High-resolution, genome-wide mapping of positive supercoiling in chromosomes

Monica S. Guo[1*], Ryo Kawamura[3], Megan Littlehale[1], John F. Marko[3,4], Michael T. Laub[1,2*]

1. Department of Biology
   Massachusetts Institute of Technology
   Cambridge, MA 02139
   USA
2. Howard Hughes Medical Institute
   Massachusetts Institute of Technology
   Cambridge, MA 02139
   USA
3. Department of Molecular Biosciences
   Northwestern University
   Evanston, IL 60208
   USA
4. Department of Physics and Astronomy
   Northwestern University
   Evanston, IL 60208
   USA

* Correspondence: laub@mit.edu, msguo@mit.edu

Abstract

Supercoiling impacts DNA replication, transcription, protein binding to DNA, and the three-dimensional organization of chromosomes. However, there are currently no methods to directly interrogate or map positive supercoils, so their distribution in genomes remains unknown. Here, we describe a method, GapR-seq, based on the chromatin immunoprecipitation of GapR, a bacterial protein that preferentially recognizes overtwisted DNA, for generating high-resolution maps of positive supercoiling. Applying this method to *E. coli* and *S. cerevisiae*, we find that positive supercoiling is widespread, associated with transcription, and particularly enriched between convergently-oriented genes, consistent with the "twin-domain" model of supercoiling. In yeast, we also find positive supercoils associated with centromeres, cohesin binding sites, autonomously replicating sites, and the borders of R-loops (DNA-RNA hybrids). Our results suggest that GapR-seq is a powerful approach, likely applicable in any organism, to investigate aspects of chromosome structure and organization not accessible by Hi-C or other existing methods.

Introduction

The DNA inside every cell can adopt a wide range of topologies. Genomic DNA can become supercoiled when the DNA duplex winds about its own axis forming a right-handed superhelix (positive supercoiling) or a left-handed superhelix (negative supercoiling). As DNA writhe can interconvert with twist, positive and negative supercoils can also manifest as over- or under-twisted DNA, respectively. Because overtwisted DNA inhibits strand melting and undertwisted DNA promotes it, DNA supercoiling can profoundly impact the binding of regulatory proteins, promoter firing dynamics, DNA replication, and chromosome architecture (Dorman and Dorman, 2016; Gilbert and Allan, 2014). Despite the importance of DNA topology, the location and distribution of supercoils in genomes remains virtually unknown.

Supercoils are introduced by the translocation of RNA polymerase. When the DNA duplex is unwound during transcription, positive supercoils occur ahead of the polymerase and negative supercoils in its wake, producing the "twin-domain" model of supercoiling (Liu and Wang, 1987; Wu et al., 1988). Supercoils can then diffuse into neighboring loci, though how far they travel and what factors restrict their movement are not well understood (Gilbert and Allan, 2014). Supercoils can also be introduced and removed by DNA topoisomerases, enzymes that transiently break and rejoin the DNA backbone (Pommier et al., 2016; Vos et al., 2011). Topoisomerase activity is essential for DNA replication, with the rapid removal of the positive supercoils ahead of the replication fork necessary to prevent replisome arrest (Postow et al., 2001). Whether supercoils are persistent in genomes or rapidly removed by topoisomerases is not known. Our understanding of how supercoiling impacts chromosome organization and function is severely limited by a lack of high-resolution methods for mapping supercoils in living cells and the inability to specifically interrogate positive supercoiling.

Chromosome conformation capture technologies such as Hi-C have dramatically altered our understanding of chromosome organization. However, Hi-C typically has a resolution of only 5-10 kb and does not capture supercoiling, which generally operates on shorter length scales (Kempfer and Pombo, 2020). Classic methods to interrogate supercoiling, e.g. ultracentrifugation of whole chromosomes or plasmid electrophoresis, only infer average supercoiling, and other methods, which rely on supercoiling-dependent promoters or recombination frequencies, have limited throughput, precluding genome-scale studies (Corless and Gilbert, 2017; Patrick Higgins, 2017). More recently, supercoiling has been measured via preferential crosslinking of psoralen derivatives to undertwisted, negatively supercoiled DNA (Achar et al., 2020; Bermúdez et al., 2010; Kouzine et al., 2013; Lal et

al., 2016; Naughton et al., 2013; Sinden et al., 1980; Teves and Henikoff, 2014). However, psoralen-based studies are still limited in resolution, as they typically rely on sparse crosslinking (1-2 per 10 kb) (Bermúdez et al., 2010; Naughton et al., 2013). Additionally, psoralen-based studies infer the presence of positive supercoils by the absence of crosslinking. Consequently, RNA polymerase, nucleosomes, DNA-binding proteins, or unwound DNA could each block psoralen intercalation and complicate the interpretation of crosslinking efficiency (Bermúdez et al., 2010; Toussaint et al., 2005; Wellinger and Sogo, 1998). These issues could impact the conclusions from a psoralen-based study suggesting that coding regions in yeast are positively supercoiled, with negatively supercoiled DNA accumulating at gene boundaries (Achar et al., 2020), a finding in apparent conflict with the twin-domain model of supercoiling.

Here, we develop a high-resolution method to probe the distribution of positive supercoils in cells. Our approach, GapR-seq, is based on chromatin immunoprecipitation (ChIP) sequencing of GapR, a bacterial protein that preferentially binds overtwisted DNA. Our previous work in the bacterium *Caulobacter crescentus* demonstrated that GapR localizes to the 3' ends of highly transcribed regions and is required, together with type II topoisomerases, to relax positively supercoiled DNA during replication (Guo et al., 2018). We showed with *in vitro* topological assays and a crystal structure that GapR likely binds overtwisted DNA (Guo et al., 2018). We now confirm, using single-molecule magnetic tweezer experiments, that GapR preferentially recognizes positively supercoiled DNA and does not interact stably with negatively supercoiled DNA. These results suggested that GapR could serve as a sensor of positive supercoils in any cell, which we tested in *Escherichia coli* and *Saccharomyces cerevisiae*. In both organisms, GapR-seq yields strong signal in intergenic regions known or expected to harbor positively supercoiled DNA, accumulating downstream of highly-transcribed regions, particularly between convergently oriented genes. In yeast, we also find positively supercoiled DNA associated with centromeres, cohesin-binding sites, and autonomously replicating sequences. GapR-seq further suggests that overtwisted DNA may be associated with the boundaries of DNA-RNA hybrids, or R-loops. Thus, taken together our work demonstrates that GapR-seq is a powerful new approach for mapping positive supercoils and investigating how they shape the structure and function of chromosomes in all kingdoms of life.

Results

GapR interacts with overtwisted, positively supercoiled DNA

We previously showed that GapR (1) binds at sites of expected positive supercoiling in *Caulobacter* cells and (2) GapR constrains overtwisted DNA in topological assays *in vitro* (Guo et al., 2018). Our crystal structure captured GapR as a dimer-of-dimers that fully encircled DNA, without any base-specific contacts and with a narrow DNA-binding cavity that should preferentially accommodate overtwisted DNA (**Fig. 1A**). These findings suggested that GapR could be a probe for positive supercoiling. Subsequently, other crystal structures of GapR in complex with DNA were solved (Huang et al., 2020; Lourenço et al., 2020; Tarry et al., 2019) and featured a larger GapR cavity (**Fig. 1A**), leading to a proposal that GapR does not have a topological preference for DNA. However, crystal structures cannot reveal whether GapR preferentially binds supercoiled DNA. Therefore, we turned to magnetic tweezers (MT) to interrogate GapR binding to single DNA molecules with controlled superhelical density (σ). Briefly, one end of an 11.4 kb dsDNA fragment was immobilized to the coverslip of a flowcell while the other end was bound to a magnetic bead (**Fig. 1B**, **S1A**). The flowcell was then placed on top of a magnet, so that rotation of the magnet introduces over- or undertwisting of the DNA; at low forces (~0.3 pN), the DNA then adopts either positive or negative supercoiling (writhe), which shortens the DNA molecule. This structural change is observed by measuring DNA length, the distance between the magnetic bead and a reference bead fixed to the coverslip (**Fig. 1B**).

We first characterized the behavior of naked DNA by measuring its length at various σ (from –0.03 to +0.03 σ), generating a "rotation-extension curve" (**Fig. 1C, S1B**). We then added GapR (at 10, 100, or 1000 nM) to the relaxed DNA and repeated the rotation-extension experiment. After introducing positive σ, we observed significant lengthening of GapR-bound DNA compared to naked DNA (**Fig. 1C, S1B-D**). These results indicate that GapR constrains the added positive σ, preventing writhing, and lengthening the DNA (**Fig. 1C, S1B-D**). At 1 μM GapR, the DNA was longest at +0.015 σ. Further increasing σ shortens the DNA because the additional positive σ cannot be constrained by GapR and converts to writhe (**Fig. 1C, S1B**). These data, together with prior *in vitro* topological assays (Guo et al., 2018), support our model that GapR binds overtwisted DNA.

We observed that DNA length sometimes increased in the presence of GapR after introducing negative σ; however, the variability in GapR-DNA length was considerably larger at negative σ than at positive

σ or compared to undertwisted naked DNA (**Fig. 1C, S1B, S1E**). Moreover, the length of GapR-bound DNAs fluctuated at negative σ as a function of time (**Fig. 1D, S1E**), with a substantially larger coefficient of variation in DNA length at negative σ compared to positive σ or naked DNA (**Fig. 1E**), indicating that GapR-DNA binding at negative σ is unstable. These behaviors were reversible and did not display hysteresis; we performed multiple rotation-extension experiments on the same GapR-bound DNA, finding that GapR-DNA stably maintained its length when overtwisted, but varied in length substantially when undertwisted (**Fig. 1F**). To our knowledge, these behaviors are unique to GapR. MT studies of other DNA-binding proteins have not reported analogous supercoiling-dependent instability in DNA length (Ding et al., 2014; Sun et al., 2013; Vlijm et al., 2017; Zorman et al., 2012). We propose that GapR rapidly diffuses along or frequently dissociates from negatively supercoiled DNA, but stably interacts with positively supercoiled DNA (**Fig. 1G**), suggesting that GapR could be used as a positive supercoil sensor.

GapR is associated with positive supercoils in *Escherichia coli*

To test if GapR could be used to monitor positive supercoiling in cells, we placed GapR-3xFLAG under tetracycline-inducible control in *E. coli*, an organism without a GapR homolog, and performed chromatin immunoprecipitation-sequencing (ChIP-seq) after inducing GapR (**Fig. S2A**). Importantly, GapR induction did not affect the growth rate of *E. coli*, alter global transcription, or the expression of known supercoiling-sensitive genes (Peter et al., 2004) (**Fig. S2B-D**). Comparing the ChIP of GapR-3xFLAG to an untagged GapR control revealed hundreds of reproducible peaks throughout the *E. coli* chromosome (**Fig. S2A, S2E**). As in *Caulobacter* (Guo et al., 2018), we found a modest correlation between GapR binding and AT-rich DNA, but AT-content alone cannot explain or predict the distribution of GapR (**Fig. S2F**).

Because positive supercoils are introduced into DNA by RNA polymerase (Liu and Wang, 1987; Wu et al., 1988), they should localize at or near the 3' ends of highly expressed genes and transcription units (**Fig. 2A**). We therefore compared our GapR ChIP and RNA-seq profiles. For example, **Fig. 2B** shows a highly expressed ribosomal protein operon where GapR binds from just inside the 3' end of *rplQ* to ~2 kb downstream (see also **Fig. S2G**). To test if this GapR binding was transcription-dependent, we treated cells with the RNA polymerase inhibitor rifampicin for 20 min before performing GapR ChIP. Consistent with our results in *Caulobacter* (Guo et al., 2018), rifampicin

largely abrogated GapR binding downstream of *rplQ* (**Fig. 2B**, see also **S2G**), suggesting that GapR binds where positive supercoils form in *E. coli*.

To quantitatively assess if GapR ChIP is associated with predicted regions of positive supercoiling genome-wide, we compared GapR binding at the 5' and 3' ends of all long (≥ 1500 bp) transcription units (TUs, i.e. genes or operons), normalized in each case to enrichment within the TU, observing significant occupancy of GapR only at the 3' ends (t-test, $p < 10^{-10}$, **Fig. S2H**). To determine if GapR binding is transcription-dependent, we calculated the change in GapR enrichment near the 5' and 3' ends of all long TUs following rifampicin treatment. The distribution of changes at 5' ends was symmetric and centered around 0, with the distribution of changes for the 3' ends significantly shifted to the right (t-test, $p < 10^{-13}$, **Fig. S2I**), indicating that GapR binding near the 3' ends of TUs is sensitive to transcription.

If GapR is recognizing transcription-dependent positive supercoiling, binding should correlate with transcriptional strength. To test this idea, we compared GapR ChIP and transcription-dependent GapR enrichment at long TUs at various expression levels (**Fig. 2C-D, S2J**, see **Methods**). GapR binding at the 5' end relative to within the TU was not dependent on expression level (t-test, $p > 0.01$ for all expression cutoffs, **Fig. 2C, S2J**). In contrast, at 3' ends, GapR binding was dependent on expression, with highly expressed TUs having significantly increased GapR occupancy relative to within the TU (t-test, $p < 10^{-13}$ for all expression cutoffs, **Fig. 2C-D, S2J**). We also ordered long TUs by expression level and plotted as a heatmap the transcription-dependent change in GapR surrounding the 5' and 3' ends. These heatmaps clearly demonstrated that GapR was enriched specifically after the termination site of highly expressed TUs, with GapR occupancy typically extending several kb downstream (**Fig. 2E**, **S2J**). In contrast, GapR binding was de-enriched at the 5' ends of and within highly expressed genes. Notably, GapR was not found at the 3' ends of all well-expressed TUs. However, when we examined exceptions further, we found that these TUs were oriented in tandem with other highly expressed genes, such that GapR accumulated at the 3' end of the downstream TU (**Fig. S3A**). Likewise, GapR enrichment at the 3' ends of poorly expressed genes (or at 5' ends) was typically attributable to the effects of a well-expressed TU on the opposite strand (**Fig. S3B**). Collectively, these analyses support the conclusion that GapR is localized to the positive supercoils produced by transcription in *E. coli*.

GapR recognizes positive supercoiling as a tetramer

While striking, our ChIP results cannot exclude the possibility that GapR is localized downstream of transcription simply because such DNA is more accessible. To control for this possibility, we sought GapR mutants that bound DNA but no longer recognize DNA topology. Previous work demonstrated that truncations in the C-terminal tetramerization domain generated constitutively dimeric GapR ($GapR^{1-76}$) (Huang et al., 2020; Lourenço et al., 2020). Because this dimeric GapR cannot encircle the DNA duplex yet retains all of the DNA binding residues of GapR, we reasoned that this variant would bind DNA without recognizing supercoiling.

To test this hypothesis, we expressed and purified dimeric $GapR^{1-76}$ (**Fig. S3C**). Electrophoretic mobility shift assays (EMSA) showed that $GapR^{1-76}$ binds DNA, albeit with lower affinity than full-length GapR (**Fig. S3D**). We then asked if $GapR^{1-76}$ binds positive supercoiling by performing a topological assay in which a circular, nicked plasmid was incubated with GapR and then subsequently treated with T4 DNA ligase to trap any supercoils constrained by GapR. Positive or negative superhelicity was assessed by 2-dimensional chloroquine electrophoresis. Whereas full-length GapR trapped positive supercoils, $GapR^{1-76}$ did not alter plasmid topology (**Fig. 3A, S3E**). We conclude that dimeric GapR binds DNA but no longer recognizes DNA topology, indicating that positive supercoiling recognition requires a tetrameric conformation.

To validate that tetrameric GapR is recognizing positive supercoiling *in vivo*, we compared the ChIP profiles of full-length GapR-3xFLAG with $GapR^{1-76}$-3xFLAG. Notably, $GapR^{1-76}$ does not bind at the 3' ends of TUs as seen with full-length GapR (**Fig. 3B, S3F**). In fact, the ChIP profiles of $GapR^{1-76}$ and full-length GapR were not correlated (**Fig. 3C**), demonstrating that full-length GapR is not simply bound to accessible DNA. Altogether, our data strongly support the idea that GapR is recognizing overtwisted, positive supercoiled DNA *in vivo*. We propose that GapR-seq provides a direct, high-resolution readout of positive supercoiling in living cells.

Positive supercoils accumulate in regions of convergent transcription

Because positive supercoils are generated downstream of translocating RNA polymerase, we hypothesized that these supercoils, and GapR, should be strongly associated with convergently-oriented TUs (**Fig. 3D**). Indeed, we found that GapR, but not $GapR^{1-76}$, was frequently enriched between convergently-oriented operons in *E. coli* (**Fig. 3E-F**). Convergently-oriented regions had higher GapR signal compared to intragenic or divergently oriented regions (t-test, $p < 10^{-9}$, **Fig. 3F,**

examples in **Fig. S3H-I**), whereas GapR$^{1\text{-}76}$ bound similarly in convergently- and divergently-oriented regions (**Fig. S3G**). Importantly, GapR enrichment between convergently oriented TUs was dependent on transcription (t-test, $p < 10^{-27}$; **Fig. 3G**).

To further validate the association between GapR, positive supercoiling, and convergently oriented TUs, we selected the ~220 genomic regions showing the highest and lowest transcription-dependent changes in GapR ChIP (see **Methods**) and asked how TUs were oriented around these regions. Regions with the highest transcription-dependent changes in GapR were highly enriched for convergently oriented TUs compared to regions with the lowest transcription-dependence or intergenic regions (Fisher's exact test, $p < 10^{-35}$ and $p < 10^{-14}$, respectively, **Fig. 3H**). Together, our *in vitro* and *in vivo* data demonstrate that GapR ChIP effectively reads out the locations of overtwisted, positive supercoiled DNA in living cells. Furthermore, our results validate the "twin-domain" model of supercoiling and reveal that persistent positive supercoils arise downstream of active transcription units and are trapped by converging RNA polymerase in bacterial cells.

GapR is associated with positive supercoiling in *S. cerevisiae*

Next, we asked if our GapR-seq method could be extended to the budding yeast, *Saccharomyces cerevisiae*, which also does not encode a GapR homolog. We integrated either *gapR* or *gapR-3xFLAG* into *S. cerevisiae* at *LEU2* under control of the *GAL1-10* promoter. We grew cells to exponential phase in the presence of raffinose to repress GapR expression, and then induced GapR for 6 hr with galactose before performing ChIP-seq. As in bacteria, (1) expression of GapR-3xFLAG did not significantly alter the transcriptional profile of yeast (**Fig. S4A**); (2) GapR-3xFLAG was reproducibly enriched at specific sites in the genome when compared to the untagged control (**Fig. S4B-C**); and (3) there was only modest correlation between GapR ChIP and local AT-content (**Fig. S4D**).

As in bacteria, we frequently observed GapR peaks at, and extending beyond, the 3' ends of most genes, with peaks almost never occurring within coding regions, and extending ~900 bp long on average, somewhat shorter than seen in *E. coli* (**Fig. 4A, S4E-F**). Also as in bacteria, GapR was significantly enriched at the 3', but not the 5', ends of genes (**Fig. 4B**). To determine if GapR is recognizing transcription-dependent positive supercoiling, we computationally compared our GapR-seq and RNA-seq profiles. We found that GapR enrichment at the 3' ends of genes was clearly correlated with transcriptional strength (t-test, $p < 10^{-25}$ for all expression cutoffs at 3' ends, **Fig. 4C-D**). Additionally, and again as found in *E. coli*, GapR was enriched at regions of convergent transcription compared to

divergent or intragenic regions (t-test, convergent vs all other regions, $p < 10^{-56}$; **Fig. 4E**). Finally, we examined the highest peaks and valleys of GapR occupancy (see **Methods**), finding that there was a significant association between GapR binding and a convergent gene orientation (Fisher's exact test, $p < 10^{-13}$, **Fig. 4F**). We observed identical results when we performed an analogous experiment for cells grown in glycerol before GapR induction (**Fig. S4F-I**). Taken altogether, our results suggest that GapR-seq identifies *S. cerevisiae* genomic regions harboring positive supercoils, and that this topology typically arises downstream of highly expressed genes and particularly between convergently oriented genes.

GapR binding in *S. cerevisiae* is responsive to transcription

To further validate that GapR is recognizing transcription-dependent positive supercoiling, we arrested cells in G1 for 2 hr with α-factor before inducing GapR. Compared to cycling cells, α-factor arrested cells upregulate genes required for mating and downregulate genes specific to S and M phases (**Fig. S5A**). Thus, upon α-factor arrest, we anticipated increased GapR enrichment at the 3' ends of upregulated genes and decreased GapR occupancy at the 3' ends of downregulated genes. Indeed, some of the largest changes in GapR-seq arose near genes known to be induced or repressed during mating such as *FIG1* or *YGP1* (**Fig. 5A, S5B**).

To quantitatively assess how GapR binding is affected by altered transcription, we first examined α-factor dependent GapR binding at the 5' and 3' ends of each gene. As anticipated, GapR occupancy increased at the 3' ends of genes induced in α-factor and modestly decreased at the 3' ends of genes repressed by α-factor (**Fig. S5C**). To better visualize these changes, we ordered genes by their change in expression in α-factor and plotted as a heatmap the change in GapR at each gene's 5' and 3' ends (**Fig. 5B**). These heatmaps showed that upon α-factor treatment, GapR binding was often substantially increased at the 3' ends of upregulated genes and decreased at the 3' ends of downregulated genes (**Fig. 5B**). Collectively, these results demonstrate that GapR-seq reveals transcription-dependent positive supercoiling in *S. cerevisiae*, as it does in *E. coli* and *C. crescentus*. Further, our data validate the "twin-domain model" in *S. cerevisiae*, revealing that persistent positive supercoils are found downstream of actively transcribed genes.

GapR binds nucleosome-free regions, but is not excluded from heterochromatin or DNase-inaccessible DNA

Unlike bacteria, yeast genomes are packaged into nucleosomes. Thus, we wanted to assess whether GapR-seq is impacted by the presence of nucleosomes and, more generally, if GapR can report on positive supercoiling in both eu- and hetero-chromatin. We first examined GapR binding in heterochromatin, such as the yeast mating cassettes, and found that that GapR can still access these relatively compacted loci (**Fig. S5D**). To directly interrogate how nucleosomes impact GapR binding, we computationally compared GapR-seq to nucleosome occupancy inferred from micrococcal nuclease footprinting (MNase-seq), in which nucleosome centers are marked by peaks in read coverage (Cutler et al., 2018). We found that nucleosomes are often in close proximity to GapR peaks (**Fig. 5C, S5D-E**), with positions of high GapR enrichment found within 200 bp of nucleosomes (**Fig. 5D, S5F**). We conclude that GapR can bind near nucleosomes and is not generally excluded from heterochromatic DNA.

