## Supplemental Figures and Tables

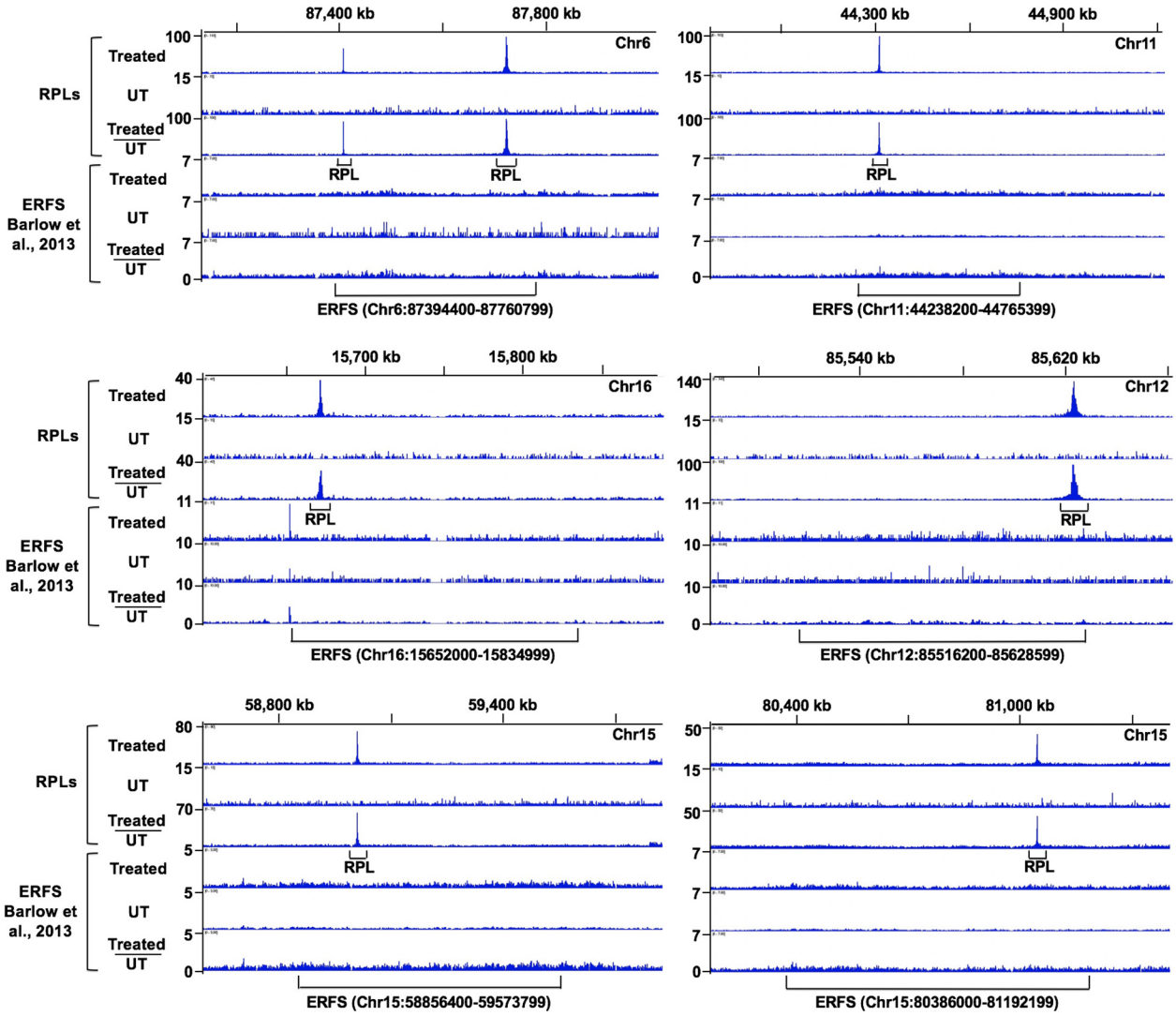
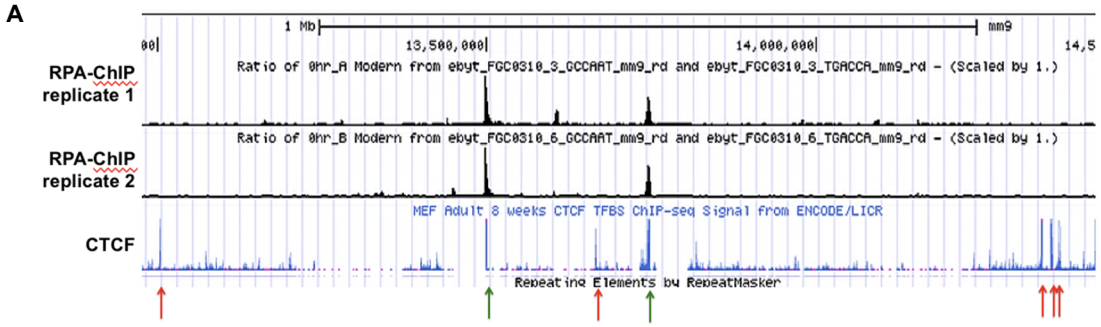



