

# Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*

Steve G. Hershman<sup>1</sup>, Qijun Chen<sup>2</sup>, Julia Y. Lee<sup>2</sup>, Marina L. Kozak<sup>2,3</sup>, Peng Yue<sup>3</sup>, Li-San Wang<sup>2,4,5</sup> and F. Brad Johnson<sup>2,3,5,\*</sup>

<sup>1</sup>College of Arts and Sciences and Vagelos Scholars Program, University of Pennsylvania, <sup>2</sup>Department of Pathology and Laboratory Medicine, <sup>3</sup>Cell and Molecular Biology Graduate Program, <sup>4</sup>Penn Center for Bioinformatics, and <sup>5</sup>Penn Institute on Aging, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Received September 9, 2007; Revised October 14, 2007; Accepted October 19, 2007

## ABSTRACT

Although well studied *in vitro*, the *in vivo* functions of G-quadruplexes (G4-DNA and G4-RNA) are only beginning to be defined. Recent studies have demonstrated enrichment for sequences with intramolecular G-quadruplex forming potential (QFP) in transcriptional promoters of humans, chickens and bacteria. Here we survey the yeast genome for QFP sequences and similarly find strong enrichment for these sequences in upstream promoter regions, as well as weaker but significant enrichment in open reading frames (ORFs). Further, four findings are consistent with roles for QFP sequences in transcriptional regulation. First, QFP is correlated with upstream promoter regions with low histone occupancy. Second, treatment of cells with *N*-methyl mesoporphyrin IX (NMM), which binds G-quadruplexes selectively *in vitro*, causes significant upregulation of loci with QFP-possessing promoters or ORFs. NMM also causes downregulation of loci connected with the function of the ribosomal DNA (rDNA), which itself has high QFP. Third, ORFs with QFP are selectively downregulated in *sgs1* mutants that lack the G4-DNA-unwinding helicase Sgs1p. Fourth, a screen for yeast mutants that enhance or suppress growth inhibition by NMM revealed enrichment for chromatin and transcriptional regulators, as well as telomere maintenance factors. These findings raise the possibility that QFP sequences form *bona fide* G-quadruplexes *in vivo* and thus regulate transcription.

## INTRODUCTION

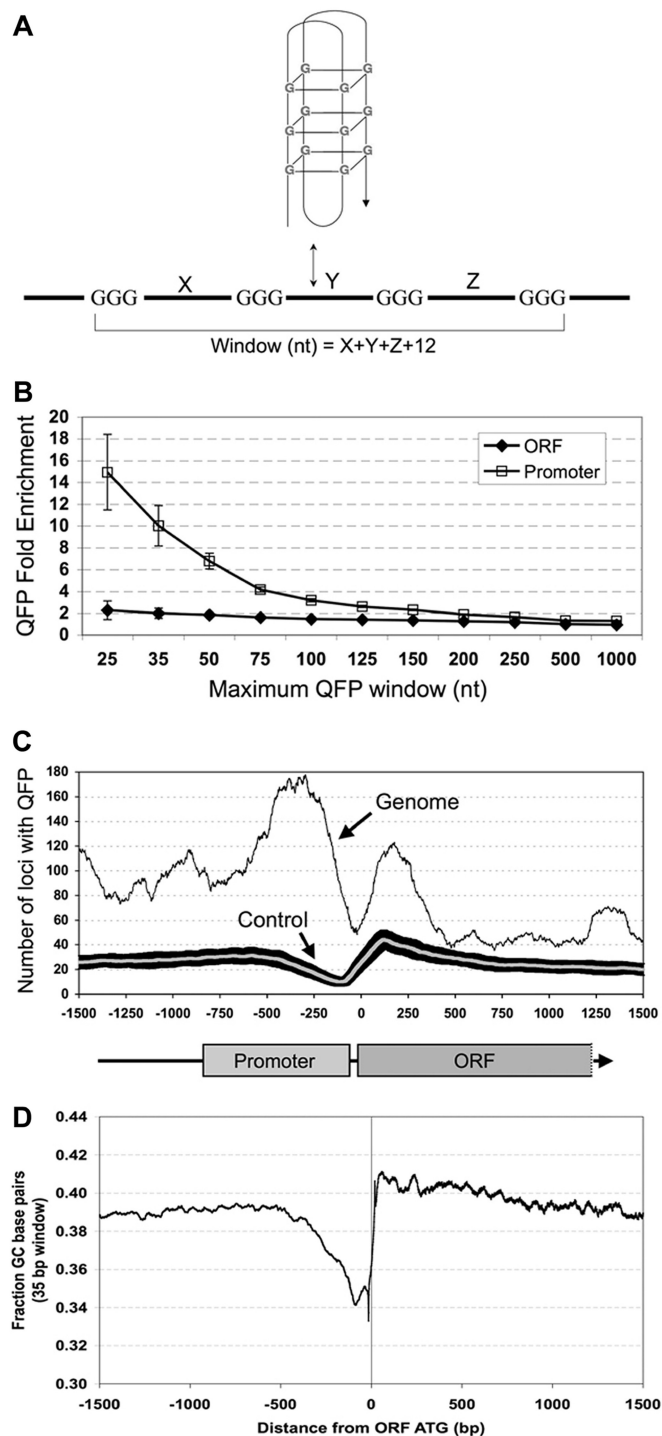
G4-DNA and G4-RNA are families of DNA and RNA structures comprising stacked arrangements of planar G-quartets that themselves comprise four Hoogsteen-bonded guanines that come from one or more nucleic acid chains (an intramolecular example is shown in Figure 1A) (1,2). G4 structures are highly stable under physiological pH and salt conditions, and a growing number of proteins that selectively bind or process them have recently been identified [reviewed in (3,4)], including the yeast Sgs1p helicase and the related human Werner and Bloom syndrome proteins, WRN and BLM (5–7).

Eukaryotic genomic loci that typically contain G-rich QFP sequences with the capacity to form G4-DNA (at least *in vitro*) include telomeres, the ribosomal DNA (rDNA), certain minisatellites and immunoglobulin (Ig) heavy chain gene switch regions (8–12). A key question is whether QFP sequences actually form G-quadruplex structures *in vivo*, or whether they are instead a marker of some other function of the sequences.

Telomeres have provided the best evidence to date for *in vivo* formation and function of G4-DNA. Telomeres usually end with a single-stranded 3' extension of the G-rich strand at the chromosome terminus; because single-stranded G-rich telomere strands readily form G4-DNA *in vitro*, these overhangs may be particularly susceptible to G4-DNA formation. The recent demonstration that G4-DNA can be detected *in vivo* at telomeres in *Stylonychia lamnae* cells (13), in a fashion dependent on the expression of the TEBP $\beta$  telomere-binding protein (9), whose *Oxytricha* homologue itself catalyzes G4-DNA formation *in vitro* (14), provides compelling evidence that G4-DNA can form *in vivo*. Further, the human POT1

\*To whom correspondence should be addressed. Tel: 215-573-5037; Fax: 215-573-6317; Email: johnsonb@mail.med.upenn.edu  
Present address:

Julia Y. Lee, Department of Biology, Saint Joseph's University, Philadelphia, PA.



**Figure 1.** Distribution of QFP within the yeast genome. (A) Intramolecular QFP was calculated by searching for sequences falling within windows of different sizes and each possessing four runs of at least three Gs. The window size for any QFP sequence was equal to  $X + Y + Z + 12$ , where  $X$ ,  $Y$  and  $Z$  are the length of the loops between each run of three Gs. One of many possible intramolecular G4-DNA folds is shown for illustrative purposes. (B) Enrichment of QFP sequences in promoters and ORFs. The number of loci with at least one QFP sequence within promoters ( $-850$  to  $-50$  compared to start of translation) or ORFs were calculated for windows of various sizes ( $X$ -axis). Fold enrichment of actual QFP ( $Y$ -axis) was then determined by dividing these frequencies by QFP frequencies that were calculated for the average of randomly generated genomes, which were themselves

protein that binds the telomere 3' overhang, disrupts G4-DNA formation *in vitro* (15). Conversely, POT1 is lost from telomeres in cultured cells treated with the G4-DNA small molecule ligand telomestatin, resulting in telomere uncapping, and suggesting that the overhang can exist either in POT1-bound or G4-DNA forms (16). Furthermore, the RTEL protein that is homologous to a *Caenorhabditis elegans* DNA helicase thought to process G4-DNA (17), is an important regulator of telomere length in mice (18). In addition, defects in telomere maintenance in cells lacking WRN, BLM or Sgs1p are widely hypothesized to result from defects in G4-DNA processing during replication or recombination, because these helicases show particularly high activity in unwinding G4-DNA substrates (3,18–20).

