YY1 Interacts with Guanine Quadruplexes to Regulate DNA Looping and Gene Expression

Lin Li[1], **Preston Williams**[1], **Wendan Ren**[2], **Michelle Y. Wang**[1], **Zi Gao**[1], **Weili Miao**[1], **Ming Huang**[3], **Jikui Song**[2], **Yinsheng Wang**[1,3,*]

[1]Department of Chemistry, University of California, Riverside, CA 92521, USA

[2]Department of Biochemistry, University of California, Riverside, CA 92521, USA

[3]Environmental Toxicology Graduate Program, University of California, Riverside, CA 92521, USA

Summary

The DNA guanine quadruplexes (G4) play important roles in multiple cellular processes, including DNA replication, transcription and maintenance of genome stability. Here, we showed that Yin Yang-1 (YY1) can bind directly to G4 structures. ChIP-Seq results revealed that YY1 binding sites overlap extensively with G4 structure loci in chromatin. We also observed that YY1's dimerization and its binding with G4 structures contribute to YY1-mediated long-range DNA looping. Displacement of YY1 from G4 structure sites disrupts substantially the YY1-mediated DNA looping. Moreover, treatment with G4-stabilizing ligands modulates the expression of not only those genes with G4 structures in their promoters, but also those associated with distal G4 structures that are brought to close proximity via YY1-mediated DNA looping. Together, we identified YY1 as a novel DNA G4-binding protein, and revealed that YY1-mediated long-range DNA looping requires its dimerization and occurs, in part, through its recognition of G4 structure.

Keywords

YY1; Guanine quadruplex; DNA-DNA interaction; Gene regulation

Introduction

The DNA guanine quadruplexes (G4) are stable secondary structures of DNA that form in guanine-rich regions of the genome.[1] Although the in vitro formation of G4 has been known for decades, only until recently have the formation and biological functions of G4 in cells

been demonstrated.[1–3] Computational analysis of potential G4-forming sequences uncovered more than 376,000 G4 sequence motifs in the human genome.[4] Additionally, chromatin immunoprecipitation-sequencing (ChIP-Seq) analysis with the use of G4 structure-specific antibodies revealed approximately 10,000 potential G4 sites in chromatin of cultured human cells.[5–7] Moreover, several DNA G4-recognition proteins have been identified, including DNMT1, hnRNP A1, nucleolin, PARP1, SLIRP, etc.[3,8–11] The large number of G4 sites detected in chromatin and the existence of multiple cellular proteins for G4 structure recognition suggest the crucial functions of G4 structures in cells. Indeed G4 structures have been shown to assume important roles in various DNA metabolic processes including replication, transcription and repair.[1–3]

Yin Yang-1 (YY1) – a zinc finger-containing transcription factor[12] – is ubiquitously expressed and modulates many crucial biological and cellular processes, including DNA replication/transcription/repair, cell proliferation/differentiation, and embryogenesis.[12–14] Recently, YY1 was found to promote the formation of enhancer-promoter loops, which involves YY1's interaction with duplex DNA harboring a consensus YY1-binding motif.[15]

Here, we identified, by using an unbiased quantitative proteomic method, YY1 as a putative G4-binding protein. We also demonstrated that YY1 is capable of interacting directly with G4 DNA in vitro and in cells, and this interaction contributes to YY1-mediated long-range DNA looping and gene expression regulation.

Results

Identification of YY1 as a candidate G4-binding protein

We recently reported a stable isotope labeling by amino acid in cell culture (SILAC)-based quantitative proteomic workflow for the discovery of G4-binding proteins, which involved the use of biotin-conjugated G4 DNA probes derived from the promoters of *KIT* and *MYC* genes and the human telomere (hTEL), along with the corresponding mutated sequences incapable of folding into G4 structures (M4).[11] The proteomic data also revealed YY1 as a candidate G4-binding protein, with the SILAC protein ratios (G4/M4 probes) being 1.92±0.11, 1.77±0.11 and 2.02±0.21 for the cKIT, cMYC and hTEL probes, respectively (Extended Data Fig. 1). Representative ESI-MS and MS/MS for a tryptic peptide derived from YY1, FAQSTNLK, are shown in Supplementary Fig. 1 and Supplementary Fig. 2.

YY1 binds directly with G4 structures in vitro

We next asked whether the interaction between YY1 and G4 structure is direct by measuring the binding affinities between recombinant human YY1 protein and G4 or M4 probes using fluorescence anisotropy (Fig. 1). To this end, we first purified GST-tagged YY1 protein and, because GST is known to form a homodimer,[16] we removed the GST tag and further purified the tag-free YY1 protein (see Materials and Methods). The results showed that recombinant YY1 binds strongly with all three G4 structures, with the K_d values being 26.7±2.2, 26.3±2.1 and 82.0±5.3 nM for G4 structures derived from cKIT, cMYC and hTEL, respectively (Fig. 1a–c). The binding affinities toward M4 probes, are at least 10-fold

lower than that toward G4 DNA (Fig. 1a–c). These results demonstrate that YY1 binds directly and strongly with G4 structures in vitro.

We also examined whether the binding between YY1 and G4 alters the folding of G4 structure. The circular dichroism (CD) spectra for the G4 DNA probes in YY1-DNA complexes are very similar to those of the free G4 DNA probes (Extended Data Fig. 2), underscoring that the interaction with YY1 does not alter the folding of G4 structures.

Pyridostatin (PDS) and 5,10,15,20-tetra-(*N*-methyl-4-pyridyl)porphyrin (TMPyP4) are small-molecule ligands that bind specifically and stabilize G4 structures.[17,18] Our results from electrophoretic mobility shift assay (EMSA) showed that both PDS and TMPyP4 could displace YY1 from G4 structure in vitro (Fig. 1d–f), further substantiating YY1's capability in binding directly with G4 structure.

Duplex DNA with a sequence motif of 5′-AANATGGC-3′ was previously identified as the consensus YY1-binding sites.[19,20] For comparison, we also measured the K_d value for YY1 in binding with a consensus motif-carrying duplex DNA, and it turned out that the binding affinity ($K_d = 11.6\pm0.8$ nM, Supplementary Fig. 3) was only slightly higher than those observed for its binding with G4 structures.

The role of YY1 zinc finger domain in binding with G4

YY1 is highly conserved in vertebrates, where its homologs in human, chimpanzee, dog, mouse, rat and chicken share > 80% amino acid sequence identity (Supplementary Fig. 4). YY1 contains a poly-histidine cluster, a REPO domain and four zinc fingers (Extended Data Fig. 3a). The zinc fingers, located in the C-terminal portion of YY1 and involved in its binding with duplex DNA, are identical among these species (Extended Data Fig. 3 and Supplementary Fig. 4), suggesting that the functions of the zinc finger domain in YY1 are highly conserved.

To assess which segment of YY1 protein is involved in its binding with G4 structure, we generated several truncated or mutated fragments of YY1 protein and measured their binding affinities towards G4 DNA (Extended Data Fig. 3 and Supplementary Fig. 5). Our results showed that YY1 fragments lacking the C-terminal zinc finger ($YY1_{1-382}$) or with a cysteine in the third zinc finger being mutated to a serine ($YY1_{C360S}$) exhibit markedly diminished binding affinities toward G4 structure (Extended Data Fig. 3b–c). On the other hand, truncated YY1 proteins without the N-terminal 230 (i.e. $YY1_{231-414}$) or 292 (i.e. $YY1_{293-414}$) amino acids display similar or slightly stronger binding affinities towards hTEL G4 structure relative to full-length YY1 (Extended Data Fig. 3d–e). Neither $YY1_{231-414}$ nor $YY1_{293-414}$, however, exhibits selective binding toward G4 over M4 (Extended Data Fig. 3d–e). Together, the above results support that the C-terminal zinc-finger domain of YY1 is required, but not sufficient, for the protein to discriminate G4 structure from single-stranded M4.

The capability of YY1 in binding to both consensus motif-containing duplex DNA and G4 DNA and the requirement of the zinc finger domain for both binding modes suggest that these two types of bindings may compete with each other. To test this, we assessed the

binding of YY1 with G4 structure or duplex DNA harboring the YY1 consensus motif in the presence or absence of different concentrations of unlabeled G4 DNA or the consensus motif-bearing duplex DNA (Supplementary Fig. 3). It turned out that both competitors could repress, in a concentration-dependent manner, YY1's binding with G4 or YY1 consensus motif-containing duplex DNA (Supplementary Fig. 3), underscoring the competitive binding of YY1 towards G4 structure and the consensus motif-containing duplex DNA.

YY1 binds with G4 structures in cells

We next investigated the interactions between YY1 with G4 structures in cells by using ChIP-Seq analysis. To this end, we employed CRISPR-Cas9 to incorporate a tandem affinity tag (3×Flag-2×Strep) to the C-terminus of endogenous YY1 protein in HEK293T cells (YY1-TAPTAG, Supplementary Fig. 6). ChIP-Seq analysis with the use of anti-Flag antibody and the YY1-TAPTAG cells led to the detection of 12,063 significant peaks (Fig. 2). An analysis of the overlap between YY1 ChIP-Seq peaks and G4 ChIP-Seq peaks (a BG4 ChIP-Seq dataset from the NCBI GEO database) showed that approximately 39% (4721 out of 12063) of YY1 ChIP-Seq peaks in our dataset overlap by at least one base pair with G4 ChIP-Seq peaks with a hypergeometric p-value of 1.184E-126 (Fig. 2a–b). Substantial overlaps (3965 and 3648 out of 12063 peaks, or 33% and 30%, respectively) were also observed after expanding the size windows of overlap analysis to 8 and 30 base pairs (Supplementary Table 1b–c), which correspond to the lengths of YY1-binding motif and G4 DNA, respectively.

It is worth noting that 3086 out of the 8955 BG4 ChIP-Seq peaks correspond to potential quadruplex-forming sequencing ($GGGN_{1-7}GGGN_{1-7}GGGN_{1-7}GGG$, with N being any of the four canonical nucleotides).[4] This parallels the previous G4 sequencing results and shows that ~70% of observed G4 sequences could not be assigned to the computationally predicted quadruplex-forming sequences.[21] This was attributed to the formation of G4 structures with long loop regions (>7 bases) and/or the involvement of non-contiguous guanine(s) in G4 formation.[21] In this vein, it is of note that the BG4 antibody employed for the ChIP-Seq analysis recognizes G4 structures in general, where it cannot differentiate parallel, anti-parallel and mixed parallel/anti-parallel configurations of G4 structures.[5]

We also analyzed statistically the peak width and distribution of signal enrichment for the overlapping peaks identified from the YY1 ChIP-Seq and BG4 ChIP-Seq results (Extended Data Fig. 4). We found that the mean widths for YY1 and BG4 ChIP-Seq peaks are very similar, and the enrichment signal summits for YY1 and BG4 ChIP-Seq peaks overlap with each other (Extended Data Fig. 4). Among the 4721 overlapping peaks in BG4 ChIP-Seq and our YY1 ChIP-Seq datasets, 2603 are located in gene promoters and 925 harbor the YY1 consensus motif (Supplementary Table 1a). High levels of overlaps were also found for the BG4 ChIP-Seq dataset and two other YY1 ChIP-Seq datasets from the NCBI GEO database (Supplementary Fig. 6c–d and Supplementary Table 1). Among the 4603 and 3186 overlapped peaks found between BG4 ChIP-Seq and the two YY1 ChIP-Seq datasets (GSM1010753 and GSM110897), 2504 and 1684 are located in gene promoters, and 1195 and 1120 contain the YY1 consensus motif, respectively (Supplementary Table 1). These

results revealed that YY1 binds with G4 structures in cells, and more than half of the overlapped peaks are located in gene promoters.

We also assessed how treatments with PDS and TMPyP4, which could bind to G4 structures in cells,[22,23] impact the interaction between YY1 with G4 structures in cells. The results from ChIP-Seq experiments with the YY1-TAPTAG cells showed that the treatment with PDS or TMPyP4 led to significantly diminished interaction of YY1 with its binding regions (Fig. 2c and Supplementary Fig. 7), with the most pronounced diminutions being observed for those YY1 ChIP-Seq peaks that overlap with BG4 ChIP-Seq peaks (Fig. 2d–e). Moreover, treatment of cells with PDS or TMPyP4 resulted in a progressive decrease in the chromatin localization of YY1 (Fig. 2f–g), where we co-treated cells with MG132, a proteasome inhibitor, to minimize protein degradation during the treatment. Together, these results lent further evidence that YY1 binds with G4 structures in cells, and this binding could be perturbed by G4-stabilizing ligands.

YY1-G4 interactions in YY1-mediated DNA looping

YY1 was recently shown to be an important regulator for long-range DNA looping.[15,24,25] We next assessed whether YY1-G4 binding functions in this process by employing HiChIP-Seq, a method developed for monitoring protein-directed chromatin conformation.[26] The results from HiChIP-Seq experiments with the use of YY1-TAPTAG cells showed that treatment with PDS or TMPyP4 led to significant reductions in YY1-mediated DNA looping (Fig. 3a and Supplementary Fig. 8). Statistical analysis revealed that the percentage of DNA looping modulated by YY1 in a distance range of 5–200 kb is markedly attenuated upon treatment with PDS or TMPyP4 (35.6%, 9.4% and 9.9% for control, PDS- and TMPyP4-treated cells, respectively. Fig. 3b). Likewise, treatment with PDS or TMPyP4 resulted in weaker interactions between the genomic regions occupied by YY1 (ChIP-Seq peaks) and their surrounding regions (Fig. 3c and Supplementary Fig. 9). These results demonstrated that YY1's role in DNA looping involves its interaction with G4 structures.

To further assess the role of G4 structure in YY1-mediated DNA looping, we asked how unwinding of G4 structure by ectopic overexpression of a G4 resolvase (i.e. BLM helicase[27]) modulates the YY1-mediated DNA-DNA interaction (Extended Data Fig. 5). In this vein, we confirmed that overexpression of BLM led to substantially diminished signal for G4 structure and YY1 binding at the selected loci (Extended Data Fig. 5a–d). In addition, HiChIP followed by real-time quantitative PCR (HiChIP-qPCR) experiment showed that the BLM helicase-mediated unwinding of G4 structure abrogated the YY1-mediated DNA looping at these sites (Extended Data Fig. 5d–e).