**Figure S1.** Overlap of RPL Peaks with ERFS Hotspots, related to Figure 2

The presence of RPL peaks within six different ERFS hotspot regions. Two combined biological replicates of aligned ATRi+aph$^{18hrs}$ RPA-ChIP Seq samples from this study are depicted in the first row (RPLs - Treated). Two aggregate biological replicates of aligned control DMSO-treated RPA-ChIP Seq samples from this study are depicted in the second row (RPLs - UT). The third row depicts the ratio of the treated over the untreated tracks using 500 bp fixed bin windows (RPLs – Treated/UT). Two combined biological replicates of aligned HU_WT_RPA-ChIP Seq samples from Barlow et al., 2013 are depicted in the fourth row (ERFS - Treated). One replicate of aligned control non-treated RPA-ChIP Seq sample from Barlow et al., 2013 is depicted in the fifth row (ERFS – UT). The sixth row depicts the ratio of the treated over the non-treated tracks using 500 bp fixed bin windows (ERFS – Treated/UT). Fastq reads from all samples were processed, filtered and aligned in the same manner. Ratio tracks were generated using bigwigCompare from deepTools. Black brackets below the tracks depict RPL peaks and ERFS hotspots.

A




| RPL peaks that overlap with CTCF binding site (mm10) | Repeat associated with RPL/CTCF binding site |
|---|---|
| chr3:60045124-60046070 | CACAG/CTGTG |
| chr5:104330061-104332393 | CACAG/CTGTG |
| chr6:137002744-137003442 | CACAG/CTGTG |
| chr6:29747650-29749580 | CACAG/CTGTG |
| chr10:11352664-11353218 | CACAG/CTGTG |
| chr13:74528418-74531130 | CACAG/CTGTG |
| chr15:38975091-38975368 | CACAG/CTGTG |
| chr17:13550990-13553240 | CACAG/CTGTG |
| chr5:25818041-25818538 | CACAG/CTGTG, CACACACAG/CTGTGTGTG |
| chr13:35038338-35039022 | CACAG/CTGTG, CACACAG/CTGTGTG |
| chr6:86329935-86334080 | CAGAGG/CCTCTG |
| chr6:87771924-87779491 | CAGAGG/CCTCTG |
| chr6:90513949-90522651 | CAGAGG/CCTCTG |
| chr12:84280554-84285121 | CAGAGG/CCTCTG |
| chr6:125414800-125419184 | AGGCAGG/CCTGCCT |
| chr9:121927770-121933091 | CAGG/CCTG |
| chr11:5169861-5170581 | CGGTGCCTGACATACAC/GTGTATGTCAGGCACCG |
| chr16:45352944-45353173 | ACAGACAGG/CCTGTCTGT |
| chr19:6813620-6815348 | TCACCATGCAGGACTTG/CAAGTCCTGCATGGTGA |
| chr2:122375449-122375692 | tRNA |
| chr4:118547249-118548098 | GGTAGAGGTGGCAGTGGGCGTGCAGGAGAAGCTGTAGGTC |
| chr7:120835161-120835853 | tRNA |
| chr11:69124069-69124890 | tRNA |
| chr18:34861029-34861890 | CCG/GGC, CAG/CTG |

B

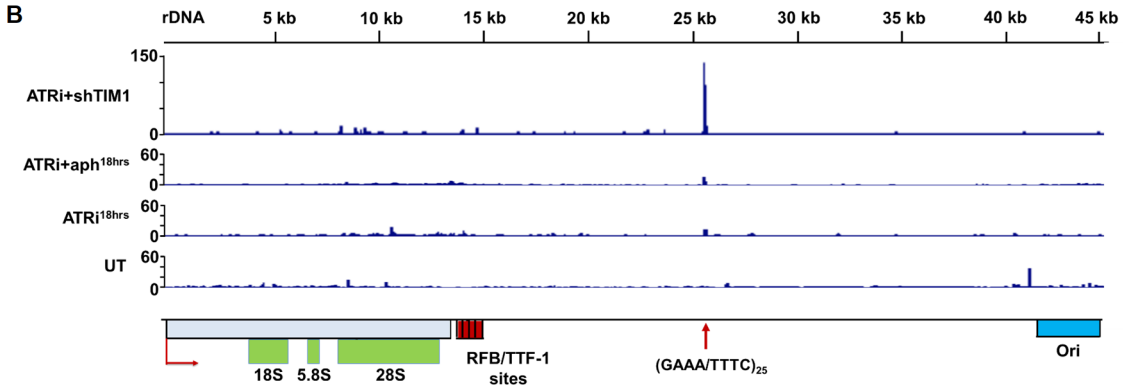



**Figure S2**. Overlap of RPL peaks with CTCF binding sites and rDNA repeats, related to Figure 2

(A) Overlap of RPL peaks from two biological replicates of ATRi+aph$^{18hrs}$ with CTCF-binding sites (track from UCSC genome browser). Green arrows denote overlap between $(CACAG)_n$-centered RPL peaks and CTCF peaks; red arrows denote CTCF peaks that do not correspond to observable RPL peaks. Table summarizes comprehensive list of RPL peaks and associated simple repeats that overlap with CTCF binding sites.