Outside of telomeres, the demonstration that the level of G4-DNA observed in human Ig class switch regions that had been transcribed in *Escherichia coli* is inversely related to the expression of the RecQ helicase, which itself unwinds G4-DNA (21), also provides evidence for G4-DNA formation *in vivo* and shows that it can be linked to transcription (12). Further, the demonstration that c-Myc expression can be inhibited by a small molecule G4-DNA ligand via a promoter QFP sequence in cultured cells suggests that G4-DNA can regulate gene expression *in vivo* (22,23). However, it remains possible in this last example that G4-DNA does not form naturally but forms only in the presence of the added ligand.

In addition to high QFP in eukaryotic telomeres, rDNA, minisatellites and Ig heavy chain gene switch regions, recent studies of the human, chicken and bacterial genomes have shown that there are numerous QFP sequences at additional loci throughout these genomes (24–27). Remarkably, genes with QFP fall into functional classes; for example, human oncogenes and tumor suppressor genes have particularly high or low QFP, respectively (28). QFP sequences occur with higher than random frequency upstream of transcriptional promoters, and at least in the case of humans, particularly at nuclease hypersensitive sites, suggesting a role in transcriptional regulation (29). It is hypothesized that separation of the base-paired strands of the DNA duplex, for example, stimulated by negative supercoiling, transcription factor binding or promoter melting associated with transcription, enables G4-DNA structures to form intramolecularly (26,29). There is indeed direct evidence that even in the

constructed using position-specific base frequencies (see Materials and Methods section). Error bars are SDs generated by randomly sampling examples of actual QFP sequences 1000 times (see Materials and Methods section). (C) Locations of QFP sequences with respect to the start of ORFs. The number of loci having QFP sequences (window = 50 nt) at each position compared with the start of translation was calculated for all ORFs and the number of loci with QFP sequences within 200 bp of each point is shown. Also shown, for comparison, is the average QFP of 100 randomly generated genomes (gray line) with SDs (black area surrounding gray line) constructed using position-specific base frequencies. Exclusion of loci close to telomeres yielded similar results (Supplementary Figures 1A and B). (D) The average GC content in a 35 bp window for all ORFs is shown for positions from  $-1500$  to  $+1500$  bp compared to the start of translation.

absence of factors favoring duplex unwinding, that the stability of the c-kit promoter quadruplex is sufficient to favor its formation over the competing duplex form, even when flanked by extensive duplex DNA (30). Examples of transcriptional regulatory proteins that bind G4-DNA with high affinity, and which are thus candidates for mediating transcriptional regulatory effects of G4-DNA, include nucleolin, MyoD, LR1 and Rap1p (3,31–33). Alternatively, or in addition, G4-DNA might exert transcriptional regulatory effects through DNA topology or chromatin structure.

Here we survey the distribution of QFP sequences in the *Saccharomyces cerevisiae* genome. We find that, similar to vertebrates and *E. coli*, QFP is found at higher than random frequency, and is particularly enriched in upstream promoter regions. Moreover, we describe associations between QFP and histone occupancy and changes in gene expression under conditions predicted to influence G4 levels or function. These novel findings are consistent with the hypothesis that G4 structures may form in yeast and regulate gene expression.

## MATERIALS AND METHODS

### *In silico* location of QFP sequences

Yeast sequences were obtained from the *Saccharomyces* Genome Database (SGD) ([http://www.yeastgenome.org/gene\\_list.shtml](http://www.yeastgenome.org/gene_list.shtml)). The QFP algorithm scanned for the presence of four repeats of at least three consecutive guanines each where the distance between the beginning of the first and end of the last repeat was less than a window of defined nucleotide length; this can be expressed as  $\text{GGGN}_X\text{GGGN}_Y\text{GGGN}_Z\text{GGG}$  where  $12 + X + Y + Z < \text{window size}$ , and is similar to the algorithm employed by Eddy and Maizels (28). We also compared our algorithm to one previously employed in analyses of the human genome ( $\text{GGGN}_{1-7}\text{GGGN}_{1-7}\text{GGGN}_{1-7}\text{GGG}$ ) (24,25), and it gave results similar to ours at a window size of 25 nt. Software, implemented in Python, is available on request. The location of each QFP sequence was defined to be its midpoint. For analyses of ORFs ( $N = 6576$ ) and associated upstream promoter sequences, all ORFs (including ORFs flagged as dubious in SGD) were used.

### Enrichment of QFP sequences in promoters and ORFs

To determine the QFP enrichment of promoters and ORFs, the number of these sequences was counted in the promoters (–50 to –850 bp with respect to the start of translation) and the ORFs (up to the first 3Kb). To generate a null model of quadruplex frequencies, promoter and ORF sequences of all genes were aligned so their ORF start positions were matched; the nucleotide frequencies for each position relative to the start of the ORF were then computed. We then used these position-specific nucleotide frequencies to generate 100 random genomes for 25 nt windows and 25 random genomes for all larger window sizes. Their average quadruplex counts and SDs were calculated; these numbers of random genomes were sufficient to generate coefficients of variation of <5% for

each window. The mean QFP counts over the random genomes were calculated as a reference to derive the enrichment (fold-increase) of QFP sequences in the real genome. Error bars (SDs) were generated by bootstrapping genes (sampling genes with replacement) (34).

To create a sliding window of the location of QFP sequences relative to the start of the ORFs, we analyzed each ORF and its promoter for quadruplexes within a window of 50 nt. We used a sliding window of 200 nt to determine the number of QFP sequences within 100 nt of each site. To assess the significance of QFP enrichment location-wise with respect to the start of ORF, 100 random genomes were generated using position-specific nucleotide frequencies as a reference, and used to compute the position-specific means and SDs.

### NMM treatment

Two independent BY4741 yeast colonies (*Mat a, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*) were inoculated in synthetic complete (SC) medium (containing 0.01% uridine rather than uracil) and each culture was then grown with or without 8 μM *N*-methyl mesoporphyrin IX (NMM; Frontier Scientific, Logan, UT, USA) overnight. Cells were rediluted to  $1 \times 10^6$  cells/ml in SC medium, maintaining growth with or without 8 μM NMM, and harvested in exponential phase after an additional 6 h of growth. RNA was extracted using the Qiagen (Valencia, CA, USA) RNeasy Mini kit according to manufacturer's instructions. Microarray hybridization and scanning were performed at the Penn Microarray Facility with the Yeast Genome 2.0 Affymetrix GeneChip Array (Santa Clara, CA, USA) using 5 μg RNA for each sample.

### Microarray hybridization and data analysis

All microarray data preparation and analyses were performed using the statistical software R (35). Standard Affymetrix quality controls were performed using the Bioconductor (36) package Simpleaffy (37). For the NMM and *sgs1* mutant experiments, we included only probesets identified as 'Present' by the MAS5.0 algorithm (38) in at least one of the four arrays for further analysis; this left us with 6272 out of 10928 and 8125 out of 9335 probesets respectively. Signal values were obtained using the *gcma* algorithm (39). Principal components analysis and hierarchical clustering (average linkage) of the Pearson's correlation coefficients for all genes revealed significant separation based on condition (NMM treatment or genotype) that far exceeded experimental variation. We used Rank Product (40), a nonparametric method designed for experiments with a small number of replicates (41), to identify differentially expressed genes, with a *P*-value cutoff of 0.05.

### Significance of gene list overlaps and gene ontology enrichment analysis

Significance of the overlap between two lists of genes differentially expressed under the experimental conditions was determined using Fisher's exact test (i.e. a hypergeometric test) (42). We computed the *P*-value for overlap between the two gene lists (one-tailed



*P*-value) using R (35). To determine the list of enriched gene ontology terms in a gene list, we used the YEAST gene ontology library from Bioconductor (36) or the GO Slim Biological Process terms from the Stanford SGD, to determine the set of annotated genes for each gene ontology term and applied Fisher's exact test to obtain one-tailed *P*-values.

### NMM genetic screen

Enhancers and suppressors of NMM growth inhibition were obtained from the viable haploid deletion collection in the strain BY4741 (Open Biosystems YSC1053). Individual deletion strains were grown in 96-well plates, together with the wild-type control, overnight in YPD with gentle agitation at 30°C. Cells were then diluted to  $5 \times 10^5$  cells/ml in SC medium in a second plate and grown for 6 h at 30°C to a density of approximately  $2 \times 10^6$  cells/ml. Fifty microlitres of such cells were transferred to each of three fresh plates containing 50 µl of 0, 12 or 24 µM NMM in SC medium. Relative cell densities were read using an Opsy MR microplate reader (Thermo Labsystems, Franklin, MA, USA) after 16 additional hours of growth. Percent growth inhibition was calculated as  $100 \times [1 - (\text{OD with NMM} / \text{OD without NMM})]$ . Replicate experiments with the wild type strain revealed an SD of 5% for growth inhibition. NMM-S strains were defined as those at least 3 SDs (i.e. 15%) more sensitive than wild-type when grown at one concentration of NMM (6 or 12 µM NMM) and at least 1 SD more sensitive than wild-type at the second NMM concentration; NMM-R strains fit the same criteria except that they were resistant to growth inhibition compared to wild-type.