We also compared GapR enrichment to DNase I hypersensitivity (DNase-seq) data, which probes general DNA accessibility (Zhong et al., 2016). Although there was some overlap between sites of GapR binding and DNase cleavage, there were many DNase-sensitive regions not bound by GapR, and many loci with high GapR ChIP that were not DNase-accessible (**Fig. 5C, S5D-E**), indicating that DNA accessibility is not predictive of GapR enrichment. We then generated heatmaps of DNase accessibility at genomic regions with the highest GapR enrichment (**Fig. 5E**), and vice versa (**Fig. S5G**). These heatmaps revealed that GapR peaks are not highly accessible to DNase and DNase-sensitive loci are not highly-enriched for GapR. Next, we examined DNase sensitivity and nucleosome binding in different transcriptional orientations. Although convergently transcribed regions have increased GapR occupancy (**Fig. 4E**), these loci are less DNase-accessible and more nucleosome-free than divergently transcribed regions (Kolmogorov-Smirnov test, DNase-seq $p < 10^{-131}$, MNase-seq $p < 10^{-71}$; **Fig. 5F)**. Taken together, these results demonstrate that GapR prefers to bind in nucleosome-free regions, but DNA supercoiling, rather than chromatin accessibility, is primarily responsible for GapR occupancy.

Positive supercoiling in yeast is associated with centromeres, pericentromeres, and cohesin

Collectively, our results show that GapR-seq maps where positive supercoils accumulate, such as the 3' ends of genes. We also asked if our GapR data captured positive supercoiling in other contexts. In

yeast, positive supercoiling has been proposed to accumulate at centromeres, with supercoiling constrained within the centromeric sequences (CEN) and stabilized by binding of the CBF3 complex and the centromeric histone H3 variant, CENP-A/Cse4 (Díaz-Ingelmo et al., 2015; Steiner and Henikoff, 2015; Verdaasdonk and Bloom, 2011). Although not highly enriched at CEN in experiments performed in raffinose, GapR accumulates at CEN upon α-factor arrest (**Fig. 6A-B, S6A**). Similarly, when grown in glycerol such that G1 phase is extended, we also observed association of GapR with CEN (**Fig. S6A**). By aligning GapR enrichment over all 16 CEN, we found that GapR occupancy was highest immediately to the 5' of CEN, upstream of the first centromere determining element and remained high ~500 bp to the 5' and 3' of CEN, with a small 3' shoulder (**Fig. 6C, S6B**). These data validate the notion, based on prior plasmid-supercoiling and *in vitro* studies, that positively supercoiled DNA is found within centromeres (Díaz-Ingelmo et al., 2015; Steiner and Henikoff, 2015).

Yeast pericentromeres are 10-30 kb cohesin-associated regions that flank centromeres (Lawrimore and Bloom, 2019). Cohesin is a chromosome organizing protein complex that mediates sister chromatid cohesion, homologous recombination, and other diverse functions by topologically linking distant loci (Moronta-Gines et al., 2019). Cohesin accumulates between convergent genes, including those that define pericentromere boundaries, and rapidly compacts positively supercoiled DNA *in vitro*, suggesting that cohesin may preferentially associate with such DNA (Glynn et al., 2004; Lengronne et al., 2004; Paldi et al., 2020; Sun et al., 2013). We investigated the relationship between positive supercoiling and cohesin localization by comparing our GapR data to previously published Scc1 (the kleisin subunit of cohesin) ChIP from cells arrested in metaphase (Paldi et al., 2020). In all media conditions, GapR was modestly to highly enriched between the convergent genes that mark pericentromere boundaries (**Fig. 6D, S6C-D**). Outside of pericentromeres, cohesin is also frequently, but not exclusively, associated with convergent genes. These cohesin-enriched regions were also bound by GapR (**Fig. S6E**), supporting the idea that cohesin binding is associated with positive supercoiling.

To systematically examine any relationship between positive supercoiling and cohesin, we generated heatmaps of GapR enrichment surrounding the 500 highest cohesin-bound regions, finding that >90% of all cohesin peaks in glycerol had significant neighboring GapR enrichment within 200 bp (GapR enrichment > $\mu + \sigma$; **Fig. 6E**). Conversely, when we examined cohesin enrichment surrounding the 500 highest GapR ChIP peaks, we found that positively supercoiled DNA was frequently, but not always, associated with strong cohesin binding (**Fig. S6F)**. Our results support the notion that positive supercoiling influences cohesin localization. More broadly, our findings (1) validate the idea that

positive supercoils are a key feature of centromeres, pericentromeres, and cohesin-binding sites, and (2) that GapR-seq reveals, with high resolution, the positions of these supercoils within the yeast genome.

Positive supercoiling is found within the rDNA locus and at autonomously replicating sequences

In addition to centromeres and cohesin-binding sites, we also observed GapR enrichment within the 150-200 tandem repeats of ribosomal DNA (rDNA) found on the right arm of Chromosome XII (**Fig. 7A**). GapR binds at two major peaks in the rDNA locus: (1) one in the unique region at the 3' end of the rDNA locus, which likely arises from transcription of the last 35S rDNA repeat, and (2) one within the rDNA, ~1600 bp upstream of the ribosomal autonomously replicating sequence (rARS) that coincides with the replication fork barrier (RFB), with an additional minor peak over the rARS (**Fig. 7A**). These two peaks manifest in all media conditions, but are most prominent in α-factor arrested and glycerol-grown cells (**Fig. 7A, S7A**), and do not result from changes in rDNA copy number (**Fig. S7A**). The RFB is a ~100 bp sequence at the 3' end of the 35S rDNA where Fob1p binds to block replisome progression and prevent collisions between the replication fork and RNA polymerase transcribing the 35S rDNA (Brewer and Fangman, 1988; Kobayashi and Horiuchi, 1996; Kobayashi, 2003). The GapR-seq signal was centered over the Fob1p binding sites (Ter1 and Ter2) within the RFB, precisely where the replication machinery would arrest, and to the 5' of rARS (**Fig. 7A, S7A**). Because of the GapR enrichment near rARS, we asked whether GapR was enriched near other ARS and found that GapR was enriched within many ARS compared to intergenic sequences in cells treated with α-factor and grown in glycerol (**Fig. S7B-C**). Given that α-factor arrest and glycerol growth both lead to an extended G1 phase, DNA replication is likely not responsible for the positive supercoiling at these regions. Instead, this accumulation could be due to transcriptional effects or proteins bound to ARS (e.g., pre-replicative complex) and RFB that act as barriers to supercoiling diffusion.

Positive supercoiling is associated with R-loops

Our α-factor and glycerol GapR-seq datasets also revealed many strong peaks associated with retrotransposable (Ty) elements, usually with highest enrichment over the terminal repeats (LTR) that flank Ty elements (**Fig. 7B, S7D)**. These peaks were especially striking in the vicinity of poorly expressed Ty elements or divergently oriented regions next to Ty elements (**Fig. 7B, S7D**) because transcription-dependent positive supercoiling does not occur in this context. The LTRs of Ty elements are associated with stable DNA-RNA hybrids (R-loops), and have been mapped by DRIP-seq, which

uses the S9.6 antibody to specifically recognize DNA-RNA hybrids (Chan et al., 2014; Hage et al., 2014; Niehrs and Luke, 2020; Wahba et al., 2016). To compare our GapR-seq to published S1-DRIP-seq data (Wahba et al., 2016), we aligned all 49 yeast Ty elements and examined R-loop formation and GapR binding. We observed two peaks of DNA-RNA hybrids centered on the LTRs of Ty elements, with the peaks of GapR centered just beyond each DRIP-seq peak (**Fig. 7C, S7E**). These results suggest that positive supercoils are associated with R-loops.

DNA-RNA hybrids also occur at telomeres, where transcription of telomeric sequences produces a long, noncoding telomeric repeat-containing RNA (TERRA) that invades telomere DNA and mediates telomere maintenance (Balk et al., 2013; Graf et al., 2017; Luke and Lingner, 2009; Niehrs and Luke, 2020). Yeast telomeres are highly repetitive so many telomere ends are incompletely sequenced, but each typically consists of a telomeric repeat and an X element, with ~50% of telomeres also containing one or more Y' elements (Louis, 1995). To assess GapR and R-loop enrichment in these regions, we assigned reads mapping to repeat sequences randomly across copies of that repeat, allowing for analysis of these repetitive sequences in aggregate. For telomeres containing Y' elements, we observed DNA-RNA hybrids coincident with the telomeric repeats and where TERRA transcription occurs (**Fig. 7D**) (Pfeiffer and Lingner, 2012). Notably, GapR is also highly enriched at these telomeres, with enrichment greatest over the telomeric repeats and remaining high, ~500 bp towards centromeres, past the DNA-RNA hybrids (**Fig. 7D, S7F-G**). Although some transcription does occur near and within Y' elements, we find that GapR enrichment is much higher in magnitude at telomeres than other transcribed regions (compare **Fig. 7D** with **Fig. 5A**), suggesting that transcription cannot fully explain GapR binding at these loci. We then examined telomeres containing X elements but not Y' elements and found that GapR and R-loops are enriched at these telomeres as well (**Fig. S7F-G**). Because R-loops occur when TERRA invades and unwinds a DNA duplex, these R-loops likely produce hyper-negatively supercoiled regions of DNA and may be accompanied by the compensatory structuring of overtwisted, positively supercoiled DNA that balances torsional stress (**Fig. S7H**). Local overtwisted DNA could also act as a barrier to prevent further expansion of R-loops (**Fig. S7H**). Taken all together, our results indicate that positive supercoils are features of many chromosomal loci in yeast. More broadly, we propose that GapR-seq is a flexible and powerful new approach for probing positive supercoiling in cells, from bacteria to eukaryotes.

Discussion

The pervasiveness, chromosomal context, and consequences of supercoiling remain poorly understood, in part because methods to map positive supercoiling *in vivo* at high resolution have been lacking. Here, we developed a method to interrogate positive supercoiling in both bacterial and eukaryotic cells using GapR, a protein sensor of overtwisted DNA. Using single-molecule magnetic tweezer experiments, we demonstrated that GapR preferentially and stably binds overtwisted DNA and does not stably interact with undertwisted DNA. Consequently, GapR localizes to overtwisted, positively supercoiled DNA in bacteria and yeast, allowing positive supercoils to be systematically identified by GapR chromatin immunoprecipitation sequencing – GapR-seq. This new method revealed that positive supercoiling is a pervasive feature of genomes, with remarkably similar patterns documented in bacteria and yeast. Positive supercoils accumulate in a transcription-dependent manner at the 3' ends of genes and are particularly enriched in regions where transcription is convergent. In yeast, GapR-seq further revealed that positive supercoils are associated with centromeres, cohesin binding sites, autonomously replicating sites (ARS), and DNA-RNA hybrids (R-loops), suggesting that positive supercoils may have regulatory or structural roles in each of these chromosomal elements.

GapR is a sensor for overtwisted DNA

Other proteins are known to interact preferentially with positively or negatively supercoiled DNA (Ding et al., 2014; Sun et al., 2013; Zorman et al., 2012), but, to our knowledge, GapR-DNA interactions are unique in that they are destabilized by negative supercoiling, leading to a significant preference for overtwisted DNA (**Fig. 1D**). Although GapR interacts with a range of DNA structures *in vitro* (Huang et al., 2020; Tarry et al., 2019), our magnetic tweezers data and previous topological assays (Guo et al., 2018) indicated that GapR interacts most stably with overtwisted conformations of DNA (**Fig. 1D**). We propose that GapR engages in cycles of sliding, dissociation, and/or rebinding when in complex with relaxed or undertwisted DNA, explaining the dynamics observed in our magnetic tweezer experiments (**Fig. 1G**).

We exploited the binding preference of GapR to develop it as a generic sensor of overtwisted DNA. There are several caveats to our GapR-seq approach. (1) Because GapR recognizes overtwisted DNA, if writhed structures (e.g., plectonemic or solenoidal supercoils) are somehow constrained from interconverting to a twisted form, GapR will not recognize these supercoils. (2) AT-rich DNA can adopt intrinsically bent structures with narrowed minor grooves that may be recognized in a

supercoiling-independent manner by GapR (Arias-Cartin et al., 2017; Guo et al., 2018; Haran and Mohanty, 2009; Huang et al., 2020; Ricci et al., 2016). In GC-rich organisms such as *Caulobacter*, the association with AT-rich DNA was more pronounced than in *E. coli* or yeast (**Fig. S3F, S4D**). (3) The dynamics of GapR-DNA exchange *in vivo* are unknown, as is whether GapR binding affects the kinetics or activity of heterologous topoisomerases. Additional studies are needed to fully understand the implications of GapR binding and the *in vivo* structures recognized by GapR.

Positive supercoiling is pervasive and recapitulates the "twin-domain" model

Transcription leads to the formation of positive and negative supercoils ahead of and behind, respectively, the transcription bubble, referred to as the "twin-domain" model (**Fig. 2A**; (Liu and Wang, 1987). The existence of transcription-dependent supercoiling *in vivo* has been confirmed indirectly in numerous ways (Nelson, 1999), including measuring transcription-dependent changes in plasmid linking number (Drlica et al., 1988; Wu et al., 1988) and interrogating the effects of topoisomerase inhibition (Khodursky et al., 2000). More recently, psoralen, which preferentially binds and crosslinks to negatively supercoiled DNA, has been used to probe supercoiling genome-wide. These studies indicated that negative supercoiling is pervasive, transcription-dependent, and enriched at promoters, consistent with the twin-domain model (Achar et al., 2020; Kouzine et al., 2013; Naughton et al., 2013; Teves and Henikoff, 2014). However, one study suggested that negative supercoils also arise downstream of transcribed genes with positive supercoils accumulating in intragenic regions regardless of transcriptional activity (Achar et al., 2020), findings at odds with the twin-domain model. In contrast, GapR-seq revealed, in both bacteria and yeast, that positive supercoiling is (1) in intergenic or transcriptionally silent regions that lie at the 3' ends of transcribed genes, (2) depleted at the 5' ends of genes, (3) transcription-dependent, with signal roughly proportional to the transcriptional activity of upstream genes, and (4) trapped by convergent transcription, all as predicted by the twin-domain model. Unlike psoralen approaches that infer positive supercoiling based on the absence of psoralen crosslinking, GapR-seq specifically and directly probes for positively supercoiled DNA.

The ability to detect positive supercoils using GapR-seq in both bacteria and yeast indicates that positive supercoils are not fully, or at least immediately, dissipated by topoisomerases *in vivo*. GapR-seq also allows mapping of positive supercoiling at high (< 1 kb) resolution, demonstrating that positive torsion appears capable of diffusing over a few kb (**Fig. S2K, S4F**). However, supercoil diffusion is limited by transcription, as GapR signal was rare within transcribed regions (**Fig. 3F, 4E**). Supercoiling may also be limited by the binding of DNA structuring proteins like nucleosomes or other complexes

(**Fig. 5D, 7A**). Finally, because the distribution of positive supercoils downstream of actively transcribed genes was consistent between bacteria and yeast, we anticipate that similar patterns are likely to be found in other organisms as well.

Positive supercoiling in chromosome organization

GapR-seq suggested an association between positive supercoiling and yeast centromeres, cohesin binding sites, ARS, and R-loops, revealing potentially significant roles for positive supercoils in genome organization and function. Interestingly, these associations were strongest in conditions where yeast were primarily in G1 phase, suggesting that active replication may clear GapR from the DNA or that rapid growth diminishes the deposition of GapR on chromosomes, dampening signal. For centromeres, prior work suggested that the intrinsic architecture and assembly of the CENP-A histone complex at centromeres leads to positive supercoiling (Díaz-Ingelmo et al., 2015). Additionally, positive supercoiling has been proposed to aid in cohesin deposition (Sun et al., 2013). Our results support these ideas and now provide insight into the precise localization of positive supercoils at these chromosomal regions (**Fig. 6**). In higher eukaryotes, cohesin is also found outside of centromeres where it accumulates at some CTCF sites to extrude loops and form topologically associated domains (TADs) (Fudenberg et al., 2017; Nora et al., 2017; Rao et al., 2017). Given the association between cohesin and positive supercoils documented here, we suggest that GapR-seq may be particularly useful in probing the contribution of positive supercoiling to TAD formation.

Positive supercoiling associated with ARS and R-loops has not been well characterized or carefully probed, again because of limited methods for mapping supercoils *in vivo*. We propose that positive supercoiling near the rARS and other ARS may result from enhanced trapping of topological stress in these regions. Positive supercoiling occurs in replication-transcription encounters (García-Muse and Aguilera, 2016), and our data raise the possibility that supercoiled DNA could be trapped between the converging replication and transcription machineries in yeast. Our results also indicate that positive supercoils occur adjacent to, but do not fully overlap with, R-loops such as the boundaries of Ty elements and telomeres (**Fig. 7B-E**). The noncoding telomeric repeat–containing RNA (TERRA) invades telomeric DNA to form R-loops and promote telomere maintenance in eukaryotes (Balk et al., 2013; Bettin et al., 2019). Because every ~10 bp captured in an R-loop represents one negative supercoil, DNA-RNA hybrids like TERRA are potential reservoirs of extreme negative superhelicity (Stolz et al., 2019). Consequently, the formation or expansion of R-loops must be balanced by compensatory overwinding elsewhere (**Fig. S7H**). Furthermore, positive supercoiling adjacent to

TERRA hybrids could be a barrier to further melting of the DNA duplex and R-loop spreading (**Fig. S7H**). Recent studies have suggested that an overabundance of telomeric R-loops causes replicative stress and increased recombination rates in human cells, with these general pathways conserved in yeast (Pan et al., 2019; Petti et al., 2019). Further work will be needed to dissect how positive supercoils arise near R-loops and how they impact genome structure and function.

In sum, our GapR-seq approach provides high-resolution, genome-wide maps of positive supercoils in both bacteria and yeast. These maps reveal the extent and distribution of both transcription-induced positive supercoils as well as supercoiling in other genomic contexts, such as centromeres and telomeres, where positive supercoils may play important roles in genome organization. We anticipate that our GapR-seq method will be easily extended to diverse bacterial and eukaryotic organisms for probing the origins and consequences of DNA torsion and understanding how DNA topology impacts gene expression, chromosome structure, and genome maintenance.

Figure Legends

Figure 1. GapR interacts stably with overtwisted, positively supercoiled DNA

(A) Comparison of GapR-DNA crystal structures. Left, 6GC8 (Guo et al., 2018); middle, 6OZX (Tarry et al., 2019); right, overlay. Diameter of 6GC8 (orange arrow) and 6OZX (grey arrow) indicated.

(B) Schematic of magnetic tweezer (MT) experiment. See also Figure S1A.

(C) Behavior of naked DNA (left), DNA incubated with 1 μM GapR (middle), and overlay (right) in a rotation-extension experiment with the corresponding DNA conformation superimposed. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment.

(D) DNA ± 1 μM GapR behavior over time from (C) under no supercoiling (σ = 0.0, left), positive supercoiling (σ = +0.03, middle), and negative supercoiling (σ = -0.03, right).

(E) Coefficient of variation of force-extension experiments of DNA ± 1 μM GapR. Data indicate mean ± SEM, n ≥ 3.

(F) Hysteresis of force-extension experiments. Traces indicate multiple rotation-extension measurements from one DNA molecule ± 1 μM GapR.

(G) Model of GapR binding to positively or negatively supercoiled DNA.

See also Figure S1 and Table S1.

Figure 2. GapR is associated with positive supercoiling in *E. coli*

(A) Positive supercoiling is generated downstream of transcription (top) and is associated with active transcription (bottom) as predicted by the "twin-domain" model.

(B) GapR ChIP profiles at a highly expressed operon. AT content (top), with AT content below the genomic average (50%) plotted in reverse. ChIP-seq (middle) of untreated (orange) or rifampicin-treated (pink) GapR-3xFLAG cells and untreated GapR cells (grey). Transcription from the forward (green) and reverse (blue) strands with the position of annotated genes indicated (bottom).

(C) Transcription-dependent change in GapR ChIP at 5' (left) or 3' (right) ends normalized by binding within the transcription unit (TU) at different expression thresholds. Student's t-test p-value shown.

(D) Examples of GapR-3xFLAG ChIP without (orange) or with (pink) rifampicin treatment. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated. Expression values are colored using the same rpkm cutoffs as in (C).

(E) Heatmap showing transcription-dependent change in GapR around 5' and 3' ends for the top and bottom 300 long TUs sorted by expression.

See also Figure S2 and Figure S3.

Figure 3. GapR recognizes DNA supercoiling and is associated with convergent transcription

(A) GapR$^{1-76}$ does not recognize DNA topology. Full-length GapR and GapR$^{1-76}$ were incubated with nicked plasmid before treatment with T4 DNA ligase and subsequent quenching, deproteinization, and electrophoresis (schematic). Gel analysis of plasmid topology with supercoiled and relaxed standards.

(B) GapR$^{1-76}$-3xFLAG ChIP (green) and GapR-3xFLAG ChIP without (orange) and with (pink) rifampicin treatment (top). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated.

(C) Correlation between GapR-3xFLAG and GapR$^{1-76}$-3xFLAG ChIP experiments.

(D) Positive supercoils are trapped by convergent transcription.

(E) ChIP of GapR$^{1-76}$-3xFLAG (green) and GapR-3xFLAG without (orange) and with (pink) rifampicin treatment at convergently oriented TUs. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated.

(F) GapR ChIP in gene bodies (dark grey), in divergent regions (blue), convergent regions (red), and where transcription is in the same orientation (purple). Overlay of divergent and convergent regions (bottom).

(G) Transcription-dependent changes in GapR plotted as in (F).

(H) Regions with high transcription-dependent change in GapR are more frequently between convergent genes. Pie charts summarize the orientation of flanking genes for all intergenic regions (top) and intergenic regions with highest (bottom left) or lowest (bottom right) transcription-dependent change in GapR.

See also Figure S3.

Figure 4. GapR is associated with positive supercoiling in yeast

(A) ChIP of *S. cerevisiae* grown in raffinose before GapR induction. AT content (top), ChIP-seq (middle) of GapR-3xFLAG (orange) or untagged GapR (grey) expressing cells. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(B) Mean GapR enrichment (GapR-3xFLAG ChIP normalized by untagged ChIP) in a 500 bp window at the 5' and 3' end of long genes. Student's t-test p-value is shown.

(C) Mean GapR enrichment at 5' and 3' ends of long genes at various transcriptional cutoffs. Student's t-test p-value is shown.

(D) Examples of GapR-3xFLAG (orange) and untagged GapR (grey) ChIP. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated. Expression values are colored using the same rpkm cutoffs as in (C).

(E) GapR is enriched between convergently oriented genes.

(F) GapR-bound regions are more frequently between convergent genes. Pie charts shown as in Fig. 3H.

See also Figure S4.

