

# The Interplay between G-quadruplex and Transcription



Nayun Kim<sup>1,\*</sup>

Department of Microbiology and Molecular Genetics, The University of Texas Health Science Center at Houston; The University of Texas Graduate School of Biomedical Sciences, Houston, TX, USA

## ARTICLE HISTORY

Received: August 19, 2017  
Revised: November 22, 2017  
Accepted: December 21, 2017

DOI:  
10.2174/0929867325666171229132619



CrossMark

**Abstract:** G4 DNA is a non-canonical DNA structure consisting of a stacked array of G-quartets held together by base pairing between guanine bases. The formation of G4 DNA requires a cluster of guanine-runs within a strand of DNA. Even though the chemistry of this remarkable DNA structure has been under investigation for decades, evidence supporting the biological relevance of G4 DNA has only begun to emerge and point to very important and conserved biological functions. This review will specifically focus on the interplay between transcription and G4 DNA and discuss two alternative but interconnected perspectives. The first part of the review will describe the evidence substantiating the intriguing idea that a shift in DNA structural conformation could be another layer of non-genetic or epigenetic regulator of gene expression and thereby an important determinant of cell fate. The second part will describe the recent genetic studies showing that those genomic loci containing G4 DNA-forming guanine-rich sequences are potential hot-spots of genome instability and that the level and orientation of transcription is critical in the materialization of genome instability associated with these sequences.

**Keywords:** G4 DNA, transcription, R-loops, Top1, supercoiling, genome stability.

## 1. INTRODUCTION

Free guanine bases in solution readily interact with each other *via* Hoogsten bonds to form a four-membered ring-like structure referred to as a G-quartet [1]. G-quadruplex or G4 DNA, comprising multiple G-quartets stabilized by stacking also readily form from single-stranded oligonucleotides in solution. The presence of various cations, such as  $K^+$ ,  $Na^+$ ,  $Ca^{2+}$ , and  $Sr^{2+}$ , facilitates the G4 DNA formation with  $K^+$  having the most stabilizing effect. Nucleosides between guanine-runs are incorporated into the structure as loops between G-quartets, and the size of the loops can determine the relative stability of various G4 DNA configurations. Sequences with potential to form G4 DNA or "G4 motifs" were first noted at the telomeres, ribosomal DNA arrays, Immunoglobulin (Ig) heavy chain

loci, Chromosomal Fragile Sites (CFSs) and G/C-rich micro- or mini-satellites. More recently, searching for sequences with at least 4 G-runs with the loop length of less or equal to 7 nt, computational analyses identified ~1,400 and ~370,000 putative G4 motifs in the *Saccharomyces cerevisiae* nuclear genome and the human genome, respectively [2, 3]. The possible biological function of G4 DNA was first inferred when the guanine-rich telomere sequences were shown to assume the secondary structure through guanine-guanine base-pairing and when many of the proteins known to be telomere-localized were shown to bind G4 DNA with high affinity [4, 5]. Whether the G4 motifs besides the telomeric repeats would form G4 DNA *in vivo* stable enough to carry out a particular cellular function has been debated for long. Here, I discuss how transcription critically impacts the conformational change of G4 motifs in the genome into the G4 DNA structure and, reciprocally, how the G4 DNA both positively and negatively regulates the level of transcription.

\*Address correspondence to this author at the Department of Microbiology and Molecular Genetics, 6431 Fannin Street MSE R262, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; Tel: 713 500 5597; Fax: 713 500 5499; E-mail: [nayun.kim@uth.tmc.edu](mailto:nayun.kim@uth.tmc.edu)

## 2. TRANSCRIPTION REGULATION BY G4 DNA

### 2.1. *In silico* Evidence of G4 as a Regulator of Transcription

Analysis of the list of the putative G4 motifs in the human genome identified *in silico* uncovered that G4 motifs are highly enriched at the promoter regions of ~20,000 human genes compared to the whole genome-wide distribution [6]. In fact, they found that 42.7% of the genes surveyed contained at least one G4 motif within 1 Kb upstream of the transcription start site (TSS). Additionally, enrichment of G4 motifs correlated with previously characterized gene regulatory elements including enhancers, conserved transcription factor binding sites, and nuclease hypersensitive sites (NHS) [6, 7]. NHS are indicative of more accessible regions of the genome and broadly mark regulatory sequences. G4 motifs are also enriched within 500 nt downstream of TSS of human genes [8]. In yeast, where the extensive annotation of ORFs and regulatory regions allow more comprehensive characterization of the distribution pattern of G4 motifs, the correlation between gene regulatory regions and the G4 motifs was identified as the most notable aspect [9]. Although the yeast genome is relatively low in GC content (38 to 39%) compared to the mammalian genomes (46% in human and 51% in mouse), G4 motifs were enriched by 6-fold at promoter regions defined as -850 to -50 relative to TSS compared to genome-wide distribution. Another smaller peak of enrichment was present in the regions within 400 nt downstream of TSS. When yeast cells were treated with the G4 ligand N-methyl mesoporphyrin IX (NMM), there was a significant up-regulation of expression for many of the genes with G4 motifs in the promoter regions. A putative role in the important cellular function such as transcriptional regulation means that the evolutionary conservation would be expected of G4 motifs. Capra *et al.* compared the genomes of *S. cerevisiae* and six other yeast species in order to determine the level of conservation of the G4 motifs in these closely related species [10]. Allowing for up to 50 nt loop size, G4 motifs identified in these genomes showed significantly higher conservation than would be expected with 34 of 552 motifs being conserved in all 7 yeast species surveyed. In addition, within a G4 motif, nucleotides at the positions where mutation would lead to the disruption of G4 DNA structure were more highly conserved than those at non-disruptive positions, indicating that these sequences were evolved to retain their capability to form the secondary structure. G4 motifs identified in this study were strongly associated with gene promoters. Overall, this suggests that the regulation of transcrip-

tion could be a function of G4 DNA conserved from yeast to human.

### 2.2. Confirmation of *In Vivo* G4 DNA Formation at Promoters

For many of the G4 motifs identified proximal to the promoter regions, the potential to fold into G4 DNA structure has been verified by *in vitro* approaches, including Dimethyl Sulfate (DMS) footprinting, Circular Dichroism (CD) and Electrophoretic Mobility Shift Assay (EMSA). However, the critical proof of *in vivo* G4 DNA formation at these same sequences has been absent until the recent identification and characterization of antibodies specific to G4 DNA [11-14]. Among these, two recombinant antibodies, HF2 and BG4, were isolated by screening the Tomlinson J library of phage-displayed single-chain variable fragment (scFv) antibodies. Both of these scFv antibodies bind specifically to G4 DNA *in vitro* and have been successfully used in the detection of cellular G4 DNA in immunohistochemistry (IHC) and/or chromatin immunoprecipitation (ChIP) analyses [13-15]. ChIP with the HF2 antibody followed by deep sequencing identified 175 HF2-bound peaks in human breast cancer cells that contain the consensus G4 motifs with the maximum loop size of 7nt [14]. In order to determine the biological consequence of the G4 DNA, the authors determined the effect of the G4-stabilizing ligand pyridostatin (PDS) on the gene expression of a set of 8 genes identified to have *in vivo* G4 DNA formation at the promoter regions by HF2-ChIP-seq experiments. Significant changes were observed for 6 of these genes, confirming that the presence of G4 DNA does regulate transcription from these promoters. Interestingly, both significant down-regulation and up-regulation of the gene expression levels by PDS-treatment were observed suggesting that the G4 DNA-mediated regulation involves both activation and repression of transcription.

BG4 antibody binds to DNA G-quadruplex with a very high affinity ( $K_d = 1$  to 2 nM) with no detectable binding to single strand RNA or DNA [13]. Human cells incubated with BG4 exhibit discreet foci that disappear upon treatment with DNase or competition with G4 DNA-forming oligos. For the mitotic spread of chromosomes, BG4 bound at telomeres are visible at each ends of the chromosome as expected with the human telomeres consisting of G-rich TTAGGG repeats. However, the majority of BG4 foci were present outside of the telomeric regions. The identity of these non-telomeric BG4 foci were recently clarified by the ChIP-seq experiment where the fragments of protein/ nu-

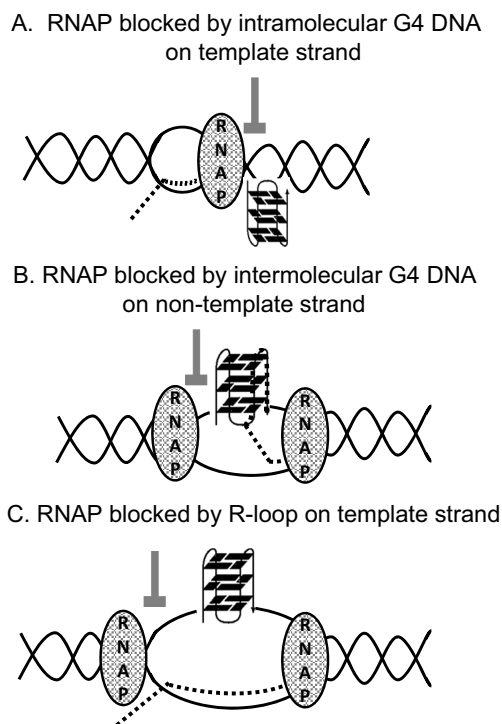
cleosome-crosslinked human chromosomes are immunoprecipitated with bead-immobilized BG4 and followed up by deep sequencing [15]. Of the ~10,000 BG4-bound peaks identified in the immortalized human keratinocytes, 87% conformed to a G4 forming sequence. However, the number of the BG4 peaks were significantly lower than would be predicted by *in silico* analysis indicating that the presence of G-run containing sequence does not constitute the sufficient condition for G4 DNA folding *in vivo* and that not all G4 motifs assume the G4 DNA conformation. This also implies that the number and distribution of G4 DNA could vary by the epigenetic state determined by the cell type and status. Surprisingly, only 21 % of the BG4 peaks fell within the strictly canonical G4 consensus sequence of  $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$  indicating that a broader definition of G4 motifs needs to be considered. Similar results were obtained when a different scFv-type antibody named D1, specific for parallel G4 DNAs, was expressed from a plasmid construct in cultured human cervical carcinoma SiHa cells [16]. The stringent G4 consensus of  $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$  comprised minority of the D1 binding sites identified by the ChIP-seq approach in this study. More of the D1 binding sites fell into a broader consensus allowing for longer loop lengths ( $G_{3+}N_{1-12}G_{3+}N_{1-12}G_{3+}N_{1-12}G_{3+}$ ). One interpretation of the overrepresentation of non-canonical G4 forming sequences is that binding of BG4 or D1 antibody potentially stabilizes very transient G4 DNA *in vivo*. This forced stabilization of G4 structures by the antibody binding could potentially lead to a high rate of false positives in future studies using this or other similar approaches. Therefore, biochemical probing to clarify the relative thermal stability of G4 DNA/ antibody complex compared to the unbound G4 DNA is a critical future direction. An alternative explanation for the non-canonical G4-forming sequences identified as BG4- or D1-binding sites is that, depending on the context, the loops larger than 7-nt can be stably accommodated in G4 DNA configurations *in vivo* and are therefore biologically relevant. In contrast to the report indicating that shorter loop lengths of <4 nt result in the greater thermal stability of the G4 DNA *in vitro* and the higher risk of genome instability *in vivo* [17], others have reported of stable G4 DNA with significantly larger loops not conforming to the definition of G4 sequence widely used in computational analysis ( $G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$ ) [18]. Additionally, one of the two GC-rich sequence motifs identified at the promoter of the human BCL2 gene formed a stable parallel G4 DNA with 13-nt loop [19].