## RESULTS

### Genomic distribution of intramolecular QFP sequences

The distribution of sequences with intramolecular QFP was determined using an algorithm requiring four runs of three or more guanines on one strand and within a window of defined size, ranging from 25 to 1000 nt (Figure 1A). Recently, application of a similar algorithm with a window of 100 nt to the human genome revealed highly non-random associations between QFP and genes in particular functional classes (28). Analysis of yeast open reading frames (ORFs) and promoters (defined as sequences between 50 and 850 nt upstream of the ORF) transcribed by RNA polymerase II revealed significant enrichment of QFP at window sizes up to ~250 nt for ORFs and beyond 1000 nt for promoters (Figure 1B). QFP sequences within the sense or anti-sense strands were summed for this analysis, and fold enrichment was calculated by comparing the frequency of QFP loci in the actual yeast genome to the average frequency of QFP loci in simulated yeast genomes generated by randomly permuting nucleotides using position-specific base frequencies (see 'Materials and Methods' section). Thus the GC contents of the real and simulated genomes are identical and the QFP enrichment in the real genome reflects sequence organization beyond simple GC content. QFP enrichment was strongest for promoters and at

**Table 1.** Numbers of ORFs and linked promoters with QFP in *S. cerevisiae*

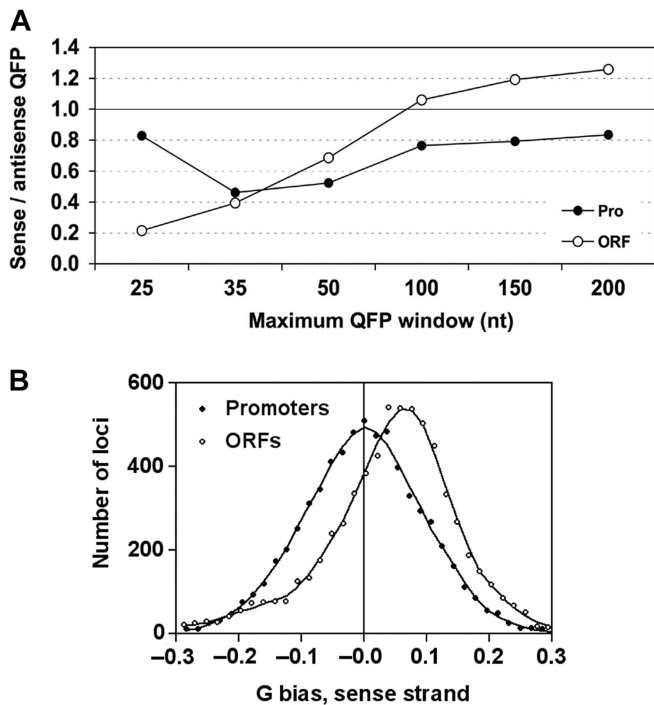
Window	ORFs		Promoters	
	All loci	Non-Tel loci <sup>a</sup>	All loci	Non-Tel loci
25	11	8	33	27
35	54	48	87	75
50	236	229	288	259
75	779	753	735	704
100	1436	1405	1254	1217
150	2807	2769	2395	2347

Number of loci = 6576. <sup>a</sup>Non-Tel loci indicates that loci near telomeres ( $N = 85$ , see Supplementary Table 3) were excluded from the analysis.

smaller window sizes, with greater than 6-fold enrichment for windows of 50 nt or less. In contrast, the QFP enrichment within ORFs was 2-fold or less, although still significant. The higher QFP of promoters compared to ORFs is striking given their similar GC contents (38.3% for promoters; 39.7%, for ORFs). The absolute numbers of ORFs and promoters containing QFP sequences for windows of different sizes are described in Table 1, and the genomic coordinates of all QFP sequences, as well as the identities of loci having promoter or ORF QFP (25 nt window), are listed in Supplementary Table 1.

To determine the location of QFP sequences with respect to the start of ORFs, the number of loci having QFP (window = 50 nt) at each nucleotide position upstream and downstream of the start of all ORFs was calculated (Figure 1C). For comparison, the locations of QFP sequences in the average of 100 simulated genomes generated in the same fashion as the controls in Figure 1A were also calculated. A broad peak of QFP was observed for real ORFs from ~50 to 850 nt upstream of the start of translation, as was a second and smaller peak spanning the first 400 nt of the ORF. These peaks were separated by a small region of low QFP centered near the start of transcription, which is typically 15–75 nt upstream of the ORF in *S. cerevisiae* (43). Again, this pattern does not correlate simply with GC content, as can be seen by comparison with the average GC content of base pair positions with respect to the start of ORFs (Figure 1D). Smaller peaks of QFP are also apparent beyond 1 kb upstream and downstream of the start of ORFs. However, because the mean intergenic distance is 536 bp and mean ORF length is 1385 bp in *S. cerevisiae* (44), sequences at these distances overlap substantially with the ORFs and promoters of adjacent loci, and the significance of these peaks is thus unclear.

When individual strands were examined, ORF sense strands showed reduced QFP when compared to their respective antisense strands at window sizes under 100 nt, and promoter sense strands showed reduced QFP compared to antisense at all window sizes (Figure 2A). QFP does not correlate simply with G content because sense and antisense strands have similar G content in promoters, but sense strands have higher G content than antisense strands in ORFs (Figure 2B). The reduced QFP

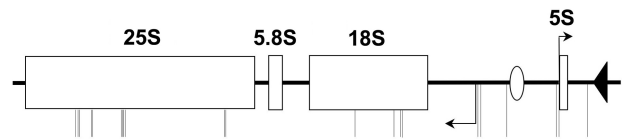


**Figure 2.** Sense versus antisense strand distributions of QFP and G-bias. (A) Ratios of QFP in the sense vs. antisense strands in promoters (filled circles) and ORFs (open circles) for windows of the indicated sizes. (B) Frequencies of G-bias on the sense strand among individual ORFs and upstream promoter regions. Shown are the number of loci with different levels of G-bias, defined as the difference on the sense strand in the number of G and C nucleotides divided by the total number of G and C nucleotides.

in the ORF sense strand may reflect selection against shorter G-quadruplexes in mRNA, which can be inhibitory to translation (45), and we note that similar strand asymmetry was observed in human exonic regions (24). In contrast, the elevated QFP in promoters raises the possibility that G-quadruplexes might have a transcriptional regulatory function.

### QFP of the rDNA and telomere repeats

The G-rich rDNA and telomere repeats of yeast were not included in the analyses above. The rDNA consists of 100–200 tandem direct repeats on chromosome XII of a 9.1 kb region that includes the transcription units encoding the 5S, 5.8S, 25S, and 18S rRNAs, as well as transcribed and non-transcribed spacer sequences (46). The 5S locus is transcribed by RNA polymerase III, while the other three rRNA genes are transcribed as a single 35S precursor transcript by RNA polymerase I, in the orientation opposite to the 5S locus. The locations of QFP sequences (window = 100 nt) within a single rDNA repeat are shown in Figure 3. There are 21 such sequences in each rDNA repeat, giving the rDNA an approximately 10-fold higher QFP density than the remainder of the genome, excluding telomere repeats. Remarkably, all 21 QFP sequences are located on the sense strand with respect to the 35S transcript, and no QFP sequences are present on the opposite strand even at a window size of



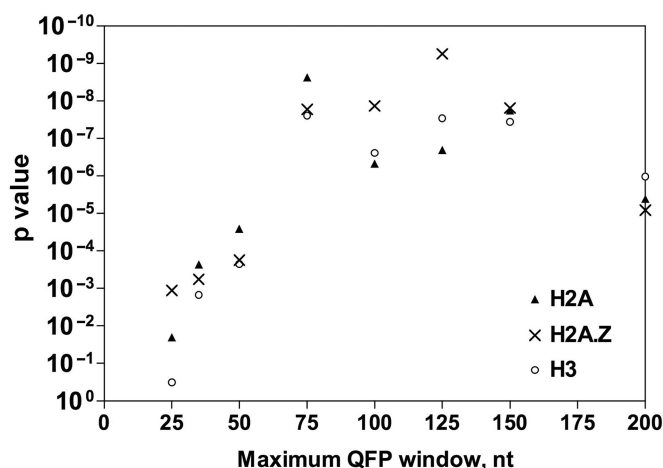
**Figure 3.** QFP distribution in the rDNA. One rDNA repeat is shown, and the center of each QFP sequence is indicated by a downward-extending vertical line. The 35S transcript, containing the 18S, 5.8S and 25S rRNAs, extends from the leftward promoter (arrow) to the left edge of the diagram, and the 5S rRNA promoter is indicated by the arrow pointed to the right. The origin of replication and replication fork block region are indicated by an open circle and closed triangle, respectively.

