The RGG Domain of the Yeast Nucleolin Nsr1 Is Required for the Genome Instability Associated with Co-transcriptionally Formed G4 DNA

Shivani Singh[1], Alexandra Berroyer[1,2], and Nayun Kim[1,2,*]

[1]Department of Microbiology and Molecular Genetics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA

[2]MD Anderson Cancer Center UT Health Graduate School of Biomedical Sciences, Houston, TX, 77030, USA

*To whom correspondence should be addressed. Tel: 01-713-500-5597; Fax: 01-713-500-5499; Email: nayun.kim@uth.tmc.edu

ABSTRACT

A significant increase in genome instability is associated with the conformational shift of a guanine-run-containing DNA strand into the four-stranded G4 DNA. Until recently, the mechanism underlying the recombination and genome rearrangements following the formation of G4 DNA *in vivo* has been difficult to elucidate but has become better clarified by the identification and functional characterization of several key G4 DNA-binding proteins. Mammalian nucleolin NCL is a highly specific G4 DNA-binding protein with a well-defined role in the transcriptional regulation of genes with associated G4 DNA-forming sequence motifs at their promoters. The consequence of the *in vivo* interaction between G4 DNA and nucleolin in respect to the genome instability has not been previously investigated. We show here that G4 DNA-binding is a conserved function in the yeast nucleolin Nsr1. Furthermore, we demonstrate that the Nsr1-G4 DNA complex formation is a major factor in inducing the genome instability associated with the co-transcriptionally formed G4 DNA in the yeast genome. The G4-associated genome instability and the G4 DNA-binding *in vivo* requires the arginine-glycine-glycine (RGG) repeats located at the C-terminus of the Nsr1 protein. Nsr1 with the deletion of RGG domain supports normal cell growth and is sufficient in *in vitro* G4 DNA binding as well as protection from the cytotoxicity of a G4 ligand TMPyP4. However, this deletion results in abrogation of *in vivo* binding to the co-transcriptionally formed G4 DNA and in severe reduction in the G4-assicated genome instability. Our data suggest that the interaction between Nsr1 with intact RGG repeats and G4 DNA impairs genome stability by precluding the access of G4-resolving proteins and obstructing replication.

1. Introduction

G-quadruplexes or G4 DNAs are unique non-canonical four-stranded DNA structures that can form from guanine-rich nucleic acid sequences [1, 2]. Four guanine molecules interact *via* Hoogsteen bonds to yield G-quartets that in turn stack on top of each other, held together by intervening loops of variable length and sequence to form the G4 DNA. The size and orientation of the loops can determine the relative stability of various G4 DNA configurations [3]. Though some computational analyses revealed that >375,000 and >1,400 G4 motifs in the human and the *Saccharomyces cerevisiae* genomes, respectively, have the quadruplex-forming potential (QFP), it is still unclear how many of these QFP sequences actually assume the stable G4 configuration *in vivo* [4, 5]. Nevertheless, it is becoming quite evident that these structural transformations potentially provide a substantial informational capacity to DNA with specific functions [1]. Recent bioinformatic studies in human, yeast and bacterial genomes have revealed that G4 motifs are not randomly located within genomes, but particularly enriched in certain functional regions, such as those in ribosomal and telomeric DNA, G-rich micro- and minisatellites and mitotic and meiotic double-strand break (DSB) sites [6-8]. Their high prevalence in and around genes particularly proximal to transcriptional start sites (TSSs) suggests a possible role as *cis*-acting regulatory elements [9-11].

The presence of G4 DNA was recently shown to impair DNA replication by hindering the progress of replicative polymerases, and the failure to resolve these structures transforms the sequence motifs into potential hotspots for genomic instability [12]. In human, the occurrence of G4 DNA motifs reportedly overlap with recombination-prone regions such as certain proto-oncogenes and the sites of frequent translocation breakpoints [13, 14]. Consistently, chromosomal translocations in the proximity of G4 motifs have been observed in a subset of leukemias and lymphomas [15]. The potential G4 DNAs additionally have been correlated with a number of human neurological diseases, such as the frontotemporal dementia (FTD), amyotrophic lateral sclerosis, Alzheimer's and fragile X syndrome [16-18]. The resolution of G4 DNA structures is thus imperative in preserving genome integrity.

The genetic tractability of *S. cerevisiae* has provided considerable insight into the mechanisms involved in maintaining stability of G-rich repetitive sequences. The instability at the G-rich human minisatellite CEB1 inserted in the *S. cerevisiae* genome was shown to be dependent on the ability of the CEB1 motif to form G4 DNA and was not observed with other tandem repeats lacking G4 DNA-forming potential [19, 20]. The G4-forming sequence derived from the guanine-run containing immunoglobulin switch Mu (Sμ) region becomes highly unstable when actively transcribed in the context of the yeast genome. Transcription conferred a critical strand bias since genome rearrangements at Sμ were elevated only when the guanine-runs were located on the non-transcribed strand [21]. The direction of replication and transcription, when in a head-on orientation, further elevated genome instability at the Sμ sequence [22]. Also at the Sμ sequence, the lack of functional topoisomerase 1 (Top1) significantly elevated various types of genome instability, likely by facilitating the structural transition of a G-rich sequence to a G4 structure due to the accumulation of negative helical stress in DNA. The loss of heterozygosity and copy number alterations (deletions and duplications), both of which are frequently observed in cancer genomes, are also elevated when the Sμ sequence was actively transcribed.

The biological functions of G4 DNA are largely dependent on the protein factors that modulate the G4-conformation and/or serve as a bridge to recruit additional protein regulators [23]. These G4 binding proteins can be classified into three functional groups: 1) telomere-related proteins, such as the shelterin complex, human CST (CTC1-STN1-TEN1), and yeast

RAP1 and Est1 [24-26]; 2) proteins that unfold and/or process the G4 structure, such as the helicases including RecQ family helicases hBLM, hWRN, ySgs1, and yPif1 [27]; and 3) proteins that stabilize G4 structures including MAZ and nucleophosmin [28, 29]. Mutations in some of these G4-interacting proteins have been linked to genetic diseases such as Werner syndrome, Fanconi anemia, and cancer [27, 30-32]. A defective BLM helicase failed to unwind G4-DNA and caused increasing recombination frequencies and a high incidence of cancer in Bloom's syndrome [33]. More recently, the co-transcriptional activator Sub1, which interacts with both G4 DNA and the G4-helicase Pif1, was shown to suppress the G4-associated genome instability by facilitating the recruitment of Pif1 helicase to co-transcriptionally formed G4 DNA structures [34].

The human Nucleolin (NCL) is a highly abundant and conserved nucleolar phosphoprotein. Its major function is in the ribosomal RNA maturation with additional roles in chromatin remodeling, transcription, and apoptosis [35]. The altered expression and subcellular localization of NCL is a common biomarker of a variety of cancers demonstrating its clinical relevance [3, 36-38]. Although initially described as a G4-RNA binding protein, recent evidences indicate that NCL preferentially and selectively binds both endogenous and exogenous G-rich sequences that can fold into G4 DNA. It is suggested that NCL acts as a chaperone to promote the correct folding of complex nucleic acid structures [39-41]. Together with the nuclear riboprotein hnRNPD, NCL forms a lymphocyte-specific complex LR1 (lipopolysaccharide responsive factor 1), which binds at the G4 DNA-forming Immunoglobulin heavy chain (IgH) switch regions [42]. The binding of NCL to the G4-forming hexanucleotide repeat expansion (HRE) $(GGGGCC)_n$ in *C9orf72* has been reported to be responsible for the initiation of molecular cascades that lead to neurodegenerative diseases [17]. NCL is composed of three main structural domains; the amino-terminal domain containing four acidic stretches was shown to induce chromatin decondensation through interaction with histone H1 [43], while the central region containing tandem RNA binding domains (RBDs) and the multiple RGG (arginine/glycine/glycine) boxes at the C-terminal domain contribute to its high-affinity interaction with G4 DNA [44, 45].

Similar to the human protein, the yeast Nucleolin Nsr1 has been demonstrated to be involved in pre-rRNA processing and ribosome production [46, 47]. While there are no significant sequence similarities in the amino-terminal regions between yNsr1 and hNCL, the C-terminal half of yNsr1, consisting of two tandemly repeated putative RNA binding domains (RBDs) and the multiple RGG (arginine/glycine/glycine) motifs, has a high sequence similarity to the carboxyl-terminal part of hNCL (37% identity in 249 amino acids) [46, 48]. Unlike the plethora of studies highlighting the NCL/G4 interaction, the role of yeast Nsr1 related to G4 DNA has not been extensively studied. In the current study, we examined whether the yeast Nsr1 has a role in the structural flux of G4-forming sequences. We identified an important biological role of Nsr1 in enhancing various types of genome rearrangements associated with the co-transcriptionally formed G4 DNA. We show that the yeast Nsr1 can bind G4 DNA both *in vitro* and *in vivo* and that the disruption of Nsr1 lowers the G4-associated recombination, while its overexpression further exacerbates instability in a dose-dependent manner in cells lacking Top1. Cells deficient in Nsr1 showed delayed growth recovery following treatment with the G4 ligand TMPyP4. The C-terminal RGG domain of Nsr1 is required to form a complex with G4 DNA *in vivo* and to promote genetic instability, but is dispensable for the resistance to TMPyP4. Our results point to an important role of Nsr1 in G4-associated genome maintenance.

2. Materials and methods

2.1. Yeast strains and plasmids

Yeast strains used for the mutation and recombination assays were derived from YPH45 (*MATa, ura3-52 ade2-101 trp1Δ1*). Construction of strains containing the *pTET-lys2-GTOP or –GBTM* constructs were previously described [21]. Gene deletions were carried out via the one-step allele replacement by amplification of loxP-flanked marker cassettes. Nsr1-expression plasmid was constructed by amplifying NSR1 ORF along with 490 nt upstream and 250 nt downstream from the yeast genomic DNA and cloning into the yeast *CEN* vector pRS316. The deletion constructs, N-Term Nsr1 and C-term Nsr1, have been previously described [49]. The expression plasmid for human Nucleolin was constructed by amplifying the NCL ORF from GFP-Nucleolin (Addgene; #28176) using primers NCL-For (CGCGGATCCATGGTGAAGCTCGCGAAGGCAG) and NCL-1XHA Rev (CCGCTCGAGCTAAGCGTAATCTGGAACATCGTATGGGTATTCAAACTTCGTCTTCTTTCC). The BamHI/ XhoI digested NCL PCR product was cloned into BamHI/ XhoI digested pGPD2 (Addgene; #43972). The expression of NCL was checked by anti HA-HRP antibody (Sigma; H6533).

