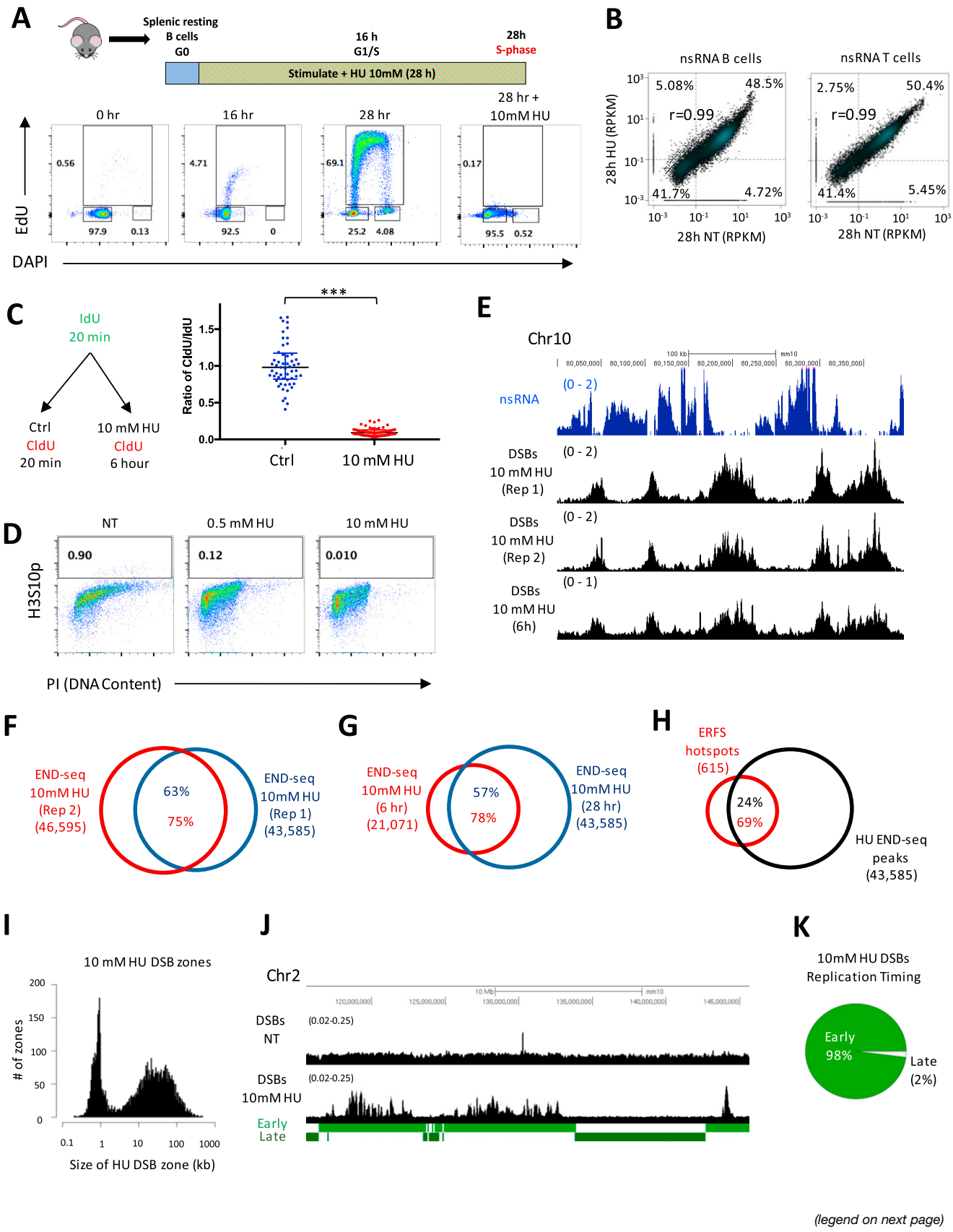
Supplemental Figures



(legend on next page)

Figure S1. Replication Origins Are Prone to Fork Collapse, Related to Figure 1

(A) Mouse splenic B cells were activated for 0 hr, 16 hr, and 28 hr (+/− 10 mM HU) then pulsed with EdU for 15 minutes to measure DNA synthesis by FACS. FACS plots show EdU incorporation (nsDNA synthesis) and DAPI (DNA content).

(B) Scatterplots for nsRNA at RefSeq genes in B cells (NT versus 10 mM HU) and T cells (NT versus 2 mM HU), all of which were activated for 28 hours.

(C) Molecular combing was performed on DNA fibers on B cells activated for 22 hours. Lengths of green-red/IdU-CldU replication tracks (n = 57 for control and 63 for 10mN HU 6h) were measured and the CldU/IdU ratio was calculated. The median and interquartile range are shown on the dot plot. Mean track lengths (kb): Ctrl IdU 49.28; Ctrl CldU 49.63; 10 mM HU IdU 41.71; 10 mM HU CldU < 3.54. ***$p < 0.0001$, Mann Whitney test.