(B) RPA-ChIP/Input ratio tracks of ATRi+shTIM1, ATRi+aph$^{18hrs}$, ATRi$^{18hrs}$ and DMSO-treated control (UT) along the rDNA sequence (500 bp bin windows). Elements of the region are noted below: rRNA transcript, replication-fork barrier (RFB), and origin of replication (Ori). RPL site is denoted by red arrow.

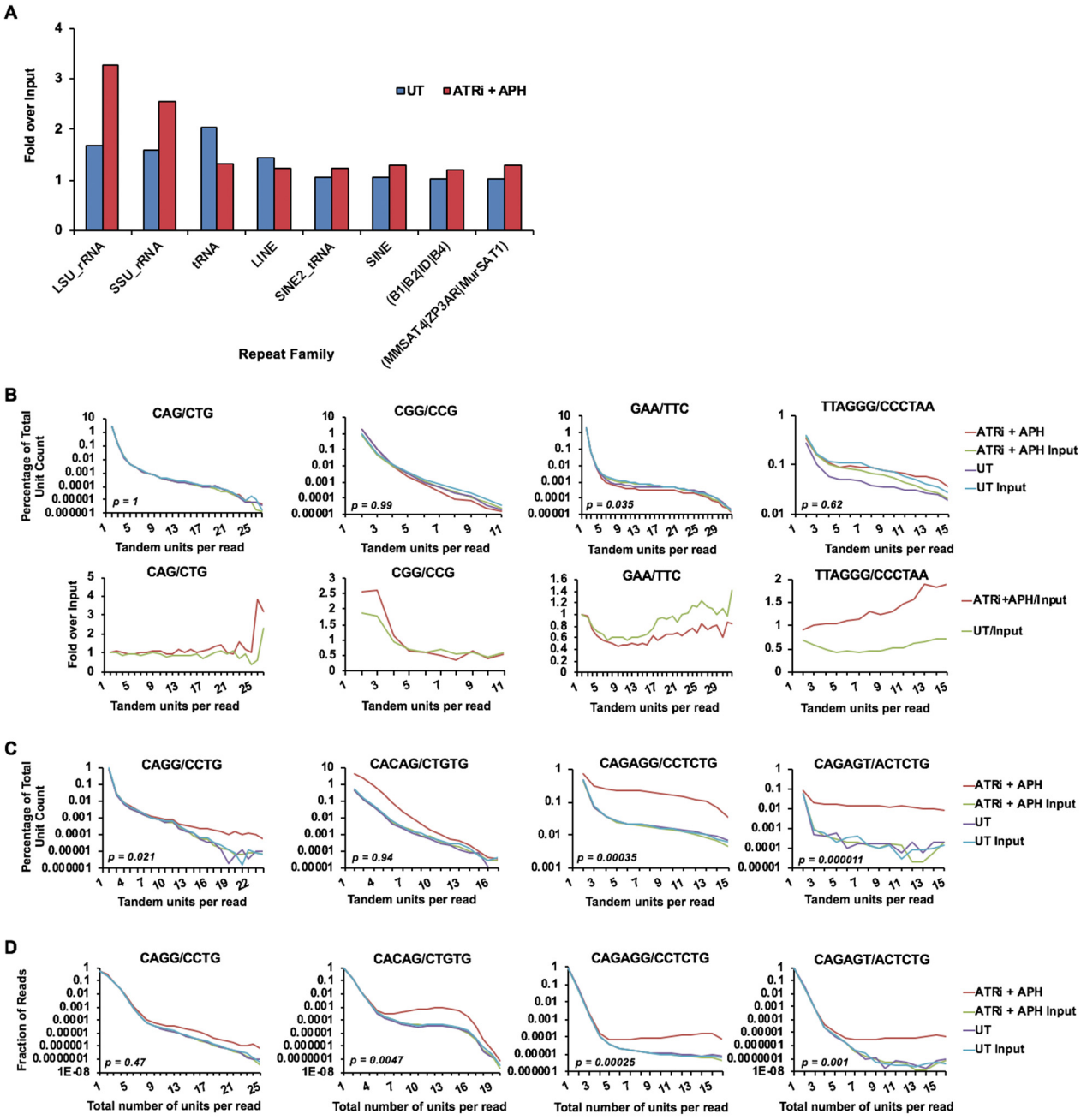



**Figure S3**. Simple repeat analysis of RPA ChIP-Seq reads, related to Figure 2

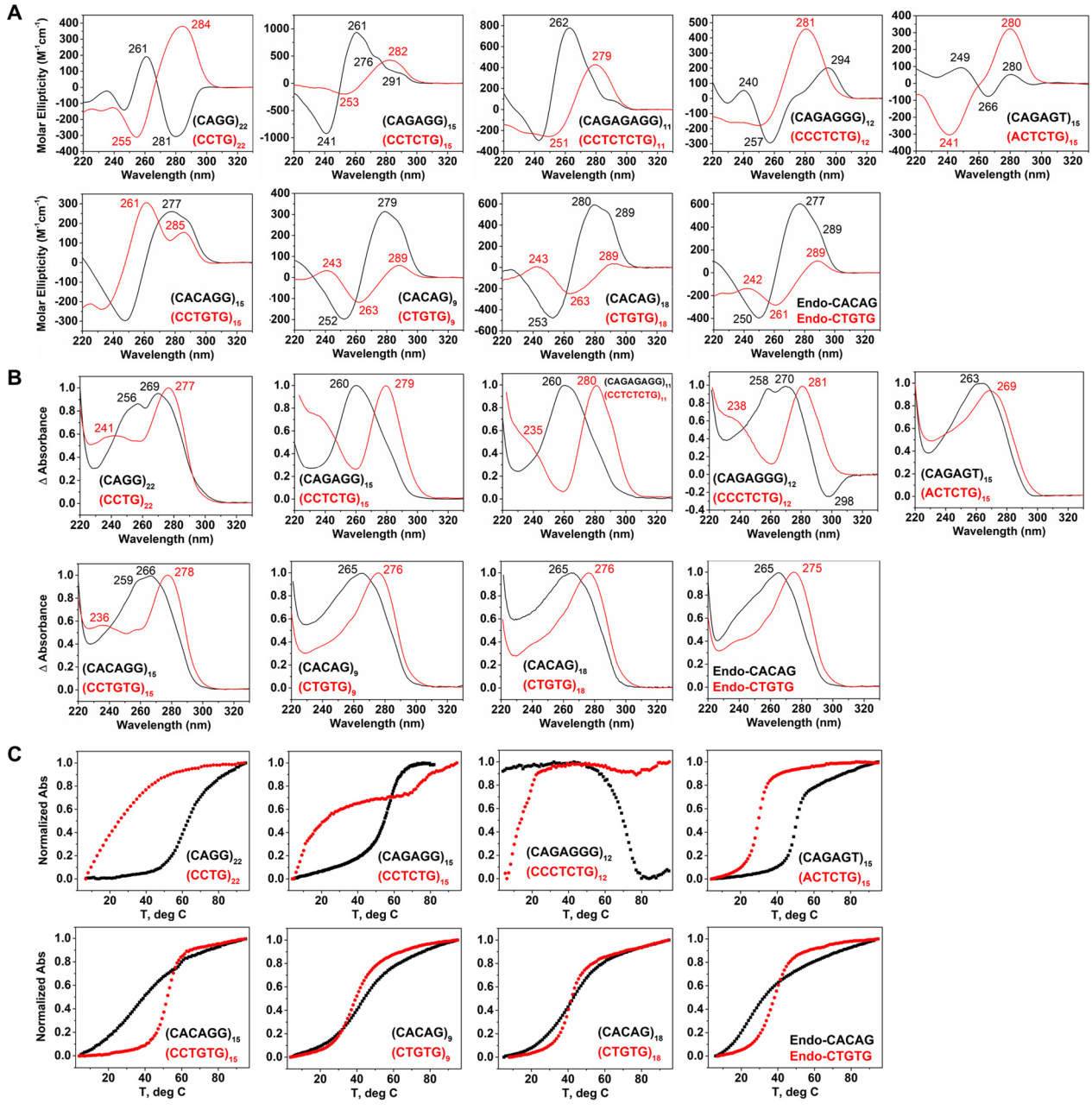
(A) Bar graph depicting the fold enrichment values of different families of complex repeats counted and averaged among replicates from each condition: ATRi+aph$^{18hrs}$ (ATRi + APH) and untreated (UT). These values are normalized by the total number of reads with at least one reported alignment.

(B) Tandem triplet repeat and telomeric repeat analysis of RPA ChIP-Seq reads from combined biological replicates of each condition: ATRi+aph$^{18hrs}$ (ATRi + APH) and untreated (UT). The occurrences of specified tandem lengths of each repeat within all sequencing reads of combined replicates were measured and graphed. In the upper panel, the x-axis depicts the range of tandem units of the repeat queried and the y-axis depicts the frequency of the occurrences as a percentage of total unit count of the repeat present in the reads of the combined replicates. At each specified tandem length, the ratio of the ATRi + APH value over its input and of the UT value over its input was graphed to depict fold over input enrichment (lower panel).

(C) Tandem repeat analysis of RPA ChIP-Seq reads from combined biological replicates of each condition: ATRi+aph$^{18hrs}$ (ATRi + APH) and DMSO-treated control (UT). The occurrences of specified tandem lengths of each repeat within all sequencing reads of combined replicates were measured and graphed. In the upper panel, the x-axis depicts the range of tandem units of the repeat queried, and the y-axis depicts the frequency of each tandem length as a percentage of the total occurrence of the repeat unit in the entire set of reads.

(D) Total repeat analysis of RPA ChIP-Seq reads from combined biological replicates of each condition: ATRi+aph$^{18hrs}$ (ATRi + APH), and untreated (UT). Sequencing reads from combined replicates were categorized by the total unit count of each repeat within a read, allowing for intervening sequence between repeats. Their frequency was calculated as the fraction of reads within the total number of reads from combined replicates that contained the specified amounts of repeat units.