We also mutated the G4-forming sequence at a specific genomic locus by using CRISPR/Cas9-based genome editing to disrupt its capability to fold into G4 structure (Fig. 4 and Supplementary Fig. 10). We found that the mutation indeed attenuated markedly the formation of G4 structure and diminished pronouncedly the binding of YY1 to the site, as revealed by BG4 ChIP-qPCR and YY1 ChIP-qPCR analyses, respectively (Fig. 4b–c). Moreover, HiChIP-qPCR results revealed that the mutation abolished the YY1-mediated DNA looping involving the site (Fig. 4d). This result substantiated the involvement of YY1-G4 structure interaction in YY1-mediated DNA looping.

Dimerization of YY1 promotes DNA-DNA interactions

YY1 protein is known to undergo dimerization, which promotes DNA-DNA interactions through the protein's binding with its consensus sequence motif in duplex DNA.[15,28] We also investigated whether YY1 dimerization enhances YY1-mediated DNA-DNA interactions in vitro. First, we confirmed, by using gel-filtration chromatography, that tag-free YY1 exists primarily as a dimer in solution (Extended Data Fig. 6). Second, our quantitative EMSA experiment revealed a 1:1 binding stoichiometry between YY1 and G4 DNA, indicating that each YY1 dimer can bind to two DNA G4 structures (Extended Data Fig. 6c–d).

We next examined, by employing a proximity ligation assay, the role of YY1 in facilitating the interactions between different DNA elements, including duplex DNA harboring a YY1 consensus motif (motif), G4 DNA (G4), and its mutated counterpart that is unable to fold into G4 structure (M4) (Extended Data Fig. 7a). The results showed that the presence of YY1 led to augmented ligation efficiencies between G4 and G4, and between G4 and motif (Extended Data Fig. 7b). Moreover, addition of a YY1 motif-containing duplex competitor to the ligation mixture led to pronounced diminutions in the efficiency of ligation of DNA elements containing G4 and/or motif (Extended Data Fig. 7b). Parallel control experiments showed that YY1 promotes the ligation between motif and motif, but not between M4 and M4, motif and M4, or G4 and M4 (Extended Data Fig. 7c). Additionally, the inclusion of a G4-stabilizing agent, PDS or TMPyP4, in the ligation mixture led to markedly diminished ligation between G4 and G4 (Extended Data Fig. 7d). These results support that the YY1's dimerization and its binding with G4 DNA enable long-range interactions between G4 DNA and another G4 DNA or a consensus motif-harboring duplex DNA.

To further examine the role of YY1's dimerization in DNA looping, we generated a separation-of-function mutant of YY1, which maintains its capability in recognition of G4 structures, but loses its ability to dimerize. Viewing that $YY1_{231-414}$ and $YY1_{293-414}$ are present as a dimer and a monomer, respectively (Extended Data Fig. 6a–b), we generated a truncated variant of YY1 with the linker region between the REPO and zinc finger domains being deleted ($YY1_{231-290}$) (Fig. 4a). We found that this truncated form of YY1 exists as a monomer in solution, while maintaining a ~ 10-fold selectivity in binding toward G4 over M4 DNA (Extended Data Fig. 8). Owing to the inherent dimerization ability of GST, GST-$YY1_{231-290}$ is present as a dimer in solution (Extended Data Fig. 8). Our proximity ligation assay results showed that the GST-tagged $YY1_{231-290}$, but not its tag-free counterpart, enhanced the ligation between two G4 DNA probes (Extended Data Fig. 8e), supporting that dimerization of YY1 is crucial for DNA-DNA interaction in vitro.

We next asked how loss of dimerization modulates YY1's ability in promoting DNA looping in cells. Our ChIP-qPCR results revealed that both $YY1_{231-290}$ and GST-$YY1_{231-290}$ are capable of binding toward G4 structures in cells (Fig. 4f); $YY1_{231-291}$, however, fails to promote DNA looping in cells, and fusion with the GST tag rescues the protein's ability in enhancing DNA looping (Fig. 4g). These results, therefore, demonstrate that dimerization of YY1 is indispensable for its function in promoting DNA looping in cells.

YY1-G4 binding regulates gene expression

YY1 was first identified as a transcription factor for its function in activation or suppression of gene expression,[15,29] and the above results showed that PDS and TMPyP4 are capable of disrupting the interactions between YY1 and DNA G4 structures (Fig. 1d–f and Fig. 2c–h). Therefore, we next asked whether genetic depletion of YY1, or treatment with PDS or TMPyP4 influences those genes that are modulated by G4 structure. For this purpose, we performed RNA-Seq analysis to assess, at the transcriptome-wide scale, the alterations in gene expression elicited by genetic depletion of YY1 and/or treatment with PDS (Fig. 5, Extended Data Fig. 9 and Supplementary Figs. 11–12). Statistical analysis showed a strong correlation between PDS-regulated and YY1-regulated genes (Pearson r > 0.67), underscoring the role of YY1-G4 structure interaction in YY1-mediated gene regulation (Fig. 5, Extended Data Fig. 9 and Supplementary Fig. 11). Moreover, the RNA-Seq results revealed that YY1 can positively or negatively regulate the expression of genes through its interaction with G4 structures.

We also validated the RNA-Seq results by assessing the mRNA expression levels of six representative genes (i.e. *MANF*, *PDHB, MYC, PGBD5, SLC25A28* and *TMEM145*) by using RT-qPCR. These six genes were chosen on the basis of the presence of G4 structures and the occupancy of YY1 in their promoter regions (Fig. 5 and Extended Data Fig. 9). The results showed that the expression levels of *MANF*, *PDHB* and *MYC* genes were markedly attenuated, whereas those of *PGBD5, SLC25A28* and *TMEM145* genes were substantially elevated in HEK293T cells upon treatment with PDS or TMPyP4, or upon shRNA-mediated knockdown of YY1 (Fig. 5, Extended Data Fig. 9 and Supplementary Fig. 12).

We also assessed whether the long-range DNA looping elicited by YY1-G4 binding plays a role in gene regulation. To this end, we monitored the mRNA expression levels of two genes, *TRMT12* and *EHD3*, which do not carry G4 structures in their promoters but are associated with G4 structure at a remote site via YY1-mediated DNA looping (Fig. 6a–b and Extended Data Fig. 10a–b). Our results showed that the mRNA expression levels of these two genes in HEK293T cells were pronouncedly diminished upon shRNA-mediated knockdown of YY1, or upon treatment with PDS or TMPyP4 (Fig. 6c and Extended Data Fig. 10c).

Next we examined whether this observation can be extended to other genes at the genome-wide scale by integrating the results obtained from YY1 ChIP-Seq, BG4 ChIP-Seq, HiChIP-Seq and RNA-Seq data. To this end, we focused on those genes whose promoters are occupied by YY1 (based on YY1 ChIP-Seq data), don't fold into G4 structures (based on BG4 ChIP-Seq data) and interact with remote sites through YY1-mediated DNA looping (based on YY1 HiChIP-Seq data). We divided these genes into two groups based on whether the remote site adopts G4 structure (based on BG4 ChIP-Seq) (Supplementary Fig. 13). We found that PDS treatment differentially modulated the expression of these two groups of genes, with more pronounced changes being observed for those genes interacting with a distal site carrying G4 structure than those without (Supplementary Fig. 13). Moreover, shRNA-mediated knockdown of YY1 abrogated the differences elicited by PDS treatment, and this is true for both the up- and down-regulated genes (Supplementary Fig. 13). These results, therefore support that YY1's role in transcriptional regulation involves, in part, the long-range DNA looping forged by its interaction with G4 structures.

Discussion

YY1 is an extensively studied transcription factor and it functions in many important biological processes,[12,13,15] where duplex DNA sequences with a consensus motif were previously identified as YY1-binding sites.[15,20] Here, we demonstrated that YY1, apart from binding to its consensus motif sequences in duplex DNA, is capable of interacting directly with G4 structures in vitro and in cells (Fig. 1 and Fig. 2). Further analysis showed that the C-terminal zinc finger domain of YY1 is indispensable for its interaction with G4 structure. Our results from the YY1 $_{231-291}$ truncation mutant showed that YY1's recognition of G4 structures also entails the N-terminal portion of the protein, though the linker region between the REPO and zinc finger domains of the protein is dispensable for this recognition (Extended Data Fig. 3 and Extended Data Fig. 8). It will be important to solve, in the future, the structure of YY1 in complex with G4 DNA so as to gain insights into the similarities and differences in YY1's recognition of G4 structure versus duplex DNA.

Spatial organization is a fundamental element in 3D genome architecture and transcription regulation; however, not much is known about how 3D genome architecture is regulated and which proteins are involved.[15,30–32] In this vein, CTCF and cohesin were shown to contribute to 3D genome organization.[30,32] Recently, YY1 was found to function in this process by enabling enhancer-promoter interactions.[15] Our results unveiled a novel molecular determinant for YY1's function in this process, i.e. through its binding with G4 structures, where displacement of YY1 from G4 structures, unwinding of G4 structures by overexpression of BLM helicase, and mutation of a G4-forming sequence using CRISPR/cas9-mediated genome editing all disrupt YY1-mediated DNA looping (Fig. 3, Extended Data Fig. 5, Extended Data Fig. 7 and Supplementary Fig. 10).

Based on our results, we proposed a mechanistic model, where the dimerization of YY1 allows for its concurrent binding of two consensus sequence motifs, two G4 structures, or one consensus sequence motif and one G4 structure, thereby enabling DNA looping. Our model is supported by results from proximity ligation assay showing that YY1 can promote the ligation between these DNA elements (Extended Data Fig. 7), from analytical gel-filtration assay revealing that YY1 forms a dimer *in vitro* (Extended Data Fig. 6), and from quantitative EMSA assay demonstrating that YY1 binds to G4 DNA at 1:1 stoichiometry (Extended Data Fig. 6). Moreover, our identification of a separation-of-function mutant of YY1 - which is defective in dimerization, but proficient in discriminating G4 structure from single-stranded DNA - and demonstration that the mutation abrogates the YY1-mediated DNA looping in cells lend further support of our mechanistic proposal (Fig. 4).

Disruption of binding between YY1 and G4 structure with the use of G4-stabilizing ligands, PDS or TMPyP4, and genetic depletion of YY1 led to altered expression of not only those genes harboring G4 structures in their promoters (Fig. 5 and Extended Data Fig. 9), but also those genes that do not carry G4 structures in their promoters but are linked with G4 structure at a remote site through YY1-mediated DNA looping (Fig. 6 and Extended Data Fig. 10). These results support that YY1-G4 interaction not only regulates the expression of proximal genes, but also distal genes through long-range DNA looping. This represents a very important mechanism for YY1-regulated gene expression.

Numerous studies have documented the association between YY1 and human cancer, and YY1 is also considered as a potential prognostic biomarker and therapeutic target.[33,34] In this vein, YY1 is overexpressed in multiple types of cancer, and its overexpression is also correlated with poor therapeutic outcomes.[33,34] In addition, YY1 is known to regulate a number of cancer-related genes.[33,34] Furthermore, G4 structures are over-represented in many cancer-related genes, and multiple studies revealed markedly increased presence of G4 structures in cancer cells,[35] indicating the involvement of G4 in cancer development. In this study, we reveal that YY1's functions in modulating gene expression and DNA looping depend, at least in part, on its binding with G4 structure. Hence, the interaction between YY1 and G4 perhaps can be exploited for the therapeutic interventions of human cancer in the future.

Together, our work unveils a novel mode of DNA recognition for a well-studied transcription factor, YY1 (i.e. in its binding with G4 DNA), and reveals the role of this type of molecular recognition in modulating long-range DNA looping and gene expression. Our study also uncovers that the G4 structure-mediated gene regulation can occur not only through G4 structures in promoter regions of genes, but also via distal G4 structures that are brought into close proximity through long-range DNA looping. Therefore, the present work illustrates new molecular determinants for DNA looping, and offers insights into the functions of YY1 and G4 structures in gene regulation.

Online Methods

Cell lines

HEK293T (293T) cells were purchased from ATCC (Manassas, VA). Cells were maintained in Dulbecco's modified Eagle's medium (DMEM, life Technologies) supplemented with 10% fetal bovine serum (FBS, Invitrogen) and 100 unit/mL penicillin/streptomycin at 37 °C in a humidified incubator with 5% CO_2.

Generation of Recombinant YY1 Proteins

The coding sequence (CDS) of human *YY1* gene was amplified by PCR and inserted into the pGEX plasmid linearized by BamHI and XhoI restriction digestion enzymes. For truncated proteins, the corresponding CDS was amplified by PCR and inserted into the pGEX plasmid using the same method. The plasmids were transformed into Rosetta (DE3) pLysS *Escherichia coli* cells and the cells cultured in LB medium at 37 °C. After OD_{600} reached approximately 0.6, the culture was cooled to 20 °C and the protein expression was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG, Sigma) at 20 °C for approximately 12 hr. The cells were then harvested and lysed in a buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 10% glycerol, and 2 mM DTT. After centrifugation, the supernatant was collected and GST-tagged proteins were purified with Glutathione Superflow Agarose (Pierce) following the manufacturer's recommended procedures. Elution was performed using a buffer containing 20 mM Tris (pH 7.5), 250 mM NaCl, 10% glycerol, 2 mM DTT, and 20 mM reduced glutathione. The GST-tagged protein was digested with 3C protease for overnight to remove the GST tag. The tag-free protein was further loaded onto the heparin column (GE Healthcare) and eluted with a NaCl gradient (0.25 – 1.0

M) in a buffer containing 20 mM Tris (pH 7.5), 10% glycerol and 2 mM DTT. The fractions containing the target protein were pooled, concentrated and further purified by size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). The proteins were quantified using Bradford Protein Assay Kit (Bio-Rad), and their purities verified by SDS-PAGE analysis and $A_{260/280}$ measurement (Supplementary Fig. 5 and Extended Data Fig. 8). The analytical gel filtration chromatography was performed using a Superdex 200 10/300 column (GE Healthcare) in a buffer containing 20 mM Tris-HCl (pH 7.5), 250 mM KCl, 10% glycerol and 5 mM DTT.