(D) FACS analysis for the mitosis marker H3S10p in murine B cells treated with 0 (NT), 0.5 mM, and 10 mM HU for 28 hours.

(E) Genome browser shots showing normalized read counts (RPM) for nsDNA and DSBs detected by END-seq in activated B cells treated with 10 mM HU for either 6 hours or 28 hours.

(F) Venn diagram showing overlap of DSB peaks determined by END-seq in activated B cells (10 mM HU, 28 hours) from independent biological replicates. Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(G) Venn diagram showing overlap of END-seq DSBs in activated B cells treated with 10 mM HU for 6 hours and 28 hours. Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(H) Venn diagram comparing 43,585 END-seq DSBs with 615 ERFS hotspots previously identified (Barlow et al., 2013). Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(I) Histogram showing size distribution for DSB zones determined by merging DSB peaks within 20kb of each other in activated B cells treated with 10 mM HU.

(J) UCSC genome browser shots showing normalized read counts (RPM) for DSBs over a large ~50 Mb region on Chr2. DSBs are shown for activated B cells either non-treated (NT) or treated with 10 mM HU for 28 hours. Replication timing is shown, determined using TimEX ratios from resting and activated B cells, highlighting DSB enrichment in early-replicating regions of the chromosome.

(K) Distribution of DSBs induced by 10 mM HU in activated B cells in early and late replicating regions. Replication timing was determined using TimEX ratios from resting and activated B cells.

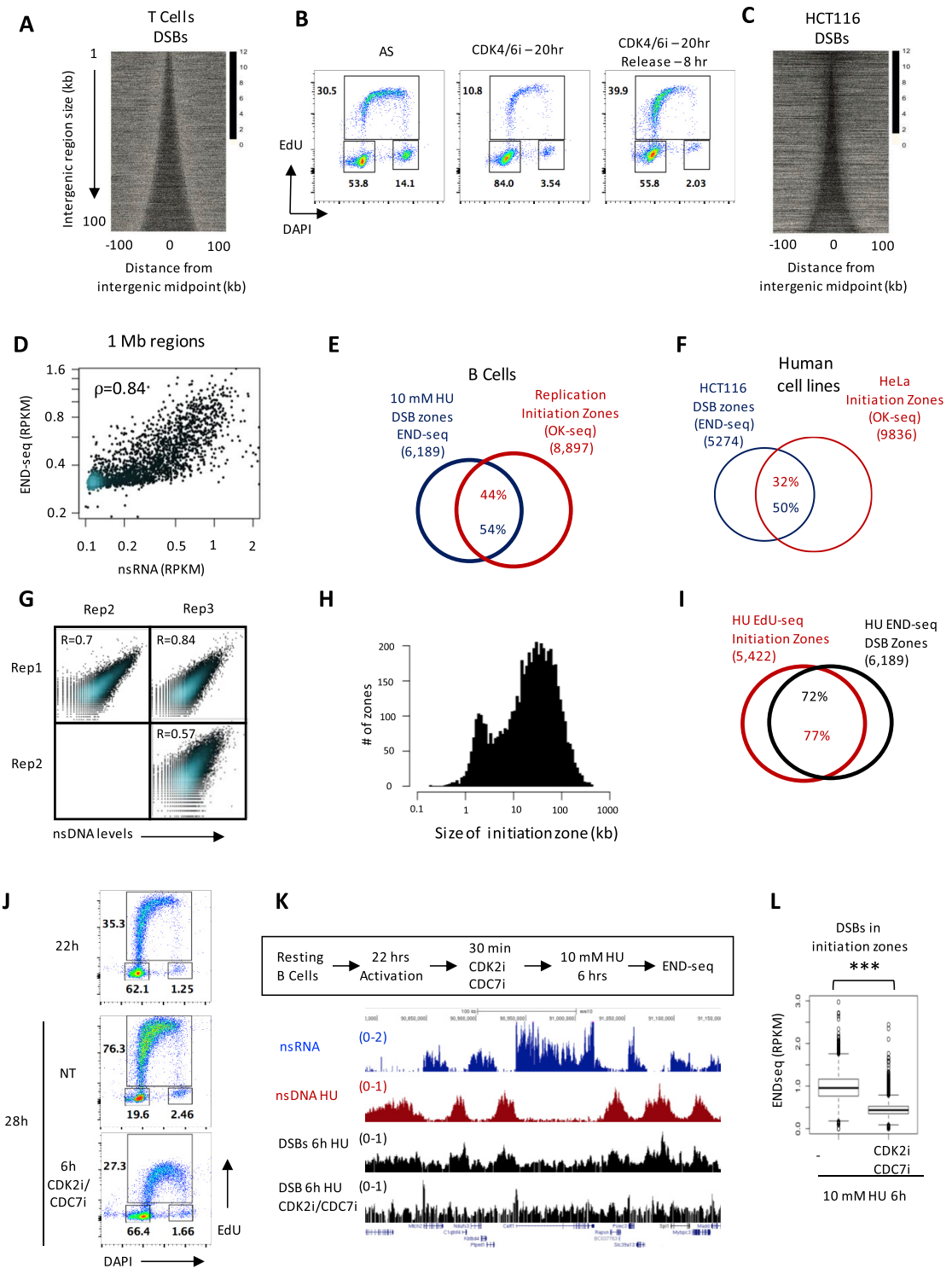


Figure S2. Relationship of Transcription and Replication Fork Collapse, Related to Figure 1