2.2. Determination of rates

Recombination rates and 95% confidence intervals were determined using the method of the median as described previously [22]. Twelve to twenty-four individual cultures were used to determine each rate and the associated 95% confidence intervals. Recombination rates are considered to be statistically different when the 95% confidence intervals, which are indicated in each graph as error bars, do not overlap. For the gross chromosomal rearrangement (GCR) assay, 5 ml cultures in YEPD medium (1% yeast extract, 2% Bacto-peptone, 2% dextrose, and 250 µg/mL adenine hemisulfate) were inoculated with single colonies and grown for 3 days at 30°C. Cells were then plated either on YEPD-agar or synthetic complete dextrose medium lacking arginine (SCD-arg) and containing canavanine (60 mg/L) and 5-Fluoroorotic acid (5-FOA; 1 g/L) [50].

2.3. Western blotting

Whole cell extracts (WCE) prepared using NaOH lysis method were separated on a 4-20% polyacrylamide gel (BioRad), and transferred using a Semi-Dry Trans-Blot Cell (BioRad). After blocking, anti-FLAG (Sigma, A8592) and anti-GAPDH (Invitrogen, MA5-15738) antibodies were used for detection. The blot was developed using West-Q Pico Dura ECL Solution (GenDepot) and the ChemiDoc MP Imaging system (BioRad). Rad53 phosphorylation was checked using Rad53 antibody (ab104232) from Abcam.

2.4. Electrophoretic mobility shift assays (EMSA)

Yeast extract preparation was carried out according to a previously described protocol [51]. For binding assays, 5´ Cy5 labeled oligos (Sigma) were resuspended in either 100 mM KCl/10 mM Tris–Cl or 100 mM LiCl/10 mM Tris–Cl, denatured at 95°C for 5 min and finally slowly cooled overnight to allow folding. Unlabeled, cold competitor oligos were treated in parallel to the labeled oligos. For binding reactions, approximately 10–15 µg of protein from each cell lysate was incubated with annealed oligos (0.2 µM) in 100 mM KCl or LiCl, 20 mM HEPES pH 7.5, 0.01% NP40, 5% glycerol, 5 mM $MgCl_2$ and incubated at RT for 15 min before running on 15% non-denaturing TBE-polyacrylamide gel. Gel images were captured using BioRad Chemidoc imager.

2.5. DNA-Dynabeads Affinity Purification of Proteins

The DNA-Dynabeads affinity purification of proteins was carried out according to the previously described protocol [52] with certain modifications. For DNA-conjugated Dynabeads preparation,

biotinylated oligonucleotides Smu-G, BRCA1-G1, Smu-C, and SS-DNA were ordered from Sigma. All oligos were incubated at 60°C overnight in the presence of 10 mM Tris pH 7.5 and 100 mM KCl and then conjugated to Streptavidin-Coupled M-280 Dynabeads® (Life Technologies) as per the manufacturer's instructions. Yeast extract was made by glass bead-mediated cell disruption in 2 ml of lysis buffer (50 mM HEPES–NaOH, pH 7.5, 300 mM KCl, 1 mM EDTA, 10% glycerol, 0.05% NP-40, 1 mM DTT, 1 mM PMSF, 1X protease inhibitor cocktail (Roche)). Following the mechanical lysis of cells with Biospec Mini-bead-beater, the cell lysate was collected in a 15 ml tube and sonicated. DNA-conjugated Dynabeads were washed once with the lysis buffer and incubated overnight with gentle inversion with the remaining yeast extract at 4 °C. The beads were washed with the lysis buffer five times and then eluted by boiling in 1XSDS-PAGE loading buffer followed by immunoblotting analysis using anti-FLAG (Sigma; Cat# A8592) or anti-Nsr1 (Clone 31C4; ThermoScientific) antibodies.

2.6. Chromatin immunoprecipitation

For Chromatin IP (ChIP), a previously described protocol was used [34] with following modifications. Anti-FLAG antibody (clone M2 - Sigma) conjugated to Protein G Dynabeads (Life Technologies) were used in pull-down. Ct values for each ChIP samples were first normalized to the corresponding input samples, and then divided by the values for the *CAN1* locus to calculate the relative fold enrichment. Primers used in the qPCR analysis were previously described [34]. P-values were calculated using student t-test.

2.7. TMPyP4 sensitivity

Six independent colonies from each indicated strain were inoculated in liquid YEPD and grown overnight at 30°C. The next day, cells were rediluted to 2×10^6 cells/ml and were grown to early log phase ($O.D._{600}$ of 0.4–0.5) in YEPD. TMPyP4 (EMD Millipore) to 150 µM final concentration or DMSO was added to each culture and grown for additional 20 h. Cells were collected by centrifugation and washed twice with water. Ten-fold serial dilutions of washed cells were spotted onto YEPD-agar plates and incubated at 30°C and pictures were taken every day for 3 days after spotting.

3. Results

3.1. Nsr1 enhances G4-mediated recombination in cells lacking Topoisomerase 1

To understand the elevated genome instability associated with G4-forming sequences, we previously developed a recombination reporter assay in the yeast model system [21]. In this reporter assay, a model G4-forming sequence from the murine immunoglobulin heavy chain switch Mu region (Sµ) is inserted into the yeast genome within the *LYS2* gene under the control of a tetracycline/doxycycline-repressible promoter (*pTET*). The Sµ sequence was inserted either in the physiological (GTOP) or in the inverted orientation (GBTM), placing the G-runs in the non-transcribed strand (NTS) or in the transcribed strand (TS) respectively. The formation of G4 DNA is favored when the G-rich strand located on the NTS is transformed into a single stranded state during transcription, freeing the guanine bases to interact with each other through Hoogsteen base-pairing. When G-runs are located in the TS, they will be occupied in base pairing with the nascent RNA strand and will not be free to fold into G4 DNA. Thus, any factors involved in the formation or stability of G4 DNA should affect the recombination occurring at the *pTET-lys2-GTOP* construct, with little to no effect on the rate of recombination at the *pTET-lys2-GBTM* construct [21, 22, 34].

To determine whether Nsr1 plays a role in the G4-mediated genome instability, we deleted the *NSR1* gene in strains containing the *pTET-lys2-GTOP or -GBTM* construct and checked the rate of recombination at this locus. In the *WT* background, the deletion of *NSR1* resulted in a slow growth phenotype of the cells as previously reported [47]. Under high transcription conditions, the rates of recombination at the *pTET-lys2-GTOP* or *-GBTM* cassette in *nsr1*Δ strains were not changed from those in *WT* strains (Figure 1A).

Topoisomerase1 was previously identified to be a crucial factor for G4 mediated genome instability [21, 22, 53]. Accumulation of negative supercoils in a *top1Δ* background was shown to promote recombination at the *pTET-lys2-GTOP* construct. In order to determine whether Nsr1 modulates G4-associated genomic instability in the absence of Top1, we deleted the *NSR1* gene in a *top1Δ* background. Under low transcription conditions, there were no statistically significant changes in the rates of recombination for both the *pTET-lys2-GBTM* and –*GTOP* constructs. However, under high transcription conditions, the rate of recombination for the *pTET-lys2-GTOP* was reduced by ~5-fold in the *top1Δ nsr1Δ* strain compared to *top1Δ.* For the *pTET-lys2-GBTM*, the rates of recombination were indistinguishable between *top1Δ* and *top1Δ nsr1Δ* backgrounds (Figure 1B). When the transcription from the *pTET* promoter was repressed by the addition of 2μg/ml doxycycline in the media, the deletion of NSR1 did not affect the rates of recombination in WT or *top1*Δ background (Figure S1).

The catalytically inactive mutant of Top1 (Top1-Y727F) with the mutation of catalytic tyrosine to phenylalanine results in complete ablation of its function in removal of supercoils, while not affecting its DNA binding activity [54]. Possibly due to its high G4-binding ability, the expression of Top1-Y727F results in recombination rates that are ~4-fold higher than in the absence of Top1 (*top1*Δ) [53]. When Nsr1 was disrupted in the *top1Y727F* backgrounds, the recombination rate at the *pTET-lys2-GTOP* cassette was significantly decreased and was comparable to the rate in the *top1*Δ strain (Figure 1C). The disruption of Nsr1 did not affect the rate of recombination at the *pTET-lys2-GBTM* cassette.

3.2. Nsr1 does not affect the G4-associated recombination in RNase H-deficient cells.

Disrupting both RNase H1 and RNase H2 in yeast leads to the accumulation of transcription-associated RNA:DNA hybrids or R-loops [55] and subsequently elevates levels of recombination for both the *pTET-lys2-GTOP and -GBTM* constructs [21]. The G loop, a higher order structure consisting of R-loop and G4 DNA, could arise from either the enhanced negative supercoils leading to G4 formation upon disruption of Top1 or due to the failure to degrade RNA leading to R-loops upon disruption of RNase Hs. Therefore, we checked whether, in addition to reducing recombination rates in a *top1*Δ strain, the deletion of Nsr1 could affect the R-loop mediated enhanced recombination rates. In the triple deletion mutants *rnh1*Δ *rnh201*Δ *nsr1*Δ*,* there was no significant alteration in the rate of recombination for the *pTET-lys2-GTOP or -GBTM* construct (Figure 1D). This data suggests that Nsr1 induced genomic instability is specific to G4 DNA formed due to the enhanced negative helical torsion under high transcription and not due to the RNA: DNA hybrid accumulation.