Fluorescence Anisotropy

Fluorescence anisotropy measurements were conducted as previously described.[11] Briefly, 5-carboxytetramethylrhodamine (TAMRA)-labeled DNA probes were annealed in a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl and 0.1 mM EDTA to form G4 structure. The binding assays were performed with 10 nM labeled DNA and the indicated concentrations of recombinant YY1 protein in a binding buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl, 10 μM ZnCl, and 1 mM DTT. After a 30-min incubation on ice, fluorescence anisotropy was measured on a Horiba QuantaMaster-400 spectrofluorometer (Photon Technology International), with the excitation and emission wavelengths being 550 and 580 nm, respectively. The instrument G factor was determined before anisotropy measurements. The data were fitted according to the following equation:

$$A_{obs} = A_o + \Delta A \times \frac{\left[DNA\right] + \left[Protein\right] + K_d - \sqrt{\left(\left[DNA\right] + \left[Protein\right] + K_d\right)^2 - 4 \times \left[DNA\right]\left[Protein\right]}}{2 \times \left[DNA\right]}$$

In the equation, [DNA] and [Protein] denote the DNA and protein concentrations, respectively. A_{obs} is the observed anisotropy value. A_0 is the anisotropy value in the absence of protein. A represents the total change in anisotropy between free and fully bound DNA, and K_d is the equilibrium dissociation constant.[36]

Electrophoretic mobility-shift assay (EMSA)

For protein-DNA binding, 25 nM TAMRA-labeled DNA was incubated with different concentrations of full-length YY1 protein, or its truncated or mutated counterparts in a binding buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl, 10 μM ZnCl, 1 mM DTT and 3% glycerol on ice for 30 min. The samples were then loaded onto a 6% polyacrylamide gel in TBE buffer (40 mM Tris-HCl, pH 8.3, 45 mM boric acid and 1 mM EDTA) at 4 °C. The samples were run at 120 V at 4 °C for 30 min. The gels were imaged with Odyssey Imaging Systems (LI-COR Biosciences).

Circular Dichroism (CD) Spectroscopy

The DNA probes (2.5 μM) were annealed in a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl and 0.1 mM EDTA to form G4 structure. The CD spectra for the DNA probes were recorded at room temperature on a Jasco-815 spectrometer (Easton, MD) in the wavelength range of 200–320 nm, and the scan rate was 1 nm/sec. For assessing how YY1 binding modulates G4 folding, the annealed DNA probes were incubated with an equal

concentration of the purified YY1 protein at room temperature for 30 min, and the CD spectra were subsequently acquired as noted above. The CD spectrum for the YY1 protein itself was then subtracted from the composite CD spectra of the YY1-G4 DNA complexes.

In vitro Proximity Ligation Assay

For proximity ligation assay, 20 nM DNA fragments containing G4 or YY1 consensus motif were incubated with 100 nM YY1 protein in a binding buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl, 10 μM ZnCl$_2$ and 1 mM DTT at 20 °C for 30 min. The mixture was subsequently diluted and the DNA fragments were ligated in a mixture containing 2 U/μL T4 DNA ligase, 2 U/μL T4 polynucleotide kinase and 1× T4 DNA ligase buffer at 30 °C for 10 min. The reactions were terminated immediately afterwards by heating at 65 °C for 5 min. The resulting mixture was diluted, and real-time quantitative PCR was employed to quantify the ligation products. The sequences for the DNA fragments and primers used in the quantitative PCR are listed in Supplementary Table 2.

Chromatin Fractionation and Western Blot

Chromatin fractionation was performed as described.[37] The cells were grown to 60–70% confluence, treated with 20 μM MG132 for 4 hr, and then treated with 20 μM PDS or 5 μM TMPyP4 at 37 °C for 12 hr. The chromatin fraction was isolated using a step-wise procedure with a cytoplasmic lysis buffer (10 mM Tris-HCl, pH 8.0, 0.34 M sucrose, 3 mM CaCl$_2$, 2 mM MgCl$_2$, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, and protease inhibitor cocktail), a nuclear lysis buffer (20 mM HEPES, pH 7.9, 1.5 mM MgCl$_2$, 1 mM EDTA, 150 mM KCl, 0.1% NP-40, 1 mM DTT, 10% glycerol, and protease inhibitor cocktail) and a chromatin isolation buffer (20 mM HEPES, pH 7.9, 1.5 mM MgCl$_2$, 150 mM KCl, 10% glycerol, protease inhibitor cocktail and 0.15 unit/μL benzonase). The proteins were again quantified by Bradford assay. After separation on a 10% SDS-PAGE, the proteins were transferred to a nitrocellulose membrane (Bio-Rad). After blocking with blotting-grade blocker (Bio-Rad), the membrane was incubated in a solution containing primary antibody and 5% BSA for 2 hr, and then incubated in a 5% blotting-grade blocker containing the HRP-conjugated secondary antibody. The Western blot signal was detected using ECL Western blotting detection reagent (Amersham). Primary antibodies used in this study included histone H3 (Cell Signaling Technology, 9715S; 1:10000), YY1 (Santa Cruz Biotechnology, SC-7341; 1:2000), Anti-Rabbit IgG Peroxidase antibody (Sigma, A0545; 1:10000), Anti-mouse IgG HRP antibody (Santa Cruz Biotechnology, SC-516102; 1:10000), and β-actin (Cell Signaling Technology, 4967S; 1:5000).

Reverse-Transcription Quantitative PCR (RT-qPCR)

RT-qPCR was conducted as previously described.[37] Total RNA was extracted with Omega Total RNA Kit I (Omega) and quantified. Reverse transcription was performed using M-MLV Reverse Transcriptase (Promega) for cDNA synthesis. RT-qPCR was carried out using iQ SYBR Green Supermix (Bio-rad) on the CFX96 RT-qPCR detection system (Bio-rad). Primers used for RT-qPCR are listed in Supplementary Table 2.

Targeted Integration and Mutation with CRISPR-Cas9

YY1-TAPTAG cell line was generated in HEK293T background using CRISPR-Cas9, following previously reported procedures.[11] The guide sequence for YY1 or G4 site was inserted into the pX330 plasmid (Addgene, Cambridge, MA, USA) which expresses hSpCas9. The guide sequence for YY1 was 5′-GAGAAGACCCTTCTCGACCA-3′. The donor plasmid for tagging YY1 (3×Flag-2×Strep tag) was synthesized (gBlock, Integrated DNA Technologies) and inserted into pUC19. For CRISPR-mediated mutation of an endogenous G4 site, the donor plasmid for the mutated G4 sequences was inserted into pUC19. After co-transfection with the two plasmids, the single-cell colonies were isolated and Western blot was performed to screen for the insertion of the tandem tag using anti-YY1 antibody (SC-7341, Santa Cruz Biotechnology). Sanger sequencing was performed to confirm the mutation of the sequence for the endogenous G4 site.

Chromatin Immunoprecipitation-Sequencing (ChIP-Seq)

ChIP experiments were conducted as previously described with a few modifications.[15,37] YY1-TAPTAG cells were cultured in DMEM medium with or without 20 μM PDS or 5 μM TMPyP4 for 12 hr before cross-linking. Approximately 2×10^6 cells were cross-linked with 1/10 volume of freshly prepared 11% formaldehyde solution at room temperature for 10 min, and quenched with 125 mM glycine for 5 min. After washing with 1× PBS for three times, the cells were resuspended in a lysis buffer I (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, and protease inhibitors cocktail) at 4 °C on a rotator for 10 min.

After centrifugation, the pellet was resuspended in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and protease inhibitor cocktail) at 4 °C for 10 min with rotation. The pellet was subsequently resuspended in the sonication buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, and protease inhibitor cocktail). Sonication was conducted using a Covaris S220 sonicator for 6 min with a peak incident power of 140, a duty cycle of 5%, and 200 cycles per burst at 4 °C. After centrifugation at 16000 g for 15 min, the supernatant was incubated with anti-Flag antibody (Cell Signaling Technology, 2368S) for 2 hr. Protein A/G Plus-Agarose (Santa Cruz) was then added and the mixture was incubated at 4 °C overnight. After washing, DNA was eluted from the beads with 100 mM NaHCO$_3$ and 1% SDS at 68 °C for 2 hr. Cross-links were subsequently reversed and proteins digested with proteinase K at 65 °C overnight. After purification, RNA in the resulting DNA samples were removed with RNase A. Finally, the DNA was purified using QIAquick PCR Purification Kit (Qiagen).

DNA sequencing library was prepared using NEBNext Ultra™ DNA Library Prep Kit for Illumina (NEB) following the manufacturer's instructions. The purified DNA libraries were subsequently quantified using Agilent 2100 Bioanalyzer and multiplexed for sequencing on a NextSeq500 Sequencing System (Illumina).

The sequencing reads of ChIP-Seq were checked with FastQC (version 0.68) and aligned to human hg19 reference genome using Bowtie2 (version 2.3.4.1) with the configuration of bowtie2-q-N 0–L 22.[37] Peak calling was performed using the Model-based Analysis of

ChIP-Seq (MACS2 2.1.1) with the following configuration: MACS2 callpeak-f BAM-g 2.7e +9–n YY1-ChIP.[38] The UCSC Genome Browser was used to visualize the mapping results. [39] The overlap of ChIP-seq peaks was analyzed using custom script. In detail, each peak in the two BED files were compared, and those peaks overlapped in the two ChIP-Seq datasets (i.e. YY1 and BG4 ChIP-Seq) by at least 1 base pair are considered overlapping peaks, following previously published procedures.[3,40] Parallel analyses were also conducted by employing the criterion that the BG4 and YY1 ChIP-Seq peaks overlap by at least 8 or 30 base pairs (Supplementary Table 1). For motif analysis, the genomic sequence of each peak was extracted, where the consensus motif of "AANATGGC" and its complementary strand of "GCCATNTT" were searched and counted. The ChIP-seq peaks were visualized using the UCSC Genome Browser on Human GRCh37/hg19.

For BG4 ChIP-qPCR experiment, chromatin was immunoprecipitated using BG4 antibody, which was purified as described,[7] following the aforementioned procedures. After purification of the BG4 antibody-enriched DNA fragments, quantitative PCR was performed using the primers listed in Supplementary Table 2.

HiChIP-Seq and HiChIP-qPCR

HiChIP was conducted following the previously described procedures.[15] Before cross-linking, the YY1-TAPTAG cells were mock-treated, or treated with 20 μM PDS or 5 μM TMPyP4 for 12 hr. Approximately 5×10^6 cells were cross-linked using freshly prepared 1% formaldehyde solution at room temperature for 10 min, and quenched with 125 mM glycine for 5 min. After washing for 3 times, the cells were incubated in HiChIP lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, and 0.2% IGEPAL CA-630 with protease inhibitors cocktail) at 4 °C for 30 min with rotation. After centrifugation, the cells were washed again with HiChIP lysis buffer and centrifuged. The resulting pellet was resuspended in 0.5% SDS and incubated at 60 °C for 10 min. To the mixture was added 1.5% Triton X-100 to quench the SDS. After incubation at 37 °C for 15 min, the chromatin was then digested with HindIII (NEB) in NEB CutSmart Buffer (NEB) at 37 °C for 6 hr. The restriction enzyme was subsequently inactivated by heating at 65 °C for 20 min. Biotin-labeled dCTP was incorporated to the restriction fragment overhangs in the Fill-in master mix (0.25 mM biotin-14-dCTP, 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP, and 50 U Klenow fragment of DNA Polymerase I in NEB CutSmart Buffer). The reaction was continued at 37 °C for 2 hr with rotation. Labeled DNA was ligated in a ligation master mix (1% Trion X-100, 0.1 mg/mL BSA, and 250 U T4 DNA ligase in T4 Ligation Buffer) at 16 °C for 4 hr with rotation.

The ligated chromatin was collected by centrifugation at 2500 g for 5 min and resuspended in ChIP sonication buffer (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and protease inhibitor cocktail). Chromatin immunoprecipitation was conducted under the same conditions as described above. After elution, RNA and proteins in the samples were removed by digestion with RNase A and proteinase K, respectively. The samples were subsequently incubated at 65 °C overnight to reverse the cross-links, and the DNA was purified from the mixture using QIAquick PCR Purification Kit (Qiagen).

After purification, the fragmented DNA was enriched using the streptavidin C1 magnetic beads (Invitrogen, 65001). The beads were washed twice with Tween washing buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) and subsequently resuspended in 2× biotin binding buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl). An equal volume of purified DNA was added, and incubated at room temperature for 15 min. The magnetic beads were then washed twice with Tween washing buffer, twice with binding buffer, and then once with TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA). The magnetic beads were collected and the DNA sequencing libraries were constructed using NEBNext Ultra™ DNA Library Prep Kit for Illumina (NEB) following the manufacturer's instructions. The purified DNA libraries were subsequently quantified using Agilent 2100 Bioanalyzer and multiplexed for sequencing on a NextSeq500 Sequencing System (Illumina).

For HiChIP-Seq data analysis, the reads were aligned to human hg19 reference genome using the Map with BWA-MEM (0.7.15.1.) tool with a gap extension penalty setting of 50 (bwa mem-A1-B4-E50-L0).[41] The interaction matrix was subsequently built with hicBuildMatrix (2.1.0) tool with restriction site being set to HindIII (hicBuildMatrix– binSize 5000–restrictionSequence AGCT).[42] The visualization of the HiChIP data was performed with hicPlotMatrix (2.1.0) tool.[42]

For HiChIP-qPCR experiment, the DNA was purified after chromatin immunoprecipitation, and was used for quantitative PCR with the primers being listed in Supplementary Table 2.

RNA-Seq and Data Analysis

Total RNA was extracted from cells with Omega Total RNA Kit I (Omega). The NEBNext® Ultra™ II RNA Library Prep Kit for Illumina was used to produce the RNA sequencing library. rRNAs were removed using the NEBNext rRNA Depletion Kit, and the resulting RNA samples were treated with DNase I to remove residual DNA. The rRNA-depleted RNA was incubated in NEBNext First Strand Synthesis Reaction Buffer by heating at 94°C for 15 min on a thermal cycler. The fragmentized RNA was reverse-transcribed using First Strand synthesis enzyme Mix (NEB) with random primers. The NEBNext Second Strand Synthesis Enzyme Mix (NEB) was used to perform the second-strand synthesis. After purification with AMPure XP Beads, the sequencing library was subsequently prepared as described above. The DNA was barcoded with NEBNext Multiplex Oligos for Illumina (NEB), and the resulting library was subjected to sequencing analysis on an Illumina HiSeq4000 insturment.

The sequencing reads of RNA-Seq were aligned to human hg38 reference genome using HISAT2 (2.1.0) with the configuration of hisat2-q-x −1 −2-S. Transcript abundance was determined using featureCounts (1.6.4) with the configuration of featureCounts-T 4-t exon-g gene_id. Differential gene expression was analyzed with DESeq2 (2.11.40.6).