(A) Heatmap for END-seq reads within intergenic regions in activated T cells treated with 2 mM HU for 28 hours, ordered by the size of the intergenic region and centered on the midpoint of the intergenic region between two active genes.

(B) FACS profiles for cell cycle synchronization of HCT116 cells pulsed with EdU. Asynchronous (AS) HCT116 cells were treated with CDK4/6i (Palbociclib, 2uM) for 20 hours to arrest in G0/G1 (middle panel), followed by a wash in fresh media for 8 hours to release into S phase (right panel).

(legend continued on next page)

(C) Heatmap for END-seq reads within intergenic regions in HCT116 cells ordered by the size of the intergenic region and centered on the midpoint of the intergenic region between two active genes. Cells were synchronized in G1 using CDK4/6 inhibitors (Palbociclib) for 20 hours and released into 10 mM HU for 8 hours.

(D) Scatterplot illustrating correlation between transcriptional activity (nsRNA RPKM) and DSB frequency (END-seq RPKM) over 1 Mb regions in activated murine B cells treated with 10 mM HU for 28 hr.

(E) Venn diagram showing overlap of 6,189 DSB zones with 8,897 initiation zones determined by OK-seq. DSB zones determined by END-seq in activated B cells treated with 10 mM HU. OK-seq was performed in proliferating B cells without HU, and replication initiation zones were predicted based on RFD profiles (STAR Methods). Overlap versus random, $p < 10^{-10}$, Fisher's exact test.

(F) Venn diagram showing overlap of 5,274 DSB zones in HCT116 cells determined by END-seq and 9,836 replication initiation zones in HeLa cells using OK-seq (Petryk et al., 2016). Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(G) Scatterplots comparing nsDNA levels from HU-EdU-seq in activated B cells from three independent replicates. Cells were treated with 10 mM HU and 20 mM EdU for 28 hours. Each dot represents a single nsDNA peak called by MACS.

(H) Histogram showing size distribution for replication initiation zones determined by HU-EdU-seq in activated B cells treated with 10 mM HU. HU-EdU-seq initiation zones were created by merging closely-spaced (less than 20kb) nsDNA peaks (see text; STAR Methods).

(I) Venn diagram showing overlap of DSB zones determined by END-seq and initiation zones mapped by HU-EdU-seq in activated B cells treated with 10 mM HU for 28 hours. Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(J) FACS plots showing EdU incorporation in murine B cells activated for the indicated time and treatment (NT versus CDK2i+CDC7i). FACS plots show EdU incorporation (DNA synthesis) and DAPI (DNA content), and gates for G0-G1, S-phase, and G2-M.

(K) Top panel: experimental flow chart, where CDK2i/CDC7i are added 30 minutes prior to HU treatment to block new origin firing and prevent fork collapse near origins upon HU treatment. Bottom panel: genome browser shots showing normalized read counts (RPM) for nsRNA, nsDNA (+ 10 mM HU, 28 hours), END-seq DSBs (+ 10 mM HU, 6 hours), and END-seq DSBs (CDK2i+CDC7i + 10 mM HU, 6 hours).

(L) Boxplot for normalized END-seq reads within 5,422 replication initiation zones detected in activated B cells treated with 10 mM HU for 6 hours ± CDK2i+CDC7i, as in Figure S2J. ***$p < 10^{10}$, Wilcoxon rank-sum test.