Figure 5. GapR binding in *S. cerevisiae* is responsive to transcription and is not restricted to open chromatin

(A) GapR enrichment at *FIG1* in raffinose without (orange) or with (green) α-factor arrest before GapR induction (top). Transcription of the forward (green) and reverse (blue) strands in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(B) Heatmap showing α-factor dependent change in GapR enrichment at the 5' and 3' ends of long TUs sorted by transcriptional change in α-factor.

(C) GapR enrichment (orange) compared to nucleosome occupancy (MNase-seq, light grey) and chromatin accessibility (DNase-seq, dark grey). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated.

(D) GapR ChIP peaks are de-enriched for nucleosomes. GapR enrichment (orange, left y-axis), MNase-seq reads (dark grey, right y-axis), mean genomic MNase-seq occupancy (dashed grey line).

(E) Heatmap of GapR enrichment (left) and DNase accessibility (right) of the 500 most GapR-enriched loci.

(F) Association between transcriptional orientation and MNase-seq and DNase-seq.

See also Figure S5.

Figure 6. Positively supercoiled DNA is associated with centromeres and cohesin

(A) GapR enrichment at CEN5 in cells without (orange) or with (green) α-factor arrest (top). Transcription of the forward (green) and reverse (blue) strands in cells grown in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(B) GapR enrichment at centromeres in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(C) GapR enrichment over all centromeres after α-factor arrest (green) or grown in glycerol (blue). Mean enrichment (solid line) with 95% confidence intervals (shaded area). Grey bar represents position of centromeres.

(D) GapR (raffinose, top) and cohesin (Scc1 enrichment, bottom) are associated with convergent genes (arrows) at pericentromere boundaries (shaded areas).

(E) Heatmaps of GapR (3 left panels) and Scc1 (right) enrichment at the 500 most Scc1-bound loci.

See also Figure S6.

Figure 7. Positive supercoiling is associated with ARS and R-loops

(A) GapR enrichment (glycerol) at the rDNA shown in two successive zoom-ins (top and bottom panel) with the replication fork barrier (RFB) and termination sequences (Ter1/Ter2) indicated.

(B) GapR enrichment (α-factor) at a Ty element (top). DNA-RNA hybrid formation by DRIP-seq (middle). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(C) Alignment of GapR and DRIP enrichment surrounding all yeast Ty elements. Data indicate mean (solid line) with 95% confidence intervals (shaded area), no enrichment (dotted line).

(D) GapR enrichment at a telomere with a Y' element (top). DRIP enrichment (middle). Transcription of the forward (green) and reverse (blue) strands with the organization of the Y' element indicated (bottom). Position given is from end of *TEL08R*.

(E) Alignment of GapR and DRIP enrichment surrounding all yeast telomeres with Y' elements as in (C). Telomeric repeats are removed from analysis.

See also Figure S7.

Supplemental Figure Legends

Figure S1. GapR binding to supercoiled DNA in a magnetic tweezer experiment, related to Figure 1.

(A) Schematic of the microscope and flowcell setup in a magnetic tweezer (MT) experiment.

(B) Behavior of naked DNA (left), DNA incubated with 1 μM GapR (middle), and overlay (right) in a typical rotation-extension experiment. Data indicate mean ± SEM, for 3 MT experiments.

(C) Behavior of DNA incubated with 10 nM GapR in a typical rotation-extension experiment. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment (left) and mean ± SEM compared to naked DNA (black), for 3 MT experiments (right).

(D) Behavior of DNA incubated with 100 nM GapR in a typical rotation-extension experiment. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment (left) and mean ± SEM compared to naked DNA (black), for 3 MT experiments (right).

(E) Rotation-extension curve (top left; data indicate mean ± SD, n = 200 at each σ) and time course measurements for a single GapR binding experiment. Time courses indicate behavior of naked DNA (black) and DNA bound by 1 μM GapR (orange) under positive supercoiling (σ = +0.03, top right), negative supercoiling (σ = -0.03, bottom left), and no supercoiling (σ = 0.0, bottom right).

Figure S2. *E. coli* GapR ChIP-seq, related to Figure 2.

(A) ChIP-seq of GapR-3xFLAG with (orange) and without induction with anhydrous tetracycline (aTc, grey), after rifampicin treatment (rif, pink), and untagged GapR with induction (blue).

(B) Growth of cells expressing GapR-3xFLAG plasmid or empty vector. Data indicate mean ± SEM, n = 6.

(C) Transcriptional profiles of cells ± GapR-3xFLAG expression.

(D) Supercoiling-sensitive genes are unaffected by GapR-3xFLAG expression. Log_2-fold change in expression (rpkm) vs genome order (excluding GapR and poorly expressed genes). Supercoiling-sensitive genes (yellow/blue) and topoisomerase genes (light green). 15 genes that change > 4-fold represent mostly proteins of unknown function or membrane proteins.

(E) Correlation between two independent GapR-3xFLAG ChIP experiments.

(F) GapR-3xFLAG ChIP versus AT content in *C. crescentus* (left) and *E. coli* (right). Mean ChIP at a given % AT (red dots). Motif from DREME (below)

(G) GapR-3xFLAG ChIP profiles. AT content (top), with AT content below the genomic average (50%) plotted in reverse. Normalized ChIP-seq (middle) of GapR-3xFLAG cells that were either untreated (orange) or rifampicin-treated (pink), and untagged GapR expressing cells that were untreated (grey). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(H) GapR accumulates at the 3' end of transcription units (TUs). GapR ChIP at 5' (grey) or 3' (orange) ends of TUs normalized by binding within the TU. Student's t-test p-value is reported.

(I) GapR binding at 3' ends depends on transcription. Transcription-dependent change in GapR ChIP at the 5' (grey) or 3' ends (orange) of TUs normalized by binding within the TU. Student's t-test p-value is reported.

(J) GapR binding at 5' (top) or 3' (bottom) ends of long TUs normalized by binding within the TU at different expression levels. Student's t-test p-value is reported.

(K) Histogram showing the length of transcription-dependent GapR binding events.

Figure S3. GapR full-length and truncation variant ChIP-seq, related to Figures 2 and 3.

(A) GapR ChIP-seq profiles in the absence (orange) and presence (pink) of the transcriptional inhibitor rifampicin (top). Transcription on the forward (green) and reverse (blue) strands with annotated genes indicated (bottom). Terminators on the top strand are indicated.

(B) GapR profiles as in (A). Shaded region shows GapR ChIP generated by the minCDE transcript that is also at the 3' end of the *ycmJ-ycgI* locus.

(C) Schematic and purification gel of full-length and tetramerization-deficient GapR$^{1-76}$.

(D) Electrophoretic mobility shift assays of full-length and GapR$^{1-76}$ binding to 210 bp DNA.

(E) Analysis of ligation products from Fig. 3A with 2D-chloroquine electrophoresis. Migration of different plasmid forms are diagrammed (left): N, nicked; R, relaxed; L, linear; (-), negatively supercoiled; (+), positively supercoiled.

(F) GapR profiles as in (A) with GapR$^{1-76}$ ChIP-seq (green).

(G) Mean GapR$^{1-76}$ ChIP in the middle of gene bodies (black), divergently transcribed regions (blue), convergently transcribed regions (red), and where transcription is in the same orientation (purple).

(H) GapR profiles as in (A) showing GapR accumulation at a representative region of convergent transcription.

(I) GapR profiles as in (A) showing GapR de-enrichment at a representative region of divergent transcription.

Figure S4. *S. cerevisiae* GapR ChIP-seq, related to Figure 4.

(A) Transcriptional profiles of cells with and without GapR-3xFLAG integrated at LEU2. The largest change in expression pattern is at the LEU2 locus.

(B) ChIP-seq profiles for GapR-3xFLAG (orange) and untagged GapR (grey) cells grown in glycerol. ChIP-seq profiles for GapR-3xFLAG cells grown in raffinose and in raffinose + α-factor. Cells were grown to OD 0.3-0.5 before addition of 2% galactose for 6 hr followed by fixation and ChIP.

(C) Correlation between two independent GapR-3xFLAG ChIP-seq experiments in raffinose.

(D) GapR-3xFLAG ChIP-seq in raffinose versus AT content. Mean enrichment at a given % AT (red dots). Motifs from DREME (below).

(E) GapR ChIP profiles in raffinose at a locus on chromosome II. AT content (top), AT content below the genomic average (64%) is plotted in reverse. Normalized ChIP-seq (middle) of GapR-3xFLAG (orange) or untagged GapR (grey) expressing cells. Transcription (bottom) from the forward (green) and reverse (blue) strands with annotated genes indicated.

(F) Histogram showing the length of GapR binding events.

(G) Mean GapR enrichment in glycerol at 5' and 3' ends of long genes at various transcriptional cutoffs. Student's t-test p-value is shown.

(H) GapR enrichment in glycerol is highest between convergently oriented genes.

(I) GapR-bound regions in glycerol are more frequently between convergent genes. Pie charts shown as in Fig. 3H.

Figure S5. GapR ChIP compared to nucleosome occupancy and chromatin accessibility, related to Figure 5.

(A) Comparison of transcription in raffinose with and without α-factor treatment.

(B) GapR enrichment at *YGP1* in raffinose without (orange) or with (green) α-factor arrest before GapR induction (top). Transcription of the forward (green) and reverse (blue) strands in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(C) α-factor dependent GapR enrichment at 5' and 3' ends for induced vs repressed genes. Student's t-test p-value is reported.

(D) GapR enrichment in raffinose (orange) compared to nucleosome occupancy measured by MNase-seq (light grey) and chromatin accessibility measured by DNase-seq (dark grey) at the yeast mating loci, HML and HMR. Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(E) GapR binding compared to nucleosome occupancy and DNase-seq at a locus on chromosome IV as in (D).

(F) Heatmap showing MNase-seq at the 500 highest GapR enriched regions.

(I) Heatmap showing GapR enrichment in raffinose at the 500 most DNAase-accessible regions.

Figure S6. GapR ChIP at centromeres, pericentromeres, and cohesin-bound regions, related to Figure 6.

(A) GapR enrichment in glycerol (blue), in raffinose (orange) and after α-factor arrest (green) at CEN3 (left) and CEN11 (right). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (below).

(B) GapR enrichment over all centromeres in raffinose. Mean enrichment (solid line) with 95% confidence intervals (shaded area). Grey bar represents position of centromeres.

(C) GapR enrichment at pericentromeres in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(D) GapR (glycerol, top) and cohesin (Scc1 enrichment, bottom) are associated with convergent genes (arrows) that mark the boundaries of pericentromeres (shaded areas).

(E) GapR and cohesin binding outside of pericentromeres as in (D).

(F) Heatmap showing GapR (top) and cohesin (bottom) enrichment for the 500 most GapR-enriched genomic regions in raffinose, after α-factor arrest, and in glycerol.

Figure S7. GapR ChIP at rDNA, autonomously replicating sequences, and R-loops, related to Figure 7.

(A) GapR ChIP with controls in the rDNA locus. GapR-3xFLAG ChIP, GapR-3xFLAG input, GapR ChIP, in raffinose (top), GapR-3xFLAG ChIP, GapR-3xFLAG input, GapR ChIP, in raffinose + α-factor (bottom).

(B) GapR enrichment at autonomously replicating sequences (ARS) in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(C) Heatmap showing GapR enrichment (glycerol) at all ARS sorted by GapR enrichment.

(D) Examples of GapR enrichment near Ty elements (purple). GapR enrichment after α-factor arrest (top), DNA-RNA hybrids by DRIP-seq (dark grey). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(E) Alignment of GapR and DRIP enrichment surrounding all yeast Ty elements. Data indicate mean (solid line) with 95% confidence intervals (shaded area).

(F) Examples of GapR enrichment near a telomere with (top) or without (bottom) a Y' element, plotted as in (D).

(G) GapR and DRIP enrichment at all telomeres with (left) and without (right) Y' elements, aligned from the end of the telomeric repeat. Data indicate mean (solid line) with 95% confidence intervals (shaded area). X element represents both the 'core' X element and any X element combinatorial repeats.

(H) Association of DNA-RNA hybrid (R-loop) forming regions and positive supercoiling.

Supplemental Table

Supplemental Table S1. Strains, plasmids, and primers.

Materials and Methods

Growth conditions and chemical treatments

E. coli strains were grown in LB (10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract) at 37 °C with shaking at 200 rpm unless noted. Antibiotics were supplemented as needed (carbenicillin: 50 μg mL^{-1} liquid / 100 μg mL^{-1} plate, and kanamycin: 30 μg mL^{-1} / 50 μg mL^{-1}). The P_{tet} promoter was induced by supplementing with 25 ng/mL anhydrous tetracycline (aTc, diluted in water) for 2 hr. For transcriptional inhibition experiments, rifampicin 300 μg/mL was added for 20 min before fixation and ChIP. *S. cerevisiae* strains were grown in YPD, YEP + 2% glycerol, YEP + 2% raffinose, or in SC -LEU as appropriate. The $P_{gal1\text{-}10}$ promoter was induced by addition of 2% galactose for 6 hr. For G1 arrest, α-factor was added to 1 μg/mL in YEP + 2% raffinose for 2 hr before addition of galactose for 6 hr before cell harvest. Optical density was measured at 600 nm using a Genesys 10 Bio Spectrophotometer or in a Synergy H1 plate reader.

Strain construction

E. coli strains were derivatives of MG1655 K-12 (ML6) or BL21(DE3). Strain ML3323 was constructed by electroporation of pKS22b-hSUMO-$GapR^{1\text{-}76}$ into BL21(DE3). Strain ML3324 was constructed by electroporation of pKS22b-hSUMO-$GapR^{1\text{-}81}$ into BL21(DE3). Strain ML3284 was constructed by electroporation of pGapR-3xFLAG into ML6. Strain ML3285 was constructed by electroporation of $pGapR^{WT}$ into ML6. Strain ML3286 was constructed by electroporation of pGapR-3xFLAG into ML6. Strain ML3325 was constructed by electroporation of $pGapR^{1\text{-}76}$-3xFLAG into ML6.

S. cerevisiae strains were derivatives of OAY470 (ML3287, gift from S. Bell lab). Strain ML3288 was constructed by transforming PmeI cut pNH605-gal1/10-$GapR^{WT}$ into ML3287 and selecting on SC -LEU plates. Strain ML3289 was constructed by transforming PmeI cut pNH605-gal1/10-GapR-3xFLAG into ML3287 and selecting on SC -LEU plates.

Plasmid construction

E. coli plasmids: Expression plasmids pKS22b-hSUMO-$GapR^{1\text{-}76}$ was constructed by amplifying *Caulobacter crescentus* genomic DNA with primers pKS22-GapR_NcoI and pKS22-GapR1-76_NotI and pKS22b-hSUMO-$GapR^{1\text{-}81}$ was constructed by amplifying *C. crescentus* genomic DNA with

primers pKS22-GapR_NcoI and pKS22-GapR1-81_NotI. These PCR products were digested with the NcoI and NotI and ligated into pKS22b digested with the same enzymes.

pGapR^{WT} was constructed by amplifying *C. crescentus* genomic DNA with primers pKVS45-gapR-f and pKVS45-gapR-r and pGapR-3xFLAG was constructed by amplifying *C. crescentus* genomic DNA hosting GapR-3xFLAG with primers pKVS45-gapR-f and GapR-pKVS_3xF_r. These PCR products were digested with the restriction enzyme BsmBI and ligated into pKVS45 digested with the same enzyme. pGapR$^{1-76}$-3xFLAG was constructed from pGapR-3xFLAG by round the horn mutagenesis using phosphorylated primers GapR1-76_l and GapR3xF_r.

Yeast Integration plasmids: pNH605-gal1/10-GapR^{WT} was constructed by performing splice-overlap-extension (SOE) PCR with a fragment containing the Gal1-10 promoter generated by PCR with primers Gal1/10_ClaI_f and Gal1/10_ClaI_r and a fragment containing GapR^{WT} generated with primers GapR_fwd and GapR_Xho_r3. The resulting SOE product was digested with ClaI and XhoI and ligated into pNH605 digested with the same restriction enzymes.

pNH605-gal1/10-GapR-3xFLAG was constructed by performing splice-overlap-extension (SOE) PCR with a fragment containing the Gal1-10 promoter generated by PCR with primers Gal1/10_ClaI_f and Gal1/10_ClaI_r and a fragment containing GapR-3xFLAG generated with primers GapR_fwd and GapR_Xho_r. The resulting SOE product was digested with ClaI and XhoI and ligated into pNH605 digested with the same restriction enzymes.

Purification of GapR

GapR was purified as reported previously (Guo et al., 2018) with the following modifications: 1L of His$_7$-SUMO-GapR expressing BL21(DE3) cells were grown in 2xYT to OD_{600} ~ 0.4-0.5 at 37 °C and induced with 0.4 mM IPTG for 18-20 hr at 18 °C. Cells were then harvested by centrifugation and resuspended in 40 mL buffer A (50 mM sodium phosphate [pH 7.5], 0.5 M NaCl, 20 mM imidazole) supplemented with a SIGMA*FAST* Protease Inhibitor Tablet (Sigma). The cell resuspension was then lysed using a Microfluidizer (15000 psi, 5 passes). The cell debris was cleared by centrifugation for 30 min at 10,000 x *g* and passed over Ni-NTA agarose resin (Qiagen) pre-equilibrated with buffer A at 4 °C. Resin was washed with buffer A, then with A containing 100 mM imidazole. GapR was then eluted with buffer A containing 500 mM imidazole. SUMO protease was then added and SUMO cleavage proceeded overnight at 4 °C with dialysis into fresh buffer A. To remove uncut His$_7$-SUMO-

GapR and SUMO protease, protein was then passed over Ni-NTA agarose resin, collecting the flowthrough, and then an additional column volume of buffer A was passed and collected from the column. The flowthrough was then diluted two-fold and then directly applied to a HiTrap Heparin HP (5mL) (GE Healthcare) column, pre-equilibrated in buffer B (50 mM sodium phosphate [pH 7.5], 10 mM NaCl). GapR was eluted with a 2-step elution (buffer B + 0.5 M NaCl and buffer B + 1.0 M NaCl), each step being 5 column volumes. 1M NaCl fractions containing GapR were collected and concentrated on an Amicon 3K Centrifugal Filter Unit (Millipore) and buffer exchanged into storage buffer (50 mM sodium phosphate pH 7.5, 200 mM NaCl, 10% (v/v) glycerol) by gel filtration using a Superdex 200 Increase 10/300 GL column (GE Healthcare). Fractions containing GapR were identified by SDS-PAGE/Coomassie staining, concentrated, snap frozen, and stored at -80°C.

GapR$^{1-76}$ was purified as with GapR with the following modifications: 0.2 L of His$_7$-SUMO-GapR$^{1-76}$ expressing BL21(DE3) cells were grown in 2xYT to OD_{600} ~ 0.4-0.5 at 37 °C and induced with 0.4 mM IPTG for 18-20 hr at 18 °C. Cells were then harvested by centrifugation and resuspended in 40 mL buffer A (50 mM sodium phosphate [pH 7.5], 0.5 M NaCl, 20 mM imidazole) supplemented with a SIGMA*FAST* Protease Inhibitor Tablet. The cell resuspension was then lysed by sonication (5 cycles of 30 sec on, 30 sec off at 40 power on a Qsonica Q700). The cell debris was cleared by centrifugation for 10 min at 10,000 x *g* followed by centrifugation for 30 min at 30,000 x g, and passed over Ni-NTA agarose resin. Resin was washed with buffer A, then eluted stepwise with 40, 60, 80, 100, and 500 mM imidazole. 80-500 mM imidazole fractions were combined and applied directly to a HiTrap Heparin HP (1 mL) (GE Healthcare) column, pre-equilibrated in buffer B (50 mM sodium phosphate [pH 7.5], 10 mM NaCl). His$_7$-SUMO-GapR truncations were washed with 5 mL 10 mM NaCl buffer B, then 2.5 mL 0.5 M NaCl buffer B, then eluted with 2.5 mL 1.0 M NaCl buffer B. 1 M NaCl fractions were dialyzed overnight in the presence of SUMO protease into fresh buffer A. To remove uncut His$_7$-SUMO-GapR$^{1-76}$, His$_7$-SUMO, and SUMO protease, dialyzed protein was passed over fresh Ni-NTA agarose resin, and washed with an additional column volume of buffer A, collecting the flowthrough throughout. The flowthrough was then applied to a HiTrap Heparin HP (1 mL) column, and processed as before. 1M NaCl fractions containing GapR$^{1-76}$ were collected and concentrated on an Amicon 3K Centrifugal Filter Unit (Millipore) and exchanged into storage buffer (50 mM sodium phosphate pH 7.5, 200 mM NaCl, 10% (v/v) glycerol) by gel filtration using a Superdex 200 Increase 10/300 GL column. Fractions containing GapR$^{1-76}$ were identified by SDS-PAGE/Coomassie staining, concentrated, snap frozen, and stored at -80°C.

Magnetic tweezer assays

Magnetic tweezer (MT) assays were performed following a previously described protocol (Giuntoli et al., 2015). Briefly, experiments were carried out using the 9702-bp plasmid pNG1175, which was linearized with SpeI and ApaI restriction enzymes (Bai et al., 2011). The linear molecule was ligated to ~900-bp PCR products carrying either biotinylated or digoxigenin-labeled nucleotides, with SpeI- and ApaI- compatible ends, respectively, resulting in a 11.4 kb long DNA fragment (leading to ~10-kb DNA tethers once attached to the flow cell).

Experiments were performed in flow cells of approximately 30 µL volume of channel were constructed by sandwiching parafilm with coverslips and glass slides. Flow cells were functionalized by incubating with 0.1 mg/mL anti-digoxigenin antibody (Roche) and 3 µm polystyrene beads (Polysciences) suspended in 1X PBS. Before an experiment, flow cells were incubated with 1% BSA in 1X PBS for overnight at 4 °C to block non-specific binding. To bind biotinylated DNA to beads, 1 ng of end-labeled pNG1175 DNA was mixed with 1 µL streptavidin-functionalized paramagnetic beads (M-270 Dynabeads, diluted 1:6 in 1X PBS with 0.1% BSA; Invitrogen) in 10 µL PBS, and incubated for 10 min at room temperature with inversion. The bead-bound DNA was then diluted with 100 µL 1X PBS and introduced into the flow cell, with the flow cell inverted upside down, and incubated for 10 min at room temperature to allow the digoxigenin-functionalized DNA ends to bind the anti-digoxigenin-functionalized coverslip. A fraction of the paramagnetic particles adsorb permanently to the cover slip, serving as reference beads to determine the position of the glass surface.