200 nt. This is similar to the enrichment for GGG sequences in the sense strand of the human rDNA (10). Eleven and four QFP sequences lie in the 25S and 18S loci, respectively, and thus are also present in their rRNA products.

The *S. cerevisiae* telomere repeats are imperfect and follow the consensus 5'-[(TG)<sub>0-6</sub>TGGGTGTG(G)]<sub>n</sub>-3' (47). The TGGGTGTG core sequence is found with an average spacing of 11 nt, and thus an average of 8 nt exists between each GGG run (47). Given an average telomere repeat tract length of 300–350 bp at the end of each chromosome, it is clear that telomeres have high QFP, consistent with earlier demonstrations that yeast telomere repeats form G4-DNA *in vitro* (33,48). Because telomere lengths and sequences are variable, the SGD does not contain full sequence data for telomere repeats. Nonetheless, we obtained many examples of telomere repeat sequences with QFP in our survey (Supplementary Table 1).

### Comparison of QFP with histone occupancy

To begin to explore the possible connection of promoter QFP with transcriptional regulation of loci transcribed by RNA polymerase II, we examined the relationship between QFP and histone occupancy. The genome-wide distributions of histones H2A, H3 and H2A.Z have been determined by chromatin immunoprecipitation (49). H2A and H3, together with H2B and H4, form canonical nucleosomes, while H2A.Z is a H2A variant found in nucleosomes that are enriched at promoters poised for transcriptional activation and is also found at subtelomeric regions where it helps prevent spreading of silent telomeric heterochromatin (49–51). For intergenic regions having the lowest occupancy (bottom 10% of all loci) for histones H2A, H3 or H2A.Z, *P*-values for overlap with promoter QFP were  $2.4 \times 10^{-9}$ ,  $2.5 \times 10^{-8}$  or  $1.7 \times 10^{-8}$ , respectively (Figure 4 and Supplementary Table 2). These *P*-values reflect significance at a window of 75 nt, but significant overlap was found for windows of 25 to >200 nt for H2A and H2A.Z and for windows of 35 to >200 nt for H3, emphasizing the robustness of the association. No correlations were apparent between promoter QFP and high levels of histone occupancy or between ORF QFP and high or low histone occupancy. Given the high QFP of telomere repeat DNA and the fact that telomeres possess distinctive chromatin including



**Figure 4.** Significance of association between promoter QFP and low histone occupancy. *P*-values for positive association between promoter QFP and intergenic fragments that are in the lowest 10% for binding by histone H2A, H2A.Z or H3 are shown.

exclusion of histones from the telomere repeat sequences and hypoacetylation of histones in subtelomeric regions (52), we considered the possibility that telomeric loci might skew the observed relationship between QFP and intergenic histone occupancy. However, exclusion of telomeric loci (Supplementary Table 3) from the analyses had little effect, with *P*-values equaling  $1.1 \times 10^{-8}$ ,  $1.2 \times 10^{-7}$  or  $3.4 \times 10^{-8}$  for overlap between promoter QFP and the bottom 10% of occupancy of H2A, H3 or H2A.Z, respectively (Supplementary Table 2; 75 nt window). The transcription factor Rap1p has been shown to be associated with low promoter nucleosome occupancy genome wide in yeast, particularly when bound in multiple copies (53). Because Rap1p duplex DNA binding sites can contain a GGG motif (54), this raised the possibility that QFP simply reflects binding of multiple Rap1p proteins. However, highly significant associations between QFP and low histone occupancy were retained after exclusion of ChIP-defined Rap1p target promoters (55) from the analyses (*P*-values:  $7.11 \times 10^{-9}$ ,  $6.85 \times 10^{-8}$  and  $1.80 \times 10^{-6}$  for H2A, H3 and H2A.Z, respectively; Supplementary Table 2; 75 nt window). Therefore, Rap1p binding does not explain the observed association. These findings suggest a possible role for QFP sequences, and perhaps *bona fide* G4-DNA, in preventing histone binding.

### Functions of genes with QFP

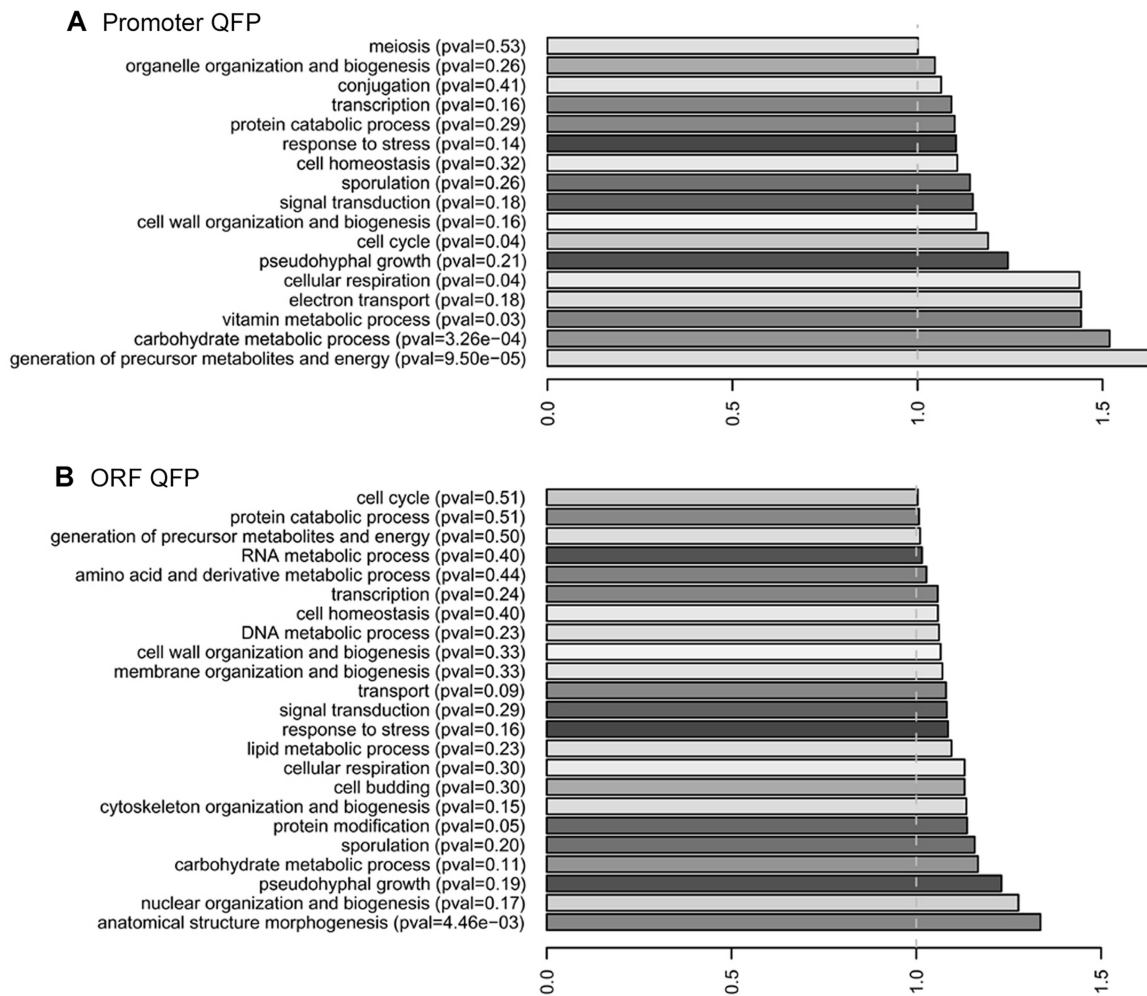
Significant correlations between QFP and gene functional categories have been observed in humans and bacteria (26,28). The enrichment of *S. cerevisiae* genes with ORF or promoter QFP within certain Gene Ontology (GO) Biological Process categories is shown in Figure 5. Full details of QFP associations with GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) categories are described in Supplementary Table 4 and Supplementary Figure 3. Notably, QFP tends to exist in promoters for genes involved in metabolism ( $P \sim 10^{-3}$  to  $5 \times 10^{-6}$ ) and ORFs connected with developmental

processes ( $P = 1.7 \times 10^{-4}$ ), similar to what has been observed in bacteria and humans, respectively (26,28). Also similar to humans and bacteria, there is some association of QFP with ORFs that regulate transcription ( $P \sim 3-6 \times 10^{-3}$ ). In addition, there is significant overlap with loci involved in mitotic cell cycle control, particularly those regulating the G1/S transition ( $P = 3.8 \times 10^{-4}$ ).