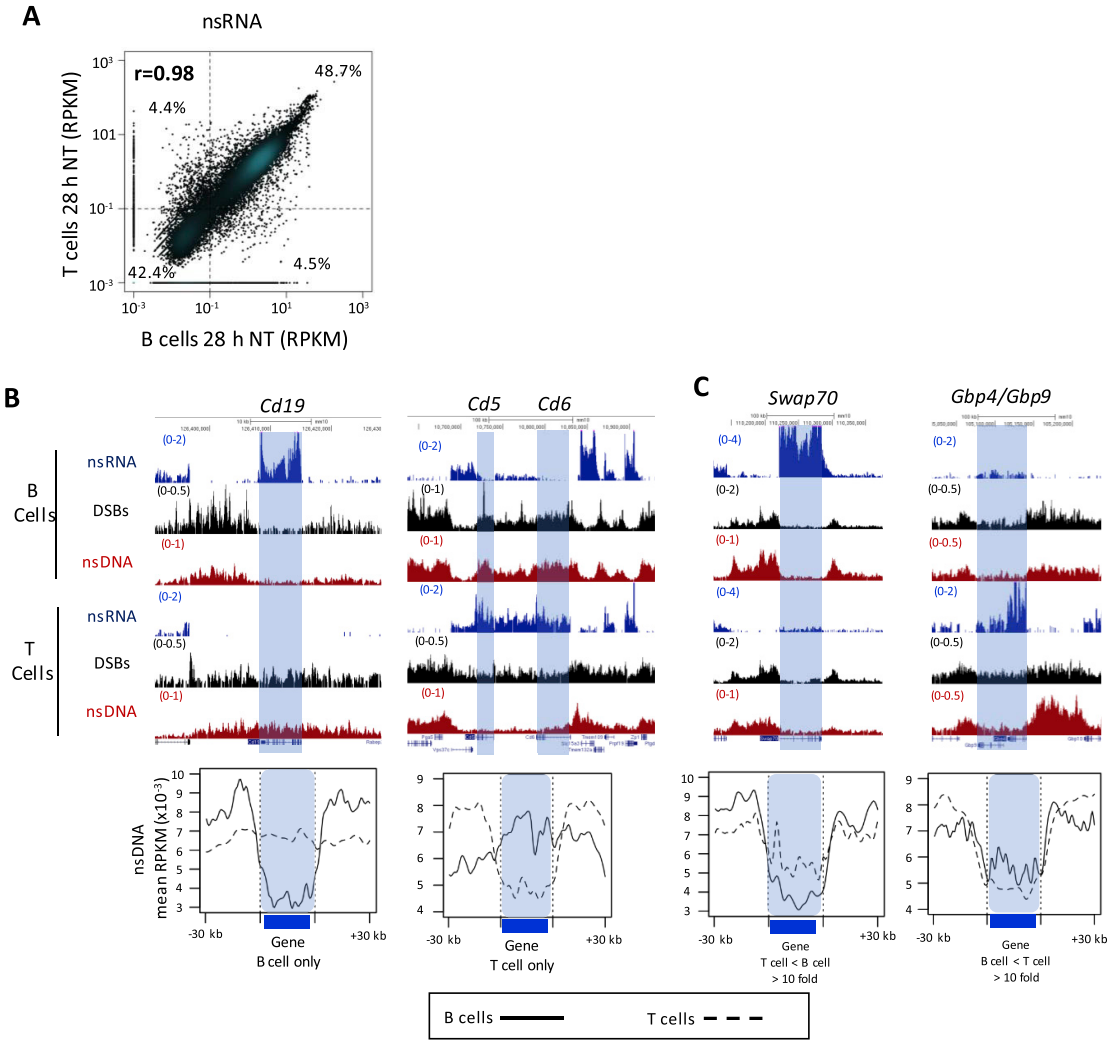


Figure S3. Replication Initiation Zones Are Prone to Replication Fork Collapse, Related to Figure 1

(A) Scatterplot comparing nsRNA (RPKM) in B cells and T cells activated for 28 hours. Upper left quadrant: Genes expressed in T cells only. Upper right quadrant: Genes expressed in both B and T cells. Lower Right quadrant: Genes expressed in B cells only. Lower left quadrant: Non-expressed in both cell types.

(B) Top panels: Genome browser shots comparing normalized read counts (RPM) for nsRNA, DSBs from END-seq, and nsDNA from HU-EdU-seq in activated B cells and T cells. *Cd19* is expressed exclusively in B cells. *Cd5* and *Cd6* are expressed exclusively in T cells. Bottom Left: Composite plots for nsDNA reads within genes expressed exclusively in B cells and 30kb flanking the gene (n = 93). Bottom Right: Composite plots for nsDNA reads within genes expressed exclusively in T cells and 30kb flanking the gene (n = 65).

(C) Top panels: Genome browser shots comparing normalized read counts (RPM) for nsRNA, DSBs from END-seq, and nsDNA from HU-EdU-seq in activated B cells and T cells treated with HU for 28 hours. *Swap70* is expressed 33-fold higher in B cells than in T cells. *Gbp4/Gbp9* are expressed 19-fold higher in T cells compared to B cells. Bottom Left: Composite plots for nsDNA within genes expressed at least 10-fold more in 0 mM HU, in B cells compared to T cells and a 30kb window flanking the gene (n = 80). Bottom Right: Composite plots for nsDNA within genes expressed at least 10-fold more in 0 mM HU, in T cells compared to B cells and a 30kb window flanking the gene (n = 31).