For (B-D), p-values were obtained at 95% confidence interval using the Kolmogorov-Smirnov test between the distributions of ATRi+aph$^{18hrs}$ and its input.




| | Lagging template | Leading template |
|---|---|---|
| 5' - $(CAGAGG)_n$ | 89% (16/18) | 11% (2/18) |
| 5' - $(CCTCTG)_n$ | 11% (2/18) | 89% (16/18) |

**Figure S4**. Biophysical parameters of enriched RPL simple repeats, related to Figure 3

(A) CD wavelength scans collected at 4 °C for repeats listed in Table S3.

(B) Thermal difference spectra, TDS, of repeats listed in Table S3.

(C) UV-vis melting curves of repeats listed in Table S3. Signal was monitored at 260 nm except for $(CAGAGGG)_{12}$ for which the signal was monitored at 295 nm (wavelength which signifies the presence G-quadruplex).

(D) Non-denaturing 12% PAGE gel of repeats listed in Table S3. The gel was run in 1X TBE supplemented with 3 mM $MgCl_2$ for 160 minutes at 150 V.

(E) Asymmetric RPL peak and repeat strand bias. Representative RPL peak depicting an absence of read alignment at the center that overlaps with tandem $(CCTCTG)_n$ repeats and an asymmetric read accumulation on either side of the repeat region. Schematic shows replication fork arriving from the side with greater read accumulation and the subsequent placement of the $(CCTCTG)_n$ repeats on the leading strand of the replication fork. Table on the right quantifies the frequency of the repeat sequence as occurring either on the leading or lagging strand of an approaching replication fork.

All samples were prepared in 10 mM lithium cacodylate pH 7.2, 100 mM KCl and 2 mM $MgCl_2$ buffer (100K2Mg buffer). Repeat number for tandem sequences was adjusted to maintain the constant length of 88-90 nucleotides.