3.3. Nsr1 promotes gross chromosomal rearrangements in the *top1*Δ backgrounds.

We previously reported that, in the absence of functional Top1, the co-transcriptionally formed G4 DNA at the *pTET-lys2-GTOP* construct leads to the increase in ectopic recombination as well as in the gross chromosomal rearrangements (GCRs) [22]. In order to measure the rates of GCR, we used a modified form of the GCR reporter system developed by Chen and Kolodner [50]. In this reporter system, the *URA3* gene was integrated into the left arm of chromosome V (CHR5) replacing the *HXT13* gene located ~8.5 kb centromere-distal to the *CAN1* gene. The *pTET-lys2-GTOP* or *-GBTM* cassette containing the Sμ G4 motif, as described before [22], is

integrated immediately centromere-proximal to *CAN1.* The loss of functional *CAN1* or *URA3* results in resistance to the drug canavanine (Can) or 5-Fluoroorotic acid (5-FOA), respectively. Thus, by simultaneous selection against two counter-selectable markers (*URA3* and *CAN1*), complex genome rearrangement can be measured. Using this modified GCR assay, it was shown that in the high transcription conditions, the disruption of Top1 leads to a significantly higher (~30-fold) GCR rate for *pTET-lys2–GTOP*, where guanine-runs are present on the NTS, compared to the *pTET-lys2–GBTM* construct where guanine-runs are on the TS. When we deleted *NSR1* in WT backgrounds, there was no change in the GCR rates for *pTET-lys2–GTOP* or *-GBTM* (Figure 2A). However, the disruption of Nsr1 in *top1Δ* backgrounds resulted in a significant, ~6-fold reduction of the GCR rates for the *pTET-lys2-GTOP* but not for *-GBTM* construct (Figure 2B). We further assessed whether the function of Nsr1 is specific to the Sμ G4-mediated genome instability or extends to other G4 motifs as well. In the GCR reporters modified to contain the G4 motif from TCF3 translocation breakpoint [56], *NSR1* deletion by itself did not alter the GCR rates. But, in a similar manner to the GCR reporter containing Sμ G4, when Nsr1 was disrupted in a *top1Δ* strain, the GCR rates were severely lowered by ~9-fold at the *pTET-lys2-GTOP(TCF3)* construct, while at the pTET-*lys2-GBTM(TCF3)* construct the GCR rates were not significantly changed (Figure 2C and 2D). Overall, these data indicate that Nsr1 promotes the instability associated with the co-transcriptionally formed G4 DNA.

3.4. RGG domain is necessary to complement for the loss of Nsr1 in *top1*Δ *nsr1*Δ strain.

For the complementation study, a full length *NSR1* gene construct along with its own promoter was cloned into the centromeric plasmid pRS316. Since the C-terminal RGG domain of Nucleolin has been shown to be important for inducing and stably binding G4 structures [40, 44], we also cloned the N-terminal 1-350 residues of Nsr1 lacking the RGG domain (Nsr1ΔRGG) (Figure 3A). The expression of both full length Nsr1 and Nsr1ΔRGG was confirmed using anti-Nsr1 antibody (Figure S2A). The ectopic expression of full length Nsr1 in the *top1Δ nsr1Δ* double mutant strain enhanced the rate of recombination for the *pTET-lys2-GTOP* by 8.5-fold, thereby resulting in the rates of recombination that were similar to a *top1Δ* transformed with the empty vector (Figure 3B and see Figure 4B). However, the expression of Nsr1ΔRGG resulted in the rate of recombination at the *pTET-lys2*-GTOP cassette that was slightly lower than the empty vector (Figure 3B). The rates of recombination of the *pTET-lys2-GBTM* cassette were not affected by either Nsr1 or Nsr1ΔRGG expression.

We next expressed the N-terminal PASK region of Nsr1 (1-171a.a.; N-term Nsr1 in Figure 3A), or the remaining C terminal region (C-Term). The N-term construct was unable to complement for the loss of Nsr1 in enhancing G4-induced recombination, with rates of recombination that overlap with the rates of recombination for the empty vector (Figure 3C). However, the C-terminal deletion construct was able to enhance the rates of recombination albeit not as robustly as the full length Nsr1. The rates of recombination for the C-terminal deletion construct were 1.8-fold higher than the rates of recombination for the empty vector. These results suggest that the RBD and RGG domains at the C-term of Nsr1 are required for promoting G4 associated instability and that the PASK domain also contributes, in part, probably by mediating protein-protein interactions.

To confirm the above results with plasmid expression constructs, we repeated the experiments by generating the C-terminal truncation mutants with the C-terminal 3XFLAG tags at the endogenous *NSR1* locus. As shown in Figure S2B, the protein expression levels were similar for the full length Nsr1, Nsr1 (1-350) and Nsr1 (1-240) as detected by anti-Nsr1 or anti-FLAG antibody. We could not detect protein for the Nsr1 (1-167) allele. We confirmed that the 3XFLAG tag at the C-terminus does not interfere with the function of Nsr1 as the recombination

rate of a *top1*Δ *Nsr1-3XFLAG* strain was similar to the *top1*Δ strain. The rates of recombination at the *pTET-lys2-GTOP* cassette in cells expressing the Nsr1 (1-167), Nsr1 (1-240) and Nsr1 (1-350) were similar to that in cells with the complete deletion of Nsr1 (*nsr1*Δ) (Figure 3D). These results further support the requirement for the RGG domain of Nsr1 in promoting G4-associated instability.

When we expressed the full length Nsr1 and Nsr1ΔRGG in *WT* strains with the normal endogenous level of Nsr1, there was no effect of overexpression of the full length Nsr1 constructs while the Nsr1ΔRGG construct reduced the rates of recombination slightly (Figure 4A). In the *top1*Δ strain, the overexpression of full length Nsr1 elevated the rates of recombination at the *pTET-lys2-GTOP* cassette by ~4.5-fold compared to vector alone (Figure 4B), consistent with the 4-fold higher Nsr1 protein levels in the *top1*Δ strain overexpressing Nsr1 (Figure S2C). When the Nsr1ΔRGG construct was expressed in the *top1*Δ strain, the rates of recombination were consistently lower than the empty vector (Figure 4B). The N-term Nsr1 (1-171) was unable to induce recombination and showed rates of recombination which were similar to the empty vector. The C-term Nsr1 (172 - 414) expression resulted in about 2-fold higher recombination rates (Figure 4C). These results suggest that Nsr1 increases recombination in a dose dependent manner, but this function requires the RGG domain.

We also tested whether the highly conserved human Nucleolin (NCL) can complement the loss of Nsr1 in yeast. C-terminal HA-tagged NCL was stably expressed from a plasmid with the yeast GPD promoter (Figure S2D) and elevated the rates of recombination at the *pTET-lys2-GTOP* cassette by ~2-fold in *top1*Δ *nsr1*Δ cells (Figure 4D). In *top1*Δ cells, the expression of NCL resulted in further 2-fold increase in the recombination at the *pTET-lys2-GTOP* indicating that the function of Nsr1 in mediating the G4-associated genome instability is conserved in the human homolog.

3.5. Nsr1 binds to the G4 DNA *in vitro* and *in vivo.*

In order to test whether the binding of Nsr1 to G4 DNA is important for the observed effect of Nsr1 in the *pTET-lys2-GTOP* and –*GBTM* constructs, we performed a gel-shift assay using the Cy5 labeled 45-mer oligonucleotide partially representing the Sμ-G sequence (SMU-G), which was previously shown to fold into G4 structures [34], and a non-G4 forming single stranded (SS) DNA as a control (Figure 5A). G4 DNA formation by the SMU-G oligo was confirmed by its binding to the G4-specific recombinant antibody BG4 in the presence of the G4-stabilizing K+ cation [57]. Whole cell extracts from *top1*Δ*, top1*Δ *nsr1-3XFLAG,* and *top1*Δ *nsr1*Δ cells were incubated with the SMU-G oligo with or without 100-fold excess of the unlabeled competitor SMU-G or SS DNA. Figure 5B shows a specific complex of protein bound to the SMU-G probe in the presence of 100 mM KCl, which disappeared in the presence of the cold unlabeled SMU-G4 competitor (band indicated by the arrow). This band was not affected by the cold unlabeled SS DNA competitor and was absent in the *top1*Δ *nsr1*Δ control cells. SMU-G oligo folded in Li+ showed a different pattern of bands when incubated with WCE from *top1*Δ cells (Figure S3A). There is a band with similar mobility as the arrow-pointed band in Figure 5B but this band (*) was competed out by the cold SMU-G and SS-DNA indicating that this complex is not dependent on the G4 DNA conformation.

Next, we performed affinity purification using four different biotinylated oligos; G-run containing SMU-G and BRCA1-G1, C-run containing SMU-C, and noG-run containing single stranded (SS)-DNA (Figure 5A). After the denaturing-slow cooling step, each of the oligos were incubated with the whole cell lysates and pulled down using the streptavidin-coupled beads. When the oligo-bound proteins were resolved by Western blot, the G4-specific antibody BG4 and the G4-binding protein Sub1 were shown to bind to SMU-G and BRCA1-G1 oligos (Figure S3B). The BG4 antibody did not show any binding to SMU-C or SS-DNA oligos, while the Sub1

protein showed a weaker interaction with the SS-DNA, and no interaction with the SMU-C oligo, validating the applicability of this approach to identify G4-binding proteins. When the four oligos were incubated with WCE from the Nsr1-expressing cells, SMU-G and BRCA-G1 oligos, but not SMU-C or SS-DNA oligos pulled down Nsr1 (Figure 5C). When we used the yeast extract from *top1*Δ *nsr1*Δ*RGG* and *top1*Δ *nsr1*Δ*RBD2,*Δ*RGG* strains, SMU-G and BRCA1-G1 but not SMU-C or SS-DNA were able to pull-down both Nsr1ΔRGG and Nsr1ΔRBD2,ΔRGG (Figure 5C). This is consistent with the previous observation that human NCL has additional G4 binding domains apart from the RGG domain. The full length Nsr1, Nsr1ΔRBD2ΔRGG, and Nsr1ΔRGG were also pulled down by the anti-NCL aptamer AS1411 (Figure S3C).