## 2.3. How does the Presence of G4 DNA Regulate Transcription?

### 2.3.1. G4 DNA as the Physical Obstacle to the RNA Polymerase Complex

Compared to the large set of indirect, mostly computational, evidence implicating G4 DNA in the gene expression regulation, studies into the mechanical basis of G4 DNA-dependent activation or repression of transcription has been quite sparse. G4 DNA as a DNA replication block has been well documented through multiple *in vitro* primer extension experiments or polymerase stop assays involving the minimal, required components of DNA polymerase activity [20-22]. Accordingly, as will be discussed below, G4 DNA-forming sequences appear as genome instability hot-spots *in vivo*. It is therefore postulated that G4 DNA can also block transcription by becoming a physical obstacle in the way of the RNA polymerase. According to *in vitro* transcription experiments with only the minimal, required proteins, a G4 motif located in the transcribed region can indeed impede the movement of RNA polymerase complex [23-26]. When the G4 motifs are present on either strand within the transcribed region, the intramolecular G4 DNA on the template strand as well as the intermolecular G4 DNA composed of the non-template DNA strand and the nascent RNA would co-transcriptionally form [27] and physically interfere with subsequent rounds of the RNA polymerase movement (Fig. 1A and 1B). Alternatively, the intramolecular G4 DNA on the non-template strand could interfere with transcription by first impeding the reannealing of two DNA strands behind the RNA polymerase complex and thereby affording a favorable condition for the formation and stabilization of RNA:DNA hybrids involving the template DNA strand (Fig. 1C) [26]. In support of this mode of transcription block, removing RNA:DNA hybrids by the addition of recombinant RNase H to the transcription reaction facilitated efficient RNA polymerase elongation past the G4 sequence in an *in vitro* transcription assay.

Another piece of evidence supporting the biological function for the G4 DNAs near TSS comes from the analysis of transcription pause sites in human T cells. RNA Pol II-mediated transcription occurs in phases of initiation, elongation, and termination [28]. Between initiation and elongation phases, Pol II transcription undergo pauses, which might serve as a check point. Such “promoter proximal pausing” can occur within 50 nt of TSS in mammalian cells and the release from this paused state into the fully elongating RNA polymerase complex is considered to be the rate-limiting step in



**Fig. (1). Mechanism of transcription block by G4 DNA.**

**A)** G4 DNA formed from the template or transcribed DNA strand becomes a physical block to the RNA polymerase complex (RNAP). Dotted line indicates the nascent transcript. **B)** An intermolecular G4 DNA formed as a hybrid between the non-transcribed DNA strand and the nascent RNA (dotted line) will block the on-coming RNAP during the next round of transcription. **C)** G4 DNA formed from the non-template or non-transcribed DNA strand stabilizes the annealing of nascent RNA (dotted line) to the template DNA strand. This R-loop will block the on-coming RNAP during the next round of transcription.

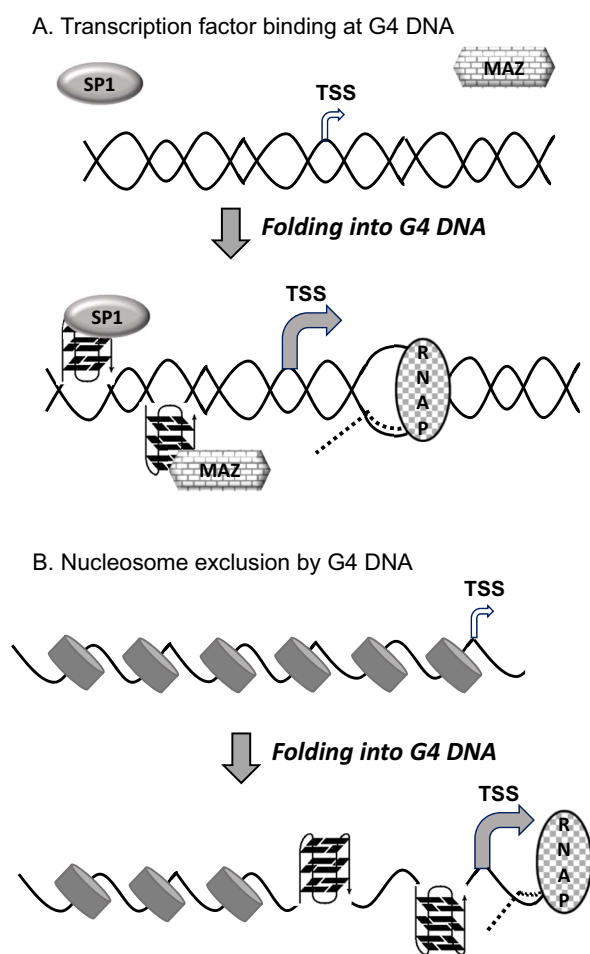
transcription. Corroborating the results of *in vitro* transcription experiments where G-runs blocked transcription by T7 RNA polymerase [24, 26], Eddy *et al.* described a correlation between the G4 motifs near TSS and the promoter-proximal transcriptional pausing [29]. Another report by Du *et al.*, however, posited a somewhat contradictory view of the role of G4 DNA located within the transcribed areas [8]. By considering ~8,000 human genes identified to contain G4 motifs proximal to TSS, Du *et al.* determined that > 5,000 genes contained at least one G4 motif within 500 nt downstream of the TSS. More of these downstream G4 motifs were present on the coding (non-template) strand and correlated with significantly higher expression levels, leading the authors to suggest a transcription-activating role for these G4 motifs. One possible mechanistic model to explain this correlation is that the G4 folding of non-

template strand keeps the template strand single-stranded, thereby, facilitating the subsequent rounds of RNA polymerization [8]. However, the argument for the transcription-stimulatory role of G4 in this study was partly based on the increased RNA polymerase occupancy near the TSS that correlate G4 motifs, which can be alternatively interpreted as the evidence of the RNA polymerase complex pausing.

### 2.3.2. G4 DNA as the Docking Site for Transcription Factors

As described above, transcription block by assembling a structural barrier can sufficiently account for the transcriptional repression by the G4 DNA located in the first intron, downstream of TSS. However, this model of transcriptional regulation by G4 DNA does not provide an explanation for how G4 DNA located 5' to TSS in the promoter regions block or enhance transcription. Entirely different perspective is required to understand the G4 DNA function as a transcription enhancer. One possible model is that G4 DNA can be a high-affinity binding site of certain transcription factors (Fig. 2A). A well-established example is the human SP1 protein, which is a zinc-finger transcription factor that is ubiquitously expressed and controls the expression of many house-keeping genes. It had been initially characterized as a typical sequence specific double strand DNA binding protein with the minimal consensus of 5'-GGGCGG-3'. However, an empirically determined SP1 binding site at the promoter of the oncogene c-KIT was shown to not contain the consensus binding sequence but instead to form the G4 structure [30]. *In vitro* binding assays were carried out to demonstrate that SP1 binds with higher affinity to the G4 DNA formed by the single strand oligonucleotide representing the SP1 binding site at the c-KIT promoter. An analysis of the actual genome-wide SP1-binding sites determined by a ChIP-seq experiment showed that ~36% of the SP1-occupied sites did not contain the 5'-GGGCGG-3' consensus binding sequence while a majority of such sites contained one or more G4 motifs. Then, SP1-dependent transactivation is not simply genetically determined (*i.e.* DNA sequence-dependent) but can significant dependent on the conformation of the DNA and thus dependent on the factors affecting the DNA conformation. Another example of G4-binding transcription factors is the Myc-associated zinc-finger protein (MAZ), which binds to the promoter of KRAS gene [31]. The promoter of KRAS gene contains a G4 forming sequence that coincides with a nuclease hypersensitivity site and overlaps the

binding site for a known transcription factor MAZ [32]. Stabilization of the G4 DNA structure favored MAZ-binding at the KRAS promoter leading to activation of transcription whereas point mutations disrupting the G4 DNA conformation led to down-regulated expression of KRAS [31]. Recently, an essential general transcription factor PC4 (hSub1), which was previously characterized as a single strand DNA binding protein [33] was also identified as a high-affinity G4 DNA binding protein. Discovery of other structure-specific transcription trans-activators could be on the way, which would shed more light on how binding of these proteins could be the mediating step in the G4 DNA-dependent activation of transcription.



**Fig. (2). Mechanism of transcription activation by G4 DNA.** **A)** Transcription factors such as SP1 and MAZ recognize and assemble onto the G4 DNA structure to activate transcription downstream. **B)** Formation of G4 DNA proximal to the Transcription Start Site (TSS) excludes the nucleosome (gray barrels) from this area, thereby activating transcription downstream. (The color version of the figure is available in the electronic copy of the article).