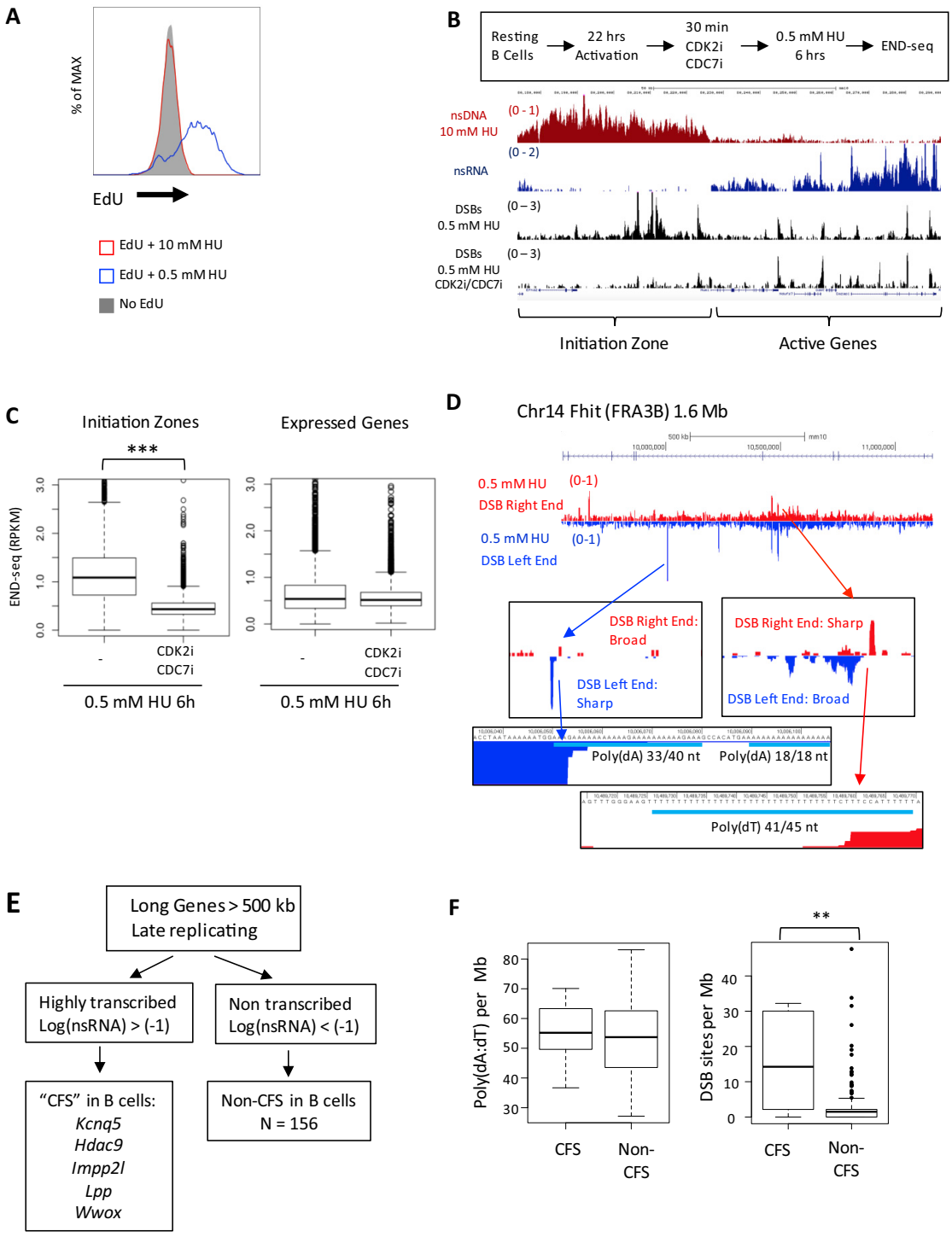


Figure S4. Progressing Replication Fork Collapse within Genes and at CFS, Related to Figure 2

(A) FACS histogram for EdU incorporation in B cells treated with 10 mM HU (red) and 0.5 mM HU (blue) for 28 hr. No EdU (negative control) is indicated in gray.

(B) (Upper panel) Activated B cells were treated with CDK2i/CDC7i, 30 minutes prior to HU, to prevent new origin firing. Replication forks established prior to CD2Ki/CDC7i treatment will continue to travel into flanking genes and collapse upon 0.5 mM HU exposure. Lower panel: UCSC genome browser shots showing normalized read counts (RPM) for nsRNA, nsDNA, and DSBs within replication initiation zones (containing nsDNA reads) and expressed gene bodies (containing nsRNA signal). DSBs within initiation zones are lost after CDK2i/CDC7i treatments, thus DSBs are due to collapse of traveling forks.

(C) Boxplot for END-seq reads within intergenic initiation zones and expressed genes detected in B cells treated with 0.5 mM HU for 6 hours ± CDK2i+CDC7i. ***$p < 10^{-10}$, Wilcoxon rank-sum test.