To determine whether the *in vitro* interaction of Nsr1 with G4 structures can be extended to *in vivo*, we measured the association of Nsr1 and G4 by using chromatin immuno-precipitation (ChIP) assay. ChIP was performed in WT and *top1Δ* cells containing either the *pTET-lys2-GTOP* or –*GBTM* construct and expressing Nsr1 with a C-terminal 3XFlag tag. Following the pull-down with αFlag antibody-conjugated beads, the qPCR analysis was done to determine the enrichment of Nsr1 at a locus 100 bp from the G4 insertion site ("G4 locus") and a locus ~3 kb away from the G4 insertion site ("3 Kb away") (Figure 5D). We observed ~2-fold enrichment of Nsr1 at the G4 locus when the G4 sequence was in the *GTOP* orientation, but not in the *GBTM* orientation in a *top1Δ* background ($p<0.0001$). Further, the enrichment of Nsr1 was significantly higher at the switch region G4 sequence insertion site than the 3' region of the *lys2* sequence 3 kb away ($p=0.0021$). Nsr1ΔRGG, which is missing the C-terminal RGG domain, however, was not significantly enriched at the G4 locus in *top1*Δ cells ($p=0.0009$ compared to full length Nsr1). These data indicate that Nsr1 specifically associates with the G4 DNA accumulated under high transcription conditions, particularly in the absence of Top1 and that this association between G4 and Nsr1 requires the RGG repeat at the C-terminus of Nsr1.

3.6. Nsr1-disruption allows for the better access by activation-induced deaminase to the single stranded DNA in the G4-prone sequence.

Activation-induced deaminase (AID) converts cytosine to uracil and is required for the somatic hypermutation (SHM) of immunoglobulin variable region genes as well as the heavy chain class switch recombination (CSR) [58]. We previously showed that the ectopic expression of human AID in the WT and *top1*Δ yeast cells resulted in a higher rate of recombination at the *pTET-lys2-GBTM* than -*GTOP* in both strain backgrounds [21]. This difference was partly attributed to the larger number of cytosines present on the single-stranded non-transcribed strand (NTS) in the *GBTM* orientation compared to the *GTOP* orientation (358 vs. 117 cytosines) since AID specifically targets the cytosines located on the single-stranded DNA. However, we postulated that NSR1-G4 DNA complex could further impede the accessibility of AID to those cytosines on the NTS in the *pTET-lys2-GTOP* construct located proximal to the G4 structure. We tested whether the cytosines present on the NTS in the GTOP-orientation could become more accessible to AID in the absence of Nsr1 protein. When we expressed hAID, the rate of recombination at the *pTET-lys2-GBTM* cassette was elevated by about 4- to 5-fold in the *WT*, *top1*Δ*,* and *top1*Δ *nsr1*Δ strains. For the *pTET-lys2-GTOP* construct*,* the rate of recombination was not changed by the expression of hAID in WT and *top1*Δ strains, but increased by about 3-fold in a *top1*Δ *nsr1*Δ strain (Table 1). This increased rate of recombination could reflect the less-restricted access of hAID to the cytosines in the absence of G4 DNA-Nsr1 complex on the NTS.

4. Discussion

Although initially described as a sequence-specific RNA-binding protein, human nucleolin (NCL) preferentially binds to G-quadruplex structures over RNA substrates and plays a crucial role in G4 metabolism [17, 39, 40]. In this study, we have explored the possibility that the yeast homolog Nsr1 also contributes to G4 DNA metabolism. As expected from its highly conserved RBD and RGG domains, we demonstrated that the physical association with G4 DNA is a conserved function in Nsr1. We performed pull-down experiments with oligonucleotides to show that multiple oligos with previously confirmed G4-forming properties (e.g. Smu-G, AS1411, and BRCA1-G1 in Figure 5A and S3C) associated with Nsr1 in yeast whole cell extract, thus demonstrating that Nsr1 is a general G4 DNA binding protein. No binding was observed when non-G4-forming oligos were used in the pull-down in further support of the specificity in the binding of G4 DNA by Nsr1 protein. Nsr1 is also significantly enriched *in vivo* at the G4 reporter construct only when the guanines are on the top, non-transcribed strand of an actively transcribed gene (i.e. *pTET-lys2-GTOP*) in a *top1Δ* background, which is the condition with the significant elevation of G4-associated recombination (Figure 5D). Nsr1-enrichment is significantly reduced ~3 kb away from the G4 motif sequence even though it is within the same transcribed unit, which further supports the specificity of Nsr1 interaction with G4 DNA. Although the *in vitro* pull-down assay showed that Nsr1ΔRGG can bind G4 DNA forming oligos, ChIP experiment showed that *in vivo* binding of Nsr1 to a G4 DNA-containing genomic locus is dependent on the presence of the RGG domain.

The disruption of Nsr1 substantially decreases the G4-associated genomic instability as manifested by the decreased rates of recombination and of GCR in *top1*Δ background by about 4- and 7-fold, respectively (Figures 1 and 2). The reduction in genome instability in the absence of Nsr1 was not specific to the switch region G-rich sequence as GCRs occurring at a different G-rich motif from the human TCF3 gene were also reduced upon deletion of *NSR1* (Figure 2D). Additionally, overexpression of Nsr1 led to further elevation of recombination specifically at the *pTET-lys2-GTOP* cassette in a *top1*Δ background (Figure 4). One interpretation of these surprising results is that the consequence of the association between Nsr1 and G4 DNA is the elevated genome instability. The genome instability associated with G4-structures should be prevented by the activity of G4-unwinding DNA helicases including yeast Sgs1 and Pif1 [59, 60]. We considered the possibility that Nsr1 can occlude DNA helicases from recognizing and then resolving G4 DNA. When we expressed hAID to induce the cytosine deamination and subsequent recombination, those cytosines present on the NTS at the G4 DNA-forming *pTET-lys2-GTOP* cassette were more susceptible in a *top1*Δ *nsr1*Δ background than in a *top1*Δ background, supporting the idea that the regions proximal to the G4 DNA are in a less accessible conformation in the presence of Nsr1. The inaccessibility generated by the association of Nsr1 with G4 DNA could obstruct the movement of replication leading to genome instability. Non-histone protein-DNA complexes forming a replication block have been previously reported for the origin of replication (ORC) complex and Rap1 [61]. Further work will be needed to confirm that the Nsr1-G4 DNA complexes create strong replication barriers in the yeast genome.

In the mammalian systems, NCL was shown to localize to the site of DNA double strand breaks (DSB) and to reorganize nucleosomes proximal to DSBs [62]. Since this role of NCL is critical in the early step of DSB repair by homologous recombination (HR), there is a possibility that the reduction in the recombination rate at the *pTET-lys2-GTOP* cassette in *nsr1*Δ yeast cells is due to the general defect in HR. This alternative explanation for the role of Nsr1 is contradicted by the fact that the *NSR1*-deletion did not affect the recombination rate of the *pTET-lys2-GBTM* construct in *top1*Δ background or the recombination rates of either constructs in a *rnh1*Δ *rnh201*Δ background (Figure 1). Also, the deletion of *NSR1* led to the reduction in GCR events (Figure 2), which normally occur through the HR-independent, *de novo* telomere

addition in haploid yeast cells [63]. Finally, in order to rule out any defect in DNA damage checkpoint response, we checked the phosphorylation state of Rad53 in the WT, *top1*Δ, *nsr1*Δ, and *top1*Δ*nsr1*Δ cells treated with the DSB-inducing drug Zeocin. After 90 mins with Zeocin, Rad53 phosphorylation was apparent in all cell types indicating that the DNA damage checkpoint response was not affected by the disruption of Nsr1 (Figure S4).

Located C-terminal to the multiple RNA binding domains (RBDs), yeast Nsr1 and human NCL contain 6 and 9 repeats of arginine-glycine-glycine or RGG motifs, respectively, which are found in proteins associated with important nucleic acid-mediated functions such as transcription, pre-mRNA splicing, DNA damage signaling and mRNA translation [64]. Other proteins with verified roles in DNA metabolism and cancer development that contain RGG motifs include Mre11 (a DSB processing enzyme), Mll4 (a histone methyltransferase), and Ews (Ewings Sarcoma protein; DNA damage response protein). Recently, the RGG-box of hnRNPA1, a member of ribonucleoproteins, was shown to bind specifically to the telomeric G4 DNA [44]. Earlier biochemical analysis showed that the RGG domain of hNCL alone can bind G4 DNA with high affinity [45]. In yeast Nsr1, RGG motifs also appear to play an important function. When we deleted the RGG domain from Nsr1, the resulting Nsr1ΔRGG construct lost the ability to complement the full-length Nsr1 in promoting G4-associated genome instability at the *pTET-lys2-GTOP* cassette either in *top1*Δ *nsr1*Δ or *top1*Δ background (Figure 3B and 4B). However, this truncation does not affect the cell growth as the complete deletion of Nsr1 (Figure S5A) or eliminate its ability to bind the G4-forming oligos Smu-G, BRCA1-G1, and AS1411 as shown in our pull-down assay (Figure 5 and S3C). Also, the sensitivity to the G4 ligand TMPyP4, which was acutely elevated in *nsr1*Δ, is not elevated in cells expressing Nsr1ΔRGG (Figure S5B). These results suggest that there is a significant difference in terms of stability and/or conformation between the G4 DNA complexed with full-length Nsr1 and Nsr1ΔRGG that would require further biophysical approaches to identify.

The elevation of genomic instability by Nsr1 and its physical association with G4 DNA suggests that Nsr1 binding leads to stabilization of the co-transcriptionally formed G4 DNA and that the Nsr1/G4 DNA forms a nucleoprotein complex that acts as a replication block. Once bound to the G4 DNA, NCL can not only act as molecular chaperone for the G4 folding but also self-associate to create a higher order complex [65]. The self-association function of NCL, as well as interaction with other proteins, is dependent on the C-terminal domain rich in RGG residues [40, 44]. Interestingly, when overexpressed from a centromeric plasmid, Nsr1 protein lacking the RGG domain (Nsr1ΔRGG) did not elevate but rather slightly reduced the rate of recombination at the *pTET-lys2-GTOP* cassette in a *top1*Δ background (Figure 4B). This unexpected result could be attributed to the possible interaction of its RBD domains with the RGG domain of the native Nsr1 protein, sequestering the native Nsr1 and thereby acting as a dominant negative mutant. The RBDs 3 and 4 and the RGG domain of NCL play a critical role in the formation of G-quadruplex structure and is essential for G-quadruplex stabilization [40]. However, the rate of recombination at the *pTET-lys2-GTOP* cassette in cells expressing these RGG/RBD-deletion constructs are reduced to a similar degree as in the *nsr1* null cells (Figure S2), strongly pointing to the absolute requirement of RGG domain in Nsr1-mediated promotion of genomic instability at G4-forming sequences.