DATA AVAILABILITY

The ChIP-Seq, HiChIP-Seq and RNA-Seq data generated in this study have been deposited into the NCBI GEO database (For ChIP-Seq and HiChIP-Seq data: GEO accession number, GSE128106. For RNA-Seq: GEO accession number, GSE142075). The ChIP-Seq data for

Ishikawa and SK-N-SH cells were obtained from NCBI GEO database with accession numbers of GSM1010753 and GSM1010897, respectively.[43] The two G4 ChIP-Seq data were obtained from NCBI GEO database with accession numbers of GSE99205 and GSE107690.[3,7] The human hg19 reference genome was downloaded from https://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips.
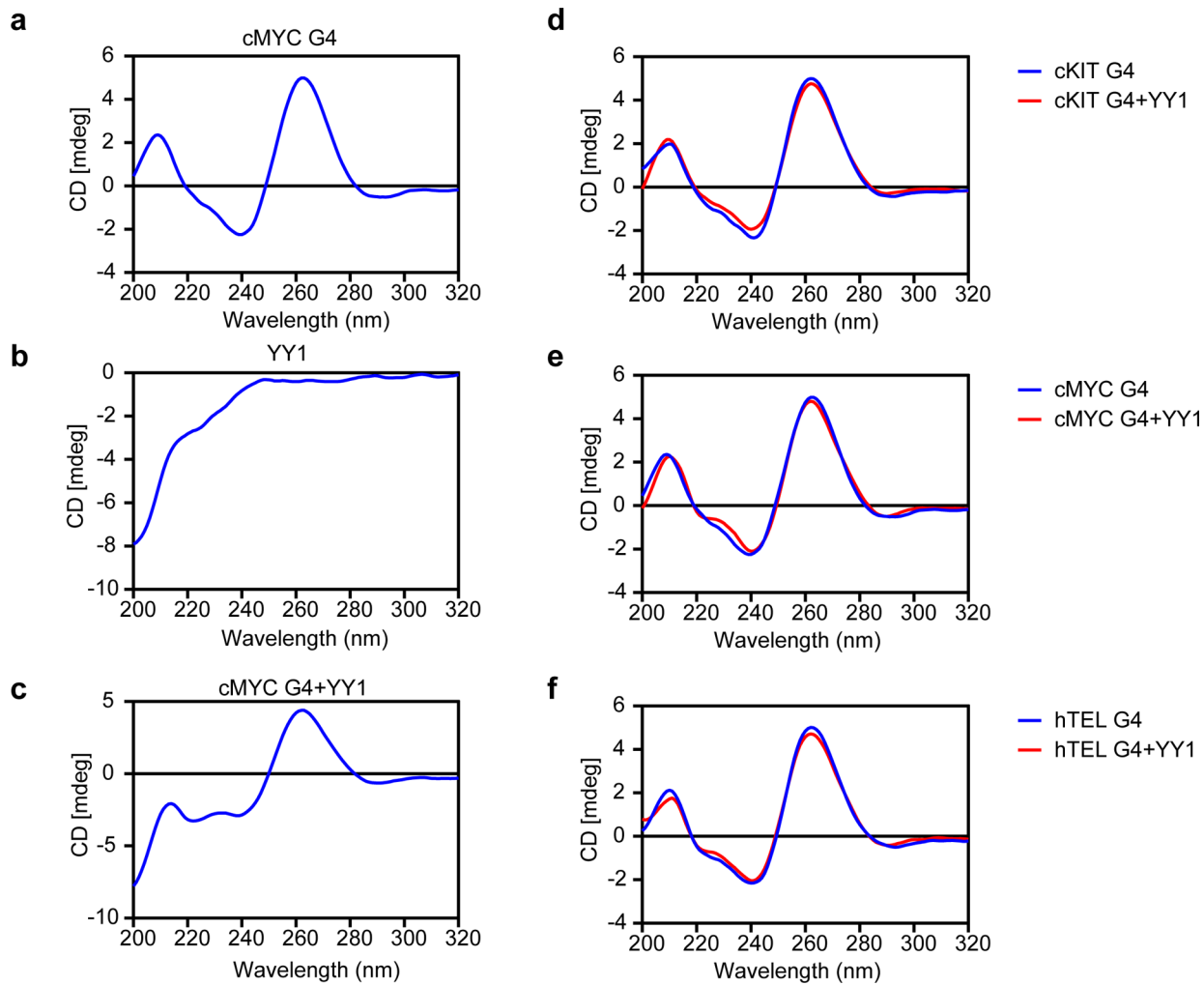
Code Availability

Custom codes used in this work are available from https://github.com/linliucr/UCR_code.

Extended Data
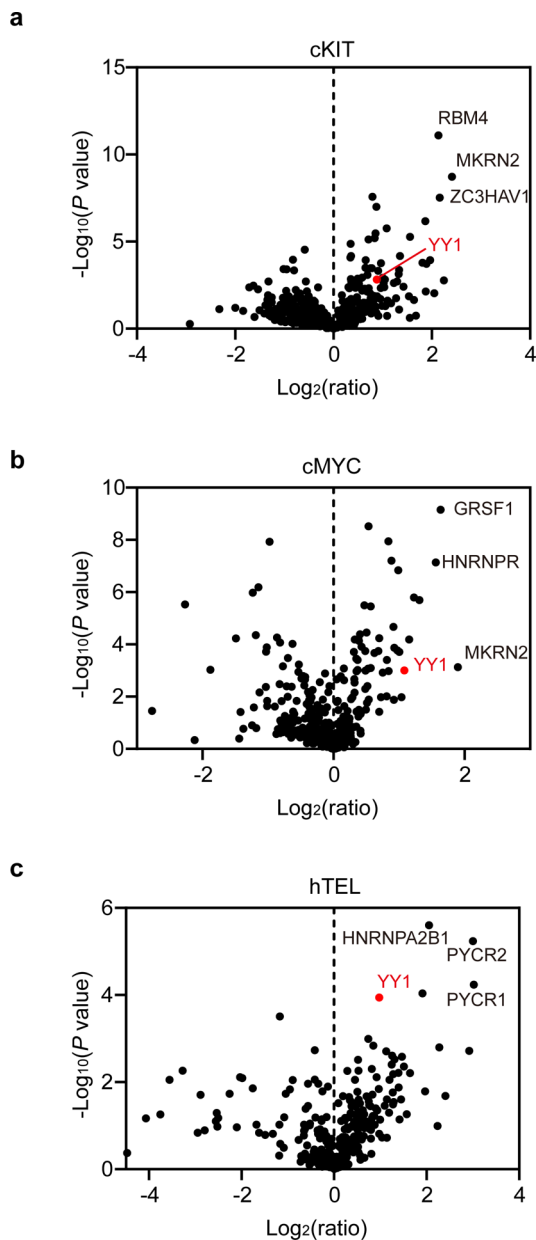
Extended Data Fig. 1. Proteome-wide identification of G quadruplex-binding proteins.

Volcano plots showing the quantification results of G quadruplex-binding proteins identified from SILAC-based interaction screening. The $-Log10(P value)$ was plotted against the log_2(ratio G4/M4). YY1 is labeled in red.

a

cMYC G4

b

YY1

c

cMYC G4+YY1

d

— cKIT G4
— cKIT G4+YY1

e

— cMYC G4
— cMYC G4+YY1

f

— hTEL G4
— hTEL G4+YY1

Extended Data Fig. 2. Circular dichroism (CD) spectra of the three G4 sequences in the absence or presence of tag-free YY1 protein.

a-c, CD spectra of cMYC G4 (**a**), YY1 (**b**) and cMYC G4-YY1 complex (**c**). **d-f**, Comparison of CD spectra for G4 probes in the presence or absence of YY1 protein. The CD spectra for the G4 probes in the presence of YY1 protein were obtained by subtracting the CD spectrum of the YY1 protein from the composite spectra of the protein-G4 DNA complexes.
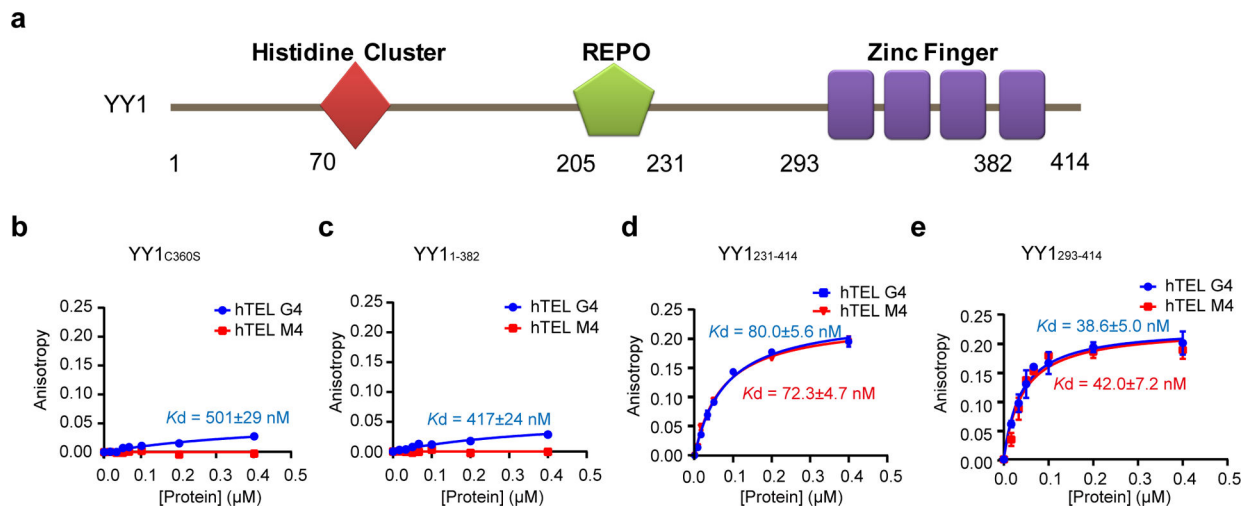
Extended Data Fig. 3. Zinc finger domain of YY1 is essential, but not sufficient for the specific binding toward G4 structure.

a, A schematic diagram depicting the domain structure of YY1 protein. **b-e**, Fluorescence anisotropy for monitoring the bindings of mutant (YY1$_{C360S}$) and truncated (YY1$_{1-382}$, YY1$_{231-414}$, and YY1$_{293-414}$) YY1 proteins with hTEL G4 and the corresponding M4 DNA probes. The data represent the mean ± S.E.M. of results obtained from 3 independent experiments.

a

b
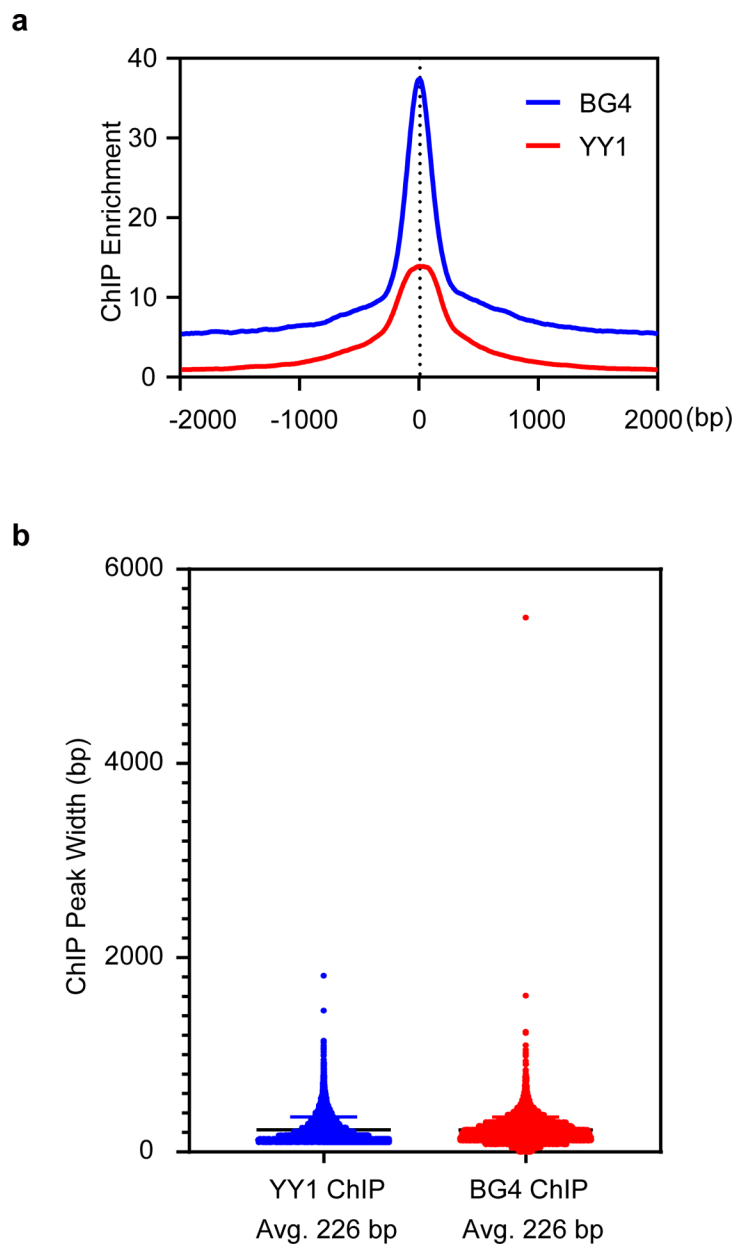
Extended Data Fig. 4. Statistical analyses of YY1 ChIP-Seq and BG4 ChIP-Seq data.
a, The ChIP-seq signal enrichment of YY1 ChIP-Seq and BG4 ChIP-Seq based on the overlapped peaks for the two datasets. The YY1 ChIP-Seq and BG4 ChIP-Seq average signal enrichments are plotted against the BG4 overlapped peaks. **b**, Analysis of peak width distribution for YY1 ChIP-seq and BG4 ChIP-seq data.
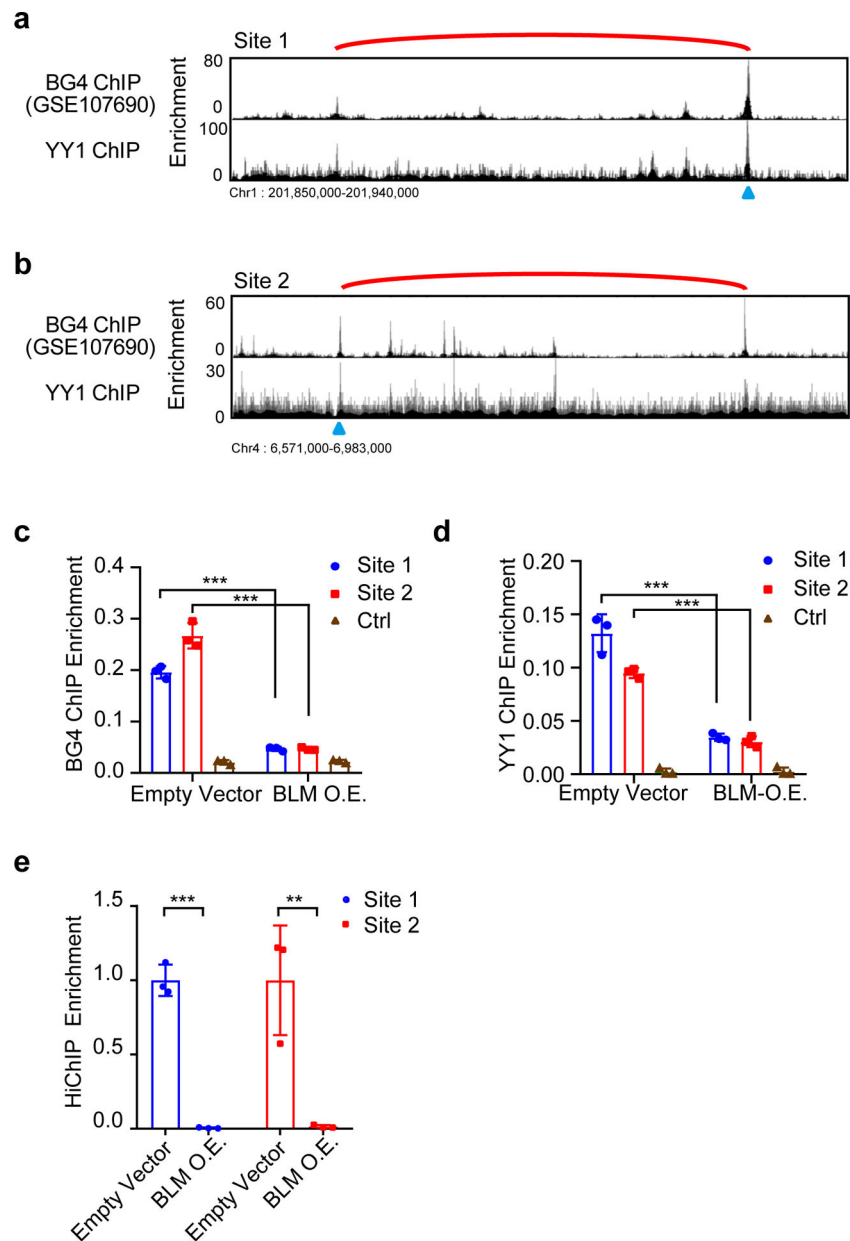
Extended Data Fig. 5. Unwinding of G4 structures with the overexpression of BLM helicase disrupts the YY1-mediated DNA looping.