Nucleolin (NCL), a highly abundant and conserved protein largely localized in the nucleolus [34], is another highly specific G4 DNA-binding protein that can serve as a transcription activator. NCL is a multifunctional protein with a major role in the ribosomal RNA maturation. In hematopoietic cells, its function is implicated in wide-range of cellular activities including B lymphocyte maintenance and regulation of apoptosis and inflammation [35]. Overexpression and mislocalization of NCL is a common biomarker of variety of cancers [36, 37]. NCL contains tandem RNA Recognition Motifs (RRMs) as well as multiple RGG (arginine/glycine/glycine) boxes at the C-terminal domain, both of which contribute to its high-affinity interaction with G4 DNA ( $K_d \cong 1$  nM) [38]. Together with hnRNP D, NCL forms a lymphocyte-specific complex LR1 (lipopolysaccharide responsive factor 1), which binds at the G4 DNA-forming Immunoglobulin heavy chain (IgH) switch regions [39]. When bound at the G4 DNA present in the promoter of the human VEGF gene, NCL functions as transcription activator [40] in a manner similar to the above-described SP1 and MAZ. However, the significance of NCL-G4 DNA interaction in the transcription process is complicated with apparently incongruous reports. NCL binds to the G4 motif that was recently identified at the Long Terminal Repeat (LTR) promoter of Human Immunodeficiency Virus-1 (HIV-1) and functions like a molecular chaperon to facilitate G4 DNA folding [41, 42]. Rather than leading to transcriptional activation like in the case of its binding to the VEGF promoter, the consequence of NCL interaction with the HIV-1 LTR promoter is to repress transcription. Overexpression of NCL in human breast epithelial cells transformed with the LTR promoter reporter construct led to the transcriptional silencing of the LTR promoter [42]. Conversely, siRNA-mediated depletion or aptamer-mediated disruption of NCL led to activation of transcription from the LTR promoter. NCL similarly functions as transcriptional repressor at the c-MYC promoter [43]. According to *in vitro* assays carried out with HeLa nuclear extract, binding of NCL to the G4 motif present proximal to the P1 promoter of c-MYC inhibits transcription and represses expression of a reporter gene under P1 promoter control *in vivo* [44, 45]. Although the mechanism of how NCL-G4 DNA complex inhibits transcription from P1 promoter both *in vivo* and *in vitro* is not entirely clear, the very close proximity of the G4 motif to the TSS (-142 to -115) leads to the speculation that this alteration in the DNA topology might interfere with the binding of other necessary transcriptional activators or impede the assembly of the transcription initiation complex.

### 2.3.3. G4 DNA and Nucleosome-free Region

A large majority of actively transcribed genes in yeast and human share a distinct nucleosome depleted region just upstream of TSS (reviewed in [46]). In addition, other regulatory regions (*i.e.* transcription factor binding sites) of active genes are relatively nucleosome-free. Comparison of the nucleosome occupied sequences in the genomes of nematode *C. elegans* and human with the computationally determined G4 motifs showed that potential G4 sequences were frequently found outside of nucleosome-bound regions [47]. For both *C. elegans* and human, this correlation between G4 motifs and the nucleosome-depleted regions were stronger when considering the subset of G4 motifs with higher expected stability (*i.e.* those with smaller loops between G-runs). Similar results were obtained when the nucleosome occupancy was compared to the G4 motifs identified in the yeast genome [48]. Nucleosome-free regions can be separated from the protein/nucleosome-dense regions of the genome by centrifugation and mapped using a technique called “formaldehyde-assisted isolation of regulatory elements with sequencing (FAIRE-seq)” [49]. In a normal human keratinocyte cell line, 98% of the G4 DNA peaks as determined by ChIP with the G4-specific BG4 antibody overlapped with the nucleosome-free regions identified by FAIRE-seq approach [15]. These results indicate that the stable formation of G4 DNA and possibly other non-canonical DNA secondary structures preclude the sharp bend in DNA required for nucleosome assembly and thus serve to locally exclude nucleosome to maintain open conformation (Fig. 2B). Nucleosome-exclusion by G4 DNA could be a mechanism of enhancing the transcription initiation and overall transcription rate. Alternatively, the correlation of G4 DNA and nucleosome depleted regions could be viewed as an indication that the unraveling of double strand DNA and folding into the G4 conformation is hindered when the DNA is compactly wrapped around a nucleosome.

## 2.4. The Functional Relevance of the G4 Motifs to the Transcriptional Regulation

### 2.4.1. G4 DNA Resolution and Transcription

The ChIP-seq experiments with the G4-structure specific antibodies HF2 and BG4 indicate that only a subset of the G4 motifs identified *in silico* assumes the G4 DNA conformation *in vivo*. Unlike the *in vitro* experiments carried out with naked single-stranded oligonucleotides in buffers, factors other than the sequence must contribute to the G4 DNA folding at a given G4 motif. If

the structural transition from the double-stranded B DNA to the tetrahelical G4 DNA with G-G base pairings is required for the transcriptional regulation, any proteins that impinge on the equilibrium of this transition are expected to affect the expression of genes with G4 motifs at promoters. Several DNA helicases have been characterized to unwind G4 DNA with higher efficiency than the canonical B DNA (reviewed in [5]). In human cells, a defect in the BLM helicase, a RecQ family 3' to 5' DNA helicase that recognizes and unwinds G4 structures [50], led to significant changes in the expression level of genes with G4 motifs near TSS [51]. This result not only reveals that the G4 DNA located at promoter regions could have a regulatory role but also suggests that the G4 DNA-dependent transcriptional dysregulation could be a pathogenic feature. Patients with Bloom syndrome, who are characterized genetically by the recessive autosomal mutations in the BLM gene, present with symptoms such as the gross developmental defect and the predisposition to leukemia and lymphoma. Similarly, in yeast cells with a deletion of the gene encoding the homolog of BLM helicase, Sgs1 (*sgs1Δ*), many genes with the promoter-proximal G4 motifs were selectively down-regulated [9]. WRN is another RecQ family 3' to 5' DNA helicase with the biochemically demonstrated capability to unwind G4 DNA structures. Loss of function mutations in WRN is responsible for the premature aging disease Werner Syndrome. Comparing the gene expression patterns in fibroblasts from Werner syndrome patients to those from healthy donors, G4 motifs were found to be highly enriched at those genes down-regulated in patient cells [52]. G4 motifs were found enriched on both the template and non-template strands, and upstream of TSS, and in the first intron. On the other hand, those genes up-regulated in Werner Syndrome patient fibroblasts were generally G4-depleted upstream of TSS. These results were corroborated by another more recent report [53]. In an independent experiment, the highly significant correlation between G4 motifs in the promoter regions and the differential gene expression levels was also made in fibroblasts from patients with Bloom or Werner syndromes but not from patients with Rothmund-Thompson syndrome [54]. Rothmund-Thompson syndrome results from a defect in RECQ4, which is a BLM/WRN-related RecQ family DNA helicase but, unlike BLM or WRN, lacks the RQC domain that directs its helicase activity to G4 DNA.

### 2.4.2. G4 DNA in Vertebrate Embryonic Development

Despite the relatively low GC content of the zebrafish genome (38.6%) compared to mammalian species such as human (46.1%) or mouse (51.2%) [55], the



computational search showed a significant enrichment of G4 forming sequences within 1000 bp upstream of TSS of zebrafish genes [56]. According to this study, those genes containing G4 motifs at the promoter proximal regions identified in zebrafish were 73% and 60% shared with human and mouse, respectively, indicating that G4 motifs in these genes were maintained and positively selected for their functional roles. Of the 926 genes identified with the promoter-proximal G4 motifs in the zebrafish genome, 120 genes were previously characterized to be involved in various pathways of development. When the G4 motifs were cloned into the reporter plasmid, G4 sequence led to stimulation of the expression of the reporter gene, which is different from the G4-dependent transcription-repression proposed by earlier work. When zebra fish embryo was injected with G4 binding ligand TMPyP4, it led to lower expression of the select G4-containing genes tested. The interaction with TMPyP4, which can stabilize or destabilize G4 structures depending on the target [57], resulted in less stable G4 structures for these particular G4 motifs according to the *in vitro* circular dichroism spectroscopic analysis [56]. This is in contrast to many other studies described in this review where TMPyP4 functions to regulate transcription of the promoter G4-containing genes by stabilizing G4 DNA conformation. Similar transcriptional down-regulation was observed when the cells were injected with the anti-sense oligos specific to the G4 motifs at the *col2a1*, *fzd5* and *nog3* genes with promoter proximal G4 motifs. These Anti-Sense Oligos (ASOs) were designed to base-pair with the G-run containing strand and thereby inhibit G4 DNA folding. Overall these results point to transcription-activation mediated by G4 DNA in zebrafish. Underscoring the significance of the G4 DNA-mediated gene regulation in the embryonic development of the fish, each ASO targeting the G4 motif present at *col2a1*, *fzd5* or *nog3* gene led to a defect in body morphology, eye development, or head morphology, respectively.

#### 2.4.3. G4 DNA in Stress Response in Plants

In plants, G4 DNAs are thought to be involved in the transcriptional response to various environmental stress conditions (reviewed in [58]). In *Arabidopsis thaliana*, a large number of genes are differentially expressed in drought conditions, and it was noted that 45% of these drought-responsive genes contain at least one G4 motif [59]. When considering the G4 motifs identified in the genome of *Zea mays* (maize) using the stringent definition of G4 motifs ( $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ ) with the maximum loop size of 7 nt, ~24 % of all

expressed genes in this plant species contained at least one G4 motif [60]. The distribution pattern of the G4 motifs within genes were similar to that described for yeast and the metazoan species; the regions proximal to TSS were significantly enriched for G4 motifs with the highest density within the 1 – 100 nt downstream of the TSS. Those G4 motifs present downstream of the TSS were highly biased to be on the template/anti-sense strand, which makes it likely that they could regulate transcription by blocking the RNA polymerase movement (Fig. 1A). In addition, the presence of multiple G4 motifs proximal to TSS correlated with those maize genes that are highly conserved in four other related pan-grass species, indicating that these G4 motifs were evolutionarily conserved due to their functional role. Many of the G4 motif-containing genes fell into the categories of energy signaling, sugar metabolism, hypoxia, and response to DNA damage, among others. More comprehensive survey of the genomes of 15 different plant species found that the consensus G4-forming sequences with runs of 3 or more consecutive guanines are very frequently found in the genomes of monocot species including *Setaria italica* (millet) and *Oryza sativa* (rice), two of the most important food crops [61]. The pattern of distribution was consistent with that found in *A. thaliana* [59]; G4 motifs were enriched at the promoter regions and relatively depleted in the coding region [61]. Moreover, the gene ontology analysis showed that many of the orthologous genes containing conserved G4 motifs were involved in several distinctive biological pathways, namely, reproductive development, ion transmembrane transport, and regulation of gene expression. Even though the empirical evidence is yet to come, according to the computational analyses described above, G4-mediated transcriptional regulation is expected to be conserved in these plant species.