The assembled flow cell was then placed on a magnetic tweezers microscope setup, consisting of a 100× 1.35 NA (Olympus) microscope objective on a piezoelectric positioner (Piezojena), with permanent neodymium magnets that are positioned using a stepper-motor-driven translator as previously described (**Fig. S1A**) (Giuntoli et al., 2015; Skoko et al., 2004). Movement of the piezoelectric positioner and the consequent force applied to the DNA is controlled by moving the magnets closer or further from the flow cell. The relative positions of the reference beads and DNA-tethered bead are measured using an algorithm that uses the degree of focus of the beads to determine their distance from the glass surface. Labview (National Instruments) is used to control the vertical (z) position and rotation of the magnets, track the x-y motions of the beads, measure the z position of the beads, and calculate the forces on the beads as previously described (Skoko et al., 2004).

At the start of each experiment, beads were tested to identify a supercoilable DNA molecule for further study (non-nicked single dsDNA that is tethered to the bead). First, the DNA molecule was rotated with the magnet to verify that the extension length of the DNA changed significantly upon rotation, indicating that the DNA is supercoilable. Next, the DNA length was measured under a variety of forces to verify that the given bead was attached by a single molecule of DNA. An initial test is that the extension length under high forces is ~2.8 μm, the expected length for 11.4 kb DNA. Next, extension and force measurements at five magnet positions were performed to verify that the apparent persistence length of the candidate molecule is as expected for a singly tethered DNA.

Experiments were performed in GapR assay buffer (40 mM sodium phosphate [pH 7.5], 100 mM NaCl). The "rotation extension" experiments were performed as follows. First, naked DNA extensions were measured against torque by rotating the magnet to twist DNA between σ = -0.03 and +0.03 at 0.3, 0.5 and 1.0 pN of magnetic forces. Next, we repeated the same series of measurements under the presence of GapR. For the rotation extension hysteresis experiment, one GapR-bound DNA was repeatedly turned from 0.0 σ to +0.03 σ down to -0.03 σ and back up to +0.03 σ in 0.005 σ steps. MT data was analyzed using Prism 7.

Growth rate analysis

Growth rate was measured using a Synergy H1 plate reader. Cells were grown overnight without inducer, diluted, and grown into mid-log OD_{600} 0.2-0.4. Cells were then diluted to OD_{600} 0.01 in 96 well plates in the presence or absence of aTc and grown for 8 hr at 37 °C with shaking at 200 rpm.

Chromatin Immunoprecipitation Sequencing (ChIP-seq)

GapR and $GapR^{1-76}$ ChIP in *E. coli* was performed as previously described for *Caulobacter* (Guo et al., 2018). Briefly, cell cultures (20 mL) were grown to OD_{600} ~0.3, diluted back to OD ~0.01, and 25 ng/mL aTc was added for induced cultures. Cells were grown for 2 hr (OD ~0.3) and then fixed by the addition of 10 mM sodium phosphate [pH 7.6] and 1% formaldehyde (final concentrations) (Sigma). When required, 25 μg/mL of rifampicin (Sigma) was added to cells for 20 minutes prior to fixation. Fixed cells were incubated at room temperature for 10 minutes and then quenched with 0.1 M glycine (Sigma) for 5 min at room temperature followed by 15 min on ice. Cells were washed three times with 1X PBS [pH 7.4] and resuspended in 500 μL of TES buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 100 mM NaCl), frozen in liquid nitrogen, and stored at -80 °C until use. Cells were then thawed and

35,000 U of Ready-Lyse (Epicentre) was added. Following 15 min incubation at RT, 500 µL of ChIP buffer (16.7 mM Tris-HCl [pH 8.1], 167 mM NaCl, 1.1% Triton X-100, 1.2 mM EDTA) containing protease inhibitors (SIGMA*FAST* Protease Inhibitor Tablets) was added. After 10 min at 37 °C, the lysates were sonicated on ice and cell debris cleared by centrifugation. Supernatant protein concentration was measured by Bradford assay (Thermo Scientific) and 500 µg of protein were diluted into 1 mL of ChIP buffer + 0.01% SDS. The diluted supernatants were pre-cleared for 1 hr at 4 °C on a rotator with 50 µL of Protein-A Dynabeads (Life Technologies) pre-blocked overnight in ChIP buffer + 0.01% SDS and 100 µg ultrapure BSA (Ambion). Beads were pelleted and 90 µL of the supernatant was removed as input DNA and stored at -80 °C, the remaining pre-cleared supernatant was incubated rotating at 4 °C overnight with 1 µL of FLAG antibody (Sigma). The immune complexes were captured for 2 hr at 4 °C with 50 µL of pre-blocked Protein-A Dynabeads. Beads were then washed consecutively at 4 °C for 15 min with 1 mL of the following buffers: low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl [pH 8.1], 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl [pH 8.1], 500 mM NaCl), LiCl wash buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl [pH 8.1] and twice with TE buffer (10 mM Tris-HCl [pH 8.1], 1 mM EDTA). Complexes were then eluted twice from the beads with 250 µL of freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$). To reverse crosslinking, 300 mM of NaCl and 2 µL of RNase A (0.5 mg / mL) (Qiagen) were added to the collective eluates which were incubated at 65 °C overnight. Samples were then incubated at 45 °C for 2 hr with 5 µL of Proteinase K (NEB) in the presence of 40 mM EDTA (pH 8.0) and 40 mM Tris-HCl [pH 6.8]. DNA from the samples was then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma) and subsequently precipitated by adding 3M sodium acetate (pH 5.2), 15 µg glycoblue (Ambion) and 1 volume of ice cold isopropanol, and stored at -20 °C overnight. DNA was pelleted and washed with 75% ethanol and resuspended in TE buffer [pH 8.0]. Input ChIP libraries were generated processing 50 µL of the yeast lysate, by reversing crosslinks, Proteinase K treatment, and phenol:choloform:isoamyl alcohol extraction as with ChIP DNA.

For *S. cerevisiae*, ChIP was performed as previously described (Neurohr et al., 2019) with the following modifications. Cells were grown in YEP + 2% glycerol to OD600 ~0.5, and 2% galactose was added for 6 hr while culture OD was maintained between OD 0.5-1.0. For α-factor arrest experiments, cells were grown in YEP + 2% raffinose to OD600 ~0.4, arrested in α-factor for 2 hr, before addition of 2% galactose for 6 hr. Cells were then fixed for 15 min at room temperature with 1% final concentration formaldehyde followed by quenching with glycine. 100 mL of cells were harvested by centrifugation

and then washed twice with 1X PBS [pH 7.4], resuspended in 1 mL FA-lysis buffer (50 mM HEPES-KOH pH7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS) and transferred into a 2 mL screw-cap eppindorf tube. Cells were pelleted quickly in a tabletop centrifuge and the supernatant was discarded by aspiration. Cells were then resuspended in 500 µL FA-lysis buffer, snap frozen in liquid nitrogen, and stored at -80 °C until use. Cells were then thawed, and FA-lysis buffer (containing SIGMA*FAST* Protease Inhibitor Tablets) and SDS was supplemented to bring up the volume to 1 mL with 0.5% SDS final concentration. 1 mL of glass beads were added and the cells were disrupted on a Fast Prep until 80-90% of cells were lysed (intensity 6.5, each 45 s cycle followed by 5 min of cooling, 5-10 cycles in total as confirmed by visual inspection). Cell debris was separated from beads by piercing the tube cap and bottom with a syringe needle, inverting the tube over a 1 mL tip in a 15 mL conical tube, and centrifuged for 1 min at 800 rpm. 9 mL of FA-lysis buffer (with protease inhibitor) was added and the lysate was ultracentrifuged in an SW41 rotor at 32,700 rpm for 20 min to pellet chromatin. The pellet was mechanically disrupted with a wooden stick and transferred to a 1.5 mL TPX microcentrifuge tube (Diagenode), resuspended in 250 µL FA-lysis buffer (with protease inhibitor), and pipetted to resuspend. Samples were sonicated at 4 °C in a Bioruptor Waterbath Sonicator (Diagenode) for 5 cycles (30 sec on, 30 sec off, high intensity), followed by further pipetting to fully resuspend chromatin. Samples were then sonicated for an additional 3 x 10 cycles (30 sec on, 30 sec off, high intensity) and quick-spun in a picofuge to recover material from tube walls every 10 cycles. After sonication, 500 µL additional FA buffer was added to the lysate and cellular debris was discarded by centrifugation at 4 °C (15 min, ~20,000 x *g*). The supernatant was transferred to a fresh Eppendorf tube, snap frozen, and stored at -80 °C. Total protein was measured by Bradford assay, samples were diluted to contain 1 mg of protein in 1 mL ChIP buffer + 0.01% SDS. Samples were then processed as with bacterial ChIP-seq, using 1 µL of anti-FLAG antibody (Sigma) for each immunoprecipitation.

ChIP sequencing libraries were built from immunoprecipitated DNA by first end repairing the DNA with 5 µL T4 DNA polymerase (NEB), 5 µL T4 PNK (NEB), and 1 µL Klenow large fragment (NEB) in 100 µL T4 DNA ligase buffer with 0.25 mM dNTPs for 30 min at room temperature. Repaired DNA was recovered by Ampure XP (Beckman Coulter) bead purification, using 100 µL beads in 300 µL 20% PEG/NaCl solution. Beads were washed twice with 80% ethanol, dried, and resuspended in 32 µL EB. The bead slurry was directly treated with 3 µL Klenow (3'→5' exo-) (NEB) in 50 µL NEB Buffer #2 with 0.2 mM ATP at 37 °C for 30 min to add 3' overhangs to DNA. Repaired DNA was recovered by Ampure XP capture and resuspended in 23 µL EB. Y-shaped adaptors were prepared by

annealing Illumina PE adapter 1 and Illumina PE adapter 2. Y-shaped adapters were added to bead slurry, and the mix was ligated in 50 µL total volume with 1.5 µL T4 DNA Ligase (NEB) for 1 hr at room temperature. Ligated library DNA was recovered and free and ligated adapters discarded using an Ampure XP capture at 0.85X volume. Library DNA was recovered by eluting Ampure beads with 33 µL 10 mM Tris-HCl [pH 8.0]. DNA libraries were amplified in 50 µL final volume with 2X KAPA HiFi Master Mix (Roche) supplemented with 5 % final concentration DMSO (Sigma) and appropriate barcoded primers. The total number of cycles was optimized for each sample to minimize the number of cycles required for library generation. Libraries were purified by two-step Ampure XP capture, by first adding 0.5X reaction volume Ampure XP and transferring the supernatant to a fresh tube to discard large fragments, followed by a second capture by adding Ampure XP to 0.82X of the original reaction volume to recover 200-500 bp amplified libraries. DNA was recovered from Ampure beads by resuspending in 20 µL 10 mM Tris-HCl [pH 8.0]. Insert size of ChIP libraries was determined to be ~250 bp on average. Paired-end sequencing of libraries was performed on either a NextSeq or a MiSeq at the MIT Bio Micro Center.

Caulobacter crescentus GapR-3xFLAG ChIP-seq was from GSE100657 (Guo et al., 2018). *S. cerevisiae* Scc1 enrichment in the presence of microtubule tension was from GSE104135 (Paldi et al., 2020).

RNA Sequencing (RNA-seq)

RNA-seq in *E. coli* was performed as previously described (Culviner and Laub, 2018). Cells were harvested immediately before ChIP-seq. 5 mL of cells were harvested by into a 5% phenol, 95% ethanol stop solution. RNA was harvested by phenol/chloroform extraction, and treated with 2 µL Turbo DNase (Invitrogen) with 5 µL SuperaseIN (Invitrogen) in 100 µL total volume at 37 °C for 20 min. RNAs were subsequently recovered by phenol/chloroform extraction. rDNA was removed using in-house protocols (Culviner et al., 2020). mRNA was subsequently fragmented RNA fragmentation reagents (Invitrogen) and cDNA was generated with random primers and Superscript III (Invitrogen). Second strand synthesis was conducted using dUTP instead of dTTP and RNase H (NEB), *E. coli* DNA ligase (NEB), and DNA Pol I (NEB) were added, followed by an incubation at 16°C for 2.5 hr. cDNA was recovered by Ampure XP (Beckman Coulter) bead purification. cDNA was then end-repaired and converted into libraries as for ChIP-seq samples. Before library amplification, the dUTP-containing second strand was digested by adding 1 µL of USER enzyme (NEB) and incubating at 37 °C for 15 minutes, followed by 95 °C for 5 minutes to inactivate USER enzyme. Libraries were generated as for

ChIP-seq samples, and paired-end sequencing was performed on a NextSeq at the MIT Bio Micro Center.

RNA-seq in *S. cerevisiae* was performed by harvesting cells immediately before ChIP-seq or in mid-log phase (OD_{600} < 1.0). 5-10 mL of cells were pelleted at 8000 rpm for 5 min and washed with 500 µL H_2O before being snap-frozen in liquid nitrogen. RNA was isolated by resuspension in 500 µL 10 mM Tris-HCl [pH 7.5], 10 mM EDTA, and 0.5% SDS followed by addition of 500 µL hot acid-phenol. Cells were then shaken in a thermomixer at 65°C at 2000 rpm for 20 min, before incubation for 5 min on ice, and phenol extraction. RNA was then isolated by phenol-chloroform extraction followed by a chloroform wash, and precipitation with isopropanol. gDNA was removed by addition of Turbo DNase and total RNA was subsequently recovered by phenol/chloroform extraction. mRNA was isolated using poly(dT) pulldown using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and processed into libraries as with *E. coli* RNA. Paired-end sequencing was performed on a NextSeq at the MIT Bio Micro Center.

Electrophoretic mobility shift assays (EMSA) and DNA topology assays

For EMSA, linear 210 bp DNA was generated by PCR and purified with PCR Purification Kits (QIAGEN). Reactions (15 mL) with indicated amounts of GapR and 210 bp DNA (40 ng) in binding buffer (40 mM sodium phosphate [pH 7.5], 100 mM NaCl, 50 mg/mL ultrapure BSA, 0.5 mM DTT) were incubated at 30°C for 60 min and then placed on ice. DNA loading buffer was added and 10 mL of the reactions were electrophoresed on 6% DNA Retardation gels (Invitrogen) at 130 V for 60 min at 4 °C. Gels were stained in SYBR Gold (Invitrogen) and imaged with a Typhoon FLA 9500 imager (GE Lifesciences).

For DNA topology assays, nicked pUC19 was generated from negatively supercoiled pUC19 (NEB) by treatment with Nt.BspQI (NEB) followed by PCR purification. Mixtures of GapR and nicked pUC19 DNA (40 ng) in 1X T4 DNA ligase buffer were incubated at 30°C for 60 min. T4 DNA ligase (NEB) was diluted to 1 U/mL in 1X T4 DNA ligase buffer and 1 U was added to reactions and incubated for an additional 2 hr at room temperature. Reactions were stopped by addition of 1% SDS and 10 mM EDTA (final concentration), and digested with Proteinase K (NEB) for 1 hr at 37°C. DNA loading buffer was added and samples electrophoresed. For one-dimensional electrophoresis, 1% TBE agarose gels were run at 130 V for 90 min and then in SYBR Gold and imaged with a Typhoon FLA 9500 imager. For two-dimensional chloroquine gels, electrophoresis was performed by first running

reactions on a 1% TBE agarose gel at 130 V for 90 min, then soaking the gel for 2 hr with shaking in 1X TBE supplemented with 4.5 mg/mL chloroquine phosphate (Santa Cruz Biotech). The gel was then turned 90° and electrophoresed in the orthogonal direction at 130 V for 60 min in 1X TBE supplemented with 4.5 mg/mL chloroquine phosphate. Chloroquine is a DNA intercalator that introduces (+) supercoils. In chloroquine, (-) supercoiled plasmids will become more relaxed, and migrate more slowly, whereas (+) supercoiled DNA will be further compacted, increasing its migration speed. The gel was washed 4 x 20 min in distilled water to remove chloroquine, stained with SYBR Gold for 2 hr, and imaged with a Typhoon FLA 9500 imager. Relaxed plasmid standard was generated with *E. coli* Topoisomerase I (NEB).

Sequencing data processing

Data analysis was performed with custom scripts in Python 3.6.9. For all histograms, a kernel density estimation (KDE) was generated, with the y-axis units indicating KDE density. All t-tests performed were two-tailed tests.

For *E. coli* ChIP, paired-end reads were mapped to the MG1655 reference genome (NC_000913.2) using bowtie2 with default parameters (Langmead and Salzberg, 2012). For *S. cerevisiae* ChIP-seq, paired-end reads were mapped to the reference genome (S288C Scer3) using bowtie2 with default parameters (Langmead and Salzberg, 2012). Once aligned, unique reads were isolated and read extension and pile-up was performed using the bedtools function genomeCoverageBed (Quinlan and Hall, 2010) and converted into wig format using custom Python scripts. The data were then smoothed by convolution with a Gaussian (m = 0, s = 250bp, x = (1000 bp, +1000 bp)) and then normalized to reads per million (rpm). Because different growth conditions (e.g. with and without α-factor arrest) led to variable rDNA copy number, experiments were normalized to total count excluding chromosome XII (containing the rDNA). Data were then smoothed over 250 bp. To generate *S. cerevisiae* GapR enrichment, a pseudocount was added to each position and the GapR-3xFLAG ChIP was normalized by the $GapR^{WT}$ ChIP (GapR-3xFLAG ChIP + 0.01)/($GapR^{WT}$ ChIP + 0.01). Scc1 ChIP-seq occupancy ratio was calculated from GSE104135 as reported (Paldi et al., 2020). Correlation between two ChIP experiments was generated by binning data every 100 nt.

For *E. coli* RNA-seq, paired-end reads were mapped to the MG1655 reference genome (NC_000913.2) and to the GapR expression plasmid using bowtie2 with default parameters (Langmead and Salzberg, 2012). Duplicated reads were filtered out and the read coverage was mapped genome by assigning each

mapped base a value of 1. To calculate mRNA expression levels, the number of reads mapped to a gene was divided by the length of the gene and normalized to yield the mean number of reads per kilobase of transcript per million sequencing reads (rpkm).

To determine if ectopic expression of GapR alters global supercoiling in *E. coli*, we compared the rpkm of all expressed genes with and without GapR expression (genes with rpkm > 20 in either condition, ~2500 genes). To examine the effects of GapR on expression of known supercoiling-sensitive genes, we compared the rpkm of genes known to be induced or repressed upon topoisomerase inhibition (Peter et al., 2004) as well as for the DNA gyrase and topo IV subunits P_{gyrA}, P_{gyrB}, P_{parC}, and P_{parE}, which have been reported to be supercoiling-sensitive in *E. coli* or other bacteria (Ferrándiz et al., 2016; Menzel and Gellert, 1987).

S. cerevisiae RNA-seq was analyzed by aligning to SacCer3 using bowtie2 with default parameters (Langmead and Salzberg, 2012). Duplicated reads were filtered out and the read coverage and rpkm values were calculated as for *E. coli*.

Identifying AT-bias and GapR-associated DNA motifs

AT content at each base pair was computed using a centered 100 bp sliding window. To identify correlations between AT content and GapR binding, AT content was plotted versus GapR ChIP at each position. To identify DNA sequence motifs enriched in GapR bound sequences, we isolated the 35 regions with highest GapR ChIP signal as was described previously in *C. crescentus* (Guo et al., 2018). For *E. coli* ChIP, we isolated regions above 0.843 rpm (regions less than 150 bp apart were merged) as input sequences, and we isolated regions below 0.30 rpm for control sequences. For *S. cerevisiae*, we isolated GapR ChIP regions above 0.376 rpm from smoothed data (regions less than 150 bp apart were merged), and we isolated regions below 0.055 rpm for control sequences. A 200 bp window centered at the maximum (or minimum, for control sequences) signal intensity of each of these regions was retrieved and submitted to DREME for sequence motif analysis (Bailey, 2011). The highest DREME motif is reported.

Assessing GapR 5' and 3' end enrichment

For *E. coli*, transcription unit (TU) annotation was taken from the Ecocyc operon annotation. GapR occupancy at the 5' and 3' ends were calculated by examining transcription units ≥ 1500 bp and determining the change in GapR per base in the 1000 bp before and after the transcriptional start site

or the transcription termination site: e.g., mean($GapR_{-1000..start}$) – mean($GapR_{start..1000}$). TUs were filtered to prevent redundancy; for divergently or convergently transcribed regions that are within 1000 bp, GapR occupancy is only calculated for one strand.

For *S. cerevisiae*, publicly available annotation datasets do not contain transcriptional start or termination site information, and only include coordinates for the coding regions of genes. Gene start and end positions were used as a proxy for transcriptional start and termination sites. GapR occupancy was determined by examining genes ≥ 1000 bp and calculating the mean normalized GapR enrichment in the 500 bp before or after the start or end of genes: e.g., mean($GapR_{end...500}$). TUs were filtered to prevent redundancy; for divergently or convergently transcribed regions that are within 500 bp, GapR occupancy was only calculated for one strand.

Assessing transcription-dependent GapR 5' and 3' end enrichment

For *E. coli*, GapR occupancy at the 5' and 3' ends were calculated by examining transcription units ≥ 1500 bp and determining the mean change in GapR in the 1000 bp before and after the transcriptional start site or the transcription termination site: e.g., (mean($GapR_{-1000..start}$)– (mean($GapR_{start..1000}$)). Transcription-dependent change in GapR occupancy at the 5' and 3' ends was calculated by examining transcription units ≥ 1500 bp and determining the mean change in GapR in the presence and absence of rifampicin in the 1000 bp before and after the transcriptional start site or the transcription termination site: e.g., (mean($GapR_{-1000..start}$) – mean($GapR+Rif_{-1000..start}$)) – (mean($GapR_{start..1000}$) – mean($GapR+Rif_{start..1000}$)). The transcriptional strength was calculated for each TU from GapR-3xFLAG induced RNA-seq data by determining the mean the number of reads mapped over each transcription unit and normalizing to yield the mean number of reads per kilobase of transcript per million sequencing reads (rpkm). Transcription unit rpkm cutoffs were chosen to isolate the highest expressing 125 and 250 TUs (>65, >25.7 rpkm), and the lowest expressing 250 TUs (< 3.284 rpkm). TUs were filtered to prevent redundancy; for divergently or convergently transcribed regions that are within 1000 bp, GapR occupancy is only calculated for one strand. To generate heatmaps of GapR enrichment at 5' and 3' ends of genes, TUs ≥ 1500 bp were sorted by expression level and the change in GapR in the presence and absence of rifampicin in 6 kb window around the transcriptional start site or the transcription termination site (e.g., $GapR_{-4000..start..2000}$ – $GapR+Rif_{-4000..start..2000}$) was plotted for the 300 highest and lowest expression TUs.

For *S. cerevisiae*, transcriptional strength was calculated similarly, except by examining transcription units ≥ 1000 bp and determining the mean GapR enrichment in the 500 bp before or after genes. Transcriptional cutoffs were chosen to isolate the highest expressing 125, 250, and 500 genes (> 455, >225.3, >110 rpkm), and the lowest expressing 500 genes (< 9.385). TUs were filtered to prevent redundancy; for divergently or convergently transcribed regions that are within 500 bp, GapR occupancy is only calculated for one strand.