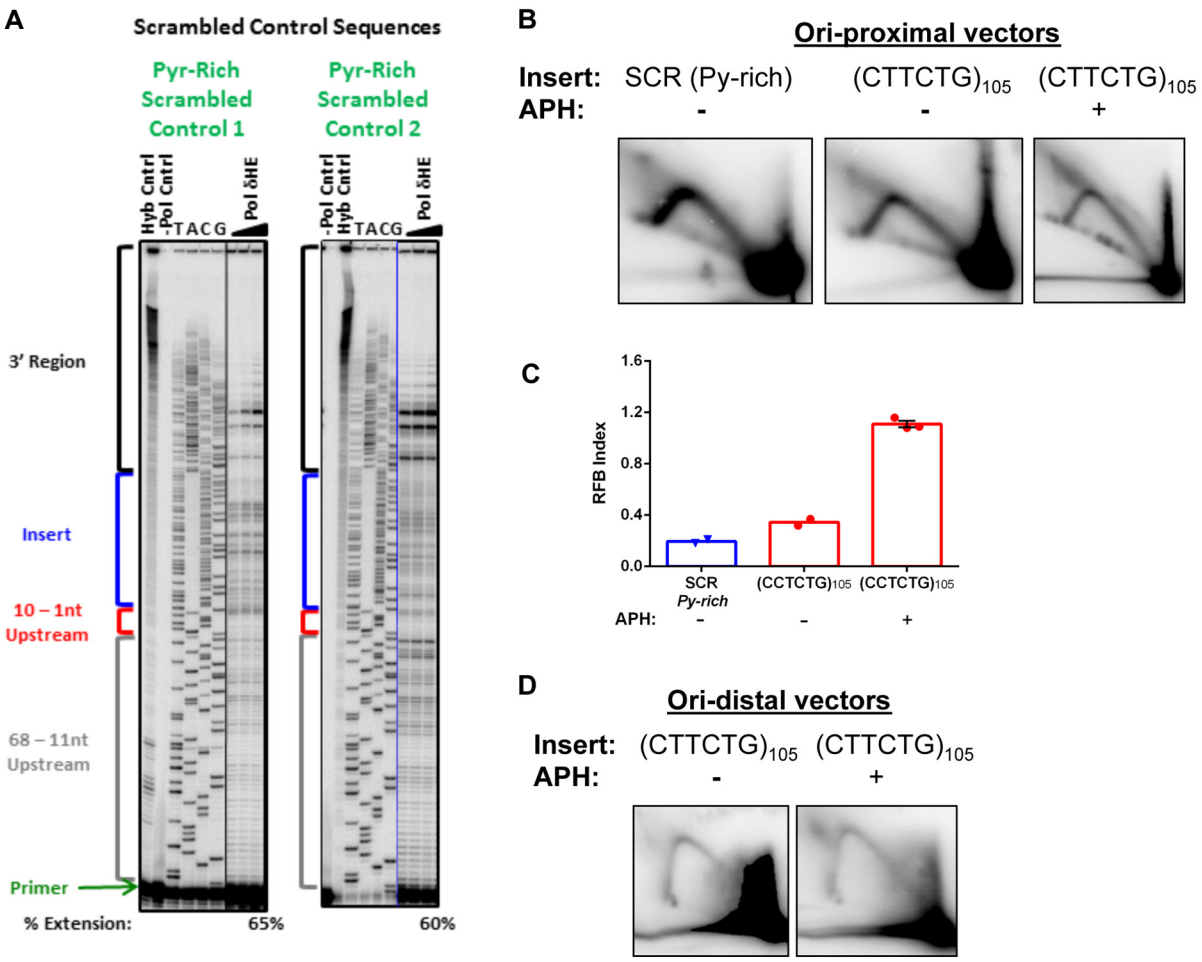



**Figure S5.** Fork-stalling through CAGAGG repeats *in vitro* and in cell culture, related to Figure 4

(A) Representative Phosphoimager gel image of Pol δHE reaction products from two separate pyrimidine-rich scrambled control insert-containing templates. Percent Extension is defined in the legend to Figure 4B and STAR Methods. Blue line indicates that Pol δHE reaction lanes were moved to the right of the sequencing lanes for consistency; see also Mendeley data.

(B) Representative 2D gels for Ori-proximal vectors. A 630 bp cassette containing $(CTTCTG)_{105}$ tandem repeats was cloned in the lagging strand orientation relative to the origin, using the vectors shown in Figure 4D. As control, a scrambled sequence of the same pyrimidine-rich nucleotide composition and length (SCR) was similarly cloned. Cells were transfected and either untreated (-) or treated with 0.6 μM aphidicolin (+) after 24 hours. Episomal DNA was isolated 48 hours after transfection. Replicated DNAs were digested with DpnI, EcoRI and Eco NI, and purified replication intermediates were separated by 2D neutral-neutral gel electrophoresis, followed by transfer to a nylon membrane and Southern hybridization using the probe indicated in Figure 4D. See also Mendeley data.

(C) Quantitation of the RFB index after ori-proximal scrambled or $(CTTCTG)_{105}$ vector replication in U2OS cells. The data are represented as mean +/- SEM, with individual data points showing the results from independent biological replicates.

(D) Representative 2D gels for (CTTCTG) repeats located distal to the origin. The experiment was carried out as described in Figure 4E. See also Mendeley data.

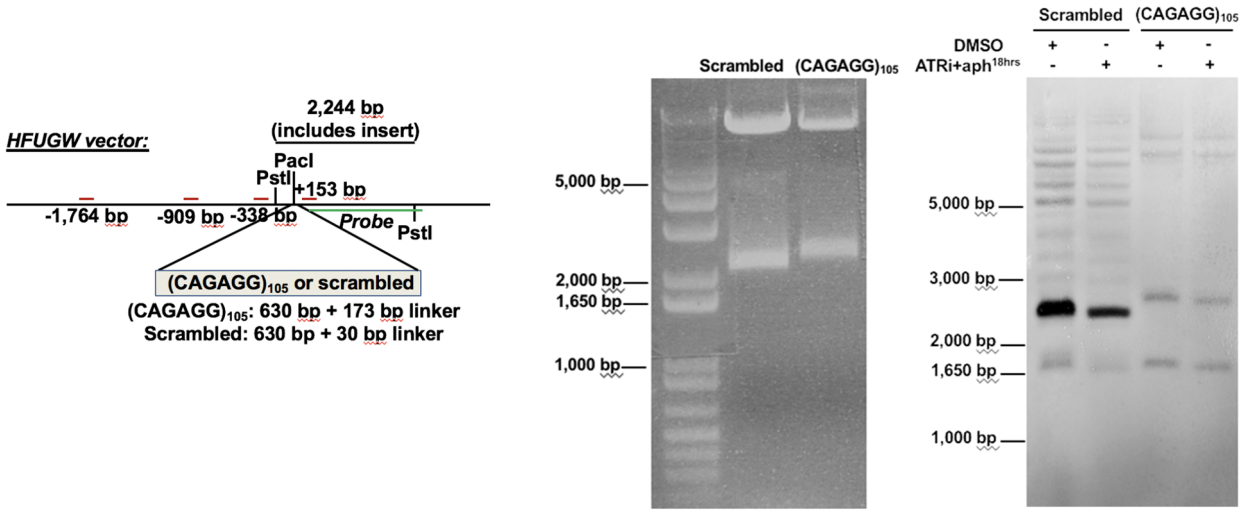



**Figure S6.** Transfection of $(CAGAGG)_{105}$ repeats into MEFs, related to Figure 6

MEF 4-3 cells stably transfected with linearized HFUGW vectors containing either a $(CAGAGG)_{105}$ or scrambled sequence in the PacI site were treated with DMSO or ATRi+aph$^{18hrs}$. Cells were collected for both BrITL and Southern analysis. Genomic DNA extracted for Southern analysis was digested with PstI, resulting in an expected band size of 2,244 bp, as shown in the schematic. The biotinylated probe used to detect the fragment is presented in green against the relevant region in the HFUGW vector. Primer sets used for BrITL qRT-PCR to detect captured breaks at different locations along the HFUGW sequence are highlighted in red. In the middle, an 0.8% TAE agarose gel depicts PstI-digested HFUGW vectors prior to transfection into cells, with insert sequences indicated above each lane. To the right, a Southern blot depicts fragments detected from the biotinylated probe in cells transfected with scrambled or $(CAGAGG)_{105}$ insert and treated with either DMSO or ATRi+aph$^{18hrs}$ (10 µg PstI-digested genomic DNA was loaded for Southern analysis).