#### 2.4.4. Significance of G4-mediated Gene Regulation and Proto-oncogenes

G4 motifs have also been identified at the promoter regions of many important proto-oncogenes (Reviewed in [62]). One example is the important proto-oncogene c-MYC, which, as described above, contains a G4 motif found proximal to the P1 promoter where the majority of transcription is initiated [43, 63]. This 27-nt sequence conforms to the strict definition of G4 forming sequence and supports the formation of very stable G4 structures with three stacked G-tetrads. Concurrent with the identification of the G4 motif at the P1 promoter, Siddiqui-Jain *et al.* demonstrated the G4-dependent regulation of P1 promoter using two

Burkitt's lymphoma cell lines with different c-MYC translocations [43]. Stabilization of G4 DNA by TMPyP4 treatment resulted in a significant down-regulation of the c-MYC transcription only in the cell line with intact P1 promoter but not in the cell line where P1 promoter is lost due to the translocation. Conversely, a single G to A mutation of this P1-associated G4 motif disrupting the G4 conformation led to elevation in c-MYC transcription. Together, these data support the model that the presence of stable G4 DNA silences the P1 promoter of c-MYC, possibly by excluding certain transcription factors from binding to this region. Multiple putative G4 motifs are identified at the promoter of KRAS, of which expression level can be repressed by treating with G4-stabilizing ligand TMPyP4 [64]. Another proto-oncogene BCL2 contains multiple G4 motifs at the promoter region [65, 66] as well as regions corresponding to the 3' UTR [67], which coincides with the major break region (MBR) where a large number of leukemia-associated interchromosomal translocations are found. Putative G4 DNA-forming sequences are also identified at the promoters of VEGF, HIF1a, Ret and c-KIT [62]. When the density of G4 motifs at 55 tumor suppressor genes and 95 proto-oncogenes were compared, a distinct pattern was uncovered. The G4-forming potential is significantly low at the tumor suppressor genes and high at the proto-oncogenes, which suggests possible evolutionary selection of G4 motifs based on the function of the genes [68].

#### 2.4.5. The Clinical Application of G4 DNA

Overall, the expression of a large number of oncogenes appears to be regulated by G4 DNA. Moreover, the transcription at these oncogenes are primarily repressed by the presence of a stable G4 DNA either at the promoter or in the 5' end of the transcribed area. This finding advanced the idea that G4-stabilizing small chemicals or ligands would be a promising tool to modulate oncogene expression to control cancer growth. TMPyP4 is a cationic porphyrin with a macrocyclic ring structure that binds to and stabilizes G4 DNA [57, 69, 70]. As already described, this compound has been extensively used to show that genes with G4 motifs in the promoter can be modulated by the addition of G4 ligand that stabilizes the otherwise transient G4 structure. Although its photosensitivity and acute cytotoxicity render TMPyP4 unsuitable for clinical application, the derivatives of this macrocyclic porphyrin are being developed as the drug targeting those cancers with overexpression of c-MYC [71]. Another G4 ligand telomestatin is a natural compound

with selective binding to G4 DNA whose anti-proliferative activity has proven effective in several different studies (reviewed in [57]). Possible clinical application of G4 binding chemicals is illustrated by the recent finding that the expression of the important DNA repair gene Brcal in post-mitotic rat neurons is down-regulated by the G4 ligand Pyridostatin (PDS) [72]. The PDS treatment also resulted in a sharp increase in DNA breaks and, compounded by the diminished repair capacity due to Brcal down-regulation, induced severe neurotoxicity. When the precise mechanism and targets of the G4-mediated gene regulation emerge through future research, the genotoxic effect and the transcriptional repression effect of various G4 ligands could be exploited to achieve synergy in eliminating cancer cells.

Another way G4 DNA is being targeted in cancer treatment is through G4 DNA binding proteins. Nucleolin (NCL), whose interaction with G4 DNA at the c-MYC promoter is described above, is often overexpressed and mislocalized in a variety of cancers [36]. In addition to c-MYC, a high affinity binding of NCL to the G4 motifs present at other oncogenes including VEGF and RET has been verified *in vitro* [73]. The prominence of NCL in cancer biology has led to multiple approaches to target this protein in anti-cancer therapeutics, some of which are currently in clinical trials [74, 75]. A 26 nt long oligo AS1411 (5'-GGTGGTGGTGGTTGTGGTGGTGGTGG-3') is an aptamer that folds into a very stable G4 DNA structure *in vitro* and interacts with the RNA binding domain and the RGG repeats of NCL with high affinity [74, 76]. AS1411 was shown to limit growth of cancer cells in cell culture experiments, animal models, and clinical trials. The mechanism by which the interaction between AS1411 and NCL leads to cytotoxic effect in cancer cells is not yet clear. Reports indicate that AS1411-treatment inhibits the proper formation of protein complexes involving NCL and the protein arginine methyltransferase 5 [77]. AS1411- and siRNA-targeted interference with NCL resulted in reduction in the BCL2 protein level in breast or glioma cancer cell lines overexpressing NCL [78, 79]. However, the effect on BCL2 level in these cases appears to be independent of the transcriptional regulation by the G4 motifs in the BCL2 promoter but rather related to the normal binding of NCL to the BCL2 mRNA that increases the mRNA half-life. Another factor that further confounds the complicated story of AS1411 is that NCL can function both as the enhancer or repressor of transcription. When AS1411 was added to the human breast cancer cells transformed with a construct designed to measure



transcription from the HIV-1 LTR promoter, transcription from LTR promoter was sharply increased, consistent with the report that LTR promoter is repressed by the binding of NCL to G4 DNA [42]. Interfering with the interaction between NCL and G4 DNA present in the promoters of several oncogenes would lead to derepression of the transcription of these genes, which seems counter to the goal of limiting proliferation of cancer cells. In order to uncover the precise mechanism of how AS1411-mediated dysfunction of NCL leads to cancer cell death will require further work that includes comprehensive examination of its effect on the oncogenes with promoter-proximal G4 motifs.

### 3. TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY AND G4 DNA

In the human genome, the number of *in vivo* G4 DNA-forming loci, as determined by the chromatin IP experiment with the G4-specific antibody BG4, was far below the number predicted by the computational sequence analysis [15]. And the number of BG4-binding genomic sites varied depending on the cell type and conditions with significantly more BG4 foci present in the immortalized HaCaT human keratinocytes compared to the normal non-immortalized keratinocytes. The number of BG4 binding sites increased when the chromatin packaging was relaxed by treating the cells with an HDAC inhibitor. The elevation in genome-wide transcription due either to transformation or HDAC-inhibitor treatment apparently shifted the equilibrium toward G4 DNA. Taken together, these data indicate that the level of transcription is an important factor regulating the transition from the canonical double stranded DNA to the G-G base-paired G4 DNA. The G-rich strand of DNA, when folded into the G4 DNA configuration, becomes a replication block *in vitro* and endangers the genome stability. G4 DNA-associated genome instability, then, will be closely tied not only to the presence of a G4 motif but to the level and orientation of transcription of such sequence.

DNA secondary structures in general and G4 DNA in particular as a cause of genome instability has been extensively reported [80]. Research into the mechanism underlying the genome instability induced by triplex or H-DNA forming GAA trinucleotide repeats, hairpin forming CAG repeats, and G4 DNA-forming G-rich sequences has accomplished many recent advances through the genetic experiments carried out in the model eukaryote, *S. cerevisiae* [81-85]. The cytotoxic effect resulting from DNA strand breaks and genome rearrangement induced by the G4 stabilizing ligands such as TMPyP4, pyridostatin or PhenDC3 has also

been extensively reported on [86-88]. In this section of the review, I will specifically focus on the role transcription plays in converting the relatively innocuous G-run containing DNA sequences into the highly unstable genomic loci.

#### 3.1. Transcription-associated Genome Instability non-B DNA Secondary Structures

Highly transcribed regions form obstacles to DNA replication and, thus, hotspots of genome instability [89, 90]. In yeast, the natural pause sites of DNA polymerase complex coincide with actively transcribed loci [91], and the elevation of transcription level leads to a corresponding elevation in mutation and recombination rates [92, 93]. The formation of non-B DNA secondary structure is a prominent element of the multi-pronged mechanism underlying this phenomenon of transcription-associated genome instability. Melting of the DNA duplex and the ensuing negative supercoiling, both of which are the obligatory elements of transcription, create conditions conducive for the folding of DNA strands into secondary structures. For G4 DNA-forming sequences, *in vitro* experiments using a G4 sequence incorporated into a circular plasmid demonstrated that the transcription-associated negative superhelical tension can facilitate the transition of guanine-rich DNA strand into the G4 conformation [94]. Definitive proof of co-transcriptional formation of G4 DNA came with the study of the G-rich immunoglobulin (Ig) switch region sequences cloned into a plasmid; a distinct bubble-like structure comprising G4 DNA was visible under the electron microscope only when the plasmid was transcribed *in vitro* by T7 RNA polymerase [95]. These bubbles consisted of RNA:DNA hybrid on one side and G4 DNA on the other side. The presence of the RNA:DNA hybrids and G4 DNA were each verified by the sensitivity to RNase H-treatment and the binding of recombinant, truncated version of high-affinity G4 binding protein NCL, respectively.