Acceleration of nucleic acid annealing by NCL depends on the self-association of multiple NCL molecules bound to nucleic acids. The nucleic acid annealing activity of NCL is localized to the two RNA binding domains (RBD3,4) and the RGG9 domain. This region is also responsible for NCL self-association [65]. The self-association of G4 DNA-bound Nsr1 molecules probably makes it a stable nucleoprotein complex. This is supported by our overexpression data wherein the increase in genome instability of a *top1*Δ strain was dependent on the dose of Nsr1 protein possibly due to the increase in the number of Nsr1 molecules

binding each G4 DNA molecule. Our *in vivo* data suggests that Nsr1 protein lacking the RGG domain might fail to promote the transcriptional-dependent annealing of G-rich sequences into a G4 structure and that the lack of RGG repeats hampers the self-association of G4 structure bound Nsr1 molecules.

In summary, we have identified a novel function of yeast Nsr1 as a G4 DNA binding protein. Nsr1 shows strong *in vitro* binding to G4 forming sequences and is specifically enriched at a co-transcriptionally formed G4 DNA *in vivo.* Moreover, we identified Nsr1 as a factor interacting with anti-NCL aptamer and modulating TMPyP4 sensitivity in yeast cells. The key finding of this work is that the interaction between Nsr1 and co-transcriptionally-formed G4 DNA results in a significant elevation in genome instability at this locus. The data presented here point to the requirement of the conserved RGG domain at the C-terminal end of Nsr1 in promoting the instability at G4 DNA. This result calls for further study into the conformational changes associated with the RGG domain in the Nsr1-G4 DNA complex that could enable such a complex to function as a replication obstacle.

Conflict of interest statement

The authors declare that there are no conflicts of interests.

Funding

This work was supported by grants from National Institutes of Health R01 GM116007 and AU1875 from Welch Foundation to N.K. A.B. is supported by R01 GM116007-S1.

Acknowldegments

Authors would like to thank Dr. A. Saiardi of University College London, London UK for the gift of plasmids with the C-term and N-term Nsr1 constructs. We also thank the members of Kim lab for discussion and critical reading of the manuscript.

References

[1] N. Maizels, L.T. Gray, The G4 genome, PLoS Genet, 9 (2013) e1003468.
[2] M.L. Bochman, K. Paeschke, V.A. Zakian, DNA secondary structures: stability and function of G-quadruplex structures, Nat Rev Genet, 13 (2012) 770-780.
[3] N. Kim, The interplay between G-quadruplex and Transcription, Curr Med Chem, (2017).
[4] J.A. Capra, K. Paeschke, M. Singh, V.A. Zakian, G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae, PLoS Comput Biol, 6 (2010) e1000861.
[5] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res, 33 (2005) 2908-2916.
[6] J. Eddy, N. Maizels, Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes, Nucleic Acids Res, 36 (2008) 1321-1333.
[7] P. Rawal, V.B. Kummarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S.K. Das, S. Chowdhury, Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation, Genome Res, 16 (2006) 644-655.
[8] A.K. Todd, M. Johnston, S. Neidle, Highly prevalent putative quadruplex sequence motifs in human DNA, Nucleic Acids Res, 33 (2005) 2901-2907.
[9] Z. Du, Y. Zhao, N. Li, Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription, Genome Res, 18 (2008) 233-241.
[10] S.G. Hershman, Q. Chen, J.Y. Lee, M.L. Kozak, P. Yue, L.S. Wang, F.B. Johnson, Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae, Nucleic Acids Res, 36 (2008) 144-156.
[11] J.F. Moruno-Manchon, E.C. Koellhoffer, J. Gopakumar, S. Hambarde, N. Kim, L.D. McCullough, A.S. Tsvetkov, The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Brca1 in neurons, Aging (Albany NY), 9 (2017) 1957-1970.
[12] P. Sarkies, C. Reams, L.J. Simpson, J.E. Sale, Epigenetic instability due to defective replication of structured DNA, Mol Cell, 40 (2010) 703-713.
[13] J. Seenisamy, E.M. Rezler, T.J. Powell, D. Tye, V. Gokhale, C.S. Joshi, A. Siddiqui-Jain, L.H. Hurley, The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4, J Am Chem Soc, 126 (2004) 8702-8709.
[14] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, 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 U S A, 99 (2002) 11593-11598.
[15] A. Bacolla, J.A. Tainer, K.M. Vasquez, D.N. Cooper, Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences, Nucleic Acids Res, 44 (2016) 5673-5688.
[16] E. Majounie, Y. Abramzon, A.E. Renton, R. Perry, S.S. Bassett, O. Pletnikova, J.C. Troncoso, J. Hardy, A.B. Singleton, B.J. Traynor, Repeat expansion in C9ORF72 in Alzheimer's disease, N Engl J Med, 366 (2012) 283-284.
[17] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A. Simko, P.G. Shaw, M.S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, C9orf72 nucleotide repeat structures initiate molecular cascades of disease, Nature, 507 (2014) 195-200.
[18] N. Maizels, G4-associated human diseases, EMBO Rep, 16 (2015) 910-922.
[19] J. Lopes, A. Piazza, R. Bermejo, B. Kriegsman, A. Colosio, M.P. Teulade-Fichou, M. Foiani, A. Nicolas, G-quadruplex-induced instability during leading-strand replication, EMBO J, 30 (2011) 4033-4046.
[20] A. Piazza, A. Serero, J.B. Boule, P. Legoix-Ne, J. Lopes, A. Nicolas, Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in Saccharomyces cerevisiae depends on G-quadruplexes or Cdc13, PLoS Genet, 8 (2012) e1003033.

[21] N. Kim, S. Jinks-Robertson, Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast, DNA Repair (Amst), 10 (2011) 953-960.
[22] P. Yadav, V. Harcy, J.L. Argueso, M. Dominska, S. Jinks-Robertson, N. Kim, Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed G-quadruplex-forming sequence, PLoS Genet, 10 (2014) e1004839.
[23] V. Brazda, L. Haronikova, J.C. Liao, M. Fojta, DNA and RNA quadruplex-binding proteins, Int J Mol Sci, 15 (2014) 17493-17517.
[24] I.M. Pedroso, W. Hayward, T.M. Fletcher, The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures, Nucleic Acids Res, 37 (2009) 1541-1554.
[25] A. Bhattacharjee, Y. Wang, J. Diao, C.M. Price, Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST, Nucleic Acids Res, 45 (2017) 12311-12324.
[26] Q.J. Li, X.J. Tong, Y.M. Duan, J.Q. Zhou, Characterization of the intramolecular G-quadruplex promoting activity of Est1, FEBS Lett, 587 (2013) 659-665.
[27] O. Mendoza, A. Bourdoncle, J.B. Boule, R.M. Brosh, Jr., J.L. Mergny, G-quadruplexes and helicases, Nucleic Acids Res, 44 (2016) 1989-2006.
[28] S. Cogoi, A.E. Shchekotikhin, L.E. Xodo, HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property, Nucleic Acids Res, 42 (2014) 8379-8388.
[29] A. Gallo, C. Lo Sterzo, M. Mori, A. Di Matteo, I. Bertini, L. Banci, M. Brunori, L. Federici, Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter, J Biol Chem, 287 (2012) 26539-26548.
[30] S.B. Cantor, D.W. Bell, S. Ganesan, E.M. Kass, R. Drapkin, S. Grossman, D.C. Wahrer, D.C. Sgroi, W.S. Lane, D.A. Haber, D.M. Livingston, BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function, Cell, 105 (2001) 149-160.
[31] S. Seal, D. Thompson, A. Renwick, A. Elliott, P. Kelly, R. Barfoot, T. Chagtai, H. Jayatilake, M. Ahmed, K. Spanova, B. North, L. McGuffog, D.G. Evans, D. Eccles, D.F. Easton, M.R. Stratton, N. Rahman, Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles, Nat Genet, 38 (2006) 1239-1241.
[32] Y. Wu, K. Shin-ya, R.M. Brosh, Jr., FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability, Mol Cell Biol, 28 (2008) 4116-4128.
[33] H. Sun, J.K. Karow, I.D. Hickson, N. Maizels, The Bloom's syndrome helicase unwinds G4 DNA, J Biol Chem, 273 (1998) 27587-27592.
[34] C.R. Lopez, S. Singh, S. Hambarde, W.C. Griffin, J. Gao, S. Chib, Y. Yu, G. Ira, K.D. Raney, N. Kim, Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA, Nucleic Acids Res, 45 (2017) 5850-5862.
[35] M.M. Tajrishi, R. Tuteja, N. Tuteja, Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus, Commun Integr Biol, 4 (2011) 267-275.
[36] C.M. Berger, X. Gaume, P. Bouvet, The roles of nucleolin subcellular localization in cancer, Biochimie, 113 (2015) 78-85.
[37] Y. Otake, S. Soundararajan, T.K. Sengupta, E.A. Kio, J.C. Smith, M. Pineda-Roman, R.K. Stuart, E.K. Spicer, D.J. Fernandes, Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA, Blood, 109 (2007) 3069-3075.
[38] Y. Satake, Y. Kuwano, T. Nishikawa, K. Fujita, S. Saijo, M. Itai, H. Tanaka, K. Nishida, K. Rokutan, Nucleolin facilitates nuclear retention of an ultraconserved region containing TRA2beta4 and accelerates colon cancer cell growth, Oncotarget, 9 (2018) 26817-26833.