Identifying GapR enriched regions

For *E. coli* GapR peaks, we isolated the top 5% of positions with greatest transcription-dependent GapR enrichment ($GapR_i$ – $GapR{+}Rif_i$ ≥ .118). The borders each GapR-bound region surrounding the enrichment peak was identified by determining where transcription-dependent GapR enrichment was above the mean + 1/3 of a standard deviation. Regions less than 150 bp apart are then merged.

For *S. cerevisiae* GapR peaks, we isolated the top 5% of positions with greatest GapR enrichment (e.g., $GapR_i$ ≥ 1.657 in raffinose). The borders each GapR-bound region surrounding the enrichment peak was identified by determining where transcription-dependent GapR enrichment was above the mean + 1/3 of a standard deviation. Regions less than 150 bp apart are then merged.

Assessing transcription-orientation dependent GapR enrichment

For *E. coli*, highly expressed TUs (≥ 17.3 rpkm, top third) were analyzed. TUs below the rpkm cutoff were discarded and assumed to be transcriptionally silent. For the remaining TUs, the regions between TUs was categorized based on whether the downstream TU is convergent, divergent, or in the same orientation. Intragenic regions < 50 bp were removed from the analysis. Mean GapR ChIP or mean transcription-dependent GapR ChIP was then calculated for each region and within TUs. The same analysis was repeated for determining the mean $GapR^{1\text{-}76}$ ChIP.

To examine transcription orientation at GapR ChIP peaks, peaks were identified as reported above. The same number of unenriched regions were identified by isolating the 7% of positions with lowest transcription-dependent GapR enrichment ($GapR_i$ – $GapR{+}Rif_i$ ≤ -0.0885). The borders of each GapR-bound region surrounding the enrichment peak was identified by determining where transcription-dependent GapR enrichment was below the mean - 1/3 of a standard deviation ($GapR_i$ – $GapR{+}Rif_i$ < -0.024). Regions less than 150 bp apart are then merged. At each GapR-enriched or unenriched region, the transcriptional propensity of the surrounding area was determined by the following procedure. First,

the mean number of reads on the forward and reverse strands was calculated for each region ± 5 kb on both sides; if the mean reads < 0.01, the region is assumed to be silent (fwd+rev < 0.01 = silent). Next, the midpoint of the enriched/unenriched region was determined, and the transcriptional strength for each strand was calculated from the midpoint to 2 kb past either end of the region. Transcriptional propensity is then assigned based on the relative transcriptional strength: $fwd_{left} > rev_{left}$ & $rev_{right} > fwd_{right}$ = convergent; $fwd_{left} < rev_{left}$ & $rev_{right} < fwd_{right}$ = divergent; other cases are assumed to be the same orientation. If the mean reads is < 0.01 within the 2 kb window, the window was expanded to 5 kb and the analysis repeated for the orientation assignment. Fisher's exact test was used to determine if GapR-enriched regions were more frequently associated with convergent transcription and de-associated with divergent transcription.

For *S. cerevisiae*, all genes were analyzed. Regions between genes was categorized based on whether the downstream gene is convergent, divergent, or in the same orientation. Mean GapR ChIP was then calculated for each region and within genes. Intragenic regions < 50 bp were removed from the analysis. To determine the transcriptional orientation at GapR enriched and unenriched regions, peaks were identified as detailed above. GapR unenriched were the 5% of positions with least GapR enrichment ($GapR_i \leq 0.626$). The borders each GapR-bound region surrounding the enrichment peak was identified by determining where transcription-dependent GapR enrichment was above the mean + 1/3 of a standard deviation. Regions less than 150 bp apart are then merged. At each GapR-enriched or unenriched region, the transcriptional propensity of the surrounding area was determined by the same procedure as with *E. coli* regions, except examining GapR ChIP.

MNase-seq and DNase-seq data processing

S. cerevisiae MNase-seq from GSM3069971 (Cutler et al., 2018) was analyzed by aligning to SacCer3 using bowtie2 with default parameters (Langmead and Salzberg, 2012). Reads were sorted and filtered with samtools (Li et al., 2009), and the center of each paired read was interpreted as the nucleosome dyad and plotted, as reported previously (Cutler et al., 2018).

S. cerevisiae DNase-seq from GSM1705337 (Zhong et al., 2016) was analyzed by aligning to SacCer3 using bowtie with the following parameters (Langmead et al., 2009) to map the first 20 base-pairs for each read: -n 2 -l 20 -3 30 -m 1 --best --strata. The position at first base-pair (5' end) of the alignment was assigned as the DNase cleavage site and given a mapped value of 1 and the total number of DNase reads were tabulated separately for the forward and reverse strands, as reported previously (Zhong et

al., 2016). For some analyses, the DNase-seq coverage was transformed by a $\log_{10}$ transform after addition of a pseudocount to each base: $\log_{10}$(DNase-seq+1).

MNase-seq and DNase-seq data analysis

To examine nucleosome occupancy and DNase hypersensitivity at GapR enriched regions, first GapR enriched regions were identified as detailed above. The regions were centered around the position of maximum GapR occupancy, and the mean GapR enrichment, MNase-seq, and DNase-seq at each base was determined for 1000-bp around the GapR peak. Correlation between GapR ChIP and DNase-seq or MNase-seq experiments was generated by binning data every 100 nt.

To identify open chromatin from DNase-seq, we isolated the top ~5% of positions with greatest DNase-seq reads (sum of the forward and reverse strand reads, DNase-seq$_i \geq 12$). The borders surrounding each DNase hypersensitive region was determined by where DNase-seq was above the mean + 1/3 of a standard deviation (DNase-seq$_i > 12.3$). Regions less than 10 bp apart are merged. To calculate GapR enrichment in open regions, the mean $\log_2$(GapR enrichment) was calculated for all DNase-seq peaks longer than 50 bp (8005 unique regions). This 50 bp length cutoff was used to ensure that the absence of GapR binding was not due to regions being shorter than a GapR binding site. For heatmaps, the maxima of the top 500 DNase-seq or GapR peaks was used as the midpoint, and the GapR, DNase-seq, or MNase-seq in a 4 kb window surrounding the peak is shown, with 10 bp binning.

To assess MNase-seq and DNase-seq at 5' and 3' ends, genes ≥ 1000 bp were examined the mean MNase-seq and $\log_{10}$(DNase-seq+1) in the 500 bp before or after the start or end of genes was calculated: e.g., mean(MNase-seq$_{end...500}$). TUs were filtered to prevent redundancy; for divergently or convergently transcribed regions that are within 500 bp, mean MNase-seq and DNase-seq was calculated for only one strand.

To assess MNase-seq and DNase-seq based on transcription orientation was performed as reported above with GapR enrichment. Briefly, all regions between genes were categorized based on whether the downstream gene is convergent, divergent, or in the same orientation. Mean MNase-seq and $\log_{10}$(DNase-seq+1) was calculated for each region and within genes. Intragenic regions < 50 bp were removed from the analysis.

Centromere, pericentromere, and cohesin analysis

The mean GapR enrichment at all centromeres was determined and compared to the mean GapR enrichment at all intergenic regions. Centromeres were aligned by their left position (oriented CDEI-CDEII-CDEIII), and the mean GapR enrichment and 95 % confidence interval at each base at all centromeres was determined for 4120-bp around the centromere.

Borders of pericentromeres were defined based on published analysis of cohesin binding and convergent genes (Paldi et al., 2020). Scc1 ChIP-seq occupancy ratio in the presence of microtubule tension was taken from GSE104135 (Paldi et al., 2020). To identify Scc1 peaks, we isolated the top 500 regions with greatest Scc1 enrichment first in the presence and then in the absence of tension. For heatmaps, the maxima of the top 500 Scc1 or the GapR peaks was used as the midpoint, and the GapR or Scc1 enrichment in a 4 kb window surrounding the peak is displayed, with 10 bp binning.

Autonomously replicating sequence analysis

The mean GapR enrichment at all autonomously replicating sequences (ARS) was determined and compared to the mean GapR enrichment at all intergenic regions. To generate heatmap of GapR enrichment at ARS, all ARS were aligned by their left position and the GapR enrichment was determined for a window -1000 to +2000 bp with 10 bp binning from this position

S1-DRIP-seq analysis

S. cerevisiae S1-DRIP-seq from SRP071346 (Wahba et al., 2016) was analyzed by aligning to SacCer3 using bowtie2 with default parameters. Genome coverage was mapped with bedtools function genomeCoverageBed and converted into wig format using custom Python scripts (Quinlan and Hall, 2010). Transposable (Ty) element locations were taken from Saccharomyces Genome Database. To generate alignment profiles of GapR at all Ty elements, and to accommodate the fact that Ty elements vary in length, each Ty element was divided into 20 bins, with the middle bin being larger or smaller to accommodate the overall size. 10 more equivalently-sized bins were then extended to either side of the ty element (~2500 bp). The mean GapR enrichment and S1-DRIP-seq with 95 % confidence intervals was then determined for all bins. At telomeres, the mean GapR enrichment and S1-DRIP-seq with 95 % confidence intervals was determined for the 1500 bp divided into 50 bp bins flanking each telomere. For Fig. S7F, GapR enrichment and S1-DRIP-seq was determined for 1500 bp divided into 50 bins starting from the first nucleotide (towards CEN) after the telomeric repeat sequence.

Acknowledgements

We thank M. Leroux, P. Culviner, D. Haakonsen, C. Tsokos, O. Rando, J. Gerton, and A. Maxwell for comments on the manuscript. We thank S. Srikant, M. Looke and G. Neurohr for strains and vital technical assistance, and members of the Laub lab for critical feedback. This work was supported by NIH grants to M.S.G. (K99GM134153), J.F.M. (U54CA193419 and U54DK107980), and to M.T.L. (R01GM082899). M.T.L. is also an Investigator of the Howard Hughes Medical Institute.

Declaration of Competing Interests

The authors declare no competing interests.

Data Availability

Datasets generated during this study are deposited at the Gene Expression Omnibus (GEO): GSE152882 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152882).

References

Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. Nature *45*, 710–715.

Arias-Cartin, R., Dobihal, G.S., Campos, M., Surovtsev, I.V., Parry, B., and Jacobs-Wagner, C. (2017). Replication fork passage drives asymmetric dynamics of a critical nucleoid-associated protein in Caulobacter. Embo J. *36*, 301–318.

Bai, H., Sun, M., Ghosh, P., Hatfull, G.F., Grindley, N.D.F., and Marko, J.F. (2011). Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. Proc. Natl. Acad. Sci. USA *108*, 7419–7424.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nat. Struct. Mol. Biol. *20*, 1199–1205.

Bermúdez, I., García-Martínez, J., Pérez-Ortín, J.E., and Roca, J. (2010). A method for genome-wide analysis of DNA helical tension by means of psoralen-DNA photobinding. Nucleic Acids Res. *38*, e182.

Bettin, N., Oss Pegorar, C., and Cusanelli, E. (2019). The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. Cells *8*.

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. Cell *55*, 637–643.

Corless, S., and Gilbert, N. (2017). Investigating DNA supercoiling in eukaryotic genomes. Brief Funct Genomics *16*, 379–389.

Culviner, P.H., and Laub, M.T. (2018). Global Analysis of the E. coli Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis. Mol. Cell *70*, 868–880.e10.

Culviner, P.H., Guegler, C.K., and Laub, M.T. (2020). A Simple, Cost-Effective, and Robust Method for rRNA Depletion in RNA-Sequencing Studies. MBio *11*.

Cutler, S., Lee, L.J., and Tsukiyama, T. (2018). Chromatin Remodeling Factors Isw2 and Ino80 Regulate Chromatin, Replication, and Copy Number of the Saccharomyces cerevisiae Ribosomal DNA Locus. Genetics *210*, 1543–1556.

Ding, Y., Manzo, C., Fulcrand, G., Leng, F., Dunlap, D., and Finzi, L. (2014). DNA supercoiling: a regulatory signal for the λ repressor. Proc. Natl. Acad. Sci. USA *111*, 15402–15407.

Díaz-Ingelmo, O., Martínez-García, B., Segura, J., Valdés, A., and Roca, J. (2015). DNA Topology and Global Architecture of Point Centromeres. Cell Rep *13*, 667–677.

Drlica, K., Franco, R.J., and Steck, T.R. (1988). Rifampin and rpoB mutations can alter DNA supercoiling in Escherichia coli. J. Bacteriol. *170*, 4983–4985.

Ferrándiz, M.-J., Martín-Galiano, A.J., Arnanz, C., Camacho-Soguero, I., Tirado-Vélez, J.-M., and la Campa, de, A.G. (2016). An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene. Nucleic Acids Res. *44*, 7292–7303.

Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L.A. (2017). Emerging Evidence of Chromosome Folding by Loop Extrusion. Cold Spring Harb. Symp. Quant. Biol. *82*, 45–55.

García-Muse, T., and Aguilera, A. (2016). Transcription–replication conflicts: how they occur and how they are resolved. Nat. Rev. Mol. Cell Biol. *17*, 553–563.

Gilbert, N., and Allan, J. (2014). Supercoiling in DNA and chromatin. Curr. Opin. Genet. Dev. *25*, 15–21.

Giuntoli, R.D., Linzer, N.B., Banigan, E.J., Sing, C.E., la Cruz, de, M.O., Graham, J.S., Johnson, R.C., and Marko, J.F. (2015). DNA-Segment-Facilitated Dissociation of Fis and NHP6A from DNA Detected via Single-Molecule Mechanical Response. J. Mol. Biol. *427*, 3123–3136.

Glynn, E.F., Megee, P.C., Yu, H.-G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). Genome-Wide Mapping of the Cohesin Complex in the Yeast Saccharomyces cerevisiae. PLoS Biol. *2*, e259–15.

Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira, M.T., and Luke, B. (2017). Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. Cell *170*, 72–85.e14.

Guo, M.S., Haakonsen, D.L., Zeng, W., Schumacher, M.A., and Laub, M.T. (2018). A Bacterial Chromosome Structuring Protein Binds Overtwisted DNA to Stimulate Type II Topoisomerases and Enable DNA Replication. Cell *175*, 583–597.e23.

Haran, T.E., and Mohanty, U. (2009). The unique structure of A-tracts and intrinsic DNA bending. Q. Rev. Biophys. *42*, 41–81.

Huang, Q., Duan, B., Dong, X., Fan, S., and Xia, B. (2020). GapR binds DNA through dynamic opening of its tetrameric interface. Nucleic Acids Res.

Kempfer, R., and Pombo, A. (2020). Methods for mapping 3D chromosome architecture. Nat. Rev. Genet. *21*, 207–226.

Khodursky, A.B., Peter, B.J., Schmid, M.B., DeRisi, J., Botstein, D., Brown, P.O., and Cozzarelli, N.R. (2000). Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. Proc. Natl. Acad. Sci. USA *97*, 9419–9424.

Kobayashi, T., and Horiuchi, T. (1996). A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. Genes Cells *1*, 465–474.

Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. Mol. Cell. Biol. *23*, 9178–9188.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat. Struct. Mol. Biol. *20*, 396–403.

Lal, A., Dhar, A., Trostel, A., Kouzine, F., Seshasayee, A.S.N., and Adhya, S. (2016). Genome scale patterns of supercoiling in a bacterial chromosome. Nat Commun *7*, 11055.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lawrimore, C.J., and Bloom, K. (2019). Common Features of the Pericentromere and Nucleolus. Genes (Basel) *10*.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573–578.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA *84*, 7024–7027.

Louis, E.J. (1995). The chromosome ends of Saccharomyces cerevisiae. Yeast *11*, 1553–1573.

Lourenço, R.F., Saurabh, S., Herrmann, J., Wakatsuki, S., and Shapiro, L. (2020). The Nucleoid-Associated Protein GapR Uses Conserved Structural Elements To Oligomerize and Bind DNA. MBio *11*.

Luke, B., and Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. Embo J. *28*, 2503–2510.

Menzel, R., and Gellert, M. (1987). Modulation of transcription by DNA supercoiling: a deletion analysis of the Escherichia coli gyrA and gyrB promoters. Proc. Natl. Acad. Sci. USA *84*, 4185–4189.

Moronta-Gines, M., van Staveren, T.R.H., and Wendt, K.S. (2019). One ring to bind them - Cohesin's interaction with chromatin fibers. Essays Biochem. *63*, 167–176.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat. Struct. Mol. Biol. *20*, 387–395.

Nelson, P. (1999). Transport of torsional stress in DNA. Proc. Natl. Acad. Sci. USA *96*, 14342–14347.

Neurohr, G.E., Terry, R.L., Lengefeld, J., Bonney, M., Brittingham, G.P., Moretto, F., Miettinen, T.P., Vaites, L.P., Soares, L.M., Paulo, J.A., et al. (2019). Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. Cell *176*, 1083–1097.e18.

Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat. Rev. Mol. Cell Biol. *21*, 167–178.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Paldi, F., Alver, B., Robertson, D., Schalbetter, S.A., Kerr, A., Kelly, D.A., Baxter, J., Neale, M.J., and Marston, A.L. (2020). Convergent genes shape budding yeast pericentromeres. Nature 1–28.

Pan, X., Chen, Y., Biju, B., Ahmed, N., Kong, J., Goldenberg, M., Huang, J., Mohan, N., Klosek, S., Parsa, K., et al. (2019). FANCM suppresses DNA replication stress at ALT telomeres by disrupting TERRA R-loops. Scientific Reports *9*, 19110–19114.

Patrick Higgins, N. (2017). Measuring In Vivo Supercoil Dynamics and Transcription Elongation Rates in Bacterial Chromosomes. Methods Mol. Biol. *1624*, 17–27.

Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O., and Cozzarelli, N.R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. Genome Biol. *5*, R87.

Petti, E., Buemi, V., Zappone, A., Schillaci, O., Broccia, P.V., Dinami, R., Matteoni, S., Benetti, R., and Schoeftner, S. (2019). SFPQ and NONO suppress RNA:DNA-hybrid-related telomere instability. Nat Commun *10*, 1001–1014.

Pfeiffer, V., and Lingner, J. (2012). TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. PLoS Genet. *8*, e1002747.

Pommier, Y., Sun, Y., Huang, S.-Y.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nat. Rev. Mol. Cell Biol. *17*, 703–721.

Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D., and Cozzarelli, N.R. (2001). Topological challenges to DNA replication: conformations at the fork. Proc. Natl. Acad. Sci. USA *98*, 8219–8226.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–320.e324.

Ricci, D.P., Melfi, M.D., Lasker, K., Dill, D.L., McAdams, H.H., and Shapiro, L. (2016). Cell cycle progression in Caulobacter requires a nucleoid-associated protein with high AT sequence recognition. Proc. Natl. Acad. Sci. USA *113*, E5952–E5961.

Sinden, R.R., Carlson, J.O., and Pettijohn, D.E. (1980). Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells: analogous measurements in insect and human cells. Cell *21*, 773–783.

Skoko, D., Wong, B., Johnson, R.C., and Marko, J.F. (2004). Micromechanical analysis of the binding of DNA-bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein complexes. Biochemistry *43*, 13867–13874.

Steiner, F.A., and Henikoff, S. (2015). Diversity in the organization of centromeric chromatin. Curr. Opin. Genet. Dev. *31*, 28–35.

Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019). Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc. Natl. Acad. Sci. USA *116*, 6260–6269.

Sun, M., Nishino, T., and Marko, J.F. (2013). The SMC1-SMC3 cohesin heterodimer structures DNA through supercoiling-dependent loop formation. Nucleic Acids Res. *41*, 6149–6160.

Tarry, M.J., Harmel, C., Taylor, J.A., Marczynski, G.T., and Schmeing, T.M. (2019). Structures of GapR reveal a central channel which could accommodate B-DNA. Scientific Reports *9*, 16679–11.

Teves, S.S., and Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. Nat. Struct. Mol. Biol. *21*, 88–94.

Toussaint, M., Levasseur, G., Tremblay, M., Paquette, M., and Conconi, A. (2005). Psoralen photocrosslinking, a tool to study the chromatin structure of RNA polymerase I--transcribed ribosomal genes. Biochem. Cell Biol. *83*, 449–459.

Vlijm, R., Kim, S.H., De Zwart, P.L., Dalal, Y., and Dekker, C. (2017). The supercoiling state of DNA determines the handedness of both H3 and CENP-A nucleosomes. Nanoscale *9*, 1862–1870.

Vos, S.M., Tretter, E.M., Schmidt, B.H., and Berger, J.M. (2011). All tangled up: how cells direct, manage and exploit topoisomerase function. Nat. Rev. Mol. Cell Biol. *12*, 827–841.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. *30*, 1327–1338.

Wellinger, R.E., and Sogo, J.M. (1998). In vivo mapping of nucleosomes using psoralen-DNA crosslinking and primer extension. Nucleic Acids Res. *26*, 1544–1545.

Wu, H.Y., Shyy, S.H., Wang, J.C., and Liu, L.F. (1988). Transcription generates positively and negatively supercoiled domains in the template. Cell *53*, 433–440.

Zhong, J., Luo, K., Winter, P.S., Crawford, G.E., Iversen, E.S., and Hartemink, A.J. (2016). Mapping nucleosome positions using DNase-seq. Genome Res. *26*, 351–364.

Zorman, S., Seitz, H., Sclavi, B., and Strick, T.R. (2012). Topological characterization of the DnaA–oriC complex using single-molecule nanomanipuation. Nucleic Acids Res. *40*, 7375–7383.