### Relationship between gene expression and factors that may impact G-quadruplexes

We next tested for associations between QFP and gene expression changes expected *a priori* to be related to G-quadruplexes. As a first test, we used Affymetrix microarrays to compare gene expression in exponentially growing yeast cultures treated with 8  $\mu$ M NMM versus untreated controls. NMM binds G4-DNA with high selectivity *in vitro*, showing no detectable binding to other nucleic acid structures, including duplex and single-stranded DNA, duplex RNA, DNA-RNA hybrids, triplex DNA and Z-DNA (7,56,57). G4-DNA ligands can stabilize quadruplexes (58,59), and NMM would be expected to behave similarly, although it has not been itself tested in this regard. NMM of 8  $\mu$ M was chosen because it inhibited the growth rate of exponentially growing cells by 25%, suggesting that the level of drug was sufficient to affect cell functions without excessive toxicity; it is also close to the 1–2  $\mu$ M  $K_d$  of NMM for G4-DNA *in vitro* and is less growth-inhibitory than levels of DNA interactive agents that have been studied in other yeast microarray studies (60). Flow cytometric analysis of NMM-treated cultures revealed an accumulation of cells in S-phase, explaining the slowed growth (Supplementary Figure 2). Duplicate treated and untreated cultures were analyzed and genes showing altered expression with  $P < 0.05$  cutoff were considered significant. Approximately 9% of genes were upregulated and 9% downregulated by NMM (Supplementary Table 5). We compared the upregulated and downregulated loci with the list of loci having QFP. Remarkably, among the four comparisons, upregulated loci were significantly associated with QFP (window = 100 nt) in both promoters ( $p = 2.2 \times 10^{-6}$ ) and ORFs ( $P = 4.9 \times 10^{-4}$ ) (Supplementary Table 6). Downregulated loci showed no QFP overlap. For ORFs, the correlation with upregulation was lost at shorter QFP windows (e.g. for window = 50 nt,  $P = 0.26$ ), but the correlation for promoters did not decline as dramatically with decreasing window size (e.g. for windows of 50 and 35 nt,  $P = 0.0015$  and 0.0034, respectively). In addition, there was a strong correlation between genes with altered expression and those found previously to be altered in the environmental stress response (ESR), a set of transcriptional responses observed following diverse environmental insults (61): the *P*-values for overlap between genes downregulated and upregulated in both the NMM and ESR sets were  $6.0 \times 10^{-26}$  and  $9.6 \times 10^{-39}$ , respectively. However, consistent with the lack of NMM binding to duplex DNA, there was no significant activation of the DNA damage signature response (60), with only one (*DUNI*) of the nine signature loci activated by NMM.

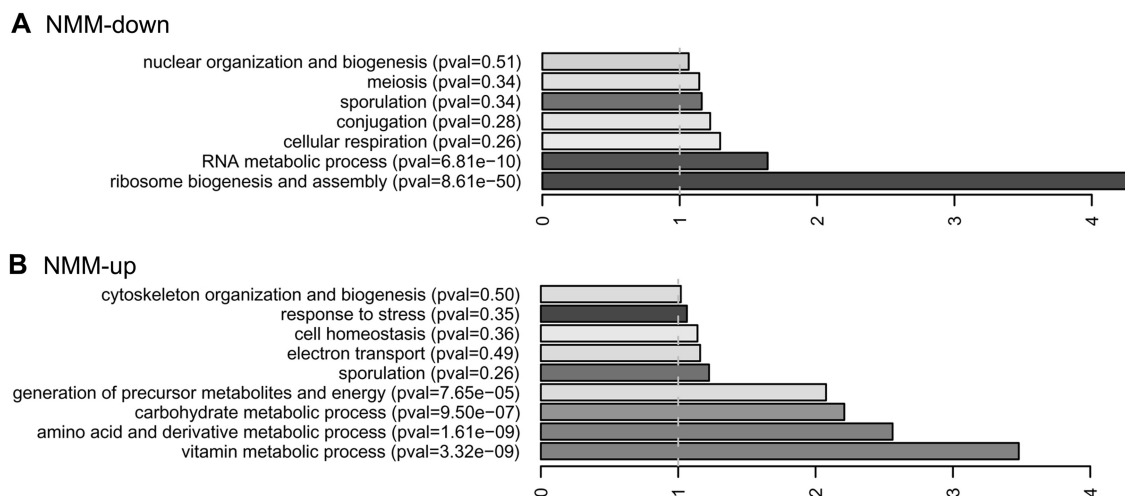




**Figure 5.** GO terms enriched for loci with promoters or ORFs with QFP. The proportion of genes in each GO category (using GO Slim terms from SGD) were determined for loci having (A) promoter or (B) ORF QFP. These proportions were then divided by the proportions for all yeast ORF loci, and enriched categories with their calculated ratios are shown. Only non-telomeric loci were analyzed, and *P*-values reflect the significance of overrepresentation.

GO categories enriched for NMM-regulated genes are illustrated in Figure 6 and Supplementary Figure 3. There is a highly significant association between NMM-downregulated genes and nucleolar function including rRNA processing and ribosome biogenesis ( $P < 10^{-47}$  for ribosome assembly and biogenesis; Figure 6 and Supplementary Table 7). As noted above, the rDNA genes and rRNAs have very high QFP and it is thus possible that G4-DNA or G4-RNA structures formed by these molecules are affected by NMM and in turn perturb nucleolar function. The downregulation of 18S and 25S transcripts detected by several independent probes on the array is consistent with this possibility (Supplementary Table 5). An alternate possibility is that growth inhibition by NMM, albeit mild, might explain the downregulation of nucleolar function because rDNA transcription is tightly coupled to growth rate (62). However, several considerations argue against this. First, the *P*-value for overlap of QFP with ribosome assembly and biogenesis is more significant than for downregulated ESR genes in general, which encompass genes downregulated by

growth inhibition ( $4 \times 10^{-48}$  versus  $6 \times 10^{-26}$ ). Second, when genes in the ribosome biogenesis and assembly GOBP group (GO:004255) were removed from the downregulated ESR gene set (which includes many genes not directly related to nucleolar function including those affecting tRNA and nucleotide biosynthesis), the significance of association between NMM treatment and downregulated ESR genes was reduced dramatically (*P*-value changed from  $6.0 \times 10^{-26}$  to  $1.7 \times 10^{-4}$ ). Thus, among downregulated ESR genes, those that affect nucleolar function appear to be the principal targets of NMM. Third, examination of genes downregulated by growth-inhibitory levels of methane methyl sulfonate (60), which like NMM causes S-phase accumulation, revealed no significant overlap with nucleolar-function genes (data not shown). We also note that, despite the accumulation of NMM-treated cells in S-phase, S-phase-regulated genes were not among those significantly affected (Supplementary Table 7), indicating that the effects of NMM extend beyond cell cycle perturbation. We conclude that the effects of NMM on expression



**Figure 6.** GO terms enriched for NMM-responsive genes. GO categories that are overrepresented among loci (A) downregulated or (B) upregulated by NMM were determined as in Figure 5, and their fold enrichments are shown.

of nucleolar-function genes are not a consequence of growth inhibition. Further, these gene expression changes, and moreover the overlap between QFP and NMM-regulation at other loci, provide support for the hypothesis that NMM affects gene expression in part by interacting with G-quadruplex targets.

As a second test for an association between QFP and gene expression changes that might be related to G-quadruplexes, we compared QFP and loci previously shown to have altered expression in *sgs1* deletion mutants (63). The Sgs1p helicase possesses a G4-DNA binding domain and unwinds G4-DNA with 10-fold higher activity than it unwinds B-DNA (7,64), and thus G4-DNA levels might be altered in *sgs1* mutants. It was not possible to predict *a priori* how QFP-related changes in gene expression in *sgs1* mutants would compare with those caused by NMM treatment. For example, *sgs1* mutants and NMM-treated cells might have higher G4-DNA levels overall, but in the latter case, the actions of the G4-DNA might be blocked by binding of NMM (e.g. via competition with G4-DNA binding factors) and thus the effects on gene expression in the two cases would be different; further, the various particular types of G-quadruplexes, or G-quadruplexes in different chromatin contexts, might be differentially affected by the two experimental manipulations. In fact, in *sgs1* mutants we observed significant associations between downregulated genes and those with QFP in their ORFs ( $P = 8.3 \times 10^{-4}$ ,  $2.3 \times 10^{-5}$  and  $2.9 \times 10^{-9}$  for 50, 75 or 100 nt windows, respectively; Supplementary Table 6). Thus, as was true for NMM treatment, the significance of association between QFP and altered gene expression increased for longer QFP windows, although the direction of regulation was different in the two cases. No significant changes were observed for promoter QFP or for upregulated loci. These data are consistent with the possibility that failure of Sgs1p to unwind ORF G-quadruplexes may inhibit gene expression, and might be related to unwinding of G4-DNA during transcription or to effects of G4-RNA

on mRNA production or stability. Interestingly, if the analysis for downregulated genes was divided into those with ORF QFP (window = 100 nt) only in the sense strand or only in the antisense strand, the association was stronger on the antisense strand ( $P = 1.4 \times 10^{-10}$ ) than on the sense strand ( $P = 1.4 \times 10^{-3}$ ), suggesting that the primary function of Sgs1p in this context may be at the DNA rather than RNA level (Supplementary Table 6).