a-b, Reads enrichment of two regions from BG4 ChIP-Seq and YY1 ChIP-Seq. The long-range DNA interactions monitored in HiChIP-PCR experiment are labeled with red arches, and the regions monitored in BG4 ChIP PCR experiments are indicated with blue triangles. **c-d**, BG4 ChIP and YY1 ChIP enrichments at the two sites were markedly diminished after ectopic overexpression of BLM helicase (BLM-O.E.). **e**, YY1-mediated DNA looping is disrupted by overexpression of BLM helicase. HiChIP-PCR quantification results of YY1-mediated DNA looping in HEK293T cells with and without the overexpression of BLM. Shown in (**c-e**) are mean ± S.E.M. of results obtained from 3 independent experiments. The p values were calculated by using two-tailed, unpaired Student's t test: **, $p < 0.01$; ***, $p < 0.001$.
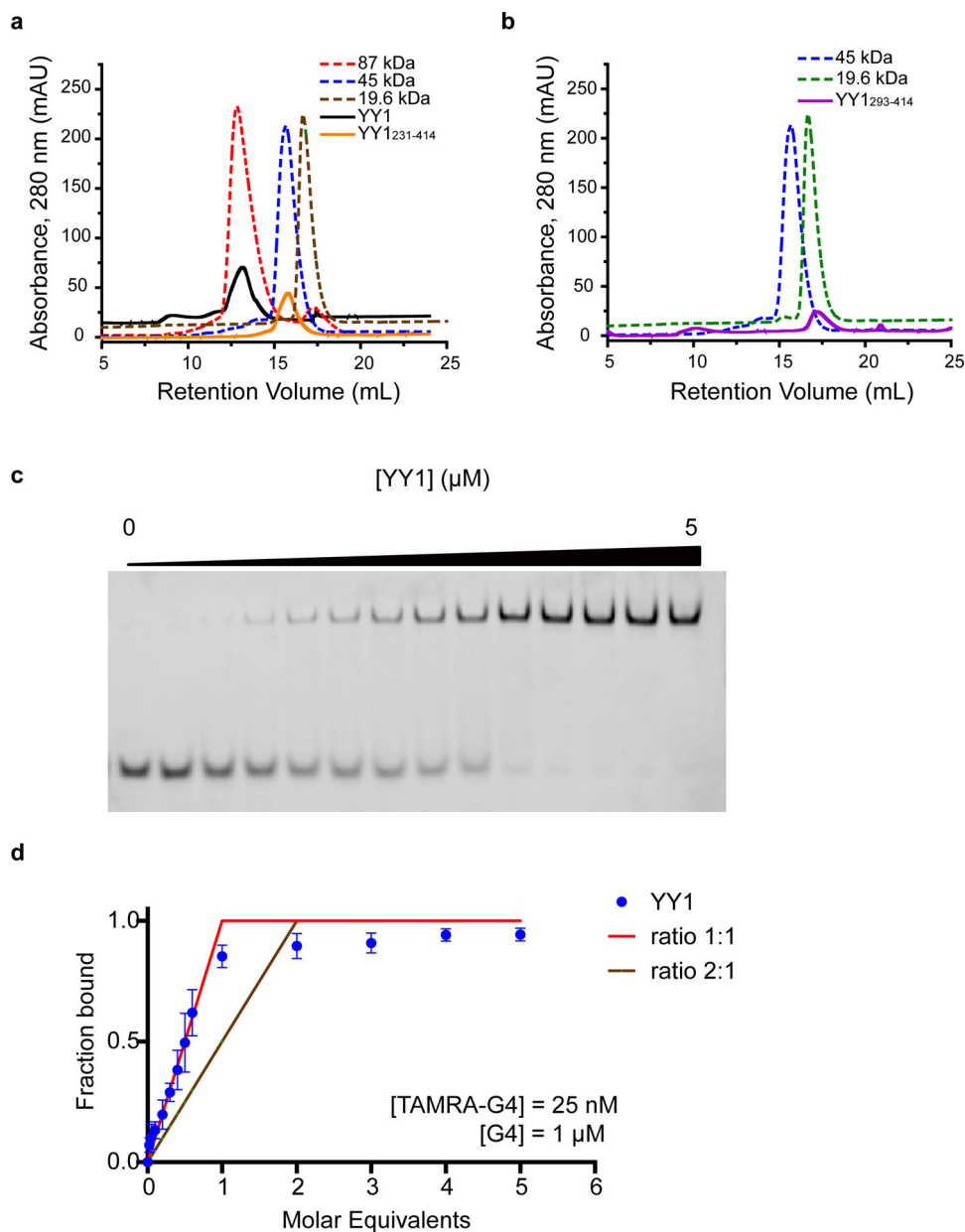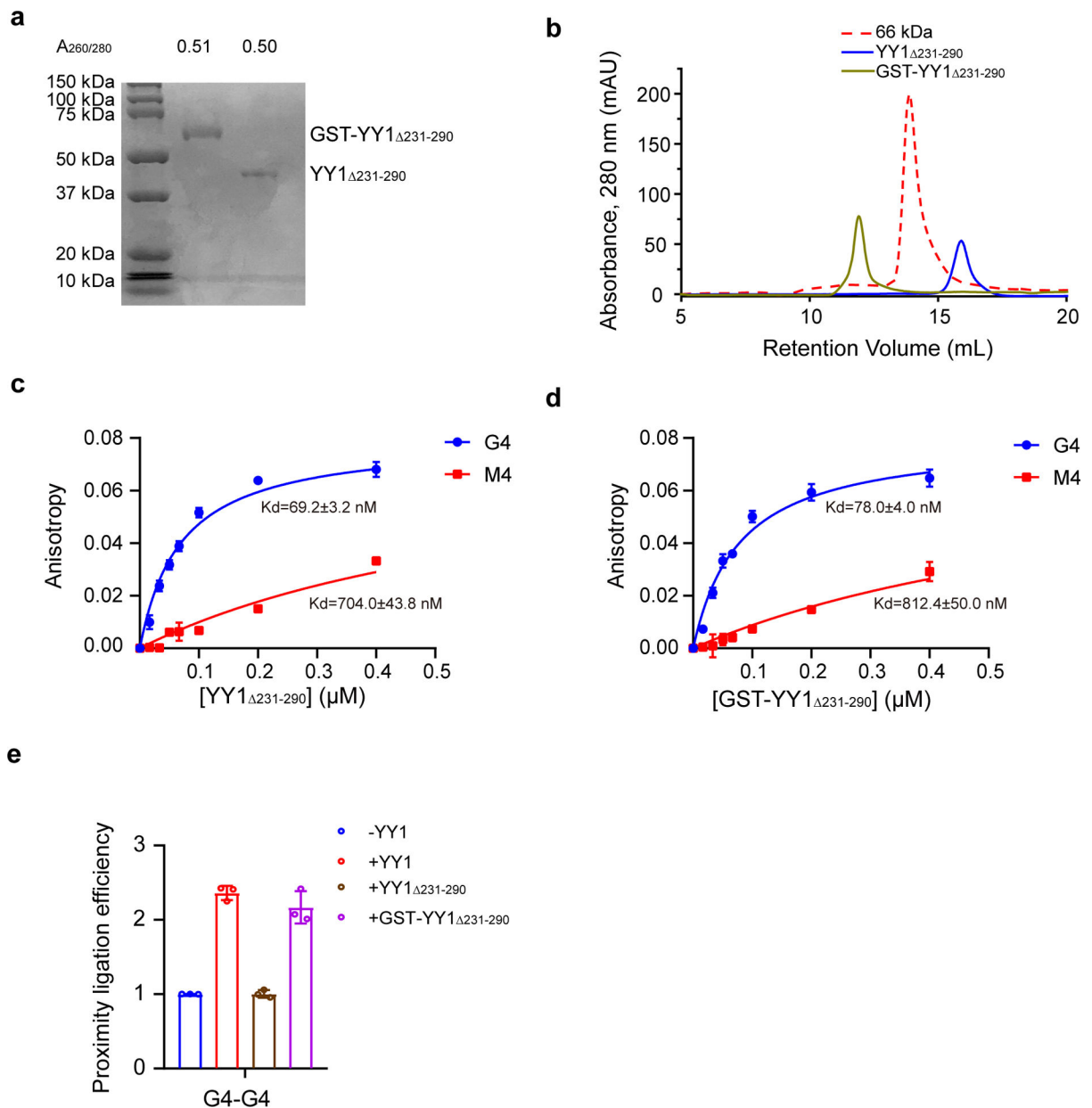
Extended Data Fig. 6. Analysis of YY1 dimerization and binding stoichiometry of the YY1-G4 DNA complex.