(legend continued on next page)

(D) UCSC genome browser shot of the *Fhit* gene illustrating the presence of DNA DSBs in activated murine B cells treated with 0.5 mM HU for 28 hr. END-seq reads are separated by strand: bottom strand sequences represent a DSB Left End (blue), while top strand sequences represent a DSB Right End (red). DSB peaks that are enriched either for bottom or top strand sequences are shown below. The DNA sequence near the sharp DSB end is shown for each break.
(E) Flow chart showing prediction of common fragile sites (CFS) in lymphocytes, based on gene size, replication timing, and transcription level.
(F) (Left panel) Boxplots showing poly(dA:dT) density within CFS and non-CFS, as determined in (E). Right panel: Boxplots showing DNA DSB density in CFS and non-CFS, as determined in (E), in B cells treated with 0.5 mM HU for 28 hours. **: Wilcoxon rank sum test; $p < 0.05$.

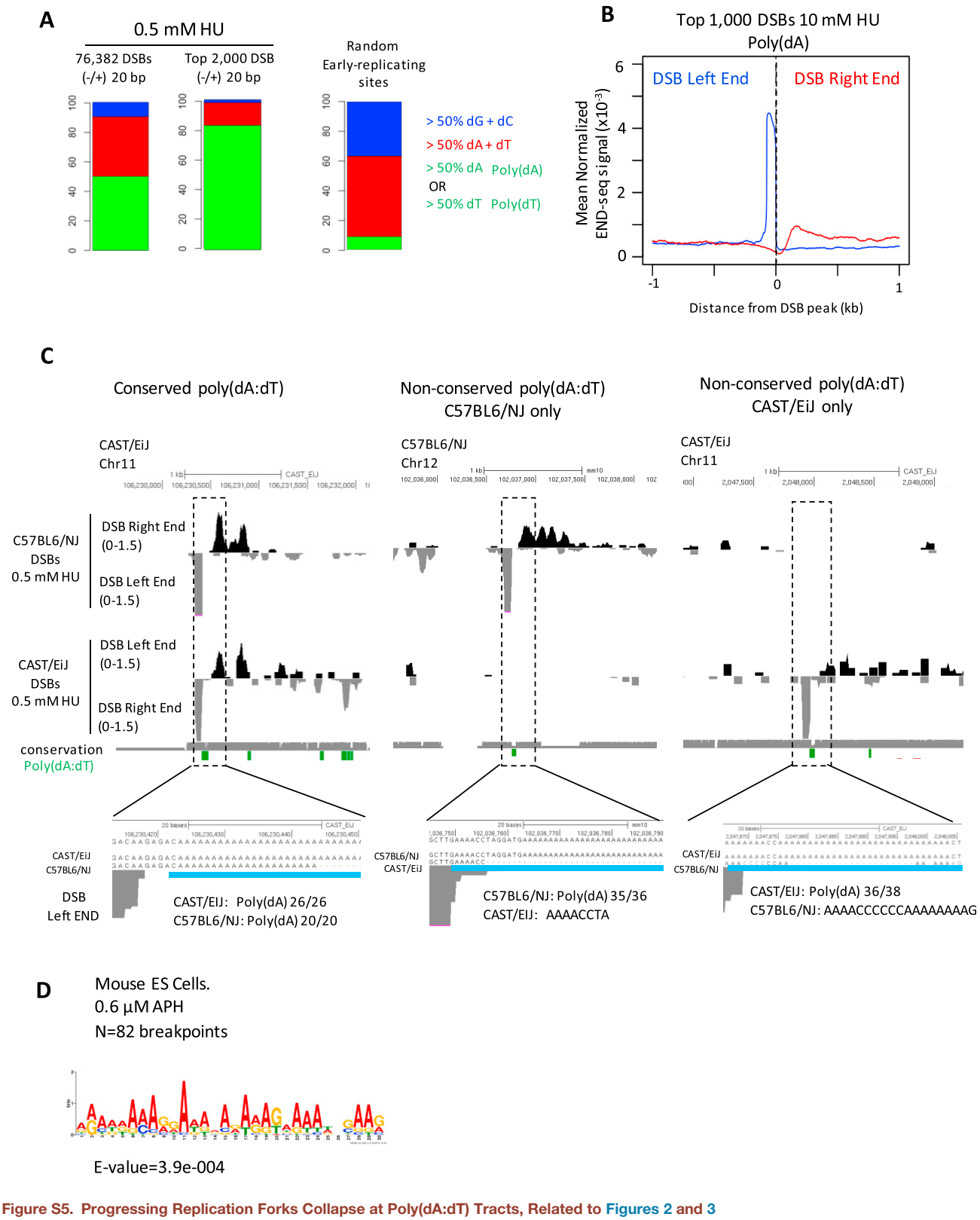


Figure S5. Progressing Replication Forks Collapse at Poly(dA:dT) Tracts, Related to Figures 2 and 3

(A) Nucleotide composition ± 20bp from the "sharp" DNA end for 76,382 DNA DSB sites in activated murine B cells treated with 0.5 mM HU for 28 hr. Middle panel: nucleotide composition ± 20bp from the "sharp" DNA end for the 2,000 most frequently broken sites in activated murine B cells treated with 0.5 mM HU for 28 hr. Right panel: for comparison, nucleotide composition of random 40 bp windows in early-replicating regions of the genome.