#### Genetic enhancers and suppressors of growth inhibition by NMM

To gain additional insight into mechanisms by which NMM affects the growth of yeast we carried out a screen of viable haploid deletion mutants for those that were hypersensitive or resistant to NMM. Wild-type and mutant cells were grown in microtiter plates with 0, 6 or 12  $\mu$ M NMM and growth was monitored by optical absorbance. Of the 5045 strains that were successfully screened, 255 mutants enhanced the sensitivity and 295 mutants suppressed the sensitivity to growth inhibition by NMM (Supplementary Table 8). GO categorization of the enhancer and suppressor mutant loci is detailed in Supplementary Table 9.

Among mutants with enhanced NMM sensitivity, GO categorization revealed the most significant enrichment for factors that acidify vacuolar pH ( $P = 1.6 \times 10^{-9}$ ), suggesting that pH regulation affects growth inhibition by NMM (Supplementary Table 10). We suspect that this effect might be unrelated to any potential G4-DNA related role of NMM. However, the second-most significant enrichment among enhanced NMM sensitivity mutants was for GO categories of chromatin remodeling ( $P = 8.9 \times 10^{-7}$ ) or modification ( $P = 2.2 \times 10^{-6}$ ), and included 24 loci, which are listed in Supplementary Table 11. These included three of the five tested members each of the SWI/SNF (*snf6*, *taf14*, *rtt102*) and RSC (*ldb7*, *npl6*, *rtt102*) nucleosome remodeling complexes (65,66), and 4 of the 12 tested members of the SAGA histone acetyltransferase complex (*sgf29*, *spt3*, *ada2*, *spt8*) (67).



They further included *htz1*, lacking the variant histone H2A.Z and *yaf9*, lacking a component of the SWR1 complex that incorporates histone H2A.Z into chromatin (68,69). Additional examples of mutants lacking factors in complexes acting in sequential steps in a chromatin modification pathway were provided by *cdc73*, *spt4*, *rad6*, *swd3* and *bre2*. Cdc73p is a member of the PAF1 complex, which is recruited to RNA polymerase II by Spt4p and is required for the activity of a Rad6p-containing complex that ubiquitinates histone H2B, which is in turn required for the activity of the COMPASS/Set1c complex that includes Swd3p and Bre2p and methylates histone H3 (70–72). Further, there were several additional NMM-hypersensitive mutants lacking factors involved in transcriptional regulation (Supplementary Table 12), including additional factors affecting elongation by RNA polymerase II (e.g. *ela1*). Thus, growth inhibition by NMM may be related to modulation of chromatin structure and transcription.

GO-categorization of mutants resistant to growth inhibition by NMM showed the most significant enrichment for factors involved in ubiquitin-dependent protein catabolism ( $P = 1.5 \times 10^{-4}$ ) (Supplementary Table 9). Remarkably, the next most significant enrichment was for telomere-maintenance ( $P = 5.8 \times 10^{-4}$ ), including 26 such factors. In fact, there was also significant enrichment among NMM-sensitive mutants for telomere maintenance factors ( $P = 9.4 \times 10^{-4}$ ), including 24 such factors, and the overlap between pooled NMM-sensitive and resistant mutants and telomere maintenance factors was highly significant ( $P = 1.1 \times 10^{-6}$ ). These factors are listed in Supplementary Table 13. The propensity for yeast telomere repeat DNA to form G4-DNA, at least *in vitro*, might explain why mutations in telomere maintenance factors modulate responses of cells to NMM. Resistance or sensitivity to NMM did not correlate with whether the mutants had longer or shorter telomeres (73), indicating that steady-state telomere length did not itself explain the effects of NMM on cell growth. We also note that there were no significant associations between NMM-sensitive or resistant mutants and the DNA replication, recombination or repair GO categories, consistent with the lack of a DNA damage gene expression signature in NMM-treated cells (Supplementary Tables 5 and 7). However, there were certain mutants in these categories that were hypersensitive to NMM (e.g. *rad50* and *mec3* mutants; Supplementary Table 8), and we can therefore not rule out an effect of NMM on a small subset of processes within these GO categories.

## DISCUSSION

Our finding that QFP sequences are enriched in the yeast genome, particularly upstream of transcriptional promoters, adds to a growing list of studies involving organisms with similar enrichment, including humans, chickens and several bacterial species (26,27,29). The association of QFP with promoters in these diverse organisms implies an evolutionarily ancient function for these sequences. Our findings provide several novel lines of evidence consistent

with possible *in vivo* roles for G-quadruplexes, particularly in transcriptional regulation. These include an association of QFP with low histone occupancy, associations between QFP and gene expression changes caused by treatment of cells with the selective G-quadruplex ligand NMM or by genetic loss of the G4-DNA-unwinding helicase Sgs1p, and connections between factors affecting sensitivity to NMM and those affecting chromatin and telomere maintenance. NMM, Sgs1p and telomeres, each have established connections to G4 DNA structures, and thus their associations with QFP-linked gene expression provide support for the idea that G4 structures can function *in vivo*; however, it remains possible that other factors explain the observed associations.

Our finding that upstream promoter regions having QFP tend to have lower occupancy of histones H2A, H3 and H2A.Z suggests that G4-DNA formation might help exclude nucleosomes from DNA. This interpretation is consistent with the recent report that QFP is 230-fold overrepresented at nuclease-hypersensitive sites near promoters in the human genome (29). Previous analyses of sequences associated with nucleosome-free regions in yeast have identified enrichment for polydA-dT tracts (74,75). These analyses did not employ search algorithms that would have detected the larger-scale QFP sequences we have identified, and further, we do not yet know if QFP sequences themselves have reduced nucleosome occupancy or if they are instead closely linked to such sites. We considered the possibility that the association might be explained by the binding of Rap1p that in itself is correlated with reduced nucleosome occupancy (53). However, our analysis proved this was not the case. While it remains possible that the binding of some other transcription factor(s) to the duplex DNA form of QFP sequences explains the association, there is as yet no obvious candidate for this factor. It is interesting that many of the deletion mutants that caused sensitivity to NMM were in chromatin proteins, including members of the RSC and SWI/SNF nucleosome remodeling complexes. One possible interpretation is that G4-DNA formation helps remove nucleosomes, but that this function is somehow blocked when G4-DNA is bound by NMM, thus placing greater reliance on remodeling complexes. Because we expect that NMM binding should stabilize G4-DNA, such blocking might be through competition with regulatory proteins that bind G4-DNA. Alternatively, alterations in nucleosome positioning caused by NMM-stimulated G4-DNA formation might require nucleosome remodeling as a compensatory response. The hypersensitivity of *htz1* mutants to NMM is consistent with each of these ideas. Upon gene activation, nucleosomes containing H2A.Z (encoded by *HTZ1*) may be more easily displaced from chromatin than those containing H2A (49), and so any difficulty caused by NMM in removing nucleosomes would be compounded in *htz1* mutants.

Among genes downregulated by NMM treatment, we observed a dramatic enrichment for genes connected with nucleolar function, including those involved in rRNA processing and ribosome assembly. This is remarkable given the high QFP of the rDNA and

its rRNA transcripts. We speculate that the downregulation of nucleolar-function genes occurs in response to a primary inhibition by NMM of rRNA expression, processing or function. We note that NMM does not alter expression of translation genes ( $P = 0.99$ , Figure 5), implying that NMM might affect rRNA prior to its involvement in protein synthesis. The various nucleolar events that lead to ribosome assembly are tightly coordinated to maximize efficiency of this energy-intensive process, and so it is not surprising that inhibition of a subset of these events (e.g. rRNA expression or processing) would result in coordinated downregulation of much of the nucleolar program (76). The presence of QFP sequences in the 25S and 18S rRNAs makes these species obvious candidates for interference by NMM, especially given the relative ease with which single-stranded RNA should form G-quadruplexes, compared with DNA where duplex species compete with the G-quadruplex isoforms. This view is consistent with our observation that NMM, which fluoresces when bound to G-quadruplexes (56), selectively stains the nucleolus of human cells in a RNase A-sensitive fashion (F.B.J. and J. Shen, unpublished observations). The downregulation of genes with nucleolar function might be a homeostatic response to reduced rRNA levels, processing or function, and need not be related to QFP within the nucleolar-function genes themselves (indeed, nucleolar-function genes are not enriched for QFP overall). In contrast, the significant genome-wide overlap between promoters and ORFs having QFP and those regulated by NMM indicates that some NMM-related gene expression changes might involve direct effects of G4 structures at certain loci. NMM may have complex and indirect effects on gene expression, but we do not currently have any specific alternative explanation to G4 structures to explain this overlap. Two aspects of this QFP-related gene regulation by NMM merit further mention. First, this regulation involved predominantly upregulation, in contrast to the downregulation of genes related to nucleolar function. NMM might activate gene expression by displacing transcriptional repressors that bind to G4-DNA or by promoting the effects of G4-DNA structures on promoter function, such as modulation of supercoiling (26). Second, the significance of this association was relatively modest ( $P = 2.2 \times 10^{-6}$  for promoter QFP). However, this is not surprising given the highly significant effect of NMM on transcription related to nucleolar function (e.g.  $P < 10^{-47}$  for ribosome assembly and biogenesis), which as we argue above may be an indirect (i.e. QFP-independent) response to QFP-dependent interference with rRNA expression or function, and thus would obscure QFP-dependent transcriptional regulation occurring at loci outside of the nucleolar category.