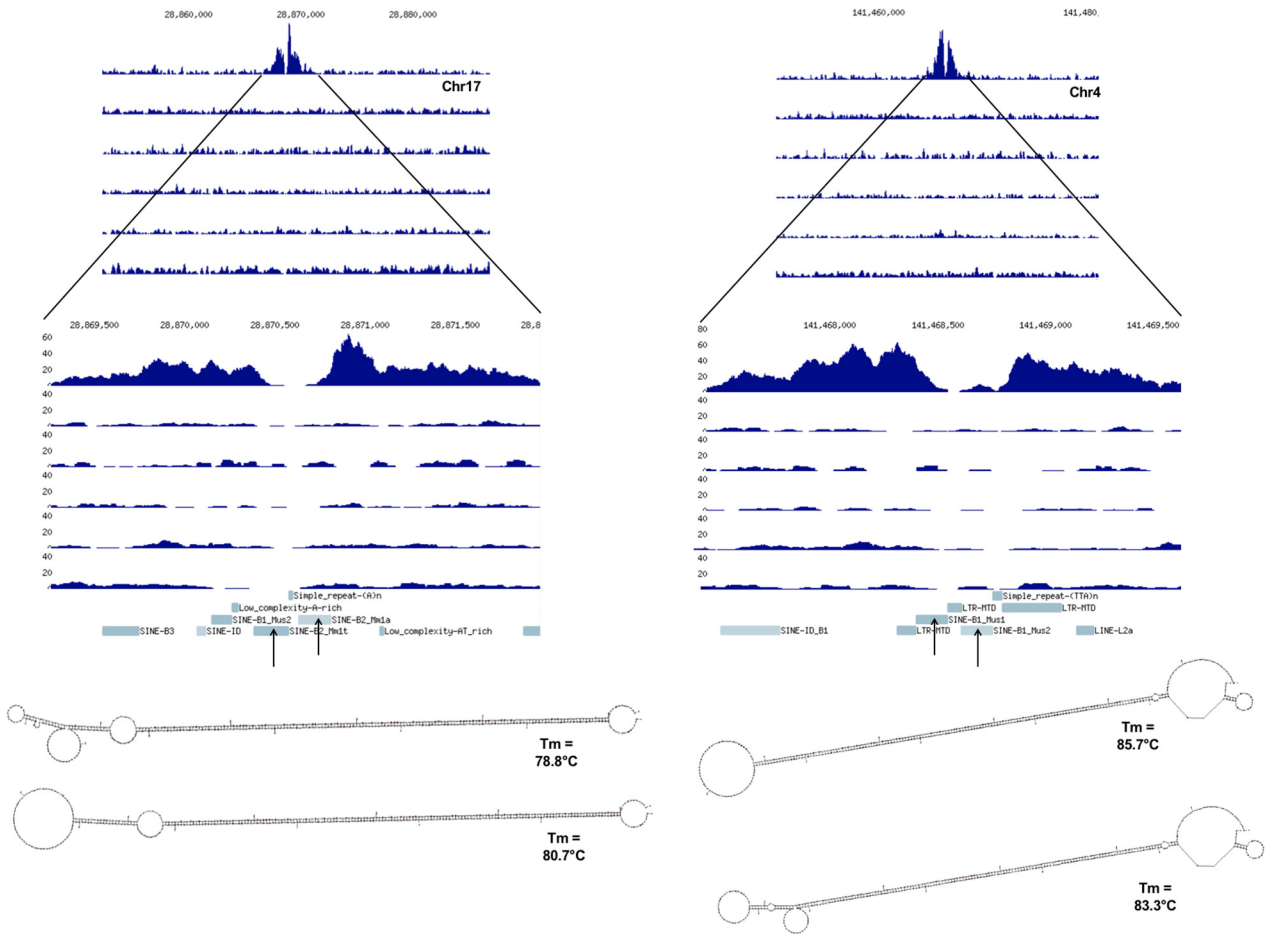



**Figure S7.** Inverted repeats identified by BrITL but not RPA ChIP-Seq in MEF 4-3 cells, related to Figure 6

ATRi+aph$^{18hrs}$ BrITL peaks at sites centered around inverted retroelement sequences. NGS coverage tracks are shown (mm9 coordinates). *Row 1*: ATRi+aph$^{18hrs}$ BrITL retrieval; *row 2*: ATRi+aph$^{18hrs}$ BrITL Input; *row 3*: DMSO BrITL retrieval; *row 4*: DMSO BrITL Input; *row 5*: ATRi+aph$^{18hrs}$ RPA ChIP; *row 6*: ATRi+aph$^{18hrs}$ RPA ChIP Input. Arrows point to the retroelements that comprise the M-fold-predicted stem-loop structures depicted below. Two possible structures are shown.