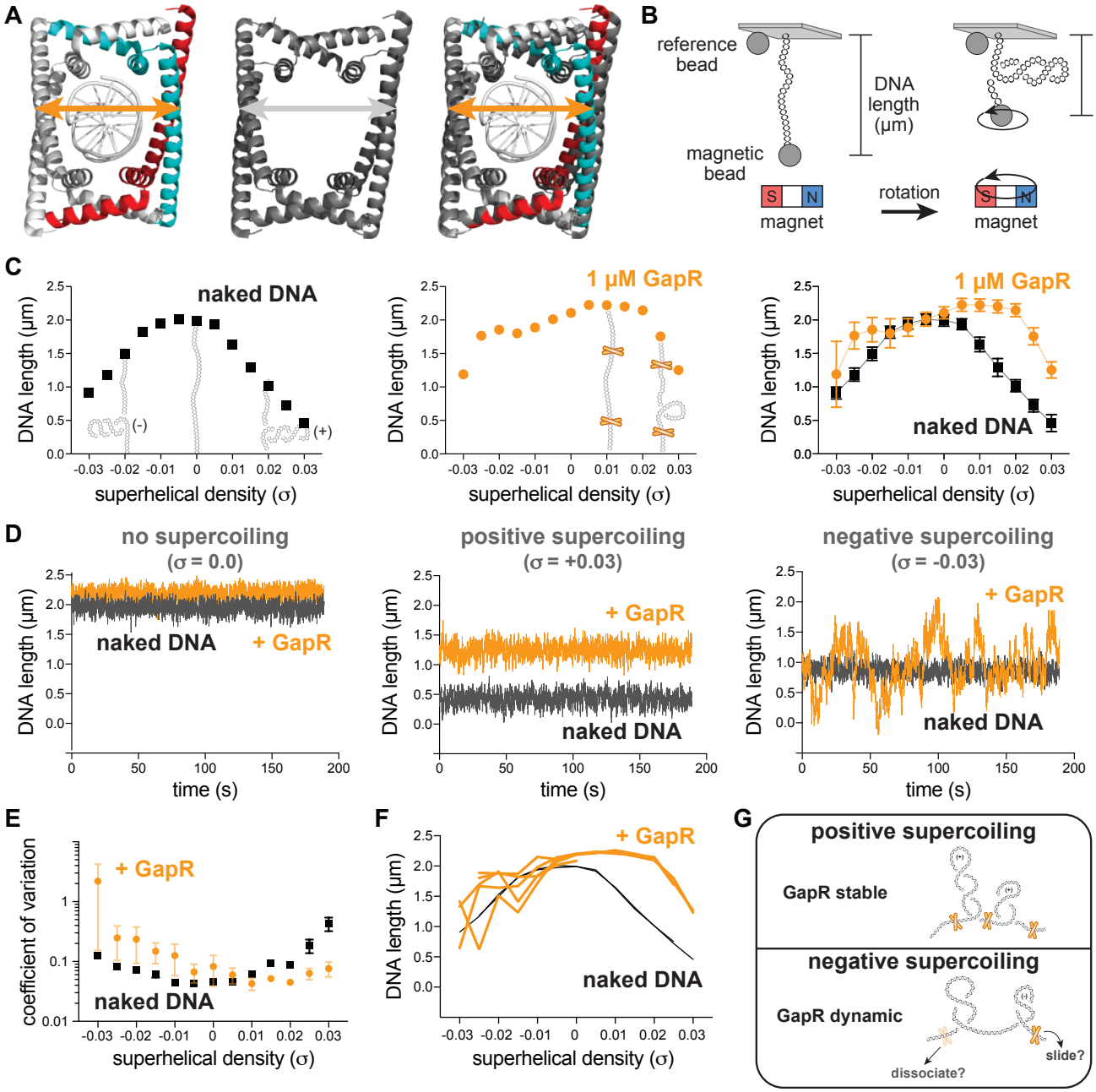

Figure 1. GapR interacts stably with overtwisted, positively supercoiled DNA

(A) Comparison of GapR-DNA crystal structures. Left, 6GC8 (Guo et al., 2018); middle, 6OZX (Tarry et al., 2019); right, overlay. Diameter of 6GC8 (orange arrow) and 6OZX (grey arrow) indicated.

(B) Schematic of magnetic tweezer (MT) experiment. See also Figure S1A.

(C) Behavior of naked DNA (left), DNA incubated with 1 μM GapR (middle), and overlay (right) in a rotation-extension experiment with the corresponding DNA conformation superimposed. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment.

(D) DNA ± 1 μM GapR behavior over time from (C) under no supercoiling (σ = 0.0, left), positive supercoiling (σ = +0.03, middle), and negative supercoiling (σ = -0.03, right).

(E) Coefficient of variation of force-extension experiments of DNA ± 1 μM GapR. Data indicate mean ± SEM, n ≥ 3.

(F) Hysteresis of force-extension experiments. Traces indicate multiple rotation-extension measurements from one DNA molecule ± 1 μM GapR.

(G) Model of GapR binding to positively or negatively supercoiled DNA.

See also Figure S1 and Table S1.

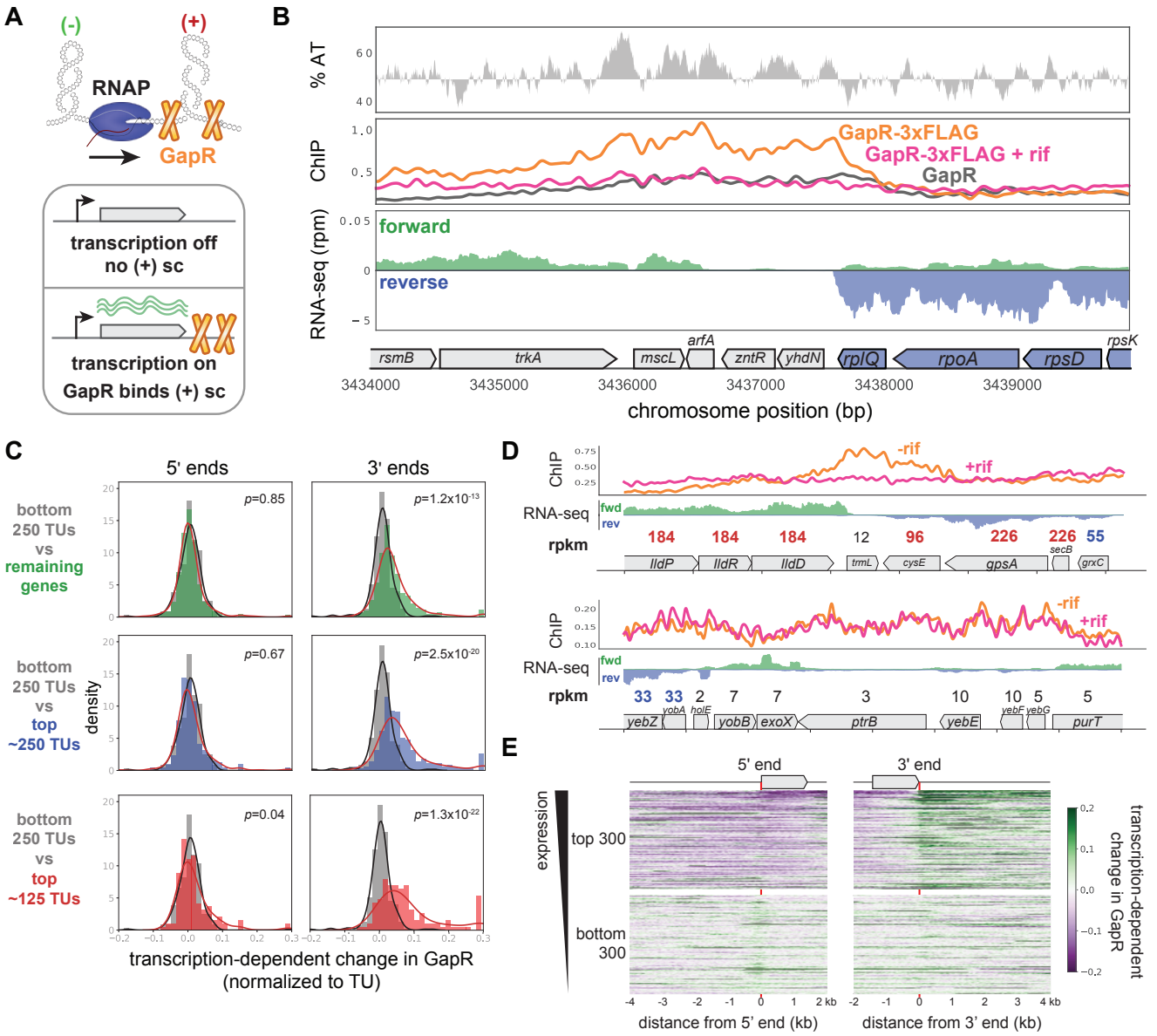

Figure 2. GapR recognizes positively supercoiled DNA in *E. coli*

(A) Positive supercoiling is generated downstream of transcription (top) and is associated with active transcription (bottom) as predicted by the "twin-domain" model.

(B) GapR ChIP profiles at a highly expressed operon. AT content (top), with AT content below the genomic average (50%) plotted in reverse. ChIP-seq (middle) of untreated (orange) or rifampicin-treated (pink) GapR-3xFLAG cells and untreated GapR cells (grey). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(C) Transcription-dependent change in GapR ChIP at 5' (left) or 3' (right) ends normalized by binding within the transcription unit (TU) at different expression thresholds. Student's t-test p-value shown.

(D) Examples of GapR-3xFLAG ChIP without (orange) or with (pink) rifampicin treatment. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated. Expression values are colored using the same rpkm cutoffs as in (**C**).

(E) Heatmap showing transcription-dependent change in GapR around 5' and 3' ends for the top and bottom 300 long TUs sorted by expression.

See also Figure S2 and Figure S3.

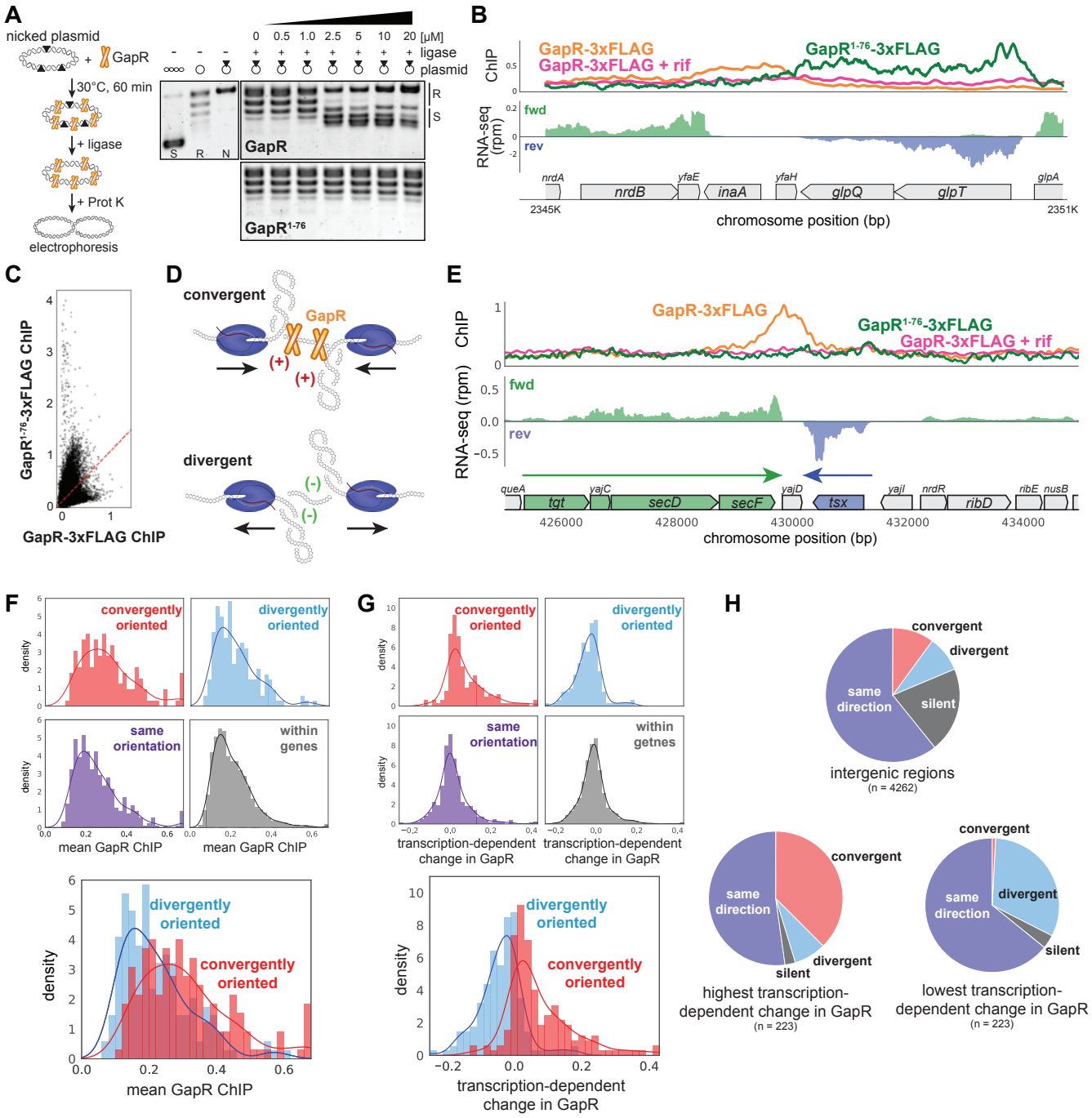

Figure 3. GapR recognizes DNA supercoiling and accumulates in regions of convergent transcription

(A) GapR$^{1\text{-}76}$ does not recognize DNA topology. Full-length GapR and GapR$^{1\text{-}76}$ were incubated with nicked plasmid before treatment with T4 DNA ligase and subsequent quenching, deproteinization, and electrophoresis (schematic). Gel analysis of plasmid topology with supercoiled and relaxed standards.

(B) GapR$^{1\text{-}76}$-3xFLAG ChIP (green) compared to GapR-3xFLAG ChIP in the absence (orange) and presence of rifampicin (pink). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated.

(C) Correlation between GapR-3xFLAG and GapR$^{1\text{-}76}$-3xFLAG ChIP experiments.

(D) Positive supercoils are trapped by convergent transcription.

(E) ChIP of GapR$^{1\text{-}76}$-3xFLAG (green) and GapR-3xFLAG without (orange) and with (pink) rifampicin treatment at convergently oriented TUs. Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated.

(F) GapR ChIP in gene bodies (dark grey), in divergent regions (blue), convergent regions (red), and where transcription is in the same orientation (purple). Overlay of divergent and convergent regions (bottom).

(G) Histogram of transcription-dependent changes in GapR plotted as in (**F**).

(H) Regions with high transcription-dependent change in GapR are more frequently between convergent genes. Pie charts summarize the orientation of flanking genes for all intergenic regions (top) and intergenic regions with highest (bottom left) or lowest (bottom right) transcription-dependent change in GapR.

See also Figure S3.

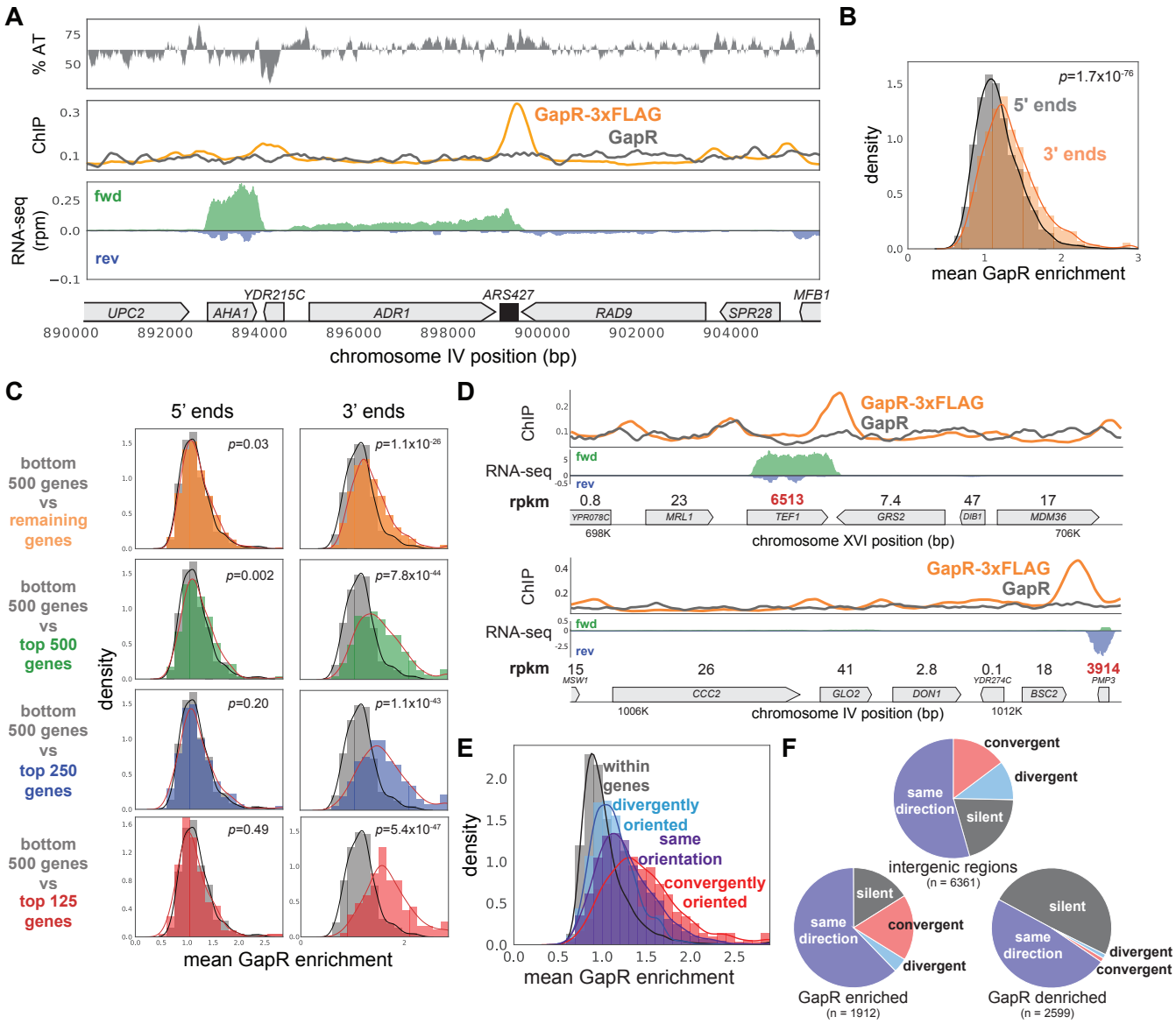

Figure 4. GapR is associated with positive supercoiling in yeast

(A) ChIP of *S. cerevisiae* grown in raffinose before GapR induction. AT content (top), ChIP-seq (middle) of GapR-3xFLAG (orange) or untagged GapR (grey) expressing cells (middle). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(B) Mean GapR enrichment (GapR-3xFLAG ChIP normalized by untagged ChIP) in a 500 bp window at the 5' and 3' end of long genes. Student's t-test p-value is shown.

(C) Mean GapR enrichment at 5' and 3' ends of long genes at various transcriptional cutoffs. Student's t-test p-value is shown.

(D) Examples of GapR-3xFLAG (orange) and untagged GapR (grey) ChIP. Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated. Expression values are colored using the same rpkm cutoffs as in (C).

(E) GapR is enriched between convergently oriented genes.

(F) GapR-bound regions are more frequently between convergent genes. Pie charts shown as in Fig. 3H.

See also Figure S4.

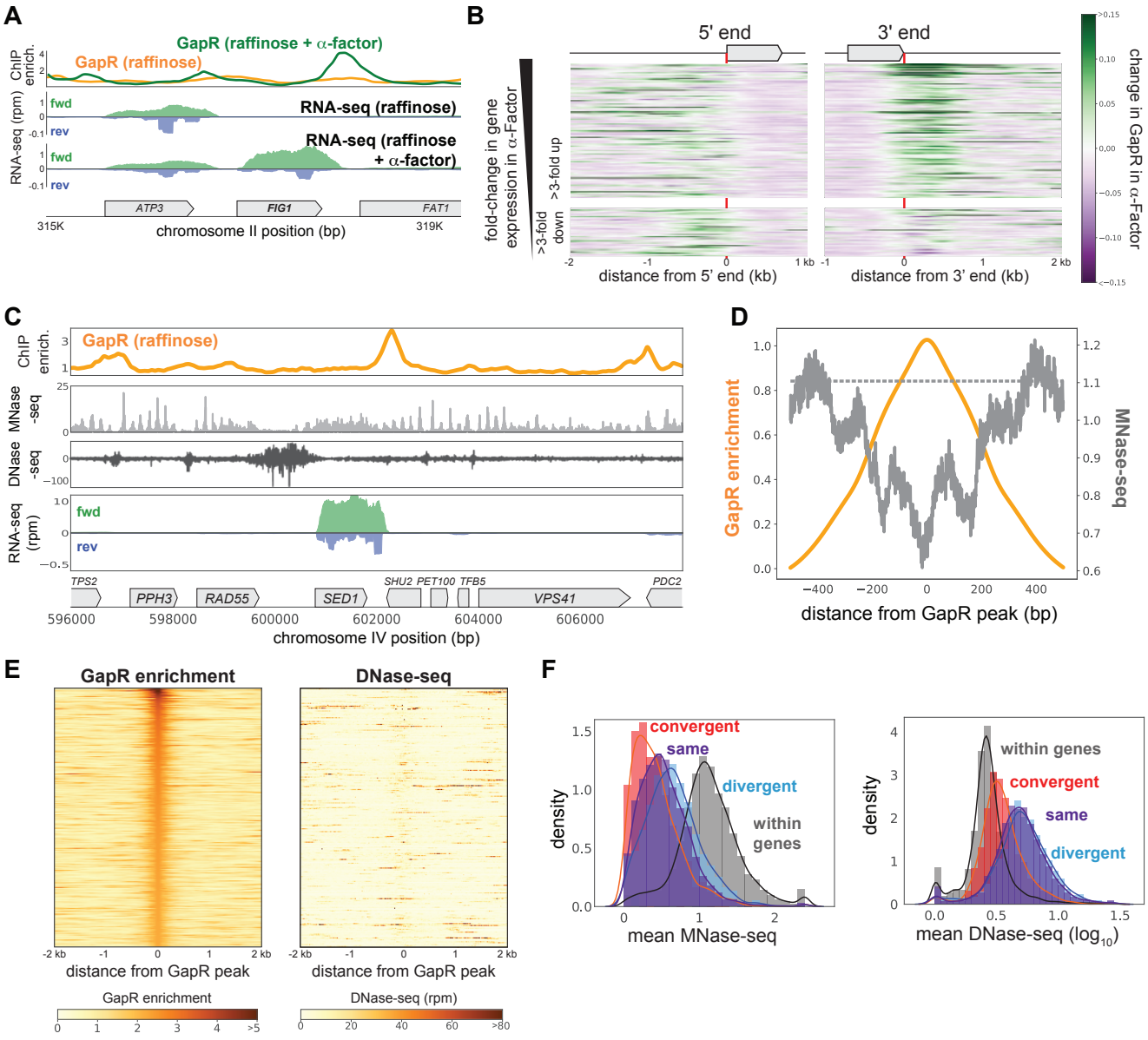

Figure 5. GapR binding in *S. cerevisiae* is responsive to transcription and is not restricted to open chromatin

(A) GapR enrichment at *FIG1* in raffinose without (orange) or with (green) α-factor arrest before GapR induction (top). Transcription of the forward (green) and reverse (blue) strands in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(B) Heatmap showing α-factor dependent change in GapR enrichment at the 5' and 3' ends of long TUs sorted by transcriptional change in α-factor.

(C) GapR enrichment (orange) compared to nucleosome occupancy (MNase-seq, light grey) and chromatin accessibility (DNase-seq, dark grey). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated.

(D) GapR ChIP peaks are de-enriched for nucleosomes. GapR enrichment (orange, left y-axis), MNase-seq reads (dark grey, right y-axis), mean genomic MNase-seq occupancy (dashed grey line).