a-b, Gel filtration chromatography revealed the dimerization of YY1 and truncated YY1$_{231-414}$, but not YY1$_{293-414}$. **c-d**, The binding stoichiometry of YY1:G4 DNA was analyzed using EMSA. The stoichiometry in binding of YY1 to G4 DNA matches with the theoretical curve in 1:1 binding stoichiometry. The data represent mean ± S.E.M. from 4 independent experiments.
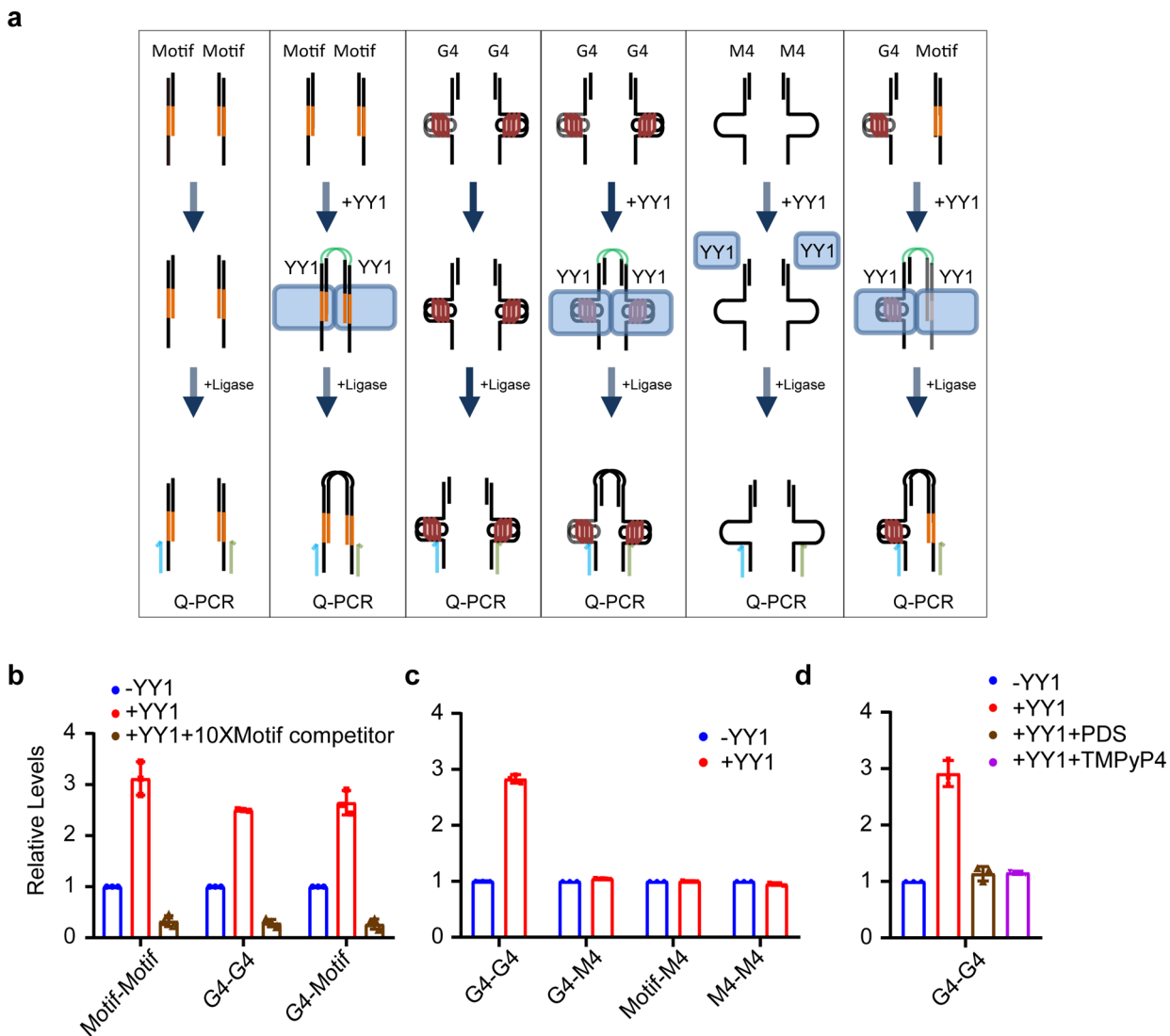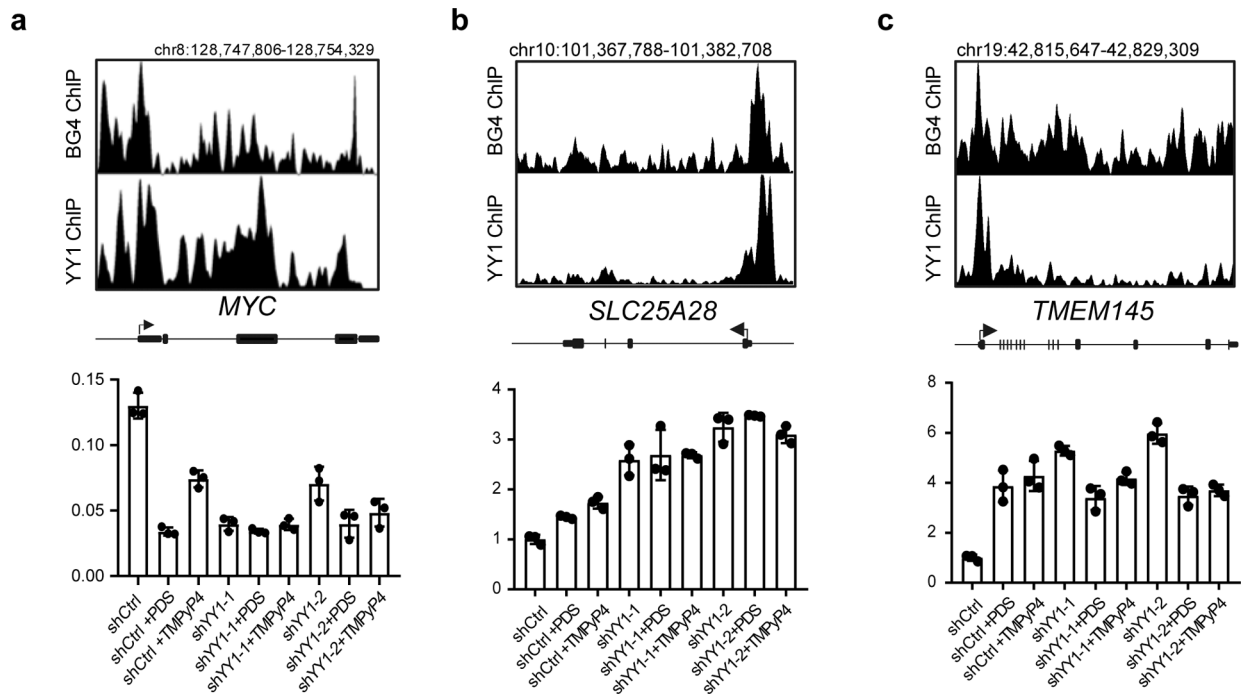
Extended Data Fig. 7. YY1 promotes interactions between DNA elements containing its consensus sequence motifs (Motif), G4 DNA (G4), or both.

a, A scheme depicting the in vitro proximity ligation assay for assessing the ability of YY1 to enhance DNA-DNA interactions involving G4 structures and/or YY1 consensus motifs. **b**, qPCR quantification results of the proximity ligation products formed between motifs, between G4 structures, and between motif and G4 structure in the presence or absence of YY1 protein. **c**, qPCR quantification results revealed the inability of YY1 in promoting the ligation between M4 (i.e. mutated sequence of G4 that can no longer fold into G4 structure) and G4, motif, or M4. **d**, qPCR quantification results of the proximity ligation products formed between G4 structures in the absence or presence of PDS or TMPyP4. The data represent mean ± S.E.M. of results from three independent experiments.

Extended Data Fig. 8. Dimerization of YY1 promotes long-range DNA looping.
a, SDS-PAGE for monitoring the purified recombinant truncated forms of YY1 that is incapable of dimerization, but able to discriminate G4 structure from single-stranded DNA. **b**, Gel filtration chromatography revealed that YY1 $_{231-290}$ (calculated monomer MW: 38.4 kDa) exists as a monomer, and GST-YY1 $_{231-290}$ is present as a dimer (calculated monomer MW: 66.3 kDa). **c-d**, Fluorescence anisotropy for monitoring the binding of YY1 $_{231-290}$ and GST-YY1 $_{231-290}$ protein with G4 structure and the corresponding mutated sequence (M4) derived from the *MYC* promoter. **e**, Proximity ligation assay showing that YY1 and GST-YY1 $_{231-290}$, but not YY1 $_{231-290}$, is capable of promoting ligation between G4 DNA sequences.

Extended Data Fig. 9. Regulation of gene expression by YY1-promoter G4 interactions.
a-c, Quantification of mRNA expression levels of *MYC* (**a**), *SLC25A28* (**b**) and *TMEM145* (**c**) genes after shRNA-mediated knockdown of YY1 and/or PDS/TMPyP4 treatment. Top of each panel shows read enrichments from BG4 ChIP-Seq and YY1 ChIP-Seq. The data represent mean ± S.E.M. of results from three independent experiments.
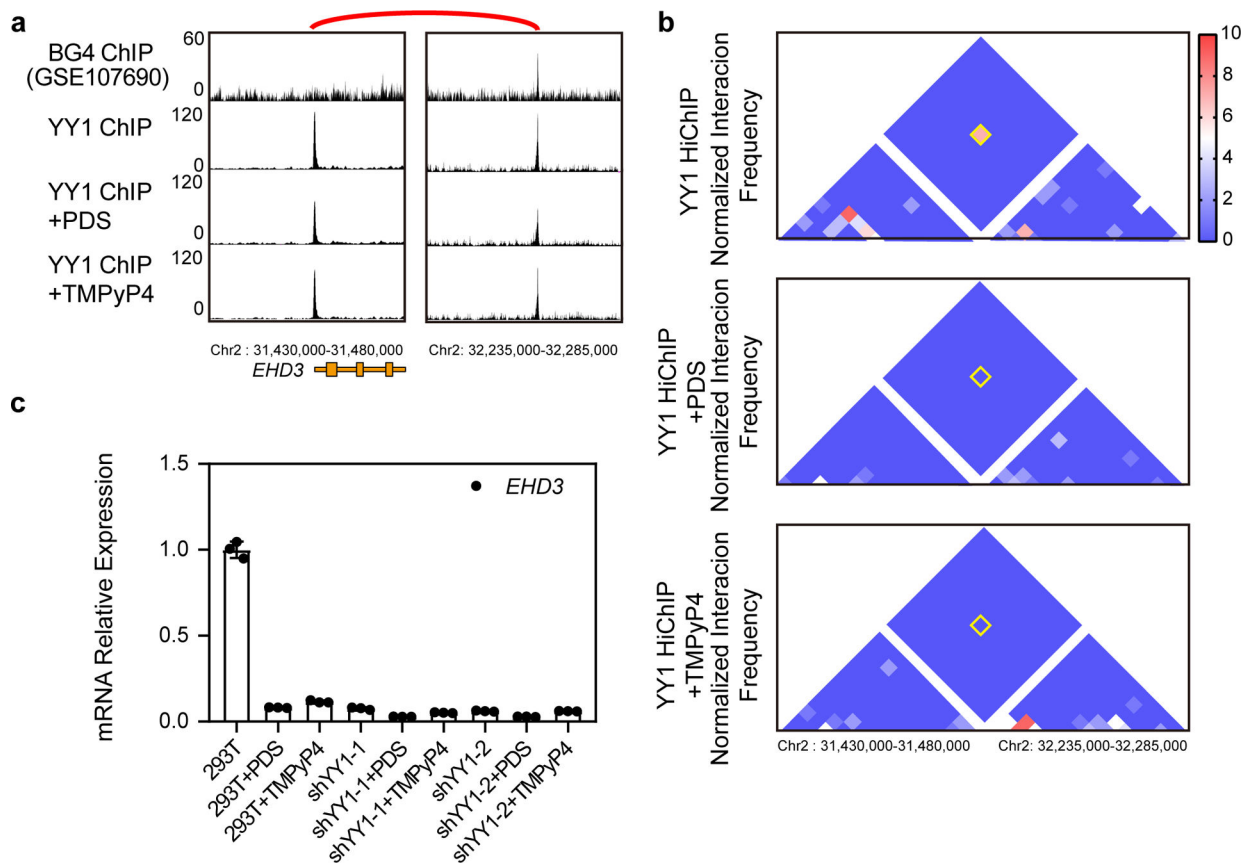
Extended Data Fig. 10. YY1-G4 binding participates in transcription regulation of *EHD3* genes through long-range DNA looping.
a, Read enrichments obtained from BG4 ChIP-seq and YY1 ChIP-Seq experiments. **b,** Normalized interaction frequency between the two sites linked with a red arch in **a** in HEK293T cells with or without PDS/TMPyP4 treatment, as captured by YY1 HiChIP. **c,** Quantification results for the mRNA expression levels of *EHD3* genes in HEK293T cells after knockdown of YY1 and/or with PDS/TMPyP4 treatment. The data represent mean ± S.E.M. of results from three independent experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Bochman ML, Paeschke K & Zakian VA DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet 13, 770–780 (2012). [PubMed: 23032257]

2. Hansel-Hertsch R, Di Antonio M & Balasubramanian S DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat. Rev. Mol. Cell Biol 18, 279–284 (2017). [PubMed: 28225080]

3. Mao SQ, et al. DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol 25, 951–957 (2018). [PubMed: 30275516]

4. Huppert JL & Balasubramanian S Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33, 2908–2916 (2005). [PubMed: 15914667]

5. Biffi G, Tannahill D, McCafferty J & Balasubramanian S Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem 5, 182–186 (2013). [PubMed: 23422559]

6. Henderson A, et al. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 42, 860–869 (2014). [PubMed: 24163102]

7. Hansel-Hertsch R, Spiegel J, Marsico G, Tannahill D & Balasubramanian S Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc 13, 551–564 (2018). [PubMed: 29470465]

8. Soldatenkov VA, Vetcher AA, Duka T & Ladame S First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. ACS Chem. Biol 3, 214–219 (2008). [PubMed: 18338862]

9. Gonzalez V, Guo K, Hurley L & Sun D Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. J. Biol. Chem 284, 23622–23635 (2009). [PubMed: 19581307]

10. Paramasivam M, et al. Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. Nucleic Acids Res. 37, 2841–2853 (2009). [PubMed: 19282454]

11. Williams P, Li L, Dong X & Wang Y Identification of SLIRP as a G quadruplex-binding protein. J. Am. Chem. Soc 139, 12426–12429 (2017). [PubMed: 28859475]

12. Gordon S, Akopyan G, Garban H & Bonavida B Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene 25, 1125–1142 (2006). [PubMed: 16314846]

13. Deng Z, Cao P, Wan MM & Sui G Yin Yang 1: a multifaceted protein beyond a transcription factor. Transcription 1, 81–84 (2010). [PubMed: 21326896]

14. Wu S, et al. A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. Nat. Struct. Mol. Biol 14, 1165–1172 (2007). [PubMed: 18026119]

15. Weintraub AS, et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171, 1573–1588 e1528 (2017). [PubMed: 29224777]

16. Maru Y, Afar DE, Witte ON & Shibuya M The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. J. Biol. Chem 271, 15353–15357 (1996). [PubMed: 8663064]

17. Han FX, Wheelhouse RT & Hurley LH Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. J. Am. Chem. Soc 121, 3561–3570 (1999).

18. Rodriguez R, et al. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J. Am. Chem. Soc 130, 15758–15759 (2008). [PubMed: 18975896]

19. Houbaviy HB, Usheva A, Shenk T & Burley SK Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. PNAS 93, 13577–13582 (1996). [PubMed: 8942976]

20. Whitfield TW, et al. Functional analysis of transcription factor binding sites in human promoters. Genome Biol. 13, R50 (2012). [PubMed: 22951020]

21. Chambers VS, et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol 33, 877–881 (2015). [PubMed: 26192317]

22. Muller S, Kumari S, Rodriguez R & Balasubramanian S Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem 2, 1095–1098 (2010). [PubMed: 21107376]

23. Siddiqui-Jain A, Grand CL, Bearss DJ & Hurley LH Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. USA 99, 11593–11598 (2002). [PubMed: 12195017]

24. Lee GR Role of YY1 in long-range chromosomal interactions regulating Th2 cytokine expression. Transcription 5, e27976 (2014). [PubMed: 25764217]

25. Liu H, et al. Yin Yang 1 is a critical regulator of B-cell development. Genes Dev. 21, 1179–1189 (2007). [PubMed: 17504937]

26. Mumbach MR, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods 13, 919–922 (2016). [PubMed: 27643841]

27. Mohaghegh P, Karow JK, Brosh RM Jr., Bohr VA & Hickson ID The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res. 29, 2843–2849 (2001). [PubMed: 11433031]

28. Lopez-Perrote A, et al. Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. J. Biol. Chem 289, 22614–22629 (2014). [PubMed: 24990942]

29. Shi Y, Seto E, Chang LS & Shenk T Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell 67, 377–388 (1991). [PubMed: 1655281]

30. Merkenschlager M & Nora EP CTCF and cohesin in genome folding and transcriptional gene regulation. Annu. Rev. Genomics Hum. Genet 17, 17–43 (2016). [PubMed: 27089971]

31. Tang Z, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell 163, 1611–1627 (2015). [PubMed: 26686651]

32. Parelho V, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422–433 (2008). [PubMed: 18237772]

33. Khachigian LM The Yin and Yang of YY1 in tumor growth and suppression. Int. J. Cancer 143, 460–465 (2018). [PubMed: 29322514]

34. Shi J, Hao A, Zhang Q & Sui G The role of YY1 in oncogenesis and its potential as a drug target in cancer therapies. Curr. Cancer Drug Targets 15, 145–157 (2015). [PubMed: 25817371]

35. Rhodes D & Lipps HJ G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 43, 8627–8637 (2015). [PubMed: 26350216]

36. Favicchio R, Dragan AI, Kneale GG & Read CM Fluorescence spectroscopy and anisotropy in the analysis of DNA-protein interactions. Methods Mol. Biol 543, 589–611 (2009). [PubMed: 19378154]

37. Li L, Miao W, Huang M, Williams P & Wang Y Integrated genomic and proteomic analyses reveal novel mechanisms of the methyltransferase SETD2 in renal cell carcinoma development. Mol. Cell. Proteomics 18, 437–447 (2019).