**Table S3.** DNA Sequences and their Biophysical Parameters, Related to Figure 3

| **Sequence** | **nt** | **TDS (nm)** | **CD: molar ellipticity, $M^{-1}cm^{-1}$ (λmax, nm)** | **Tm, °C heating** | **ΔH, kJ/mol heating** |
|---|---|---|---|---|---|
| 5'- $(CAGG)_{22}$ -3' | 88 | 269<br>256 | -324 ± 20 (280.6)<br>182 ± 52 (261) | 60.6 ± 0.3 | 182 ± 18 |
| 5'- $(CCTG)_{22}$ - 3' | 88 | 277<br>241 | 375 ± 5 (284)<br>-317 ± 3 (254.6) | < 4 | - |
| 5'- $(CACAG)_9$-3' | 45 | 265 | -198 ± 4 (252)<br>317 ± 3 (279) | 41.5 ± 0.2 | 123 ± 10 |
| 5'- $(CTGTG)_9$ -3' | 45 | 276 | 61 ± 1 (289)<br>37 ± 3 (242.5)<br>-120.4 ± 1 (263) | 38.2 ± 0.2 | 159 ± 8 |
| 5'- $(CACAG)_{18}$ -3' | 90 | 265 | 571 ± 25 (289)<br>654 ± 43 (280)<br>-498 ± 54 (253) | 41.2 ± 0.8 | 116 ± 6 |
| 5'- $(CTGTG)_{18}$ -3' | 90 | 276 | 88 ± 4 (290.7)<br>96 ± 30 (242)<br>-292 ± 17 (264) | 41.3 ± 0.5 | 262 ± 31 |
| 5'- $(CAGAGT)_{15}$ -3' | 90 | 262.5 | 42 ± 16 (280)<br>-113 ± 41 (265)<br>89 ± 14 (249) | 49.7 ± 0.3 | 394 ± 5 |
| 5'- $(ACTCTG)_{15}$ -3' | 90 | 269 | 331 ± 12 (280)<br>-312 ± 55 (242) | 29.9 ± 0.3 | 326 ± 7 |
| 5'-$(CACAGG)_{15}$ -3' | 90 | 266<br>259 sh | 255 ± 4 (277.5)<br>-318 ± 26 (248) | 35.8 ± 1.2 | 65.7 ± 7.4 |
| 5'-$(CCTGTG)_{15}$ -3' | 90 | 278<br>236 | 158 ± 9 (285.3)<br>317 ± 8 (261.3) | 51.4 ± 0.8 | 282.9 ± 9 |
| 5'-$(CAGAGAGG)_{11}$ -3' | 88 | 260 | 728 ± 68 (262.5)<br>373 ± 12 (276) | 46.6 ± 0.6 | 209.7 ± 6.6 |
| 5'-$(CCTCTCTG)_{11}$ -3' | 88 | 280<br>235 sh | 444 ± 22 (279)<br>-252 ± 9 (251.5) | ~11 | - |
| 5'-$(CAGAGG)_{15}$-3' | 90 | | 1097 ± 69 (261)<br>-686 ± 12 (241)<br>shoulder (276)<br>shoulder (291) | 55.6 ± 0.9 | 223 ± 13 |
| 5'-$(CCTCTG)_{15}$ -3' | 90 | 279 | 444 ± 28 (282.5)<br>-200 ± 7 (253) | No transition | - |
| 5'-$(CAGAGGG)_{12}$-3' | 84 | 298<br>270<br>258 | 218 ± 15 (294)<br>-272 ± 15 (258)<br>71 ± 40 (241) | 72.0 ± 1.1 | 186 ± 4 |
| 5'-$(CCCTCTG)_{12}$-3' | 84 | 278<br>238 sh | 468 ± 14 (281) | ~15 | - |
| Endo-CAGAG* | 90 | 266 | 663 ± 39 (277)<br>-412 ± 24 (250)<br>419 ± 16 (sh 289) | 24.5 ± 0.3 | 106 ± 10 |
| Endo-CTGTG** | 90 | 275 | 110 ± 14 (288)<br>-160 ± 19 (261)<br>-50 ± 19 (242) | 37.6 ± 0.3 | 200 ± 13 |

* sequence: 5'-AGCACAGACAGCACAGACAGCACAGCACAGTACAGCACAGCACAGCACAGCACAGA CAGCACAGCACAGACAGCACAGCAAAGCACAGAC-3'

** sequence: 5'-GTCTGTGCTTTGCTGTGCTGTCTGTGCTGTGCTGTCTGTGCTGTGCTGTGCTGTG CTGTACTGTGCTGTGCTGTCTGTGCTGTCTGTGCT-3'

TDS – thermal difference spectra; CD – circular dichroism; for both CD and TDS maxima and minima are reported.

Repeat number for tandem sequences was adjusted to maintain the constant length of 88-90 nt.

Peak molar ellipticity values, melting temperatures, and enthalpies of unfolding were determined as the arithmetic mean of at least three consistent trials. Errors reported for these values correspond to one standard deviation from the mean.