[39] V. Gonzalez, K. Guo, L. Hurley, D. Sun, Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein, J Biol Chem, 284 (2009) 23622-23635.
[40] V. Gonzalez, L.H. Hurley, The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, Biochemistry, 49 (2010) 9706-9714.
[41] E. Tosoni, I. Frasson, M. Scalabrin, R. Perrone, E. Butovskaya, M. Nadai, G. Palu, D. Fabris, S.N. Richter, Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription, Nucleic Acids Res, 43 (2015) 8884-8897.
[42] L.A. Dempsey, H. Sun, L.A. Hanakahi, N. Maizels, G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination, J Biol Chem, 274 (1999) 1066-1071.
[43] M.S. Erard, P. Belenguer, M. Caizergues-Ferrer, A. Pantaloni, F. Amalric, A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1, Eur J Biochem, 175 (1988) 525-530.
[44] M. Ghosh, M. Singh, RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain, Nucleic Acids Res, (2018).
[45] L.A. Hanakahi, H. Sun, N. Maizels, High affinity interactions of nucleolin with G-G-paired rDNA, J Biol Chem, 274 (1999) 15908-15912.
[46] K. Kondo, M. Inouye, Yeast NSR1 protein that has structural similarity to mammalian nucleolin is involved in pre-rRNA processing, J Biol Chem, 267 (1992) 16252-16258.
[47] W.C. Lee, D. Zabetakis, T. Melese, NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits, Mol Cell Biol, 12 (1992) 3865-3871.
[48] W.C. Lee, Z.X. Xue, T. Melese, The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs, J Cell Biol, 113 (1991) 1-12.
[49] C. Azevedo, T. Livermore, A. Saiardi, Protein polyphosphorylation of lysine residues by inorganic polyphosphate, Mol Cell, 58 (2015) 71-82.
[50] C. Chen, R.D. Kolodner, Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants, Nat Genet, 23 (1999) 81-85.
[51] C.R. Lopez, A. Ribes-Zamora, S.M. Indiviglio, C.L. Williams, S. Haricharan, A.A. Bertuch, Ku must load directly onto the chromosome end in order to mediate its telomeric functions, PLoS Genet, 7 (2011) e1002233.
[52] J. Gao, B.L. Zybailov, A.K. Byrd, W.C. Griffin, S. Chib, S.G. Mackintosh, A.J. Tackett, K.D. Raney, Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA, Chem Commun (Camb), 51 (2015) 7242-7244.
[53] P. Yadav, N. Owiti, N. Kim, The role of topoisomerase I in suppressing genome instability associated with a highly transcribed guanine-rich sequence is not restricted to preventing RNA:DNA hybrid accumulation, Nucleic Acids Res, 44 (2016) 718-729.
[54] M.D. Megonigal, J. Fertala, M.A. Bjornsti, Alterations in the catalytic activity of yeast DNA topoisomerase I result in cell cycle arrest and cell death, J Biol Chem, 272 (1997) 12801-12808.
[55] L. Wahba, J.D. Amon, D. Koshland, M. Vuica-Ross, RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability, Mol Cell, 44 (2011) 978-988.
[56] J.D. Williams, S. Fleetwood, A. Berroyer, N. Kim, E.D. Larson, Sites of instability in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures *in vitro*, Front Genet, 6 (2015) 177.
[57] R. Hansel-Hertsch, D. Beraldi, S.V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, S. Balasubramanian, G-quadruplex structures mark human regulatory chromatin, Nat Genet, 48 (2016) 1267-1272.

[58] J.U. Peled, F.L. Kuang, M.D. Iglesias-Ussel, S. Roa, S.L. Kalis, M.F. Goodman, M.D. Scharff, The biochemistry of somatic hypermutation, Annu Rev Immunol, 26 (2008) 481-511.
[59] M.D. Huber, D.C. Lee, N. Maizels, G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition, Nucleic Acids Res, 30 (2002) 3954-3961.
[60] K. Paeschke, M.L. Bochman, P.D. Garcia, P. Cejka, K.L. Friedman, S.C. Kowalczykowski, V.A. Zakian, Pif1 family helicases suppress genome instability at G-quadruplex motifs, Nature, 497 (2013) 458-462.
[61] A.S. Ivessa, B.A. Lenzmeier, J.B. Bessler, L.K. Goudsouzian, S.L. Schnakenberg, V.A. Zakian, The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes, Mol Cell, 12 (2003) 1525-1536.
[62] M. Goldstein, F.A. Derheimer, J. Tait-Mulder, M.B. Kastan, Nucleolin mediates nucleosome disruption critical for DNA double-strand break repair, Proc Natl Acad Sci U S A, 110 (2013) 16874-16879.
[63] V. Pennaneach, C.D. Putnam, R.D. Kolodner, Chromosome healing by de novo telomere addition in Saccharomyces cerevisiae, Mol Microbiol, 59 (2006) 1357-1368.
[64] P. Thandapani, T.R. O'Connor, T.L. Bailey, S. Richard, Defining the RGG/RG motif, Mol Cell, 50 (2013) 613-623.
[65] L.A. Hanakahi, Z. Bu, N. Maizels, The C-terminal domain of nucleolin accelerates nucleic acid annealing, Biochemistry, 39 (2000) 15493-15499.

Table 1: Effect of hAID expression on G4-associated recombination rates.

	Genotype	Recombination rate x 10^{-8} (CI*)		Fold Change: hAID/Vector
		Empty vector	hAID	
pTET-lys2-GTOP	*WT*	20.4 (15.1-27.9)	31.7 (16.2-60.2)	1.5
	top1Δ	220 (177-322)	219 (137-661)	0.99
	top1Δnsr1Δ	20.6 (13.7-27.8)	54.1 (46.3-65.8)	2.6
pTET-lys2-GBTM	*WT*	12.7 (8.0-17.1)	59.1 (44.7-84.5)	4.6
	top1Δ	11.4 (8.92-15.3)	45.6 (38.8-51.4)	4.0
	top1Δnsr1Δ	10.4 (8.3-15.5)	56.7 (29.7-150)	5.4

* 95% confidence interval

FIGURE LEGENDS

Figure 1. The rates of G4-associated recombination in *nsr1*Δ strains.

Guanine-runs are on the non-transcribed, top strand in single stranded state in *pTET-lys2-GTOP* cassette and on the transcribed strand annealed to the nascent RNA in the *pTET-lys2–GBTM* cassette. All graphs are showing the rates of recombination ($\times 10^{-8}$). Error bars indicate 95% confidence intervals. (**A**) Recombination rates in *WT* backgrounds. (**B**) Recombination rates in *top1*Δ backgrounds. (**C**) Recombination rates in *top1Y727F* backgrounds. (**D**) Recombination rates in *rnh1*Δ *rnh201*Δ backgrounds.

Figure 2. Rates of Gross Chromosomal Rearrangement (GCR) in *nsr1*Δ strains.

All graphs are showing the rates of 5-FOA^r/CAN^r or GCRs ($\times 10^{-8}$). The rates of GCRs occurring at CHR5 containing the *pTET-lys2-GTOP* or *–GBTM* cassette are determined by the method of median. 95% confidence intervals are indicated by the error bars. GCR rates in strains with the reporter cassette containing the Sμ-G4 in (**A**) WT backgrounds (**B**) *top1*Δ backgrounds. GCR rates in strains with the reporter cassette containing TCF3-G4 in (**C**) WT backgrounds (**D**) *top1*Δ backgrounds.

Figure 3. Complementation of NSR1-deletion.

(**A**) Domain organization of Nsr1. The deletion constructs used for complementation are indicated. **(B)** and **(C)** Graphs are showing the rates of recombination ($\times 10^{-8}$). Error bars indicate 95% confidence intervals. The *top1*Δ *nsr1*Δ double mutant was transformed with empty vector (+*Vec*), Nsr1-expression plasmid (+*Nsr1*) or the indicated deletion plasmids followed by fluctuation analysis to determine the rates of recombination. **(D)** The graph shows the rates of recombination ($\times 10^{-8}$) at the *pTET-lys2-GTOP* cassette in the *top1*Δ background with the indicated WT or mutant *NSR1* alleles at its endogenous genomic locus. All *NSR1* alleles include the C-terminal 3XFlag tag. Error bars indicate 95% confidence intervals.

Figure 4. Overexpression of Nsr1 or hNCL.

Overexpression of the full length or various deletion constructs of Nsr1 in **(A)** WT background, and **(B** and **C)** *top1*Δ background. **(D)** The *top1*Δ *nsr1*Δ or *top1*Δ yeast cells were transformed with empty vector (+*Vec*), hNCL-expression plasmid (+*NCL*) followed by fluctuation analysis to determine the rates of recombination. All graphs are showing the rates of recombination ($\times 10^{-8}$). Error bars indicate 95% confidence intervals.

Figure 5. Nsr1 binds to G4 DNA *in vitro and in vivo*.

(**A**) Sequence of the oligos used in EMSA and pulldown assays. **(B)** Electrophoretic mobility shift assay. Cy5-conjugated Smu-G was incubated with WCE containing Nsr1 or without Nsr1, in the presence of 100mM KCl. Where indicated, a cold competitor oligo were added in 100-fold excess of the Cy5-Smu-G. **(C)** For pull-down assays, the indicated, biotin-conjugated oligos were incubated with total yeast lysate from *top1Δ* cells expressing the C-terminal 3XFLAG tagged version of full length Nsr1, Nsr1ΔRGG, or Nsr1ΔRBD2,ΔRGG. αFLAG antibody was used to detect the respective proteins. **(D)** ChIP was carried out with αFlag antibody and chromatin fractions prepared from yeast cells with indicated genetic backgrounds. 5´BGL and STP primers were previously described and anneal within *LYS2* ORF at ~100 and ~3000 bp from the Sμ sequence insertion site, respectively. All values are based on at least six independent samples with the exception of the no-tag samples (N=3) and the standard deviations are indicated by error bars. P-values shown were calculated using student t-test.

Figure 1.