38. Feng J, Liu T, Qin B, Zhang Y & Liu XS Identifying ChIP-seq enrichment using MACS. Nat Protoc. 7, 1728–1740 (2012). [PubMed: 22936215]

39. Kent WJ, et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002). [PubMed: 12045153]

40. Matsuda K, et al. ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. Development 144, 1948–1958 (2017). [PubMed: 28455373]

41. Li H Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997 (2013).

42. Wolff J, et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 46, W11–W16 (2018). [PubMed: 29901812]

43. Gertz J, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol. Cell 52, 25–36 (2013) [PubMed: 24076218]
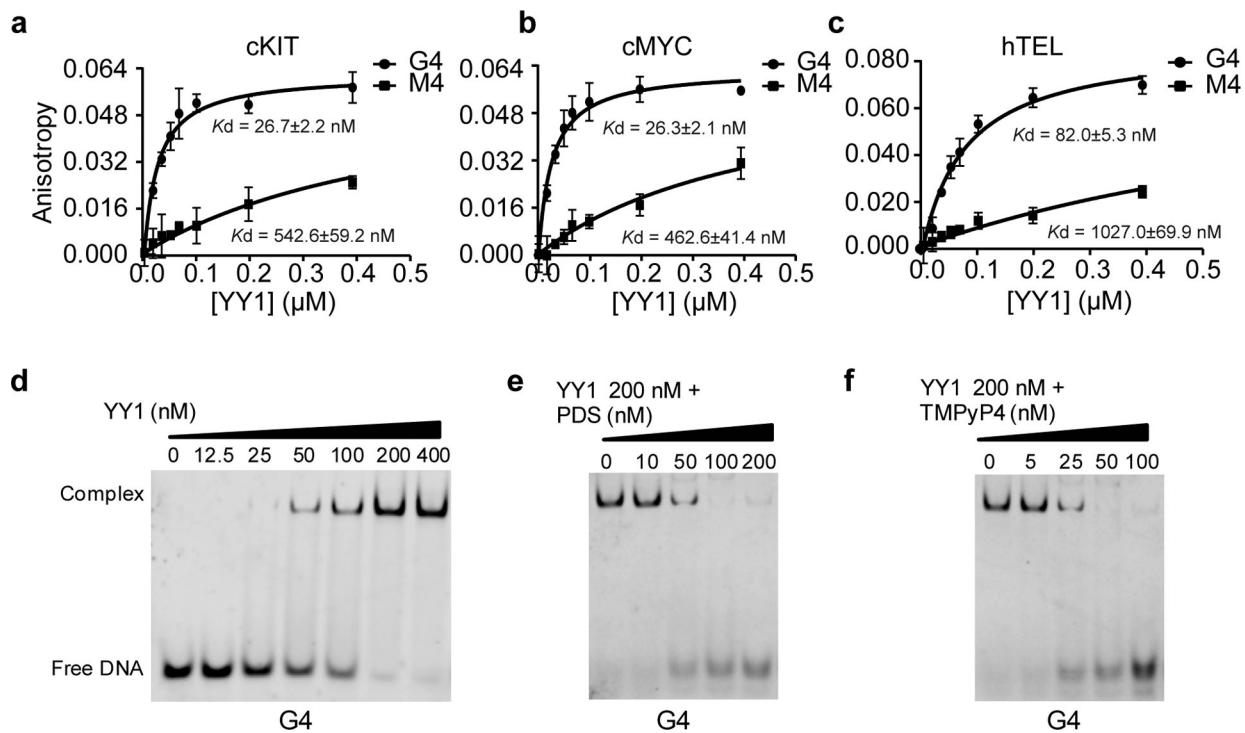
Fig. 1.
Tag-free YY1 binds selectively with G quadruplex DNA. **a-c,** Fluorescence anisotropy for monitoring the binding of YY1 protein with G4 structures and the corresponding mutated sequences derived from the promoters of *KIT* (**a**) and *MYC* (**b**) genes, or the human telomere (**c**). The data in (**a-c**) represent mean ± S.E.M. from 3 independent experiments. **d-f,** EMSA results showing the binding of YY1 towards G4 structure without treatment (**d**) or upon treatment with different concentrations of PDS (**e**) or TMPyP4 (**f**). The EMSA experiments were performed independently for three times with similar results.
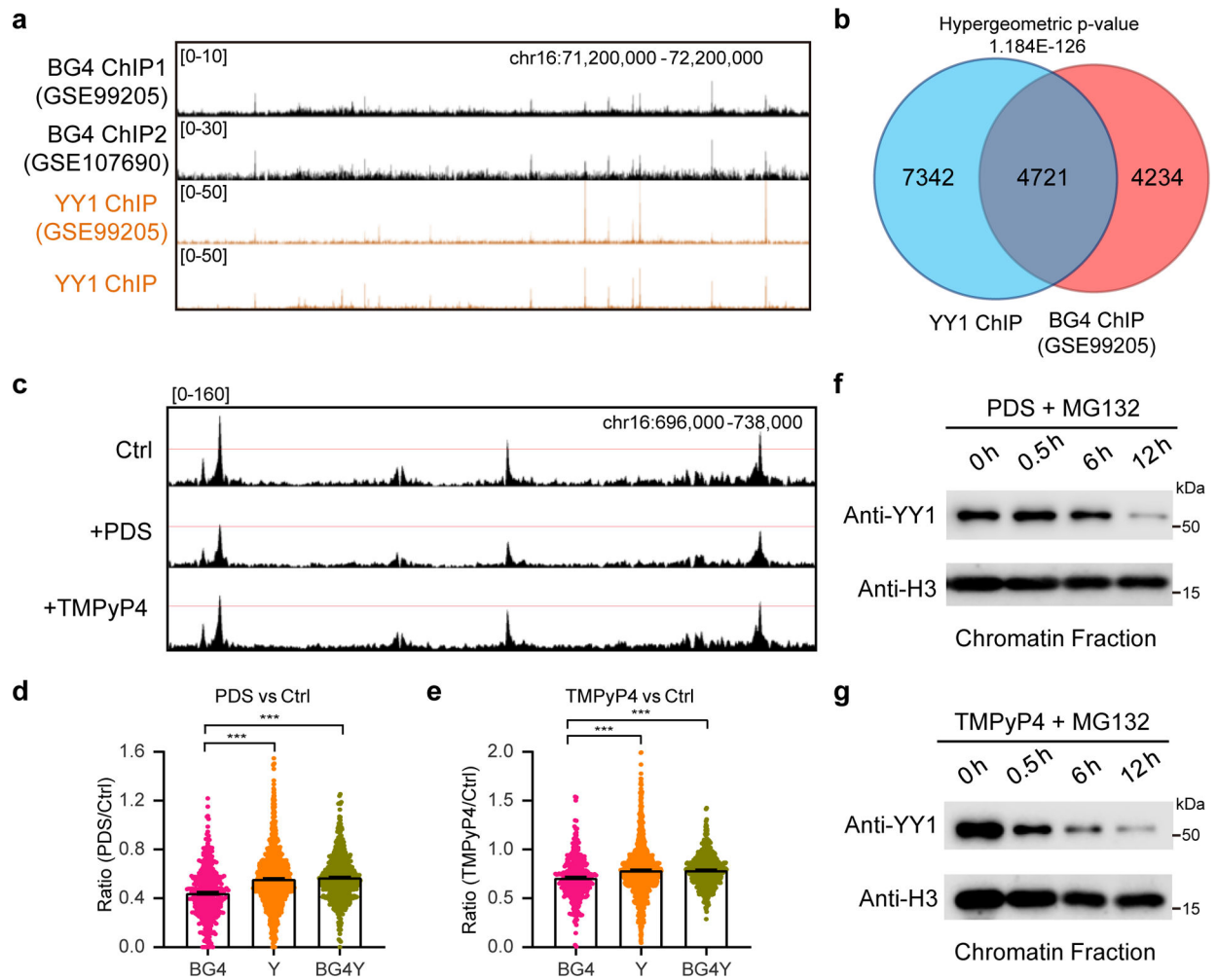
Fig. 2.
YY1 interacts with G4 structure in cells. **a,** Comparisons of YY1 ChIP-Seq data with BG4 ChIP-Seq data. **b,** A Venn diagram showing the overlap between YY1 ChIP-Seq and BG4 ChIP-Seq peaks. **c,** Comparison of YY1 ChIP-Seq data acquired for cells with or without PDS/TMPyP4 treatment. The *p* values were calculated by using the hypergeometric test. **d-e,** Enrichment ratios of YY1 ChIP-Seq peaks containing different motifs after treatment with PDS or TMPyP4. BG4, peaks overlapped with BG4 ChIP-Seq peaks; Y, peaks containing YY1 consensus motif; BG4Y, peaks containing both YY1 consensus motif and overlapped with BG4 ChIP-Seq. The data in (**d-e**) represent mean ± S.E.M. In **d**, n = 370, 1194 and 638 for BG4, Y, and BG4Y, respectively; In **e**, n = 378, 1198 and 639 for BG4, Y and BG4Y, respectively. In **d-e**, the two-tailed Student's *t* test was used to calculate the *p* values. ***, *p* < 0.001. *p* = 3.1E-15 for Y, and *p* = 1.6E-20 for BG4Y in **d**; *p* = 2.0E-6 for Y, and *p* = 5.0E-11 for BG4Y in **e**. **f-g,** Chromatin localization of YY1 after PDS (**f**) or TMPyP4 (**g**) treatment, where the experiments were performed independently for three times with similar results.
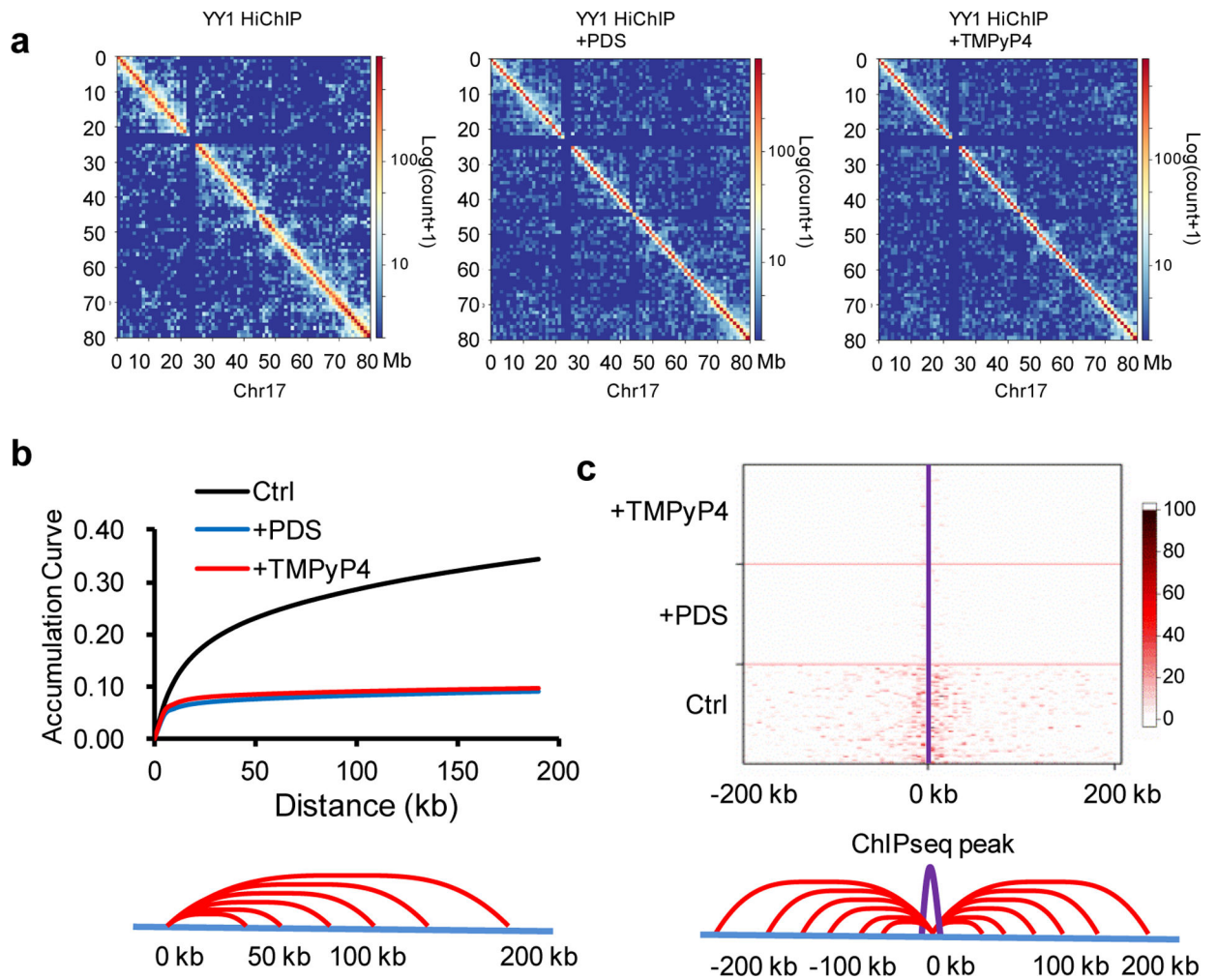
Fig. 3.