(legend continued on next page)

(B) Composite plot showing asymmetric DSBs in B cells treated with 10 mM HU for 28 hours. The 1,000 most frequently broken poly(dA) sites in 0.5 mM HU within replication initiation zones were assessed for breakage.
(C) Genome browser shots showing examples of conserved and non-conserved poly(dA:dT) tracts between CAST/EiJ and C57BL6/NJ mouse strains. END-seq reads are separated by strand: bottom strand sequences represent a DSB Left End (gray)), while top strand sequences represent a DSB Right End (black). Conservation between strains, poly(dA:dT) tracts, and sequence at the DSB left end are shown.
(D) Poly(dA)-enriched MEME motif analyzed at CNV breakpoints in mouse ES cells treated with 0.6 μM Aphidicolin for 72 hours (Arlt et al., 2012).

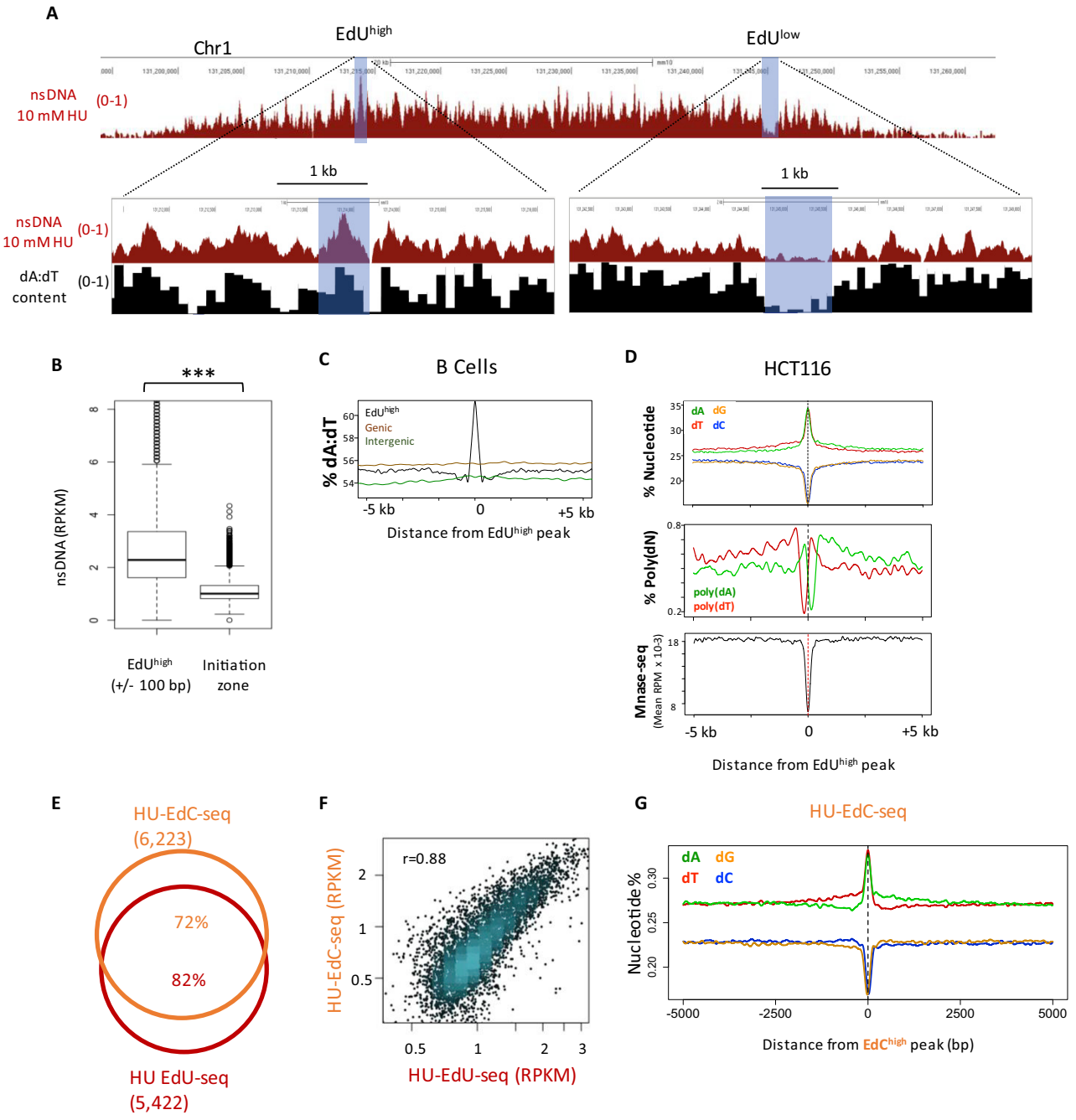


Figure S6. Features of DNA Replication Initiation Sites, Related to Figure 7

(A) UCSC genome browser shots showing normalized read counts (RPM) for nsDNA and dA:dT content within a replication initiation zone in murine B cells treated with 10 mM HU for 28 hours. EdU^{high} and EdU^{low} peaks are associated with high or low dA:dT content, respectively.