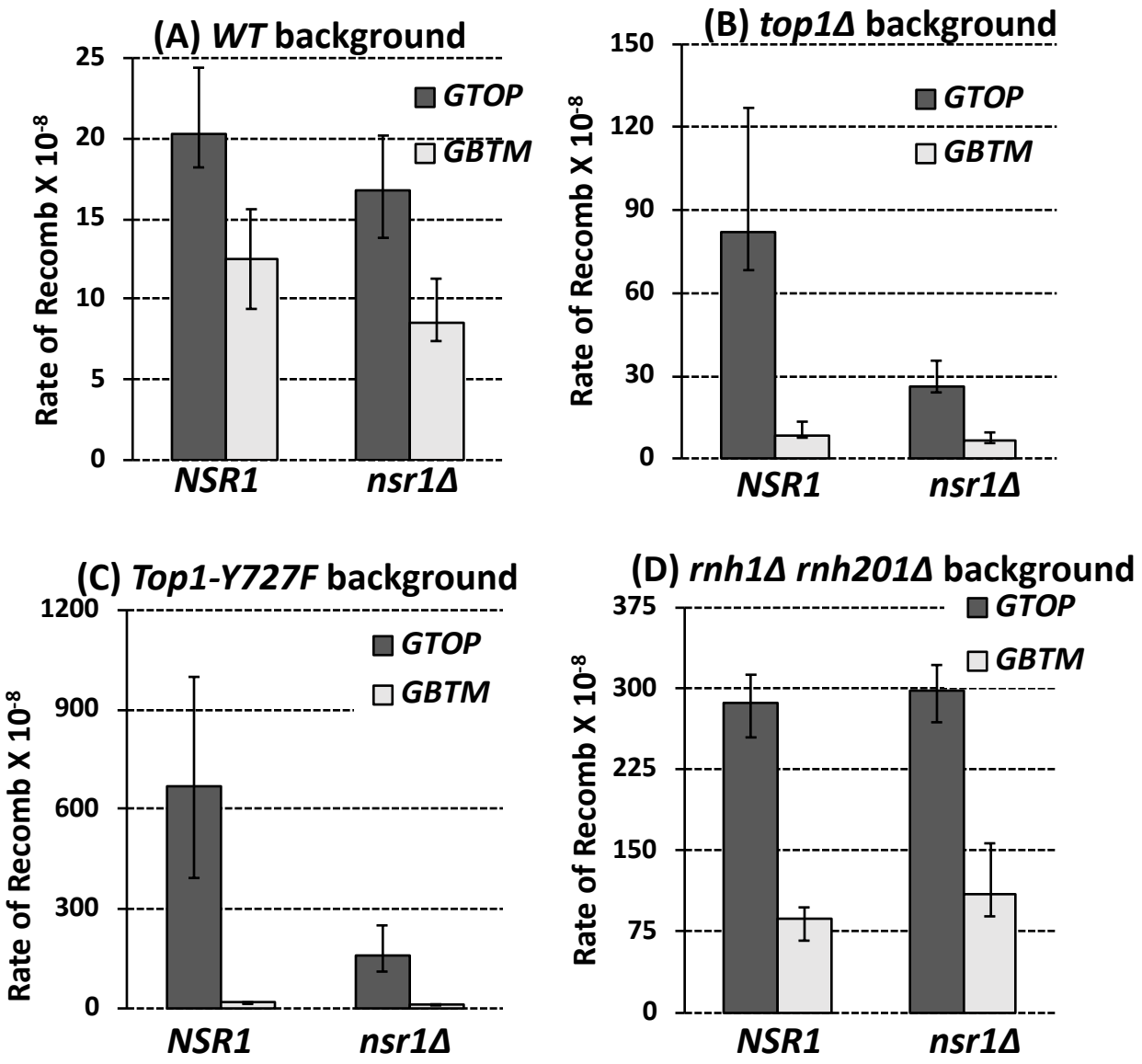

(A) WT background
Rate of Recomb X 10-8
25
20
15
10
5
0
GTOP
GBTM
NSR1
nsr1Δ
(B) top1Δ background
Rate of Recomb X 10-8
150
120
90
60
30
0
GTOP
GBTM
NSR1
nsr1Δ
(C) Top1-Y727F background
Rate of Recomb X 10-8
1200
900
600
300
0
GTOP
GBTM
NSR1
nsr1Δ
(D) rnh1Δ rnh201Δ background
Rate of Recomb X 10-8
375
300
225
150
75
0
GTOP
GBTM
NSR1
nsr1Δ

Figure 2.

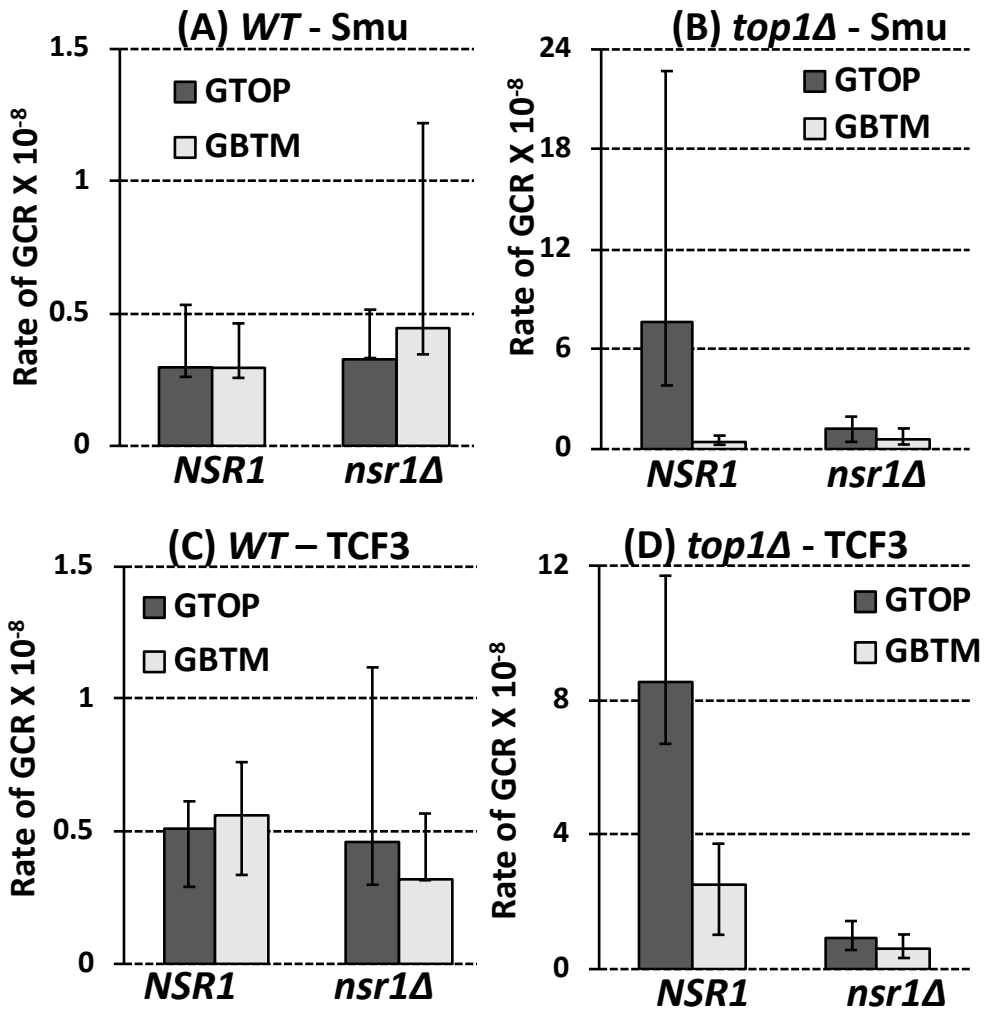

(A) WT - Smu
(B) top1Δ - Smu
(C) WT – TCF3
(D) top1Δ - TCF3
GTOP
GBTM
Rate of GCR X 10-8
NSR1
nsr1Δ
1.5
1
0.5
0
24
18
12
6
0
12
8
4
0

Figure 3.