Disruption of YY1's binding with G4 structure attenuates YY1-mediated DNA looping. **a,** HiChIP interaction map of chromosome 17 in 293T cells that were untreated (left), or treated with PDS (middle) or TMPyP4 (right). **b,** Statistical analysis of HiChIP interaction distance of 5–200 kb after PDS or TMPyP4 treatment. **c,** A heat map depicting the interactions between YY1 ChIP-Seq peaks and the surrounding region (−200 kb to 200 kb) after PDS or TMPyP4 treatment.
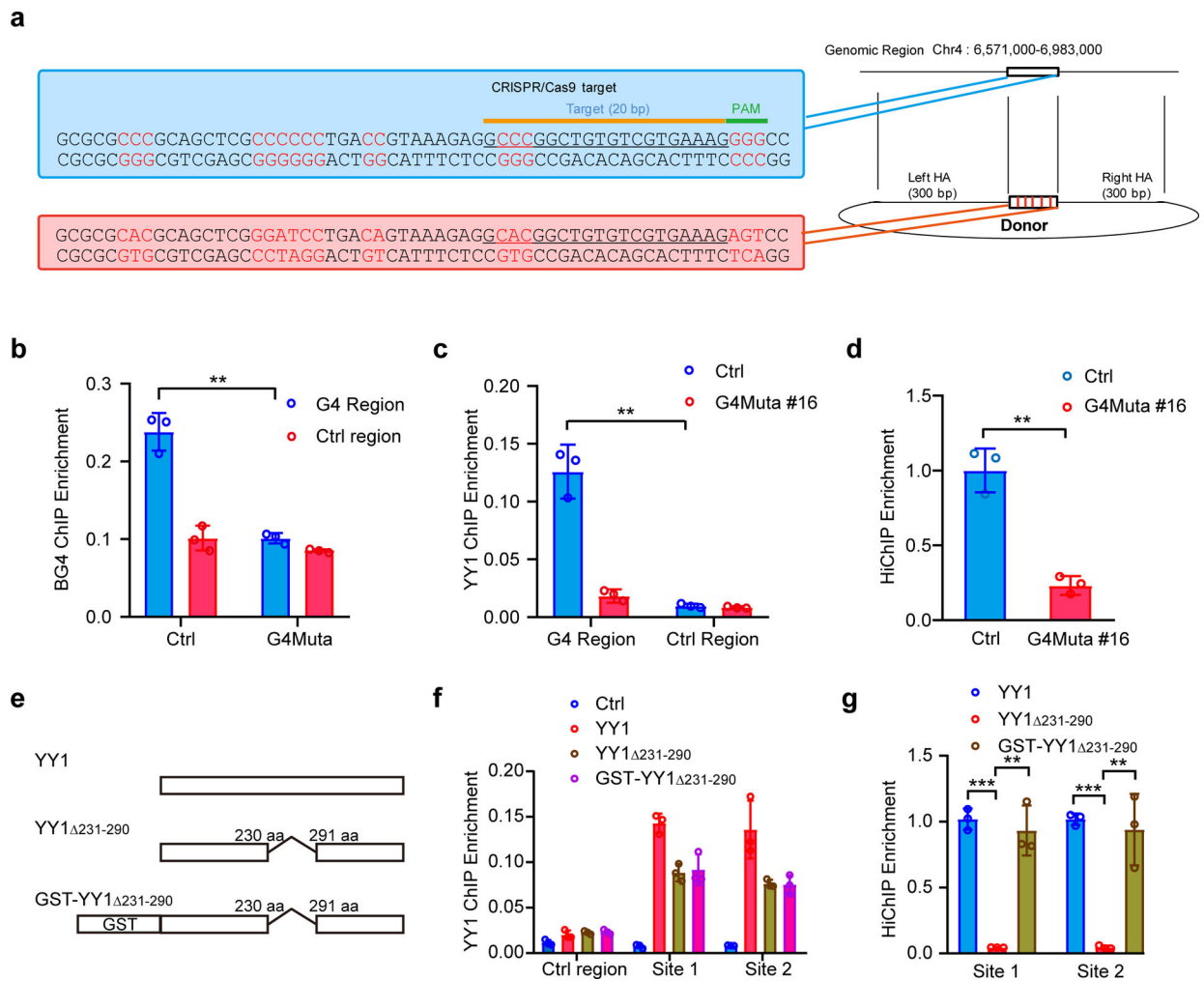
Fig. 4.

YY1-G4 interaction and dimerization of YY1 promote long-range DNA looping. **a,** A schematic diagram showing the design of CRISPR/Cas9-based editing of an endogenous G4-forming sequence. **b-c,** BG4 ChIP and YY1 ChIP enrichments of the site were substantially attenuated after mutation of the G4 sequence. **d,** YY1-mediated DNA looping was disrupted by mutation of the G4 sequence. Shown are the HiChIP-qPCR quantification results of YY1-mediated DNA looping in HEK293T cells with or without mutation of G4 sequence. **e,** A schematic diagram depicting the truncated YY1 protein that is defective in dimerization. **f,** ChIP enrichments for YY1 $_{231-290}$ and GST-YY1 $_{231-290}$ at Sites 1 and 2. Sites 1 and 2 are the same sites as described in Extended Data Fig. 5. **g,** HiChIP-qPCR quantification results of YY1 $_{231-290}$- and GST-YY1 $_{231-290}$-mediated DNA looping. The data represent mean ± S.E.M. of results obtained from 3 independent experiments. The p values were calculated by using two-tailed Student's t test: **, $p < 0.01$. ***, $p < 0.001$. $p = 0.0012$ in **b**; $p = 0.0015$ in **c**; $p = 0.0011$ in **d**; $p = 0.000031$ and 0.0013 for YY1 and GST-YY1 $_{231-290}$, respectively, for site 1, and $p = 0.0000034$ and $p = 0.0047$ for YY1 and GST-YY1 $_{231-290}$, respectively, for site 2 in **g**.
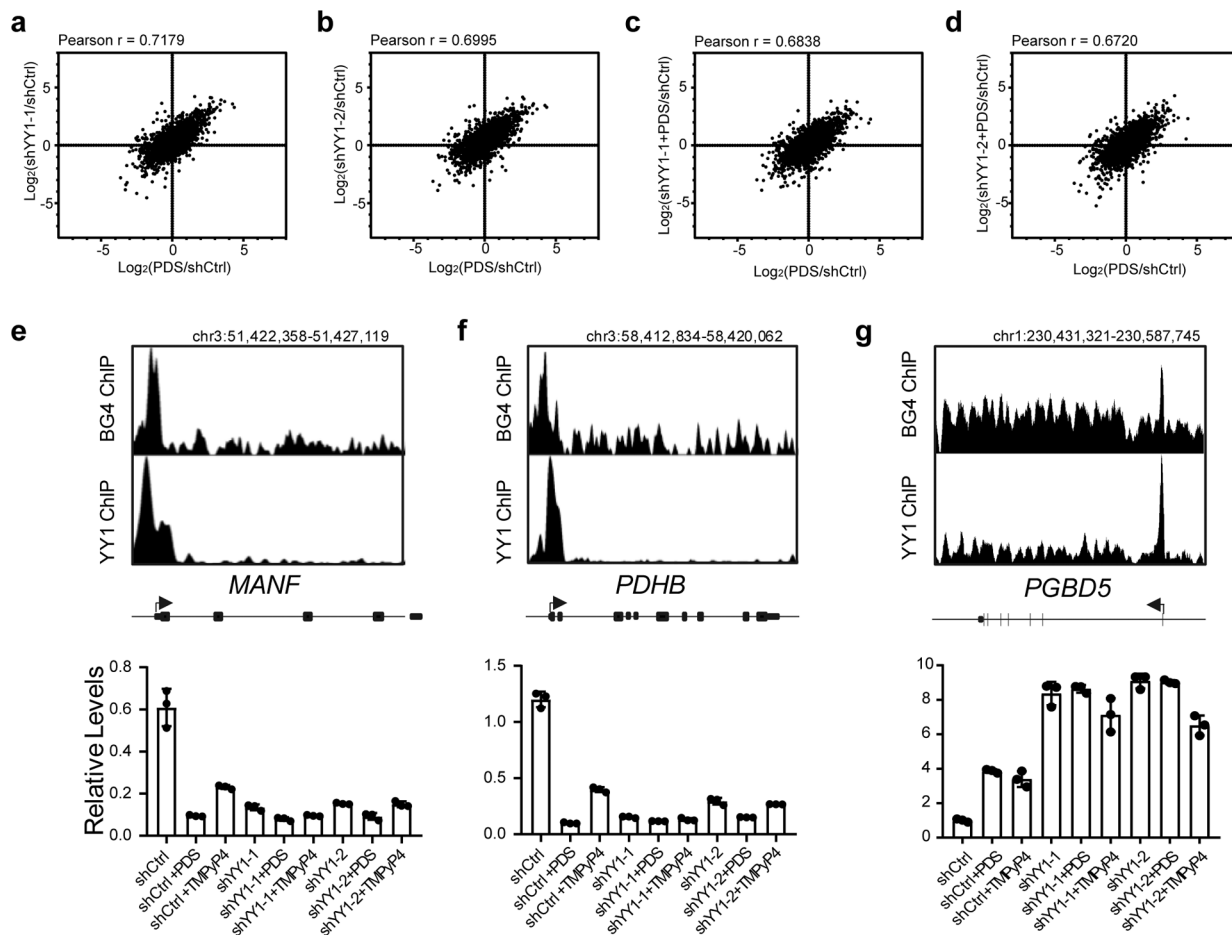
Fig. 5.
Regulation of gene expression by YY1-promoter G4 interactions. **a-d,** Dot plot showing the correlation of gene expression between PDS-treated and shYY1-treated or shYY1+PDS-treated cells. Shown are average ratios of RNA-seq results from two biological replicates. n = 14707 genes for Pearson correlation calculation in **a-d**. **e-g** Quantification of mRNA expression levels of *MANF* (**e**), *PDHB* (**f**) and *PGBD5* (**g**) genes in HEK293T cells with shRNA-mediated knockdown of YY1 and/or after PDS/TMPyP4 treatment. Top of each panel shows read enrichments from BG4 ChIP-Seq and YY1 ChIP-Seq. The data represent mean ± S.E.M. of results from three independent experiments.
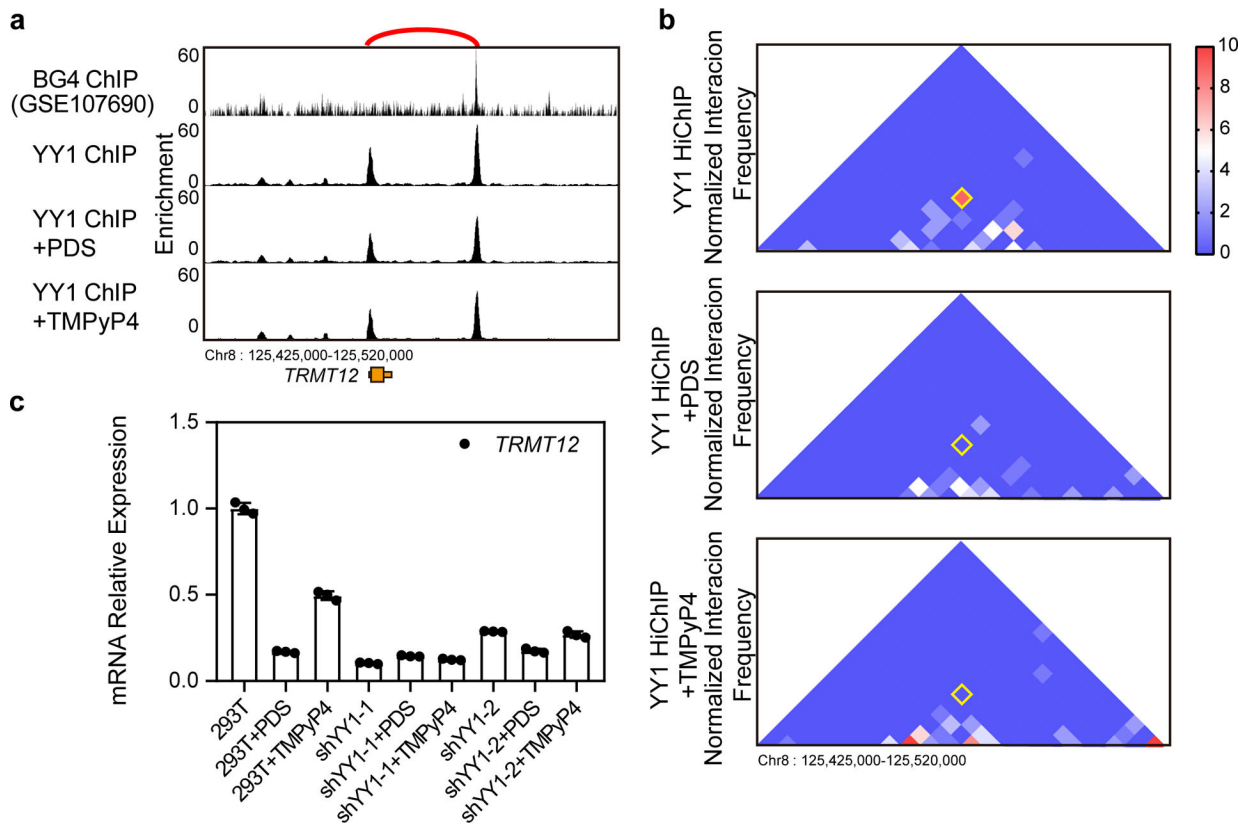
Fig. 6.
YY1-G4 binding participates in transcription regulation of the *TRMT12* gene through long-range DNA looping. **a,** Read enrichments obtained from BG4 ChIP-seq and YY1 ChIP-Seq results. **b,** Normalized interaction frequency involving the sites linked with a red arch in **a** in HEK293T cells with or without PDS or TMPyP4 treatment, as captured by YY1 HiChIP. **c,** Quantification of mRNA expression levels of *TRMT12* in HEK293T cells after knockdown of YY1 and/or with PDS/TMPyP4 treatment. The data represent mean ± S.E.M. of results from three independent experiments.