#### 3.2. G4 DNA as the Genome Instability Hotspot

Even with the plethora of *in vitro* studies of G4 DNA folding and its stability, the question of whether there is indeed the presence of G4 DNA *in vivo* stable enough to disrupt DNA replication has long been debated; but supporting evidence is fast emerging. In 2002, it was reported that a mutation in *C. elegans* Dog-1 gene, which encodes a DEAH DNA helicase related to the mammalian FANCD1, resulted in highly elevated deletions throughout the genome [96]. The loci affected by the Dog-1 mutation uniformly contained guanine runs capable of folding into G4 DNA,

tained guanine runs capable of folding into G4 DNA, suggesting the model where the failure to resolve G4 DNA structures *in vivo* leads to a significant consequence for the genome maintenance. In the human genome, besides the telomeres, ribosomal DNA arrays, chromosomal fragile sites, and G/C-rich micro- or mini-satellites, G4 DNA are enriched at the chromosomal translocation break points associated with various types of cancers. A recent survey of 19,947 translocations and 46,365 deletions in cancer genomes showed a highly significant enrichment of G4 sequences within 500 bases of the breakpoints [97], confirming the previous conclusion that G4 DNA is a major contributing factor to oncogenic transformation [98]. For blood cancers, nearly 70% of the genes involved in recurrent chromosomal rearrangements contain potential G4 DNA sequences.

The BCL2 gene is involved in several different type of chromosomal translocations, most prominent of which is t(14;18) frequently associated with follicular lymphoma. A large majority of the breaks in the BCL2 gene clusters within a 150-nt region corresponding to the 3'UTR; this region is referred to as the major break point region or MBR. The G-run containing sequence identified within MBR (5'-GAAGGAG GGCAGGAGGGCTCTGGGTGGGTCTGT-3') assumes the G4 DNA conformation *in vitro* and, when cloned into a plasmid, stalls Taq DNA polymerase [67]. For HOX11 gene, which is involved in t(10,14) translocations found in T cell leukemia, two G4 DNA-forming sequences (5'-GCGCGAGGGAGGGGAGGGGAGGG GGAGAGG-3' and 5'-AGAAGGGGGAGGGGAGGG AGAGAGG GGGCGCCG-3') capable of stalling replication and transcription *in vitro* were confirmed at the breakpoint cluster regions [99]. Within the TCF3 (E2A) gene, involved in the common t(1;19) translocations associated with acute lymphoblastic leukemia, the sites of translocation break points coincide with a high density of G4 motifs [21]. Multiple G4 motifs identified at this break point region vary in the putative loop size with varying potential to form stable G4 DNA. Several oligonucleotides representing the TCF3-specific G4 motifs form G4 DNA were shown to block primer extension by a purified recombinant DNA polymerase *in vitro*. When a 470 bp fragment consisting of the TCF3 translocation breakpoint region encompassing multiple G4 motifs was integrated into the yeast genome, it induced chromosomal rearrangements in the transcription-dependent manner.

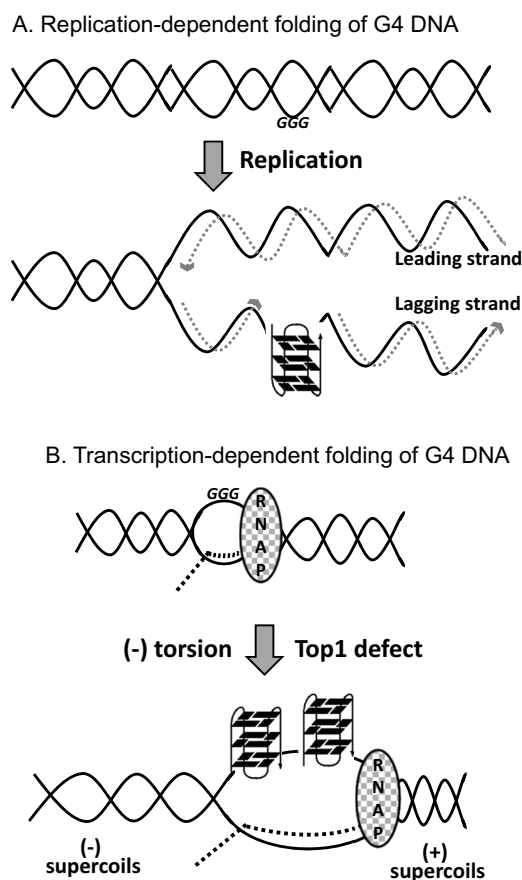
### 3.2.1. Instability at a G-rich Sequence is Elevated in a Transcription-dependent Manner

Unlike the single-stranded oligo DNA in solution used to demonstrate G4 DNA formation, DNA in the

genome is in a stable double-stranded configuration maintained by the hydrogen bonds between two complementary strands. The most significant obstacle to the G-G base pairings that hold the G quartets together in a G4 DNA is the G-C base pairing in the double strand DNA. The predominant model has been that G4 DNA folding is favored during DNA replication when the necessary single strand DNA becomes transiently present through the strand separation that precedes synthesis (Fig. 3A) [13]. Others, including my own work, have been focused on showing whether transcription can provide the opportunity for the G4 DNA formation and thereby exacerbates genome instability (Fig. 3B). The guanine-rich non-transcribed strand (NTS) of the repetitive, G/C-rich immunoglobulin heavy-chain (IgH) switch region in mammalian genomes assembles into G4 DNAs when transcribed [95]. A 770-bp fragment of mouse Immunoglobulin (Ig) switch Mu region sequence (S $\mu$ ) that contains ~20 repeats of (GAGCT)<sub>n</sub> GGGGT motif was integrated into the yeast genome and the rate of gene conversion initiating at this construct was measured when it is transcribed or silenced [100]. This fragment was inserted either in the physiological (-GTOP) or in the inverted orientation (-GBTM), placing the guanine runs in the non-transcribed strand (NTS) or in the transcribed strand (TS), respectively. Activated transcription stimulated gene conversion events initiating at the S $\mu$  fragment up to 7-fold with a clear strand bias. A significantly greater increase in recombination rate was observed when the guanine-rich DNA strand was on the NTS as in the GTOP orientation and therefore transiently single-stranded during transcription. When located on the TS, the guanine-runs would be engaged in base-pairing with the nascent RNA strand rather than being available for the G-G interaction necessary for G4 DNA folding. Similar genome instability dependent on the level and orientation of the transcription was observed when a 470 bp fragment consisting of the TCF3 translocation breakpoint region with multiple G4 motifs were integrated into the yeast genome [21]. This strand bias and dependence on active transcription strongly indicates that transcription stimulates genomic instability at G4 motifs by generating single strand DNA predisposed to fold into the DNA secondary structure.

### 3.2.2. Top1 and G4 DNA-associated Genome Instability

The impact of transcription on the dynamics of the folding of G4 DNA is illustrated by the importance of Top1 in the locus-specific elevation of genome instability at G4 motifs. Top1 is a type IB topoisomerase



**Fig. (3). Mechanism of *in vivo* G4 DNA formation. A)** The DNA strand separation during DNA replication exposes the G4 motif (*GGG*) on the lagging strand template and allows the intrastrand G-G base pairing leading to G4 DNA formation. Newly synthesized DNA strands are shown as dotted gray lines with the arrowhead indicating the 3' end. **B)** The DNA strand separation and the negative torsional stress produced during transcription allows the intrastrand G-G base pairing allowing the G4 motif (*GGG*) located on the non-transcribed DNA strand to fold into the G4 DNA conformation. Transcribed DNA strand is annealed to the nascent RNA (dotted black line). The formation of stable and extended RNA:DNA hybrids, fostered by accumulation of negative helical torsion (*i.e.* when Top1 is disrupted), increases the chance of G4 DNA formation on the non-transcribed DNA strand.

with a highly conserved role in removing topological stress produced during transcription [101, 102]. Transcription necessitates the strand separation that results in both the positive and negative supercoils ahead and behind of the RNA polymerase complex, respectively [103]. Eukaryotic Top1 is largely responsible for preventing such transcription-associated torsional stress. Upon accumulation of torsional stress associated with transcription, the equilibrium between the DNA con-

formations can shift toward alternative structures like G4 DNA as well as other non-B DNA such as hairpins, cruciform DNA, and triplex DNA (H-DNA) (reviewed in [104]). Although several reports noted that genome instability in general is not severely affected in yeast cells lacking Top1 [105-107], for the genomic loci prone to forming DNA secondary structures, the stability is acutely hampered by the loss of Top1 activity. In addition to the DNA secondary structures, the accumulation of negative torsional stress behind the RNA polymerase complex can lead to the prolonged and stable association of the nascent RNA with the template DNA strand leading to RNA:DNA hybrids or R-loop formation [108, 109]. While the extensive and stable R-loop formation affords an opportunity for the non-transcribed strand to assume non-B secondary structures, the sequestration of the non-transcribed strand in a stable, non-B DNA conformation makes it inaccessible for pairing with the complementary, transcribed strand and thus supports R-loop stabilization. R-loops and non-B DNA are each favored by the negative helical tension in DNA produced by transcription and, in turn, mutually promote stabilization of each other [89]. Accordingly, Top1-regulated maintenance of the topological homeostasis in transcribed regions is most consequential when there is the potential to form non-B structures such as G4 DNA.