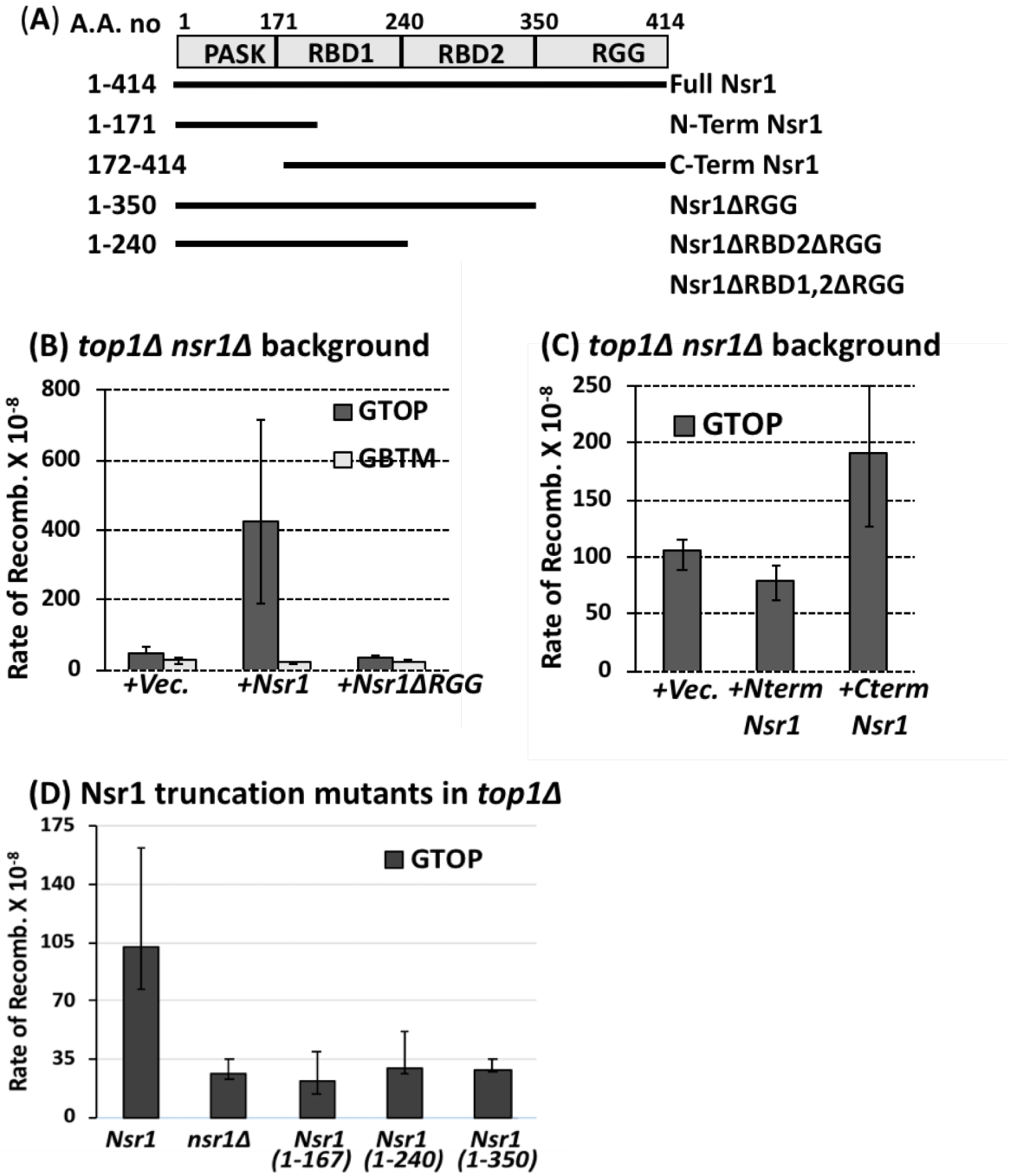

(A) A.A. no
1
171
240
350
414
PASK
RBD1
RBD2
RGG
1-414
Full Nsr1
1-171
N-Term Nsr1
172-414
C-Term Nsr1
1-350
Nsr1ΔRGG
1-240
Nsr1ΔRBD2ΔRGG
Nsr1ΔRBD1,2ΔRGG
(B) top1Δ nsr1Δ background
Rate of Recomb. X 10-8
800
600
400
200
0
GTOP
GBTM
+Vec.
+Nsr1
+Nsr1ΔRGG
(C) top1Δ nsr1Δ background
Rate of Recomb. X 10-8
250
200
150
100
50
0
GTOP
+Vec.
+Nterm Nsr1
+Cterm Nsr1
(D) Nsr1 truncation mutants in top1Δ
Rate of Recomb. X 10-8
175
140
105
70
35
0
GTOP
Nsr1
nsr1Δ
Nsr1 (1-167)
Nsr1 (1-240)
Nsr1 (1-350)

Figure 4

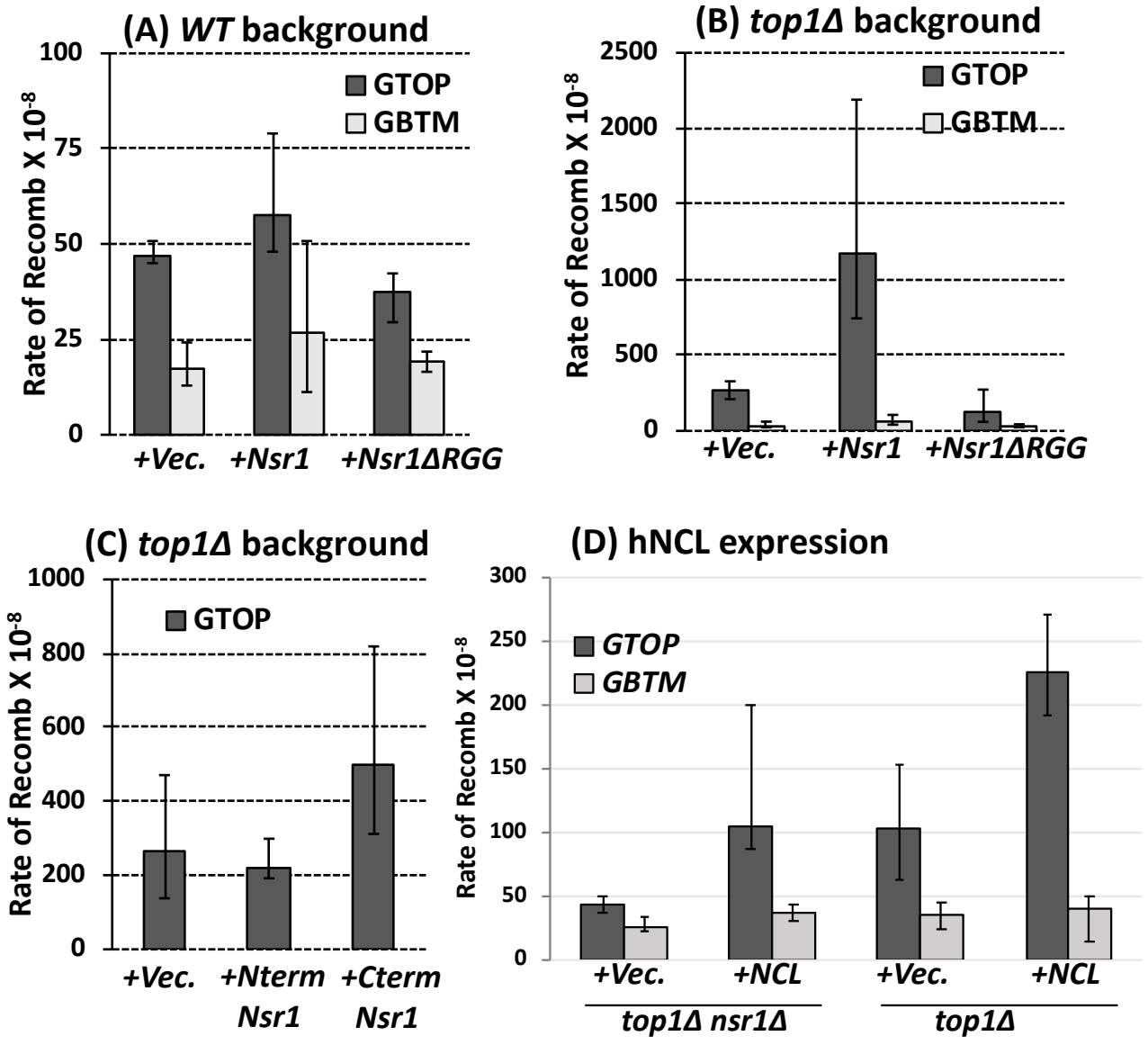
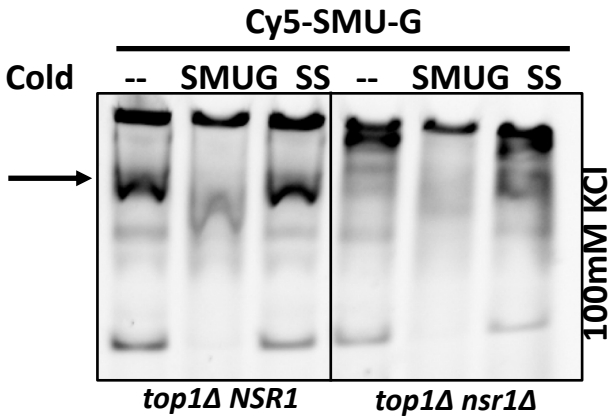
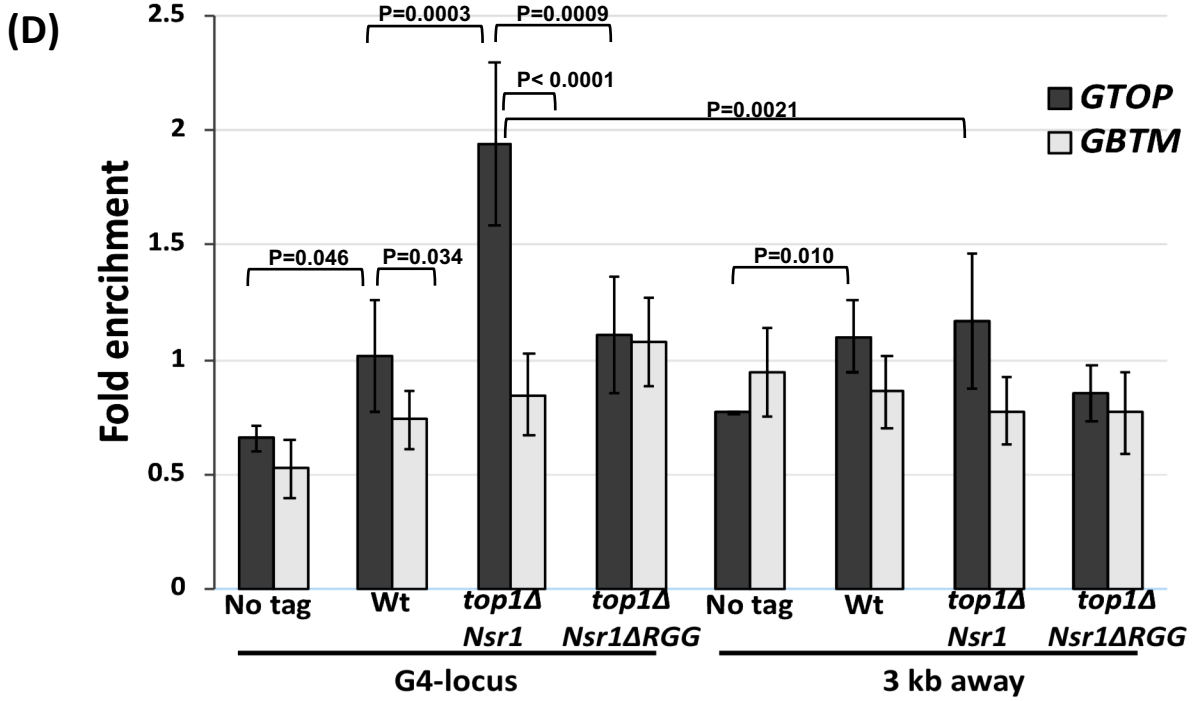

(A) WT background
Rate of Recomb X 10-8
100
75
50
25
0
GTOP
GBTM
+Vec.
+Nsr1
+Nsr1ΔRGG
(B) top1Δ background
Rate of Recomb X 10-8
2500
2000
1500
1000
500
0
GTOP
GBTM
+Vec.
+Nsr1
+Nsr1ΔRGG
(C) top1Δ background
Rate of Recomb X 10-8
1000
800
600
400
200
0
GTOP
+Vec.
+Nterm Nsr1
+Cterm Nsr1
(D) hNCL expression
Rate of Recomb X 10-8
300
250
200
150
100
50
0
GTOP
GBTM
+Vec.
+NCL
+Vec.
+NCL
top1Δ nsr1Δ
top1Δ

Figure 5.

(A)

SMU-G : GAGCTGGGGTGAGCTGGGCTGAGCTGGGGTGAGCTGGGCTGAGCT

BRCA1-G1: AGGGCTCTCCCTTGGGGGGGGGGGGCAGGAAGGGA

SMU-C : AGCTCAGCCCAGCTCACCCCAGCTCAGCCCAGCTCACCCCAGCTC

SS-DNA: GCACGCGTATCTTTTTGGCGCAGGTG