Another remarkable aspect of QFP in the rDNA is that it is restricted to one DNA strand. Because the fork block region of each rDNA repeat ensures that most replication occurs in the same direction as 35S rRNA transcription, it follows that the strand with QFP will usually be replicated by the lagging strand DNA replication machinery (77). Strikingly, this is the same arrangement that is present at telomeres, where the G-rich strands run 5' to 3' toward the chromosome termini and are thus copied by lagging

strand synthesis emanating from subtelomeric origins. This may facilitate replication because if G4-DNA causes a block to lagging strand synthesis it should still be possible to continue replication fork progression beginning with the next Okazaki fragment and later complete synthesis across the gap after G4 unwinding. Consistent with the idea that G4-DNA can cause replication difficulties is the observation that human cells lacking the WRN DNA helicase, which like its Sgs1p homologue has a predilection for G4-DNA substrates (5–7,78), suffer selectively from loss of the telomere strand copied by lagging strand synthesis (19). Interestingly, *S. pombe* telomere repeats cause replication pausing independent of their orientation with respect to leading/lagging strand synthesis (79). Thus, it is possible that G4-DNA might be a challenging substrate for both the leading and lagging strand machinery, but as we suggest, copying via lagging strand synthesis may provide more opportunities for completion of proper replication.

In addition to an association between QFP and altered gene expression following NMM treatment, we observed a significant association between QFP in ORFs and downregulated genes in *sgs1* mutants. Because this association was particularly robust for QFP in the antisense (i.e. template) strand, we speculate that Sgs1p might serve to unwind G4-DNA ahead of the advancing RNA polymerase and thus promote transcription. While the most prominent defects in cells lacking the function of RecQ-family helicases like Sgs1p appear to be related to their well-established roles in DNA recombination and replication (80), there have also been several reports of transcriptional abnormalities in cells lacking RecQ-family helicases (63,81–84), and our findings are consistent with a direct effect of Sgs1p on transcription.

In single-stranded oligonucleotide models tested *in vitro*, shorter loop lengths between G runs generally lead to G4-DNA with higher stability (85,86). However, it is not yet possible to predict G-quadruplex stability for any possible set of loop lengths and sequences, nor is it clear what factors will determine stability of G-quadruplexes *in vivo*, where chromatin and sequence-specific nucleic acid binding factors are very likely to modulate their formation and disassembly. We observed more enrichment for QFP sequences in yeast promoters at shorter window sizes, corresponding to shorter average loop length. In contrast, we observed increasing significance with larger window sizes for associations between QFP and altered gene expression following *sgs1* mutation or treatment with NMM. While some of the increased significance is a simple statistical manifestation of the greater number of QFP loci at larger windows, it may also reflect a genuine preferential association between G4-structures with longer loops and those that are regulated by factors like NMM and Sgs1p. Perhaps the formation of short-loop G4 structures is relatively insensitive to such factors, while the formation of long-loop G4 structures may be more subject to regulation.

We emphasize that until biophysical and additional genetic studies are carried out on the identified QFP sequences, if and under what circumstances they may form *bona fide* G4 structures *in vivo* is uncertain. It is, for

example, possible that G4 structures are relatively long-lived, unwound only to allow polymerases to pass through them; alternatively they may be transient and yet have persistent effects on chromatin that are mediated by other factors. Our findings using NMM indicate that G-quadruplexes might mediate effects of this compound at telomeres, the rDNA and on the expression of loci throughout the genome. It will be important to investigate these effects further, because they indicate that the many small molecule G4-DNA ligands recently developed as potential cancer chemotherapeutic agents that target telomeres (87) may impact cell biology beyond the intended effects at telomeres. Our findings connecting QFP with reduced histone occupancy and with inhibition of gene expression in *sgs1* mutants imply that G4 structures might also form naturally. Yeast provides an ideal system for testing these ideas, particularly because of the relative ease with which nucleotide-level modifications of genomic DNA can be made. For example, it should be possible to systematically mutate individual Gs in QFP sequences to disrupt any possible G4-DNA formation and then test the effects on gene expression or histone occupancy. Our findings should provide impetus for these challenging but important studies.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

## ACKNOWLEDGEMENTS

We thank Don Baldwin and the Penn Microarray Facility, and John Tobias and the Penn Bioinformatics Core, for help with the NMM microarray and analyses, Robert Wilson for sharing the yeast deletion library, Rebecca Fry for sharing raw microarray data files for the *sgs1* experiments, Brian Brunk for help with preliminary analyses of QFP sequences, and members of the Johnson lab and Eric Brown, Nina Luning Prak and Shelley Berger for discussions and comments on the manuscript. S.G.H. is a Vagelos Scholar and was also supported by a Penn Genomics Institute internship. Funding support was provided by National Institutes of Health (5R01AG021521 to F.B.J.). Funding to pay the Open Access publication charges for this article was provided by the Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine.

*Conflict of interest statement.* None declared.

## REFERENCES

- Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Maizels,N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
- Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.*, **12**, 4336–4351.
- Mohaghegh,P., Karow,J.K., Brosh,R.M.Jr., Bohr,V.A.Jr and Hickson,I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.*, **29**, 2843–2849.
- Fry,M. and Loeb,L.A. (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CG)n. *J. Biol. Chem.*, **274**, 12797–12802.
- Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.
- Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
- Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
- Hanakahi,L.A., Sun,H. and Maizels,N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **274**, 15908–15912.
- Weitzmann,M.N., Woodford,K.J. and Usdin,K. (1997) DNA secondary structures and the evolution of hypervariable tandem arrays. *J. Biol. Chem.*, **272**, 9517–9523.
- Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes. Dev.*, **18**, 1618–1629.
- Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Pluckthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
- Fang,G. and Cech,T.R. (1993) The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell*, **74**, 875–885.
- Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.
- Gomez,D., O'Donohue,M.F., Wenner,T., Douarre,C., Macadre,J., Koebel,P., Giraud-Panis,M.J., Kaplan,H., Kolkes,A. *et al.* (2006) The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.*, **66**, 6908–6912.
- Cheung,I., Schertzer,M., Rose,A. and Lansford,P.M. (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.*, **31**, 405–409.
- Ding,H., Schertzer,M., Wu,X., Gertsenstein,M., Selig,S., Kammori,M., Pourvali,R., Poon,S., Vulto,I. *et al.* (2004) Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell*, **117**, 873–886.
- Crabbe,L., Verdun,R.E., Haggblom,C.I. and Karlseder,J. (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science*, **306**, 1951–1953.
- Du,X., Shen,J., Kugan,N., Furth,E.E., Lombard,D.B., Cheung,C., Pak,S., Luo,G., Pignolo,R.J. *et al.* (2004) Telomere shortening exposes functions for the mouse Werner and Bloom syndrome genes. *Mol. Cell. Biol.*, **24**, 8437–8446.
- Wu,X. and Maizels,N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res.*, **29**, 1765–1771.
- Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) 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.
- Hurley,L.H., Von Hoff,D.D., Siddiqui-Jain,A. and Yang,D. (2006) Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Semin. Oncol.*, **33**, 498–512.
- Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
- Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
- Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006)



- Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
27. Du,Z., Kong,P., Gao,Y. and Li,N. (2007) Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun.*, **354**, 1067–1070.
  28. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
  29. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
  30. Shirude,P.S., Okumus,B., Ying,L., Ha,T. and Balasubramanian,S. (2007) Single-molecule conformational analysis of G-Quadruplex formation in the promoter DNA duplex of the proto-oncogene C-Kit. *J. Am. Chem. Soc.*, **129**, 7484–7485.
  31. Etzioni,S., Yafe,A., Khateb,S., Weisman-Shomer,P., Bengal,E. and Fry,M. (2005) Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *J. Biol. Chem.*, **280**, 26805–26812.
  32. Dempsey,L.A., Sun,H., Hanakahi,L.A. and Maizels,N. (1999) 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**, 1066–1071.
  33. Giraldo,R. and Rhodes,D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *Embo J.*, **13**, 2411–2420.
  34. Ewens,W.J. and Grant,G.R. (2001) *Statistical Methods in Bioinformatics: An Introduction*. Springer-Verlag, New York.
  35. R Development Core Team. (2007). R Foundation for Statistical Computing.
  36. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
  37. Wilson,C.L. and Miller,C.J. (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, **21**, 3683–3685.
  38. Affymetrix. (2001). Microarray Suite 5.0 User's Guide.
  39. Wu,Z. and Irizarry,R.A. (2005) Stochastic models inspired by hybridization theory for short oligonucleotide arrays. *J. Comput. Biol.*, **12**, 882–893.
  40. Hong,F., Breitling,R., McEntee,C.W., Wittner,B.S., Nemhauser,J.L. and Chory,J. (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*, **22**, 2825–2827.
  41. Breitling,R. and Herzyk,P. (2005) Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data. *J. Bioinform. Comput. Biol.*, **3**, 1171–1189.
  42. Conover,W.J. (1999) *Practical Nonparametric Statistics*, 3rd edn. John Wiley and Sons, New York, USA.
  43. Zhang,Z. and Dietrich,F.S. (2005) Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res.*, **33**, 2838–2851.
  44. Hurowitz,E.H. and Brown,P.O. (2003) Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. *Genome Biol.*, **5**, R2.
  45. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
  46. Warner,J.R. (1989) Synthesis of ribosomes in *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **53**, 256–271.
  47. Forstemann,K. and Lingner,J. (2001) Molecular basis for telomere repeat divergence in budding yeast. *Mol. Cell Biol.*, **21**, 7277–7286.
  48. Patel,P.K. and Hosur,R.V. (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res.*, **27**, 2457–2464.
  49. Zhang,H., Roberts,D.N. and Cairns,B.R. (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, **123**, 219–231.
  50. Li,B., Pattenden,S.G., Lee,D., Gutierrez,J., Chen,J., Seidel,C., Gerton,J. and Workman,J.L. (2005) Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl Acad. Sci. USA*, **102**, 18385–18390.
  51. Babiarz,J.E., Halley,J.E. and Rine,J. (2006) Telomeric heterochromatin boundaries require NuA4-dependent acetylation of histone variant H2A.Z in *Saccharomyces cerevisiae*. *Genes. Dev.*, **20**, 700–710.
  52. Tham,W.H. and Zakian,V.A. (2002) Transcriptional silencing at *Saccharomyces* telomeres: implications for other organisms. *Oncogene*, **21**, 512–521.
  53. Bernstein,B.E., Liu,C.L., Humphrey,E.L., Perlstein,E.O. and Schreiber,S.L. (2004) Global nucleosome occupancy in yeast. *Genome Biol.*, **5**, R62.
  54. Pina,B., Fernandez-Larrea,J., Garcia-Reyero,N. and Idrissi,F.Z. (2003) The different (sur)faces of Rap1p. *Mol. Genet. Genomics*, **268**, 791–798.
  55. Lee,T.I., Rinaldi,N.J., Robert,F., Odom,D.T., Bar-Joseph,Z., Gerber,G.K., Hannett,N.M., Harbison,C.T., Thompson,C.M. *et al.* (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
  56. Arthanari,H., Basu,S., Kawano,T.L. and Bolton,P.H. (1998) Fluorescent dyes specific for quadruplex DNA. *Nucleic Acids Res.*, **26**, 3724–3728.
  57. Ren,J. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.
  58. Han,H., Cliff,C.L. and Hurley,L.H. (1999) Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry*, **38**, 6981–6986.
  59. De Cian,A. and Mergny,J.L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res.*, **35**, 2483–2493.
  60. Gasch,A.P., Huang,M., Metzner,S., Botstein,D., Elledge,S.J. and Brown,P.O. (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell.*, **12**, 2987–3003.
  61. Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D. and Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell.*, **11**, 4241–4257.
  62. Fahy,D., Conconi,A. and Smerdon,M.J. (2005) Rapid changes in transcription and chromatin structure of ribosomal genes in yeast during growth phase transitions. *Exp. Cell Res.*, **305**, 365–373.
  63. Fry,R.C., Sambandan,T.G. and Rha,C. (2003) DNA damage and stress transcripts in *Saccharomyces cerevisiae* mutant sgs1. *Mech. Ageing Dev.*, **124**, 839–846.
  64. Huber,M.D., Duquette,M.L., Shiels,J.C. and Maizels,N. (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.*, **358**, 1071–1080.
  65. van Vugt,J.J., Ranes,M., Campsteijn,C. and Logie,C. (2007) The ins and outs of ATP-dependent chromatin remodeling in budding yeast: biophysical and proteomic perspectives. *Biochim. Biophys. Acta*, **1769**, 153–171.
  66. Li,B., Carey,M. and Workman,J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.
  67. Timmers,H.T. and Tora,L. (2005) SAGA unveiled. *Trends Biochem. Sci.*, **30**, 7–10.
  68. Zhang,H., Richardson,D.O., Roberts,D.N., Utley,R., Erdjument-Bromage,H., Tempst,P., Cote,J. and Cairns,B.R. (2004) The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres. *Mol. Cell Biol.*, **24**, 9424–9436.
  69. Meneghini,M.D., Wu,M. and Madhani,H.D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell*, **112**, 725–736.
  70. Qiu,H., Hu,C., Wong,C.M. and Hinnebusch,A.G. (2006) The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. *Mol. Cell Biol.*, **26**, 3135–3148.
  71. Dehe,P.M. and Geli,V. (2006) The multiple faces of Set1. *Biochem. Cell. Biol.*, **84**, 536–548.
  72. Kao,C.F., Hillyer,C., Tsukuda,T., Henry,K., Berger,S. and Osley,M.A. (2004) Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. *Genes. Dev.*, **18**, 184–195.
  73. Askree,S.H., Yehuda,T., Smolikov,S., Gurevich,R., Hawk,J., Coker,C., Krauskopf,A., Kupiec,M. and McEachern,M.J. (2004) A genome-wide screen for *Saccharomyces cerevisiae* deletion

- mutants that affect telomere length. *Proc. Natl Acad. Sci. USA*, **101**, 8658–8663.
74. Yuan,G.C., Liu,Y.J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**, 626–630.
  75. Giresi,P.G., Gupta,M. and Lieb,J.D. (2006) Regulation of nucleosome stability as a mediator of chromatin function. *Curr. Opin. Genet. Dev.*, **16**, 171–176.
  76. Granneman,S. and Baserga,S.J. (2005) Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. *Curr. Opin. Cell Biol.*, **17**, 281–286.
  77. Linskens,M.H. and Huberman,J.A. (1988) Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **8**, 4927–4935.
  78. Li,J.L., Harrison,R.J., Reszka,A.P., Brosh,R.M.Jr, Bohr,V.A., Neidle,S. and Hickson,I.D. (2001) Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry*, **40**, 15194–15202.
  79. Miller,K.M., Rog,O. and Cooper,J.P. (2006) Semi-conservative DNA replication through telomeres requires Taz1. *Nature*, **440**, 824–828.
  80. Kudlow,B.A., Kennedy,B.K. and Monnat,R.J.Jr. (2007) Werner and Hutchinson-Gilford progeria syndromes: mechanistic basis of human progeroid diseases. *Nat. Rev. Mol. Cell Biol.*, **8**, 394–404.
  81. Garkavtsev,I.V., Kley,N., Grigorian,I.A. and Gudkov,A.V. (2001) The Bloom syndrome protein interacts and cooperates with p53 in regulation of transcription and cell growth control. *Oncogene*, **20**, 8276–8280.
  82. Kyng,K.J. and Bohr,V.A. (2005) Gene expression and DNA repair in progeroid syndromes and human aging. *Ageing Res. Rev.*, **4**, 579–602.
  83. Kyng,K.J., May,A., Kolvraa,S. and Bohr,V.A. (2003) Gene expression profiling in Werner syndrome closely resembles that of normal aging. *Proc. Natl Acad. Sci. USA*, **100**, 12259–12264.
  84. Balajee,A.S., Machwe,A., May,A., Gray,M.D., Oshima,J., Martin,G.M., Nehlin,J.O., Brosh,R., Orren,D.K. *et al.* (1999) The Werner syndrome protein is involved in RNA polymerase II transcription. *Mol. Biol. Cell*, **10**, 2655–2668.
  85. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
  86. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
  87. De Cian,A., Lacroix,L., Douarre,C., Temime-Smaali,N., Trentesaux,C., Riou,J.F. and Mergny,J.L. (2007) Targeting telomeres and telomerase. *Biochimie* doi:10.1016/j.biochi.2007.07.011.