When a sequence with multiple runs of guanines is transcribed, the failure to remove the transcription-associated helical stress shifts the equilibrium toward G4 DNA formation, and the subsequent elevation in genome instability is anticipated. In budding yeast, the recombination initiating at a guanine-run containing sequence was elevated when transcribed [100] and significantly exacerbated in the absence of Top1, supporting the model of topology-driven transcription-associated instability [21, 110, 111]. The activation of transcription and disruption of Top1 stimulated recombination only when the guanine runs were on the non-transcribed (non-template) DNA strand; the effect of Top1-loss was minimal when the direction of transcription was reversed so that the guanine runs were on the transcribed (template) strand, likely in stable base-pairing with the nascent transcript [21, 100]. At the actively transcribed G4 sequence, the overexpression of *E. coli* topA, but not of the gyrase, reduced recombination [111]. The bacterial topA and gyrase each removes only negative and positive helical tensions, respectively. So the above results indicate that it is the Top1-catalyzed removal of negative supercoils that prevents the formation of G4 DNA. Further supporting the model where Top1 reduces the adverse effect of G4

nificantly increased yeast cell sensitivity to the G4-stabilizing ligand TMPyP4 [112]. In addition to indirectly suppressing the G4 DNA formation by relieving transcription-associated negative supercoils, Top1 can bind to G4 DNA with high affinity [113-115]. This property of Top1 explains why a mutation at the catalytic tyrosine (Y727F of yeast Top1) elevates G4-associated genome instability more than does complete deletion of the *TOP1* gene [111]. Because the Top1-Y727F protein retains its DNA binding function, its unexpected deleterious effect with regard to G4 stability may be due to the high-affinity interaction between Top1 and G4 DNA. Top1 is an historically important target of anti-cancer therapy, and thus defining how the G4 binding properties of Top1 affects the genome instability is an important issue to be addressed. One of the interesting directions of further research suggested by the specific interaction of Top1 with G4 DNA is the feasibility of developing G4-forming oligonucleotides as Top1 inhibitors that might be effective in a combinatorial therapy with G4 ligands.

Co-transcriptionally formed G4 DNA is subject to both positive and negative regulation through interaction with additional factors. First, those factors involved in resolving the G4 DNA structure to restore the double-stranded B DNA configuration will effectively suppress genome instability induced by the co-transcriptionally formed G4 DNA. DNA helicases that can efficiently unwind G4 DNA are canonical examples of the negative regulators of G4 DNA [5]. Sub1, a general transcription factor and single strand DNA binding protein, is a newly identified G4 DNA binding factor in yeast that specifically interacts with co-transcriptionally formed G4 DNA and helps to suppress genome instability [33, 112]. The human homolog of Sub1, formerly referred to as PC4 and now as hSub1, functionally complements the loss of yeast Sub1; it suppresses the elevated recombination at the highly transcribed G4 sequence [112]. Physical interaction with G4 DNA as well as interaction with the G4 DNA-unwinding DNA helicase Pif1 is required for Sub1 to suppress the genome instability associated with the actively transcribed G4 sequence, leading us to hypothesize that the main role of Sub1 is to stimulate the recruitment of Pif1 helicase to the co-transcriptionally formed G4 DNA. For the aforementioned Top1-Y727F mutant protein, the interaction with co-transcriptionally formed G4 DNA produces the reverse effect compared to Sub1 and aggravates genome instability, possibly by stabilizing the G4 structure and creating a strong protein-DNA complex as a road block to DNA polym-

erases [111]. Similarly, for the ubiquitous G4 DNA binding protein nucleolin, the aggravated genome instability at the actively transcribed G4 motif results upon interaction with G4 DNA (S. Singh and N. Kim, unpublished results).

### 3.3. G4 DNA and Immunoglobulin Class Switch Recombination

The transformation of G-rich sequence into G4 DNA plays an important role in the critical molecular changes required for B lymphocyte development. Immunoglobulin (Ig) genes encoding antigen receptor proteins or antibodies undergo several molecular rearrangements during the course of B lymphocyte development: V(D)J Recombination, Somatic Hypermutation (SHM) and Class Switch Recombination (CSR) [116]. CSR is activated at the cellular level by the antibody encounter with a cognate antigen and is initiated at the molecular level by AID (Activation-Induced cytosine Deaminase)-catalyzed conversion of cytosine to uracil in switch-region sequences (*e.g.*, S $\mu$ , S $\gamma$ , S $\alpha$ ) [117]. The overall outcome of CSR is the recombination between two switch sequences along with deletion of the sequence in between. Functionally, CSR changes the constant region domain of antibody molecules that interacts with downstream effectors, while the antigen-recognizing variable region remains unaltered. A defect in CSR (*e.g.* due to mutated AID) manifests as a hyper-IgM syndrome characterized by acute susceptibility to opportunistic infections [118, 119]. In addition to being required for CSR, Ig switch region sequences are involved in recurrent chromosomal translocations observed in a large fraction of human multiple myelomas and Burkitt's lymphomas [120, 121].

There is a germline promoter upstream of each switch-region sequence [122], and its activation produces sterile (non-protein coding) transcripts required for CSR. Transcription of the switch regions but not the specific germline promoters are important for CSR since transcription from heterologous promoters is sufficient to support CSR [123]. Elevation in the topological strain due to the activated transcription through the switch regions serves two putative functions: first, to more efficiently target AID, a single strand DNA-specific deaminase, to cytosines on both strands of the switch region and second, to increase R-loop formation and stabilize the single-stranded character of the non-transcribed strand. Extensive R-loop formation at activated switch regions in B lymphocytes as well as during *in vitro* transcription has been reported [124-126]. Switch-region sequences, although unique for each re-

gions, are invariably guanine-rich and have the potential to form G4 DNA, with the G-rich sequences located asymmetrically on the non-transcribed strand [95, 127]. When these sequences are transcribed either *in vitro* or in *E. coli* cells, complex structures consisting of non-transcribed strand G4 DNA and a hybrid between the transcribed strand and the nascent RNA are visible by electron microscopy [95]. Transcription of the switch regions in antigen-activated B cells will also likely promote both R-loop and G4-DNA formation. Recent reports of the involvement of Top1 in regulating CSR effectively corroborate the model where the transcription-driven topological stress is necessary for CSR. These studies showed that Top1 expression is down-regulated in B lymphocytes undergoing CSR [128], and further depletion of Top1 using siRNA results in significantly elevated CSR [129]. As described earlier, the disruption in Top1 activity, combined with active transcription from the switch region germline promoters, would lead to even greater accumulation of negative torsional stress that stimulates the formation of R-loop and G4 DNA. However, whether the co-transcriptionally formed G4 DNA at the switch region simply functions to stabilize the R-loop or serves a more significant role in CSR is still under dispute. One proposed function for G4 DNA in CSR process is to serve as a platform for proteins necessary for the synopsis of two switch regions [130]. MutS $\alpha$  complex, which normally functions in the recognition of mismatched DNA base pairs, associates with the switch region in activated B cells through G4 DNA-binding and bridges the interaction of two G4 DNA-containing DNA fragments proximal to each other.

### 3.4. G4 DNA and the Antigenic Variation in Microbial Pathogens

Co-transcriptional formation of G4 DNA is thought to play an important role in the host-pathogen interaction. Several microorganisms have been shown to take an advantage of the programmed recombination to promote continual changes in the amino acid sequence of certain surface exposed immunogenic proteins. This process, termed antigenic variation (Av), in a way invoking the recombination/mutagenesis at Ig loci resulting in the antibody diversification, results in the evasion of recognition by the host adaptive immune system. In the human pathogen *Neisseria gonorrhoeae*, the single expressed cassette encoding the pilin protein (pilE) is modified through the RecA-dependent gene-conversion recombination using multiple silent pseudo pilin cassettes located upstream [131, 132]. The genetic tractability of this organism allowed the identification

of a 16-nucleotide long G4 motif (5' GGGTGG GTTGGGTGGG 3') located just upstream of the transcription start site for pilE that is required for the Av-recombination to proceed. Mutations that would disrupt the putative G4 DNA structure severely diminished the Av efficiency indicating that its capability to fold into a G4 DNA is required for the recombination to occur. Further work showed that the transcription of the region containing this G4 motif, specifically in the orientation that would place the G-runs on the non-transcribed strand, is also necessary for the efficient Av [132]. The mechanism proposed is very similar to the Ig CSR; transcription in that particular orientation would generate the RNA:DNA hybrid on the C-rich transcribed strand, which in turn facilitates the folding of the G-rich non-transcribed strand into the G4 DNA structure. Subsequently, recombination is thought to initiate by the yet uncharacterized step that could involve a G4 DNA-specific endonuclease.

In *Borrelia burgdorferi*, the pathogenic bacteria causing Lyme disease, the sequence of a surface-exposed protein of unknown function VlsE is modified while the bacteria are maintained in the blood stream of the mammalian host [133]. This modification and resulting immune evasion appears to be very similar in molecular mechanism to what has been uncovered in *N. gonorrhoeae*; a single expressed cassette is modified through one or more rounds of gene conversion recombination using one of the many silent cassettes as the template [134, 135]. Furthermore, putative G4 motifs were identified within the transcribed VlsE gene in the regions flanking the recombination-prone variable region [136]. The G4 motifs are also present in the VlsE gene of two other related *Borrelia* species with distinct strand bias for the G-runs on the non-transcribed strand. Although genetic or molecular evidence demonstrating its requirement in the recombination process is yet to come forward, the conservation and the strand bias of the G4 motifs in VlsE gene strongly suggest the role of co-transcriptionally formed G4 DNA in mediating Av-driving recombination process in the *Borrelia* species.

### CONCLUSION AND PERSPECTIVES

Among the various non-B DNA structures that are recently being investigated as the significant instigators of genome instability, G4 DNA is uniquely implicated as a key element of the deliberate, regulated recombination programs such as the vertebrate Ig CSR and bacterial Av. As reviewed here, the findings from the yeast genetic assays utilizing the G4 forming sequences embedded in the recombination reporter constructs, to-

gether with the molecular details uncovered regarding the mechanisms CSR and Av recombination processes, revealed the critical role of transcription in transforming G4 motifs into genome instability hotspot. G-run containing sequences manifest as genome instability hotspots when the local topological stress, generated by transcription and exacerbated upon the disruption of Top1 activity, stimulates the folding of G4 DNA structure as well as the accumulation of R-loops. These steps starting from transcription activation to the changes in the genome, deleterious or advantageous to cells, are likely to have significant clinical implications. For genetic changes underlying the cellular transformation into tumors, the role of transcription in initiating the chromosomal translocations at the recurrent breakpoints containing G4 motifs, with the exception of Ig switch regions, has not yet been clarified and should be further investigated. The link between transcription and genome instability is also relevant for those neurological disorders linked to the unstable repetitive DNA sequences. For Huntington's disease, the role of transcription in elevating the pathogenic expansion of the hairpin forming (CAG)<sub>n</sub> trinucleotide repeats is strongly substantiated by experiments carried out in cultured cells [137, 138]. Similar mechanism is expected for the expansion of (CGG)<sub>n</sub> trinucleotide repeats associated with Fragile X syndrome [139] or (GGGGCC)<sub>n</sub> hexanucleotide repeats associated with Amyotrophic Lateral Sclerosis (ALS) [140]. Both (CGG)<sub>n</sub> and (GGGGCC)<sub>n</sub> repeats presumably form G4 DNA *in vivo*.