(E) Heatmap of GapR enrichment (left) and DNase accessibility (right) of the 500 most GapR-enriched loci.

(F) Association between transcriptional orientation and MNase-seq and DNase-seq.

See also Figure S5.

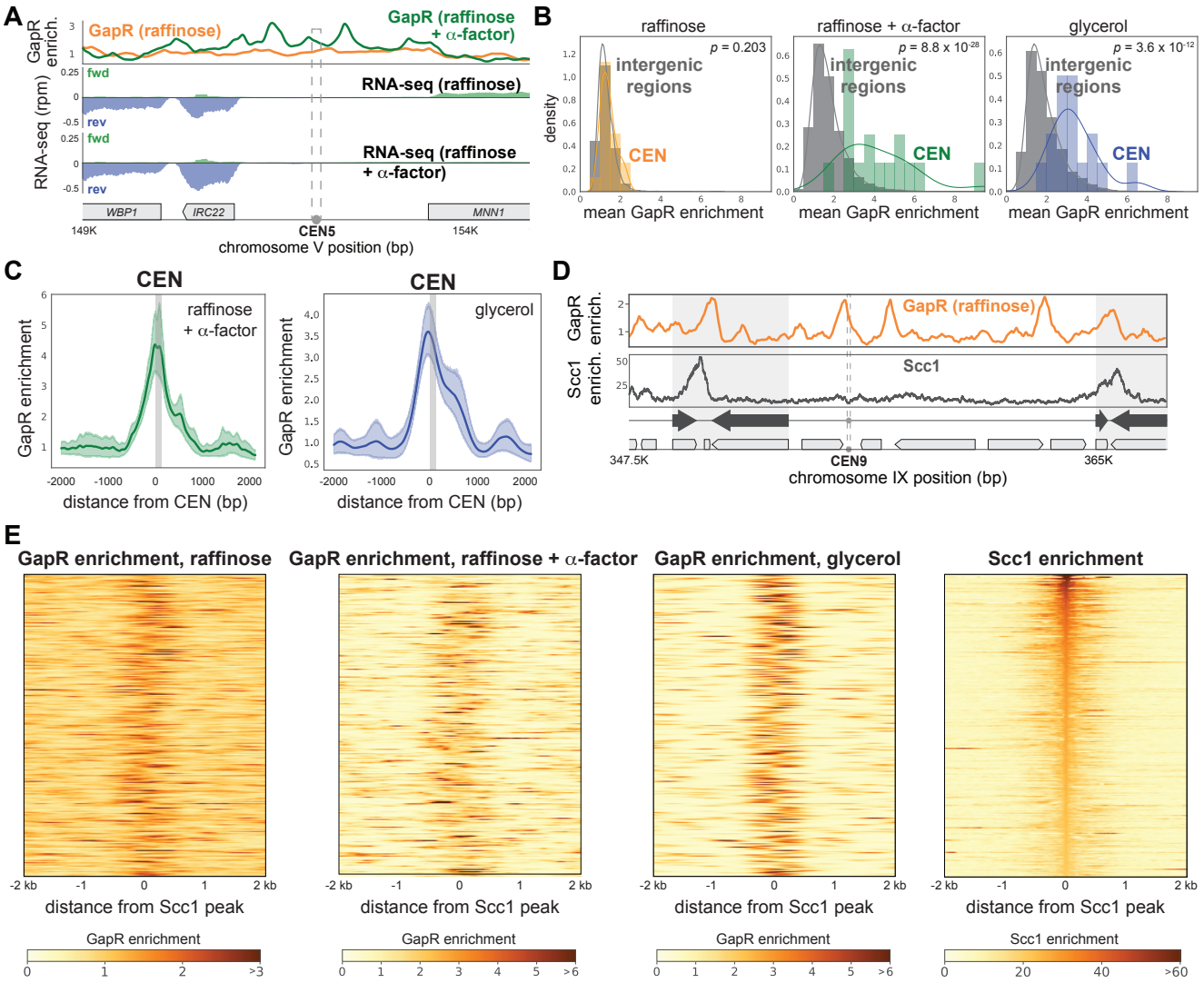

Figure 6. Positively supercoiled DNA is associated with centromeres and cohesin

(A) GapR enrichment at CEN5 in cells without (orange) or with (green) α-factor arrest (top). Transcription of the forward (green) and reverse (blue) strands in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(B) GapR enrichment at centromeres in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(C) GapR enrichment over all centromeres after α-factor arrest (green) or grown in glycerol (blue). Mean enrichment (solid line) with 95% confidence intervals (shaded area). Grey bar represents position of centromeres.

(D) GapR (raffinose, top) and cohesin (Scc1 enrichment, bottom) are associated with convergent genes (arrows) at pericentromere boundaries (shaded areas).

(E) Heatmaps of GapR (3 left panels) and Scc1 (right) enrichment at the 500 most Scc1-enriched loci.

See also Figure S6.

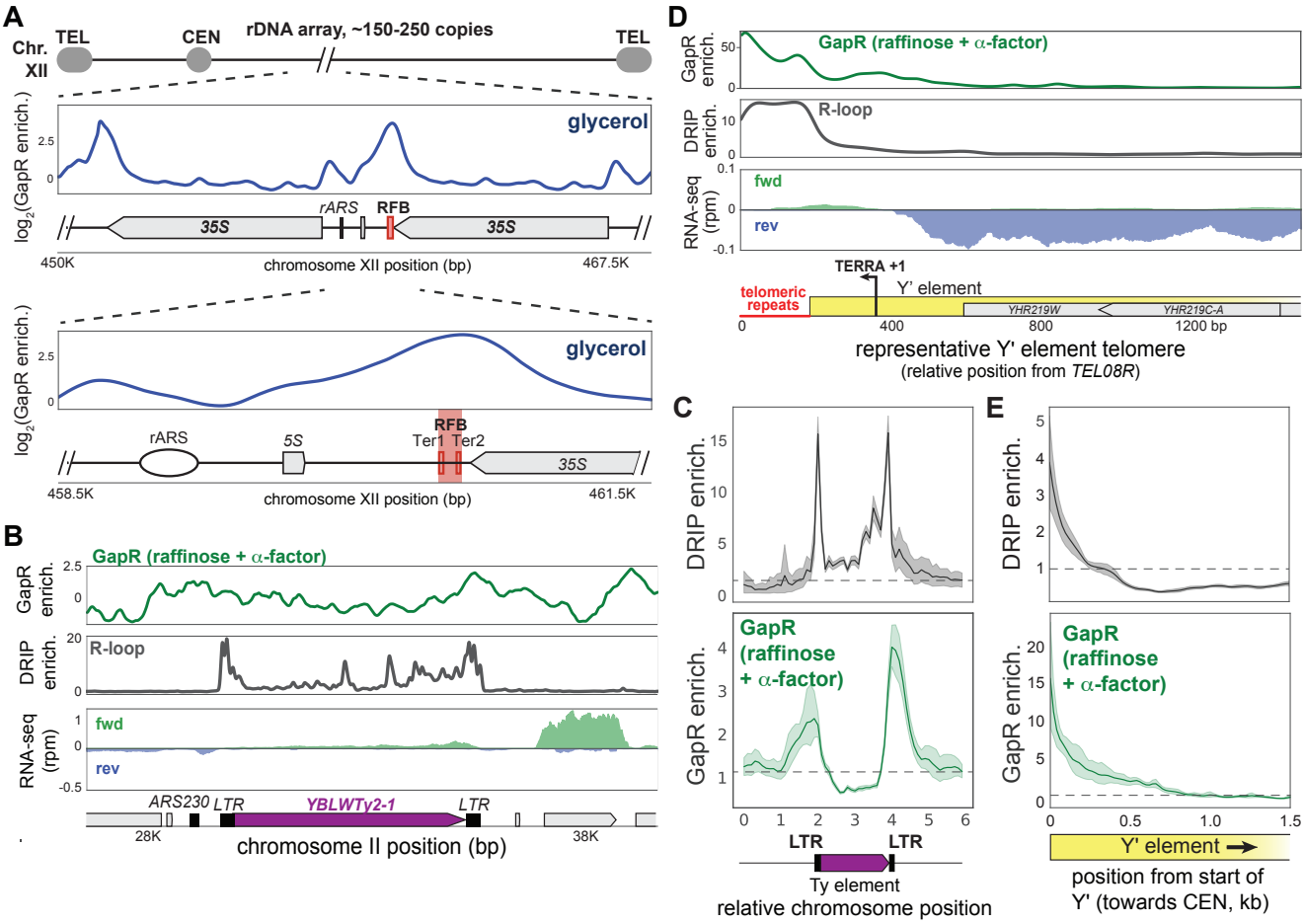

Figure 7. Positive supercoiling is associated with ARS and R-loops

(A) GapR enrichment (glycerol) at the rDNA shown in two successive zoom-ins (top and bottom panel) with the replication fork barrier (RFB) and termination sequences (Ter1/Ter2) indicated.

(B) GapR enrichment (α-factor) at a Ty element (top). DNA-RNA hybrid formation by DRIP-seq (middle). Transcription of the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(C) Alignment of GapR and DRIP enrichment surrounding all yeast Ty elements. Data indicate mean (solid line) with 95% confidence intervals (shaded area), no enrichment (dotted line).

(D) GapR enrichment at a telomere with a Y' element (top). DRIP enrichment (middle). Transcription of the forward (green) and reverse (blue) strands with the organization of the Y' element indicated (bottom). Position given is from end of *TEL08R*.

(E) Alignment of GapR and DRIP enrichment surrounding all yeast telomeres with Y' elements as in (**C**). Telomeric repeats are removed from analysis.

See also Figure S7.

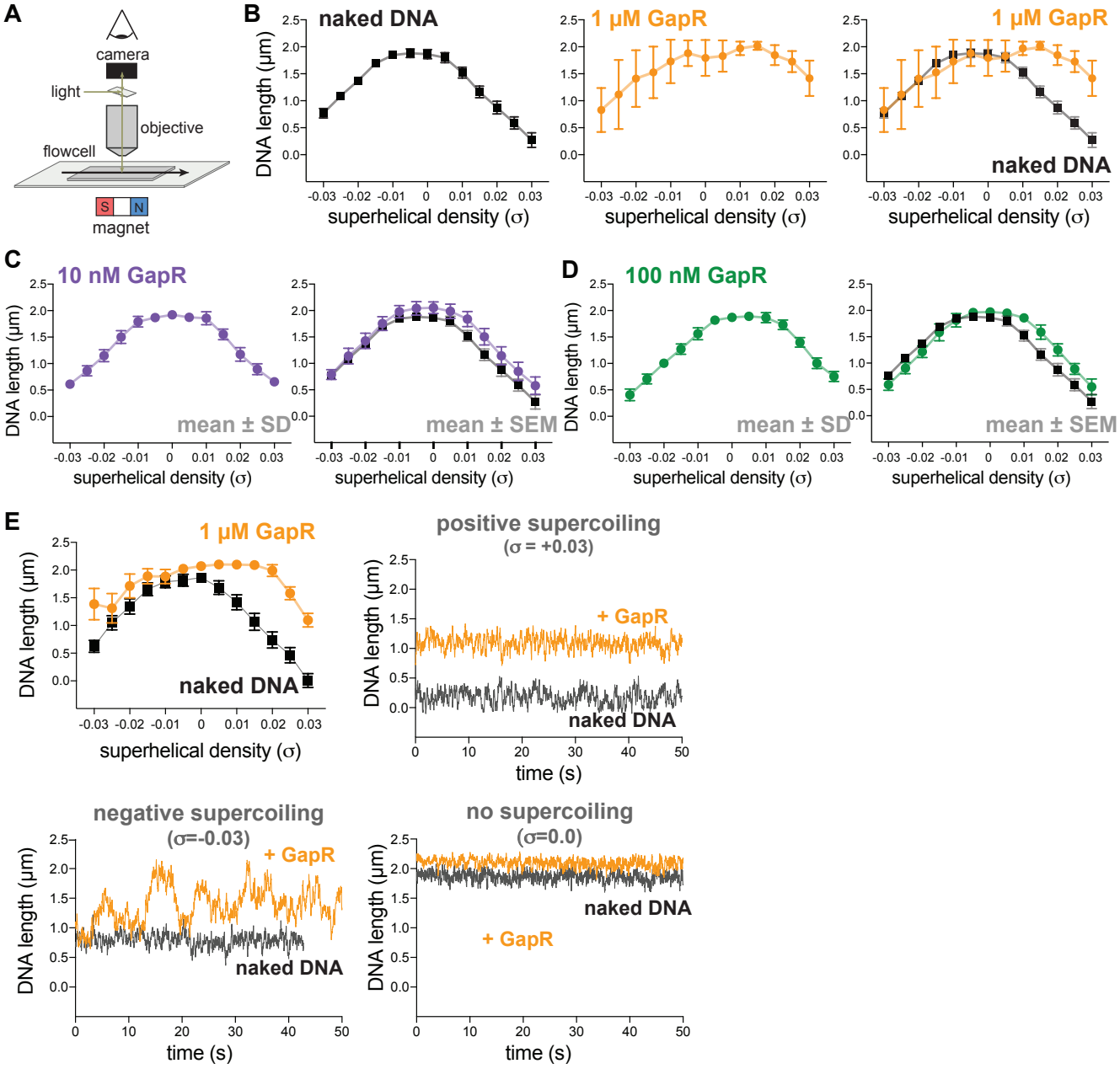

Figure S1. GapR binding to supercoiled DNA in a magnetic tweezer experiment, related to Figure 1.

(A) Schematic of the microscope and flowcell setup in a magnetic tweezer (MT) experiment.

(B) Behavior of naked DNA (left), DNA incubated with 1 μM GapR (middle), and overlay (right) in a typical rotation-extension experiment. Data indicate mean ± SEM, for 3 MT experiments.

(C) Behavior of DNA incubated with 10 nM GapR in a typical rotation-extension experiment. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment (left) and mean ± SEM compared to naked DNA (black), for 3 MT experiments (right).

(D) Behavior of DNA incubated with 100 nM GapR in a typical rotation-extension experiment. Data indicate mean ± SD, n = 200 at each σ, in a single MT experiment (left) and mean ± SEM compared to naked DNA (black), for 3 MT experiments (right).

(E) Rotation-extension curve (top left; data indicate mean ± SD, n = 200 at each σ) and time course measurements for a single GapR binding experiment. Time courses indicate behavior of naked DNA (black) and DNA bound by 1 μM GapR (orange) under positive supercoiling (σ = +0.03, top right), negative supercoiling (σ = -0.03, bottom left), and no supercoiling (σ = 0.0, bottom right).

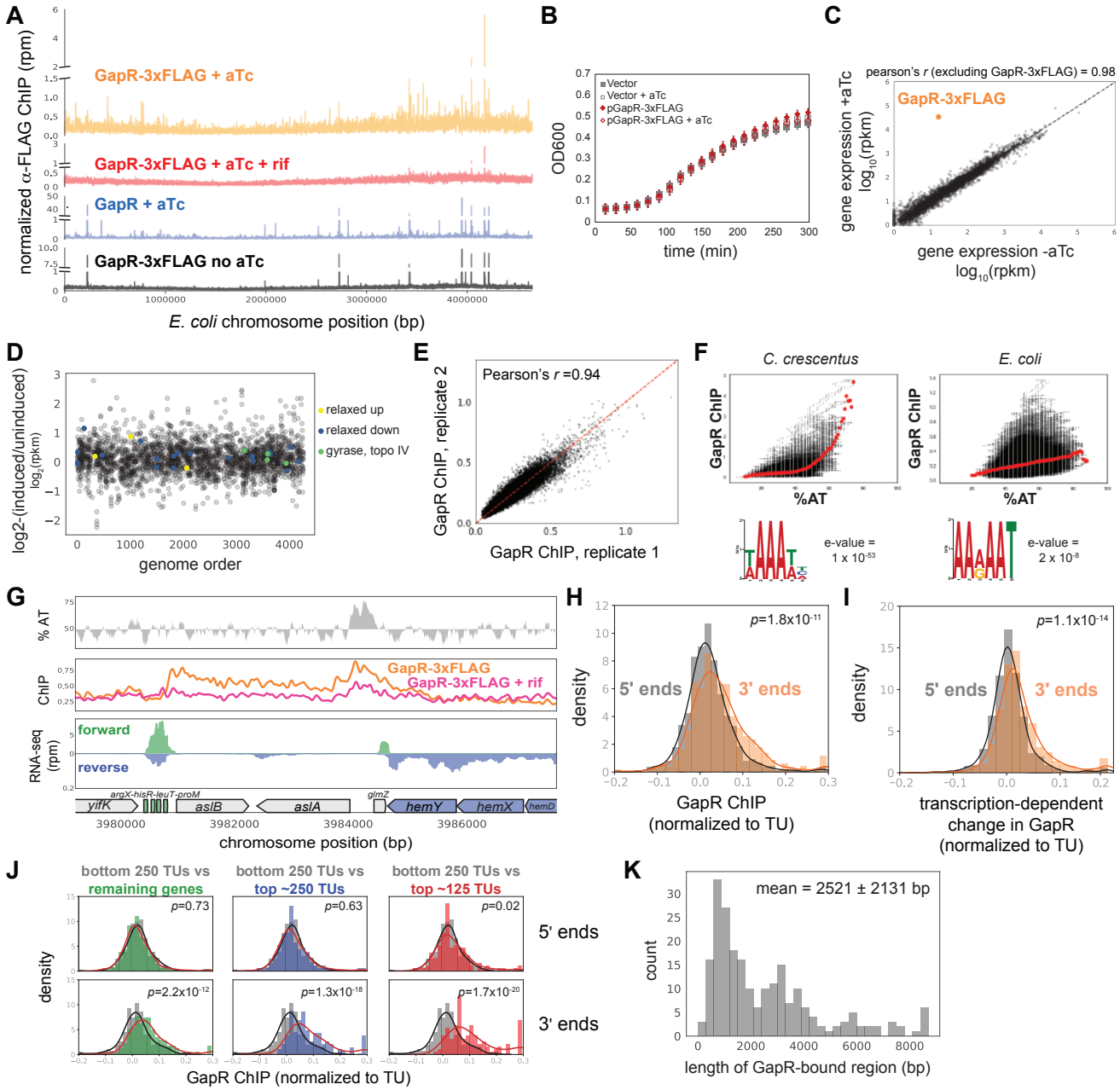

Figure S2. *E. coli* GapR ChIP-seq, related to Figure 2.

(A) ChIP-seq of GapR-3xFLAG with (orange) and without induction with anhydrous tetracycline (aTc, grey), after rifampicin treatment (rif, pink), and untagged GapR with induction (blue).

(B) Growth of cells expressing GapR-3xFLAG plasmid or empty vector. Data indicate mean ± SEM, n = 6.

(C) Transcriptional profiles of cells ± GapR-3xFLAG expression.

(D) Supercoiling-sensitive genes are unaffected by GapR-3xFLAG expression. Log_2-fold change in expression (rpkm) vs genome order (excluding GapR and poorly expressed genes). Supercoiling-sensitive genes (yellow/blue) and topoisomerase genes (light green). 15 genes that change > 4-fold represent mostly proteins of unknown function or membrane proteins.

(E) Correlation between two independent GapR-3xFLAG ChIP experiments.

(F) GapR-3xFLAG ChIP versus AT content in *C. crescentus* (left) and *E. coli* (right). Mean ChIP at a given % AT (red dots). Motif from DREME (below)

(G) GapR-3xFLAG ChIP profiles. AT content (top), with AT content below the genomic average (50%) plotted in reverse. Normalized ChIP-seq (middle) of GapR-3xFLAG cells that were either untreated (orange) or rifampicin-treated (pink), and untagged GapR expressing cells that were untreated (grey). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(H) GapR accumulates at the 3' end of transcription units (TUs). GapR ChIP at 5' (grey) or 3' (orange) ends of TUs normalized by binding within the TU. Student's t-test p-value is reported.

(I) GapR binding at 3' ends depends on transcription. Transcription-dependent change in GapR ChIP at the 5' (grey) or 3' ends (orange) of TUs normalized by binding within the TU. Student's t-test p-value is reported.

(J) GapR binding at 5' (top) or 3' (bottom) ends of long TUs normalized by binding within the TU at different expression levels. Student's t-test p-value is reported.

(K) Histogram showing the length of transcription-dependent GapR binding events.

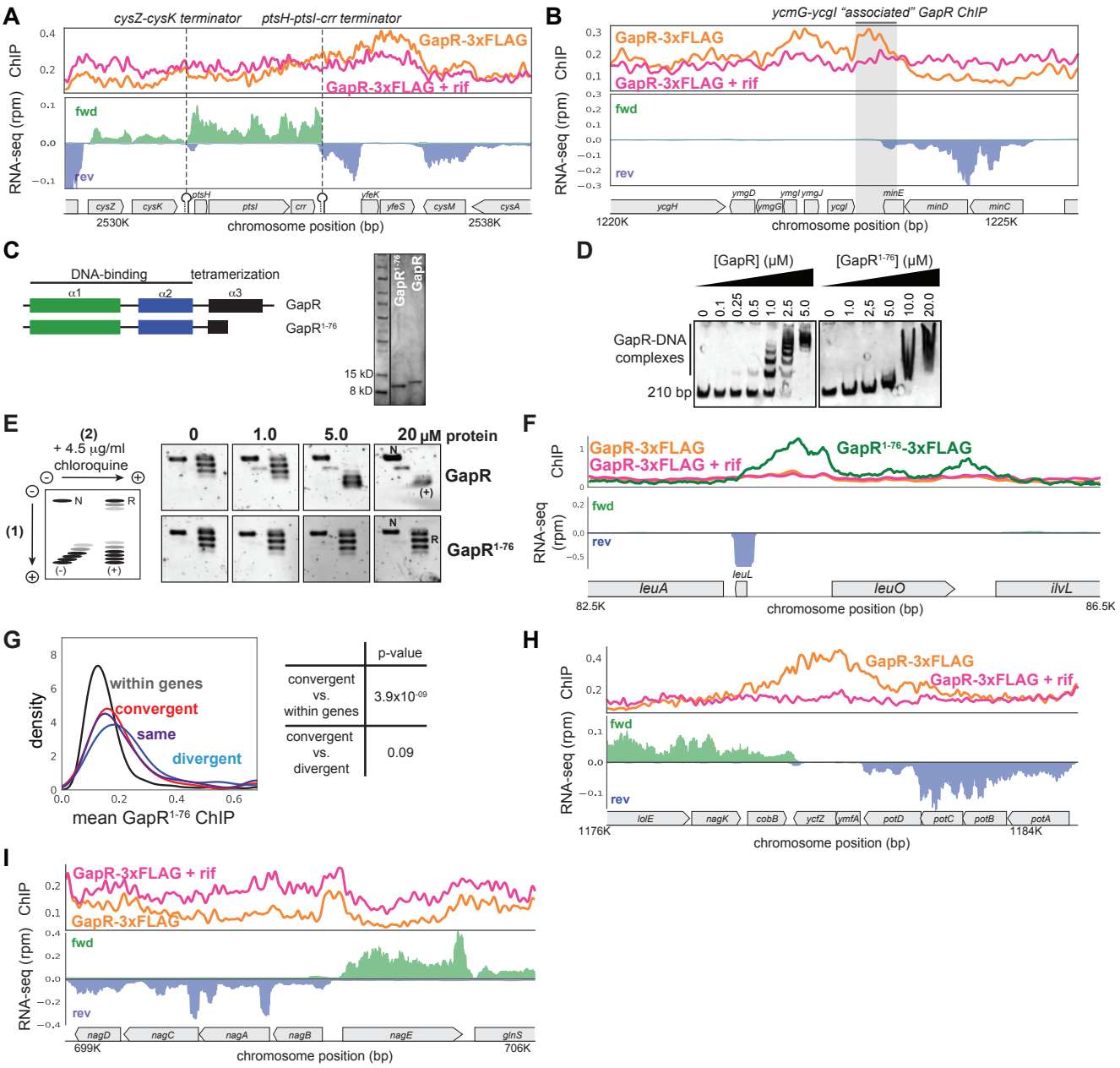

Figure S3. GapR full-length and truncation variant ChIP-seq, related to Figures 2 and 3.