(B) Boxplot showing EdU^{high} peaks are efficient sites of replication initiation compared to the whole replication initiation zone. EdU^{high} peaks (+/− 100 bp of summit) contain 2.2-fold more nsDNA than the zones on average. ***$p < 10^{-10}$, Wilcoxon rank-sum test.

(C) Composite plot for dA:dT content surrounding 21,527 EdU^{high} peaks in activated B cells, showing enrichment over randomly selected genic (brown) and intergenic (green) sequences within early-replicating regions.

(D) Composite plots for nucleotide frequency, polynucleotide frequency, and nucleosome occupancy around EdU^{high} peaks in HCT116 cells synchronized in G1 and released into 10 mM HU + EdU for 8 hours. MNase-seq data obtained from HCT116 cells (Guzman and D'Orso, 2017).

(E) Venn diagram showing overlap between replication initiation zones determined using HU-EdC-seq and HU-EdU-seq in activated B cells. Overlap versus random, $p < 10^{-10}$, Fisher's Exact test.

(F) Scatterplot comparing nsDNA levels from HU-EdC-seq and HU-EdU-seq within initiation zones determined by HU-EdU-seq in activated B cells.

(G) Composite plot for nucleotide frequency around EdC^{high} peaks determined by HU-EdC-seq in activated B cells treated with 10 mM HU for 28 hours.

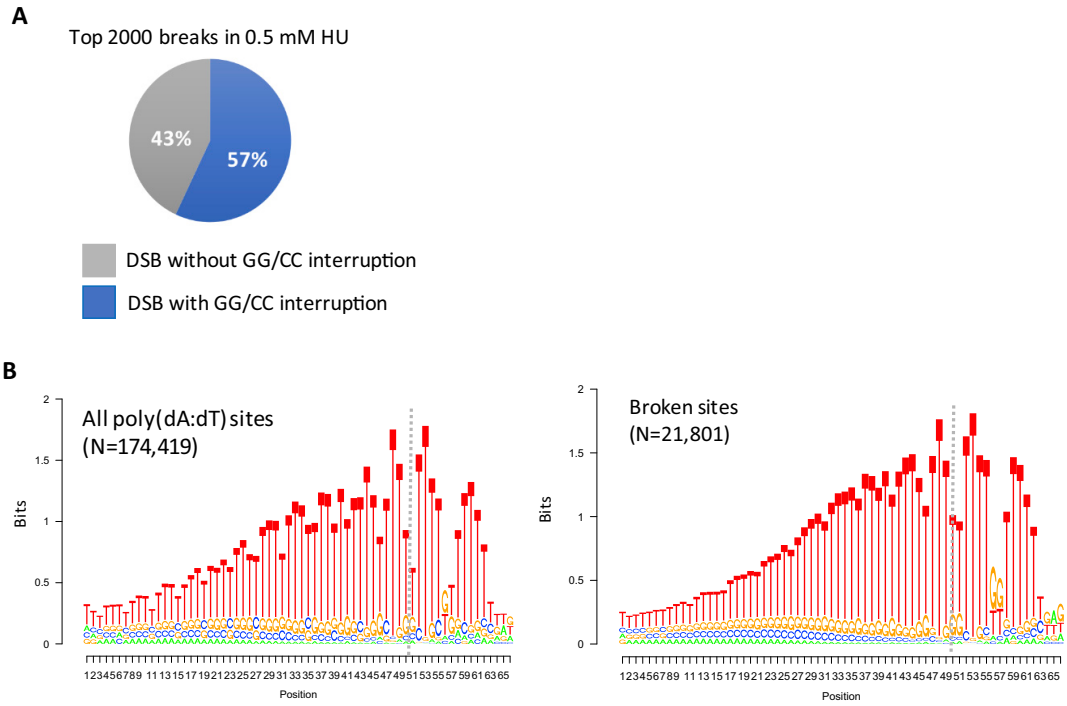


Figure S7. Polymerase Kappa signature at poly(dA:dT) tracts, Related to Figure 2

(A) Analysis of top 2,000 DSBs detected by END-seq in 0.5 mM HU at poly(dA:dT) tracts, scanning for CC/GG dinucleotide interruptions in the poly(dA:dT) motif within 10bp of the sharp DSB.

(B) Position weight matrix generated from DSB sequences in 0.5 mM HU was used to identify all similar sequences in the mouse genome, which were classified into "Non-broken sites" (left panel) or "Broken sites" (right panel) based on DSB peak calling from B cell treated with 0.5 mM HU. Sites that are broken in 0.5 mM HU (right panel) contain a strong CC/GG interruption within the motif compared to non-broken sites (left panel), implicating the CC/GG signature in chromosome fragility.