Independent of the key role played by transcription in G4 DNA-mediated genome instability, the role of G4 DNA as a significant regulator of transcription has been under intense scrutiny. Until recently, this very interesting concept of gene expression differentially regulated when the double-strand guanine-run containing DNA transitions to the G-G base-paired G4 DNA was chiefly based on the computational analyses of genome sequences. *In silico* analyses consistently demonstrated that the G4-forming sequences are notably enriched proximal to gene promoters and transcription start sites. Additional support for the concept came from studying the effect of G4 ligands or G4-resolving helicases in modulating transcription. Many unresolved questions regarding the mechanistic details of the non-genetic change in the DNA conformation leading to the changes in gene expression still remain and are complicated by some inconsistent results. Important remaining questions with clinical implications include whether G4 DNA regulates transcription directly (*e.g. via* nucleosome exclusion) or indirectly (*e.g. via* inter-

action with transcription repressor/activator). In this review, two distinct approaches to exploiting the role of G4 in transcriptional regulation for clinical application were described. First, the plethora of small-molecule G4-binding ligands variably affect the genomic stability and transcription of the G4 motif-containing loci. Better understanding of the exact targets of these chemicals will potentially lead to the improved anti-cancer therapy synergistically combining the down-regulation of DNA repair and the elevation of DNA breaks. Second, targeting of proteins such as nucleolin and topoisomerase I using G-rich aptamers exploits the high affinity binding of these factors to G4 DNA and will benefit from further study into the binding specificities and the down-stream functions of G4-bound nucleolin or topoisomerase I. Latest advances in the field, such as the new tools to investigate G4 DNA *in vivo* (*i.e.* G4-specific antibodies) and the identification of novel G4 DNA-interacting proteins, are expected to soon lead to more answers and more advances in clinical application of G4 DNA.

#### LIST OF ABBREVIATIONS

AID	=	Activation-Induced Cytosine Deaminase
Av	=	Antigenic Variation
ChIP	=	Chromatin Immunoprecipitation
CSR	=	Class-Switch Recombination
DSB	=	Double-Strand Break
G4	=	Guanine Quadruplex
GCR	=	Gross Chromosomal Rearrangement
RNAP	=	RNA Polymerase
Top1	=	Topoisomerase I
TSS	=	Transcription Start Site

#### CONSENT FOR PUBLICATION

Not applicable.

#### FUNDING

Work in the NK lab was supported by grants from the Welch Foundation (AU1875) and the National Institutes of Health (R01GM116007).

#### CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

#### ACKNOWLEDGEMENTS

Declared none.

## REFERENCES

- [1] Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **1988**, *334*(6180), 364-366. [http://dx.doi.org/10.1038/334364a0] [PMID: 3393228]
- [2] Todd, A.K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **2005**, *33*(9), 2901-2907. [http://dx.doi.org/10.1093/nar/gki553] [PMID: 15914666]
- [3] Huppert, J.L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **2005**, *33*(9), 2908-2916. [http://dx.doi.org/10.1093/nar/gki609] [PMID: 15914667]
- [4] Henderson, E.; Hardin, C.C.; Walk, S.K.; Tinoco, I., Jr; Blackburn, E.H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **1987**, *51*(6), 899-908. [http://dx.doi.org/10.1016/0092-8674(87)90577-0] [PMID: 3690664]
- [5] Fry, M. Tetraplex DNA and its interacting proteins, Frontiers in bio-science. *J. Virt. Lib.*, **2007**, *12*, 4336-4351.
- [6] Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **2007**, *35*(2), 406-413. [http://dx.doi.org/10.1093/nar/gkl1057] [PMID: 17169996]
- [7] Du, Z.; Zhao, Y.; Li, N. Genome-wide colonization of gene regulatory elements by G4 DNA motifs. *Nucleic Acids Res.*, **2009**, *37*(20), 6784-6798. [http://dx.doi.org/10.1093/nar/gkp710] [PMID: 19759215]
- [8] Du, Z.; Zhao, Y.; Li, N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **2008**, *18*(2), 233-241. [http://dx.doi.org/10.1101/gr.6905408] [PMID: 18096746]
- [9] Hershman, S.G.; Chen, Q.; Lee, J.Y.; Kozak, M.L.; Yue, P.; Wang, L.S.; Johnson, F.B. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **2008**, *36*(1), 144-156. [http://dx.doi.org/10.1093/nar/gkm986] [PMID: 17999996]
- [10] Capra, J.A.; Paeschke, K.; Singh, M.; Zakian, V.A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLOS Comput. Biol.*, **2010**, *6*(7), e1000861. [http://dx.doi.org/10.1371/journal.pcbi.1000861] [PMID: 20676380]
- [11] Henderson, A.; Wu, Y.; Huang, Y.C.; Chavez, E.A.; Platt, J.; Johnson, F.B.; Brosh, R.M., Jr; Sen, D.; Lansdorp, P.M. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **2014**, *42*(2), 860-869. [http://dx.doi.org/10.1093/nar/gkt957] [PMID: 24163102]
- [12] Fernando, H.; Rodriguez, R.; Balasubramanian, S. Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry*, **2008**, *47*(36), 9365-9371. [http://dx.doi.org/10.1021/bi800983u] [PMID: 18702511]
- [13] Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **2013**, *5*(3), 182-186. [http://dx.doi.org/10.1038/nchem.1548] [PMID: 23422559]
- [14] Lam, E.Y.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **2013**, *4*, 1796. [http://dx.doi.org/10.1038/ncomms2792] [PMID: 23653208]
- [15] Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **2016**, *48*(10), 1267-1272. [http://dx.doi.org/10.1038/ng.3662] [PMID: 27618450]
- [16] Liu, H.Y.; Zhao, Q.; Zhang, T.P.; Wu, Y.; Xiong, Y.X.; Wang, S.K.; Ge, Y.L.; He, J.H.; Lv, P.; Ou, T.M.; Tan, J.H.; Li, D.; Gu, L.Q.; Ren, J.; Zhao, Y.; Huang, Z.S. Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. *Cell Chem. Biol.*, **2016**, *23*(10), 1261-1270. [http://dx.doi.org/10.1016/j.chembiol.2016.08.013] [PMID: 27693060]
- [17] Piazza, A.; Adrian, M.; Samazan, F.; Heddi, B.; Hamon, F.; Serero, A.; Lopes, J.; Teulade-Fichou, M.P.; Phan, A.T.; Nicolas, A. Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.*, **2015**, *34*(12), 1718-1734. [http://dx.doi.org/10.15252/embj.201490702] [PMID: 25956747]
- [18] Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **2010**, *38*(21), 7858-7868. [http://dx.doi.org/10.1093/nar/gkq639] [PMID: 20660477]
- [19] Agrawal, P.; Lin, C.; Mathad, R.I.; Carver, M.; Yang, D. The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K<sup>+</sup> solution. *J. Am. Chem. Soc.*, **2014**, *136*(5), 1750-1753. [http://dx.doi.org/10.1021/ja4118945] [PMID: 24450880]
- [20] Sun, D.; Hurley, L.H. Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.*, **2010**, *608*, 65-79. [http://dx.doi.org/10.1007/978-1-59745-363-9\_5] [PMID: 20012416]
- [21] Williams, J.D.; Fleetwood, S.; Berroyer, A.; Kim, N.; Larson, E.D. Sites of instability in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures *in vitro*. *Front. Genet.*, **2015**, *6*, 177. [http://dx.doi.org/10.3389/fgene.2015.00177] [PMID: 26029241]
- [22] Kumari, R.; Nambiar, M.; Shanbagh, S.; Raghavan, S.C. Detection of G-quadruplex DNA using primer extension as a tool. *PLoS One*, **2015**, *10*(3), e0119722. [http://dx.doi.org/10.1371/journal.pone.0119722] [PMID: 25799152]
- [23] Belotserkovskii, B.P.; Neil, A.J.; Saleh, S.S.; Shin, J.H.; Mirkin, S.M.; Hanawalt, P.C. Transcription blockage by homopurine DNA sequences: Role of sequence composition and single-strand breaks. *Nucleic Acids Res.*, **2013**, *41*(3), 1817-1828. [http://dx.doi.org/10.1093/nar/gks1333] [PMID: 23275544]
- [24] Belotserkovskii, B.P.; Liu, R.; Tornaletti, S.; Krasilnikova, M.M.; Mirkin, S.M.; Hanawalt, P.C. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. USA*, **2010**, *107*(29), 12816-12821. [http://dx.doi.org/10.1073/pnas.1007580107] [PMID: 20616059]
- [25] Broxson, C.; Beckett, J.; Tornaletti, S. Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene. *Biochemistry*, **2011**, *50*(19), 4162-4172. [http://dx.doi.org/10.1021/bi2002136] [PMID: 21469677]
- [26] Belotserkovskii, B.P.; Soo Shin, J.H.; Hanawalt, P.C. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res.*, **2017**, *45*(11), 6589-6599.



