly TSSs

Figure S5

High TMP - negative supercoils

Low TMP
**Supplemental Figure S5.** Comparison of supercoiling profiles and GC content at TSSs of human and fly cell lines reveal common regulatory themes.

(A) Supercoiling is quantified by the ratio of nuclear DNA with incorporation of 4,5',8-trimethylpsoralen (TMP) versus naked genomic DNA with TMP incorporation. High TMP incorporation indicates negative supercoils. To create supercoiling plots around TSSs, the ratio of TMP reads in nuclear DNA of cells versus naked genomic DNA was normalized to the average signal around each TSS in a 6 kb region.

(B) The average supercoiling profile at TSSs of 13,226 highly expressed genes in human mammary epithelial MCF10A cells shows a peak of positive supercoils, with an average spread of 3 kb.

(C) Average GC content around TSSs of highly expressed human genes in (B). The extent of high GC content at TSSs corresponds to the propagation of positive supercoils at these TSSs.

(D) Analysis of supercoiling at 13,883 TSSs in *Drosophila melanogaster* S2 cells reveals a peak of negative supercoils with an average spread of 1.2 kb.

(E) Average GC content around TSSs of S2 cells shown in (D) shows that AT-rich regions at TSSs marked by a sharp transition to GC-rich regions delineate the boundary of supercoiling.
Figure S6

A SDC-3 peaks in Embryo

- enhancer: 622
- promoter: 682
- exon: 153
- other: 391

B SDC-3 peaks in promoters

C SDC-3 peaks in enhancers

Supercoils at DCC binding sites

WT

DCC Mutant

BP

Number of SDC-3 peaks in promoters

Number of SDC-3 peaks in enhancers
Supplemental Figure S6. Supercoiling profiles at secondary, non-autonomous DCC binding sites, located primarily in promoters and enhancers, correlate with transcriptional status of the sites and do not change upon disruption of DCC binding.

(A) The pie chart classifies 1848 different secondary binding sites of DCC subunit SDC-3 by location in enhancers, promoters, other ATAC-seq peaks, and exons.

(B) (C) Heatmaps of supercoils are plotted at 50 bp resolution in the 2 kb region around secondary DCC binding sites of wild-type (replicate 1) and DCC-mutant embryos (sdc-2 mutant, replicate 1). DCC binding sites are grouped by their presence in promoters (B) or enhancers (C).
DCC mutant embryo, rep 1
WT Embryo, rep 1

WT Embryo, rep 2
DCC mutant embryo, rep 1
DCC mutant embryo, rep 2

TAD boundary
DCC-dependent TAD boundary
Simple repeats
**Supplemental Figure S7.** The dosage compensation complex does not create supercoiling domains on X-Chromosomes.

For each X-Chromosome and each autosome, BP-seq signal for wild-type and DCC mutant embryos was averaged over 1 kb sliding windows and plotted. TAD boundaries and percentage of simple repeats are shown under each chromosome. Chromosome regions enriched in simple repeats exhibit higher variation in BP signal between experiments, likely due to variable amplification during library preparation. *C. elegans* lacks large-scale supercoiling domains on any chromosome, and TAD boundaries are not enriched for supercoils.
SS-seq - a method to map single-stranded DNA.

Panel A: DNA digested with S1 nuclease (left), and modified with KMnO4 (right). KMnO4 modifies unpaired bases.

Panel B: SS-seq peak locations. Exposed ends ligated to biotinylated hairpin adapter.

Panel C: Predicted SS DNA sites.

Panel D: Observed SS DNA sites.

Panel E: Distribution of the nucleosome distances around predicted and observed SS DNA sites.

Panel F: Observed SS sites.

Figure S8
Supplemental Figure S8. SS-seq signal is decoupled from transcription but is strongly correlated with nucleosome positioning.

(A) Method to map single-stranded DNA and other non-B forms of DNA. When KMnO₄ enters purified nuclei, it modifies unpaired pyrimidine bases and thereby prevents their reannealing, creating stable SS DNA. Purified DNA is digested with the single-strand-specific S1 nuclease, and exposed DNA ends are ligated to biotinylated hairpin adapters. Following DNA sonication to 300 bp, biotinylated DNA ends are enriched on streptavidin beads, and a second set of adapters is ligated. After digestion of adapter hairpins, the library is amplified and sequenced. The same procedure is applied to naked genomic DNA as a control. The distribution of the difference of counts from KMnO₄-treated nuclear DNA and genomic control DNA is calculated genome wide.

(B) Pie chart shows the distribution and number of SS-seq peaks in exons, introns, and extragenic regions.

(C) (D) Predicted (C) and Observed (D) SS DNA sites were classified by their presence in low, medium, or highly expressed genes or in extragenic sites. Cumulative SS-seq signal in a 100 bp window around the center of each SS-seq peak was calculated. Box plots show distribution of cumulative SS-seq signal for peaks coincident with each gene expression category or extragenic sites.

(E) (F) Symmetric nucleosomes occur at a higher frequency in observed SS-seq peaks than in predicted SS DNA sites. Shown are 2D histograms of distances between nucleosomes nearest to the right or left of SS DNA in predicted (E) and observed (F) sites.
Figure S9

A  Embryo, rep 1

B  Embryo, rep 2

C  GC content around TASs for 9918 genes with TASs ≥ 50 bp from TSSs

D  GC content around TESs for 9918 genes with TASs ≥ 50 bp from TSSs
**Supplemental Figure S9.** SS-seq signal scales with AT content at TESs and TASs, but not with levels of transcription.

(A) (B) Average SS-seq profiles are plotted at 50 bp resolution in the 2 kb regions surrounding TASs and TESs for high (3307 genes) and low (3306 genes) expression genes.

(C) (D) Heatmap of GC content in the 2 kb region around TASs (C) and TESs (D) of 9,918 genes reveals a minor increase in AT content at TASs and a robust increase at TESs.
Pearson correlation between pairwise combinations of BP-seq datasets in 500 bp windows around TSSs of the 4223 genes analyzed in Figure 2

A

Wild-type embryo, rep 1

Pearson = 0.42

Wild-type embryo, rep 2

B

DCC mutant embryo, rep 1

Pearson = 0.5

DCC mutant embryo, rep 2

C

Wild-type embryo, rep 2

Pearson = 0.65

DCC mutant embryo, rep 2

D

Starved L1, rep 1

Pearson = 0.7

Starved L1, rep 2
Supplemental Figure S10. Pearson’s correlation between BP-seq datasets of wild-type or DCC mutant embryos and starved L1 larvae. (A)-(D) Total BP signal was calculated in 500 bp windows around TSSs of highly transcribed genes. The correlation between samples is indicated by the Pearson's correlation coefficient.
Pearson correlation between pairwise combinations of SS-seq datasets in 100 bp windows around TSSs of 4246 genes analyzed in Figure 6A and around TASs and corresponding TESs for 9918 genes with TASs ≥ 50 bp from TSSs

Figure S11
**Supplemental Figure S11.** Pearson’s correlation between SS-seq datasets.

Total SS-seq signal was calculated in a 100 bp window around TSSs of highly transcribed genes and around TASs and TESs of genes with TASs at least 50 bp away from TSSs. The correlation between samples is indicated by the Pearson's correlation coefficient.