(A) GapR ChIP-seq profiles in the absence (orange) and presence (pink) of the transcriptional inhibitor rifampicin (top). Transcription on the forward (green) and reverse (blue) strands with annotated genes indicated (bottom). Terminators on the top strand are indicated.

(B) GapR profiles as in (**A**). Shaded region shows GapR ChIP generated by the *minCDE* transcript that is also at the 3' end of the *ycmJ-ycgI* locus.

(C) Schematic and purification gel of full-length GapR and tetramerization-deficient GapR$^{1-76}$.

(D) Electrophoretic mobility shift assays of full-length and GapR$^{1-76}$ binding to 210 bp DNA.

(E) Analysis of ligation products from Fig. 3A with 2D-chloroquine electrophoresis. Migration of different plasmid forms are diagrammed (left): N, nicked; R, relaxed; L, linear; (-), negatively supercoiled; (+), positively supercoiled.

(F) GapR profiles as in (**A**) with GapR$^{1-76}$ ChIP-seq (green).

(G) Mean GapR$^{1-76}$ ChIP in the middle of gene bodies (black), divergently transcribed regions (blue), convergently transcribed regions (red), and where transcription is in the same orientation (purple).

(H) GapR profiles as in (**A**) showing GapR accumulation at a representative region of convergent transcription.

(I) GapR profiles as in (**A**) showing GapR de-enrichment at a representative region of divergent transcription.

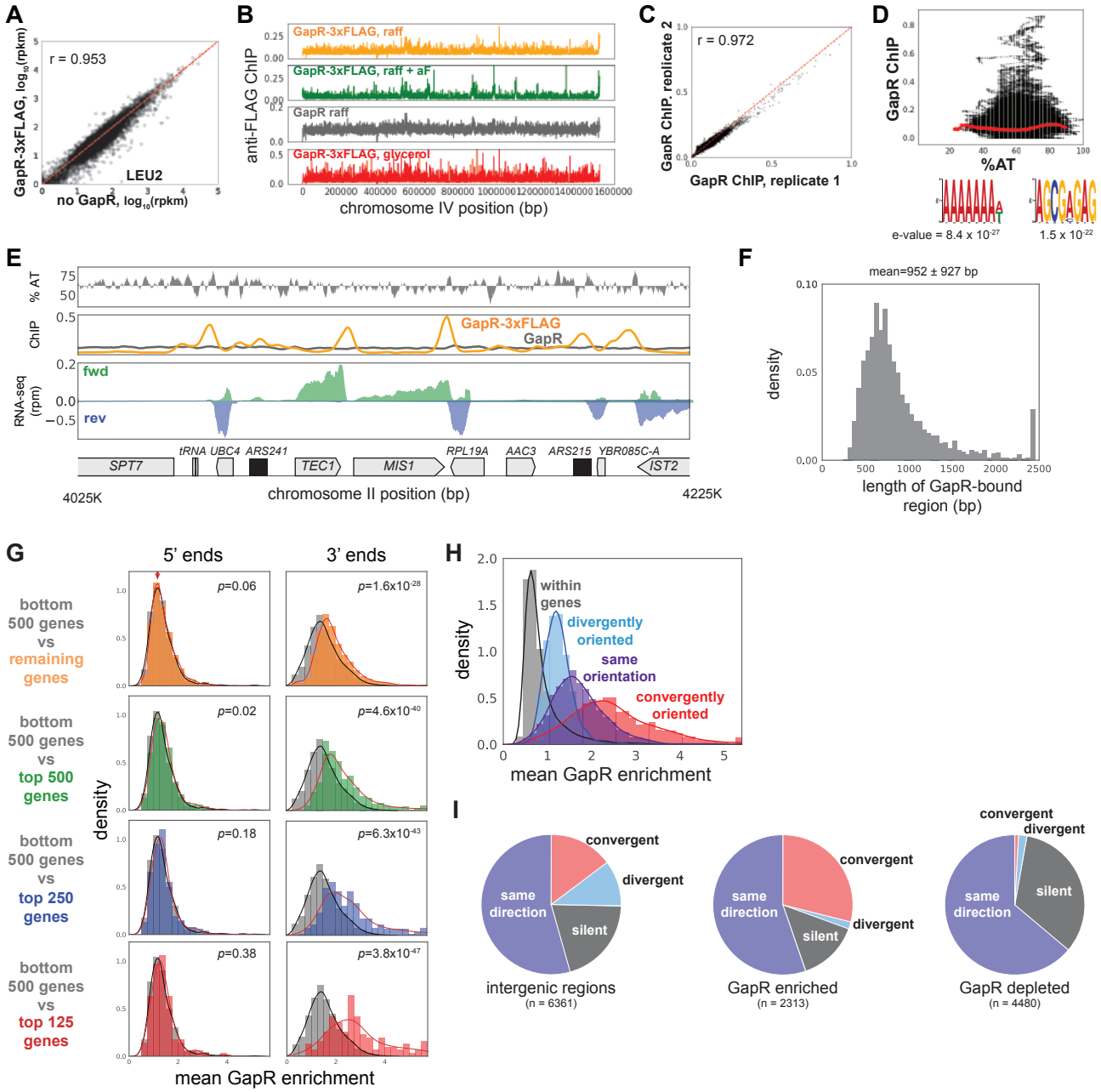

Figure S4. *S. cerevisiae* GapR ChIP-seq, related to Figure 4.

(A) Transcriptional profiles of cells with and without GapR-3xFLAG integrated at LEU2. The largest change in expression pattern is at the LEU2 locus.

(B) ChIP-seq profiles for GapR-3xFLAG (orange) and untagged GapR (grey) cells grown in glycerol. ChIP-seq profiles for GapR-3xFLAG cells grown in raffinose and in raffinose + α-factor. Cells were grown to OD 0.3-0.5 before addition of 2% galactose for 6 hr followed by fixation and ChIP.

(C) Correlation between two independent GapR-3xFLAG ChIP-seq experiments in raffinose.

(D) GapR-3xFLAG ChIP-seq in raffinose versus AT content. Mean enrichment at a given % AT (red dots). Motifs from DREME (below).

(E) GapR ChIP profiles in raffinose at a locus on chromosome II. AT content (top), AT content below the genomic average (64%) is plotted in reverse. Normalized ChIP-seq (middle) of GapR-3xFLAG (orange) or untagged GapR (grey) expressing cells. Transcription (bottom) from the forward (green) and reverse (blue) strands with annotated genes indicated.

(F) Histogram showing the length of GapR binding events.

(G) Mean GapR enrichment in glycerol at 5' and 3' ends of long genes at various transcriptional cutoffs. Student's t-test p-value is shown.

(H) GapR enrichment in glycerol is highest between convergently oriented genes.

(I) GapR-bound regions in glycerol are more frequently between convergent genes. Pie charts shown as in Fig. 3H.

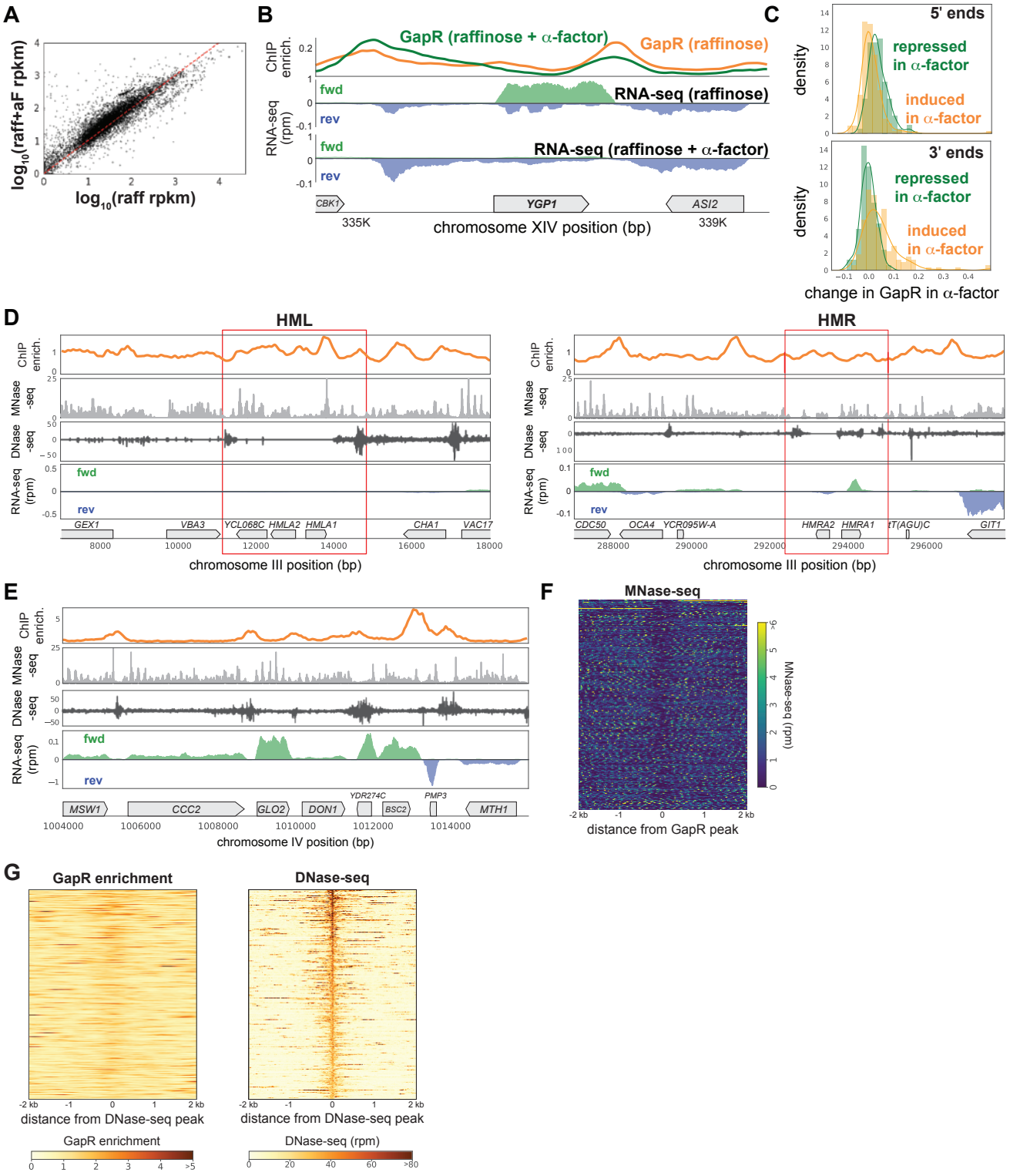

Figure S5. GapR ChIP compared to nucleosome occupancy and chromatin accessibility, related to Figure 5.

(A) Comparison of transcription in raffinose with and without α-factor treatment.

(B) GapR enrichment at *YGP1* in raffinose without (orange) or with (green) α-factor arrest before GapR induction (top). Transcription of the forward (green) and reverse (blue) strands in raffinose without (2nd panel) or with (3rd panel) α-factor arrest with annotated genes indicated.

(C) α-factor dependent GapR enrichment at 5' and 3' ends for induced vs repressed genes. Student's t-test p-value is reported.

(D) GapR enrichment in raffinose (orange) compared to nucleosome occupancy measured by MNase-seq (light grey) and chromatin accessibility measured by DNase-seq (dark grey) at the yeast mating loci, HML and HMR. Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(E) GapR binding compared to nucleosome occupancy and DNase-seq at a locus on chromosome IV as in (**D**).

(F) Heatmap showing MNase-seq at the 500 highest GapR enriched regions.

(I) Heatmap showing GapR enrichment in raffinose at the 500 most DNAase-accessible regions.

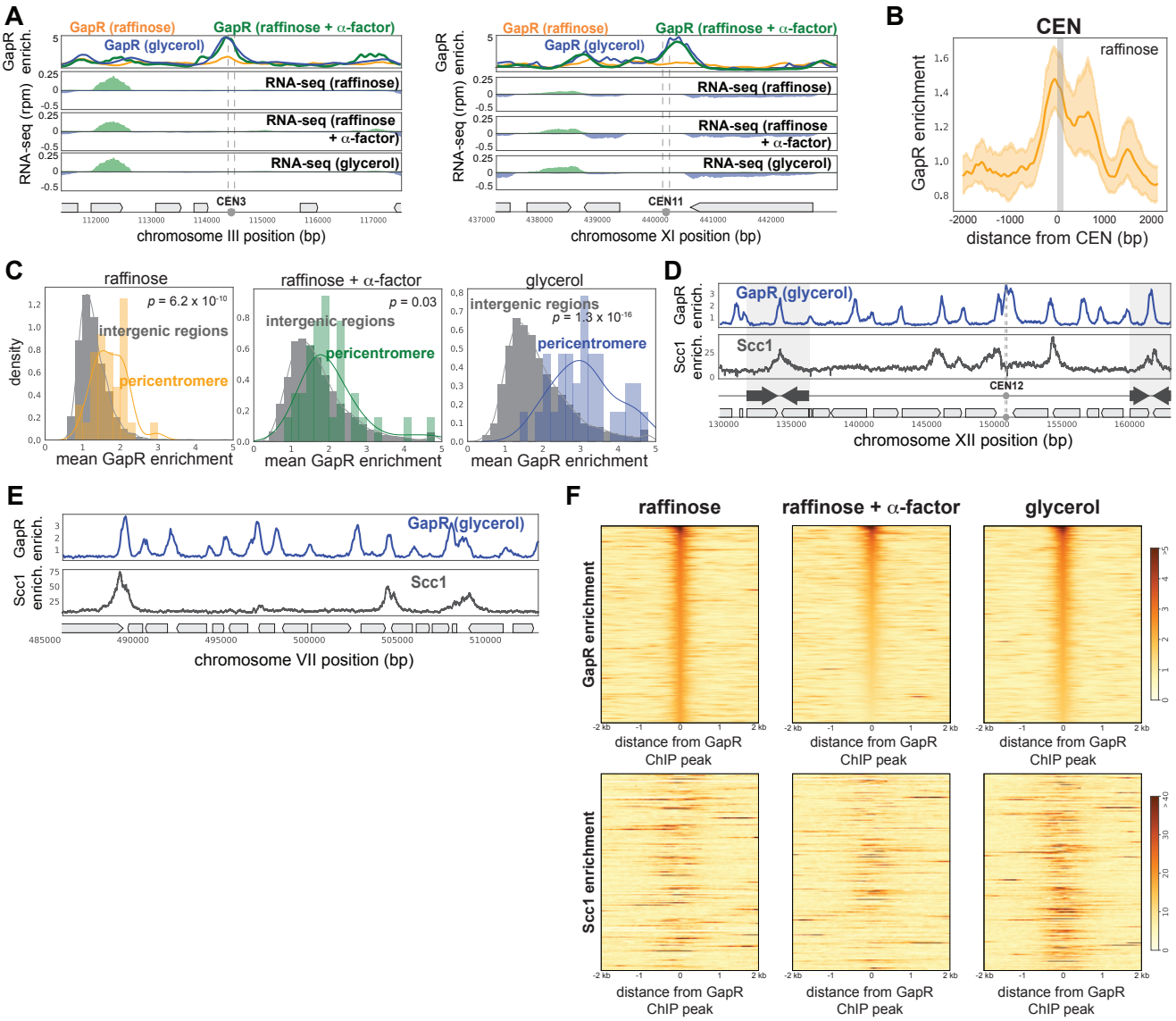

Figure S6. GapR ChIP at centromeres, pericentromeres, and cohesin-bound regions, related to Figure 6.

(A) GapR enrichment (top) in glycerol (blue), in raffinose (orange) and after a-factor arrest (green) at CEN3 (left) and CEN11 (right). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(B) GapR enrichment over all centromeres in raffinose. Mean enrichment (solid line) with 95% confidence intervals (shaded area). Grey bar represents position of centromeres.

(C) GapR enrichment at pericentromeres in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(D) GapR (glycerol, top) and cohesin (Scc1 enrichment, bottom) are associated with convergent genes (arrows) that mark the boundaries of pericentromeres (shaded areas).

(E) GapR and cohesin binding outside of pericentromeres as in (D).

(F) Heatmap showing GapR (top) and cohesin (bottom) enrichment for the 500 most GapR-enriched genomic regions in raffinose, after α-factor arrest, and in glycerol.

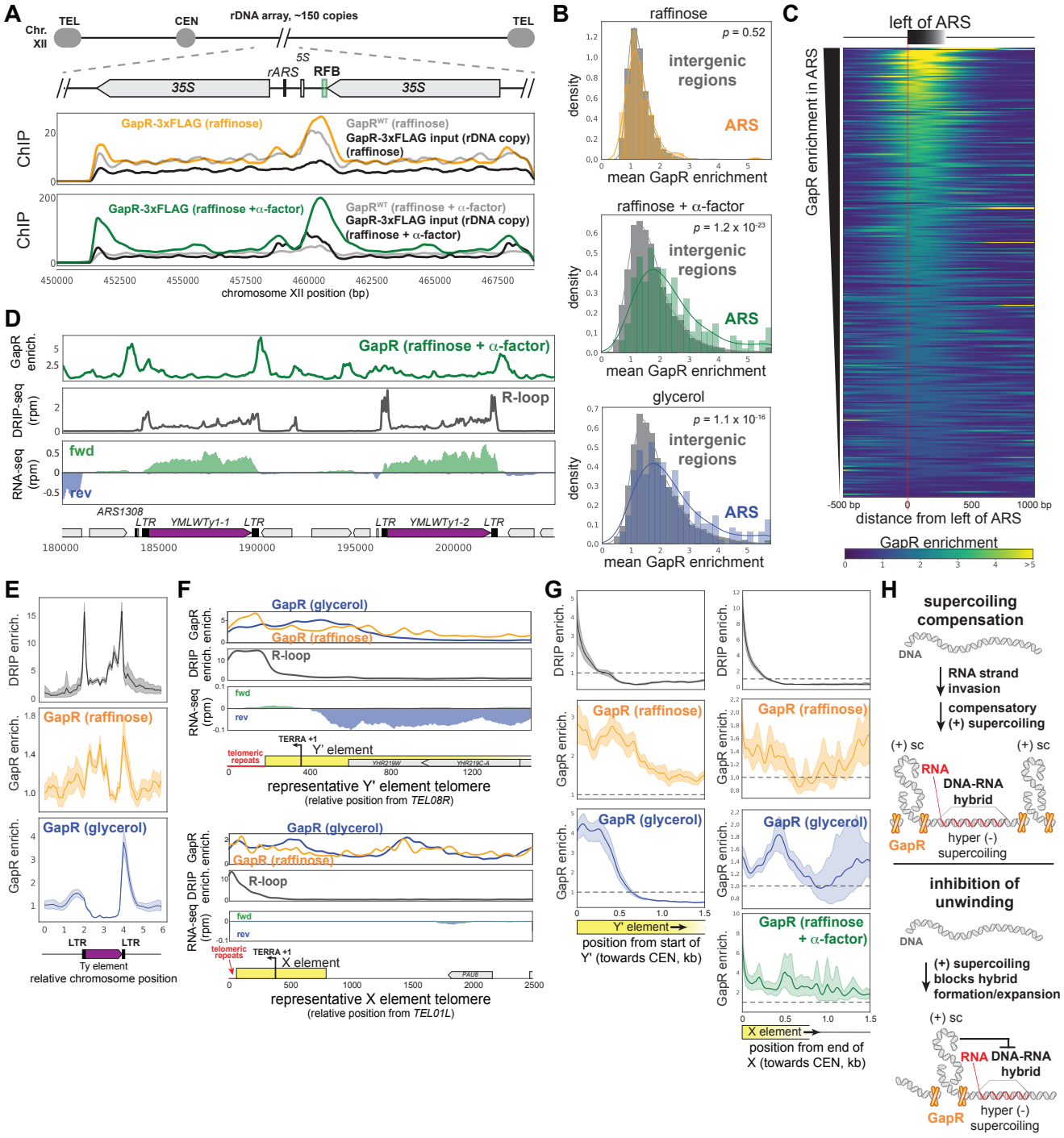

Figure S7. GapR ChIP at rDNA, autonomously replicating sequences, and R-loops, related to Figure 7.

(A) GapR ChIP with controls in the rDNA locus. GapR-3xFLAG ChIP, GapR-3xFLAG input, GapR ChIP in raffinose (top), GapR-3xFLAG ChIP, GapR-3xFLAG input, GapR ChIP in raffinose + α-factor (bottom).

(B) GapR enrichment at autonomously replicating sequences (ARS) in cells grown in raffinose, after α-factor arrest, and grown in glycerol. Student's t-test p-value is shown.

(C) Heatmap showing GapR enrichment (glycerol) at all ARS sorted by GapR enrichment.

(D) Examples of GapR enrichment near Ty elements (purple). GapR enrichment after α–factor arrest (top), DNA-RNA hybrids by DRIP-seq (dark grey). Transcription from the forward (green) and reverse (blue) strands with annotated genes indicated (bottom).

(E) Alignment of GapR and DRIP enrichment surrounding all yeast Ty elements. Data indicate mean (solid line) with 95% confidence intervals (shaded area).

(F) Examples of GapR enrichment near a telomere with (top) or without (bottom) a Y' element, plotted as in (**D**).

(G) GapR and DRIP enrichment at all telomeres with (left) and without (right) Y' elements, aligned from the end of the telomeric repeat. Data indicate mean (solid line) with 95% confidence intervals (shaded area). X element represents both the 'core' X element and any X element combinatorial repeats.

(H) Association of DNA-RNA hybrid (R-loop) forming regions and positive supercoiling.