- [27] [\[http://dx.doi.org/10.1093/nar/gkx403\]](http://dx.doi.org/10.1093/nar/gkx403) [PMID: 28498974] Zheng, K.W.; Xiao, S.; Liu, J.Q.; Zhang, J.Y.; Hao, Y.H.; Tan, Z. Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional *cis* element for transcription control. *Nucleic Acids Res.*, **2013**, *41*(10), 5533-5541.
- [28] [\[http://dx.doi.org/10.1093/nar/gkt264\]](http://dx.doi.org/10.1093/nar/gkt264) [PMID: 23585281] Gilmour, D.S. Promoter proximal pausing on genes in metazoans. *Chromosoma*, **2009**, *118*(1), 1-10. [\[http://dx.doi.org/10.1007/s00412-008-0182-4\]](http://dx.doi.org/10.1007/s00412-008-0182-4) [PMID: 18830703]
- [29] Eddy, J.; Vallur, A.C.; Varma, S.; Liu, H.; Reinhold, W.C.; Pommier, Y.; Maizels, N. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.*, **2011**, *39*(12), 4975-4983. [\[http://dx.doi.org/10.1093/nar/gkr079\]](http://dx.doi.org/10.1093/nar/gkr079) [PMID: 21371997]
- [30] Raiber, E.A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 *in vitro*. *Nucleic Acids Res.*, **2012**, *40*(4), 1499-1508. [\[http://dx.doi.org/10.1093/nar/gkr882\]](http://dx.doi.org/10.1093/nar/gkr882) [PMID: 22021377]
- [31] Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K.K.; Xodo, L.E. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.*, **2010**, *285*(29), 22003-22016. [\[http://dx.doi.org/10.1074/jbc.M110.101923\]](http://dx.doi.org/10.1074/jbc.M110.101923) [PMID: 20457603]
- [32] Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L.A.; Mergny, J.L.; Salgado, G.F. High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem.*, **2017**, *292*(19), 8082-8091. [\[http://dx.doi.org/10.1074/jbc.M117.781906\]](http://dx.doi.org/10.1074/jbc.M117.781906) [PMID: 28330874]
- [33] Gao, J.; Zybailov, B.L.; Byrd, A.K.; Griffin, W.C.; Chib, S.; Mackintosh, S.G.; Tackett, A.J.; Raney, K.D. Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA. *Chem. Commun. (Camb.)*, **2015**, *51*(33), 7242-7244. [\[http://dx.doi.org/10.1039/C5CC00742A\]](http://dx.doi.org/10.1039/C5CC00742A) [PMID: 25813861]
- [34] Tajrishi, M.M.; Tuteja, R.; Tuteja, N. Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. *Commun. Integr. Biol.*, **2011**, *4*(3), 267-275. [\[http://dx.doi.org/10.4161/cib.4.3.14884\]](http://dx.doi.org/10.4161/cib.4.3.14884) [PMID: 21980556]
- [35] Yuan, J.; Muljo, S.A. Exploring the RNA world in hematopoietic cells through the lens of RNA-binding proteins. *Immunol. Rev.*, **2013**, *253*(1), 290-303. [\[http://dx.doi.org/10.1111/imr.12048\]](http://dx.doi.org/10.1111/imr.12048) [PMID: 23550653]
- [36] Berger, C.M.; Gaume, X.; Bouvet, P. The roles of nucleolin subcellular localization in cancer. *Biochimie*, **2015**, *113*, 78-85. [\[http://dx.doi.org/10.1016/j.biochi.2015.03.023\]](http://dx.doi.org/10.1016/j.biochi.2015.03.023) [PMID: 25866190]
- [37] Otake, Y.; Soundararajan, S.; Sengupta, T.K.; Kio, E.A.; Smith, J.C.; Pineda-Roman, M.; Stuart, R.K.; Spicer, E.K.; Fernandes, D.J. Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. *Blood*, **2007**, *109*(7), 3069-3075. [PMID: 17179226]
- [38] Hanakahi, L.A.; Sun, H.; Maizels, N. High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **1999**, *274*(22), 15908-15912. [\[http://dx.doi.org/10.1074/jbc.274.22.15908\]](http://dx.doi.org/10.1074/jbc.274.22.15908) [PMID: 10336496]
- [39] Dempsey, L.A.; Sun, H.; Hanakahi, L.A.; Maizels, N. 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.*, **1999**, *274*(2), 1066-1071. [\[http://dx.doi.org/10.1074/jbc.274.2.1066\]](http://dx.doi.org/10.1074/jbc.274.2.1066) [PMID: 9873052]
- [40] Uribe, D.J.; Guo, K.; Shin, Y.J.; Sun, D. Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures. *Biochemistry*, **2011**, *50*(18), 3796-3806. [\[http://dx.doi.org/10.1021/bi101633b\]](http://dx.doi.org/10.1021/bi101633b) [PMID: 21466159]
- [41] Perrone, R.; Nadai, M.; Frasson, I.; Poe, J.A.; Butovskaya, E.; Smithgall, T.E.; Palumbo, M.; Palù, G.; Richter, S.N. A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.*, **2013**, *56*(16), 6521-6530. [\[http://dx.doi.org/10.1021/jm400914r\]](http://dx.doi.org/10.1021/jm400914r) [PMID: 23865750]
- [42] Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palù, G.; Fabris, D.; Richter, S.N. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.*, **2015**, *43*(18), 8884-8897. [\[http://dx.doi.org/10.1093/nar/gkv897\]](http://dx.doi.org/10.1093/nar/gkv897) [PMID: 26354862]
- [43] Siddiqui-Jain, A.; Grand, C.L.; Bearss, D.J.; Hurley, L.H. 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*, **2002**, *99*(18), 11593-11598. [\[http://dx.doi.org/10.1073/pnas.182256799\]](http://dx.doi.org/10.1073/pnas.182256799) [PMID: 12195017]
- [44] González, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **2009**, *284*(35), 23622-23635. [\[http://dx.doi.org/10.1074/jbc.M109.018028\]](http://dx.doi.org/10.1074/jbc.M109.018028) [PMID: 19581307]
- [45] González, V.; Hurley, L.H. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry*, **2010**, *49*(45), 9706-9714. [\[http://dx.doi.org/10.1021/bi100509s\]](http://dx.doi.org/10.1021/bi100509s) [PMID: 20932061]
- [46] Bai, L.; Morozov, A.V. Gene regulation by nucleosome positioning. *Trends Genet.*, **2010**, *26*(11), 476-483. [\[http://dx.doi.org/10.1016/j.tig.2010.08.003\]](http://dx.doi.org/10.1016/j.tig.2010.08.003) [PMID: 20832136]
- [47] Wong, H.M.; Huppert, J.L. Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.*, **2009**, *5*(12), 1713-1719. [\[http://dx.doi.org/10.1039/b905848f\]](http://dx.doi.org/10.1039/b905848f) [PMID: 19585004]
- [48] Halder, K.; Halder, R.; Chowdhury, S. Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. *Mol. Biosyst.*, **2009**, *5*(12), 1703-1712. [\[http://dx.doi.org/10.1039/b905132e\]](http://dx.doi.org/10.1039/b905132e) [PMID: 19587895]
- [49] Hogan, G.J.; Lee, C.K.; Lieb, J.D. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.*, **2006**, *2*(9), e158. [\[http://dx.doi.org/10.1371/journal.pgen.0020158\]](http://dx.doi.org/10.1371/journal.pgen.0020158) [PMID: 17002501]
- [50] Sun, H.; Karow, J.K.; Hickson, I.D.; Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **1998**, *273*(42), 27587-27592. [\[http://dx.doi.org/10.1074/jbc.273.42.27587\]](http://dx.doi.org/10.1074/jbc.273.42.27587) [PMID: 9765292]
- [51] Nguyen, G.H.; Tang, W.; Robles, A.I.; Beyer, R.P.; Gray, L.T.; Welsh, J.A.; Schetter, A.J.; Kumamoto, K.; Wang, X.W.; Hickson, I.D.; Maizels, N.; Monnat, R.J., Jr; Harris,

- C.C. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci. USA*, **2014**, *111*(27), 9905-9910. [http://dx.doi.org/10.1073/pnas.1404807111] [PMID: 24958861]
- [52] Tang, W.; Robles, A.I.; Beyer, R.P.; Gray, L.T.; Nguyen, G.H.; Oshima, J.; Maizels, N.; Harris, C.C.; Monnat, R.J., Jr The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum. Mol. Genet.*, **2016**, *25*(10), 2060-2069. [http://dx.doi.org/10.1093/hmg/ddw079] [PMID: 26984941]
- [53] Smestad, J.A.; Maher, L.J., III Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. *BMC Med. Genet.*, **2015**, *16*, 91. [http://dx.doi.org/10.1186/s12881-015-0236-4] [PMID: 26449372]
- [54] Johnson, J.E.; Cao, K.; Ryvkin, P.; Wang, L.S.; Johnson, F.B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.*, **2010**, *38*(4), 1114-1122. [http://dx.doi.org/10.1093/nar/gkp1103] [PMID: 19966276]
- [55] Romiguier, J.; Ranwez, V.; Douzery, E.J.; Galtier, N. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res.*, **2010**, *20*(8), 1001-1009. [http://dx.doi.org/10.1101/gr.104372.109] [PMID: 20530252]
- [56] David, A.P.; Margarit, E.; Domizi, P.; Banchio, C.; Armas, P.; Calcaterra, N.B. G-quadruplexes as novel *cis*-elements controlling transcription during embryonic development. *Nucleic Acids Res.*, **2016**, *44*(9), 4163-4173. [http://dx.doi.org/10.1093/nar/gkw011] [PMID: 26773060]
- [57] Monchaud, D.; Granzhan, A.; Saettel, N.; Guédin, A.; Mergny, J.L.; Teulade-Fichou, M.P. "One ring to bind them all"-part I: the efficiency of the macrocyclic scaffold for g-quadruplex DNA recognition. *J. Nucleic Acids*, **2010**, *2010*, pii. 525862. [http://dx.doi.org/10.4061/2010/525862] [PMID: 20725629]
- [58] Yadav, V.; Hemansi, ; Kim, N.; Tuteja, N.; Yadav, P. G Quadruplex in plants: A ubiquitous regulatory element and its biological relevance. *Front. Plant Sci.*, **2017**, *8*, 1163. [http://dx.doi.org/10.3389/fpls.2017.01163] [PMID: 28725233]
- [59] Mullen, M.A.; Olson, K.J.; Dallaire, P.; Major, F.; Assmann, S.M.; Bevilacqua, P.C. RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res.*, **2010**, *38*(22), 8149-8163. [http://dx.doi.org/10.1093/nar/gkq804] [PMID: 20860998]
- [60] Andorf, C.M.; Kopylov, M.; Dobbs, D.; Koch, K.E.; Stroupe, M.E.; Lawrence, C.J.; Bass, H.W. G-quadruplex (G4) motifs in the maize (*Zea mays L.*) genome are enriched at specific locations in thousands of genes coupled to energy status, hypoxia, low sugar, and nutrient deprivation. *J. Genet. Genomics*, **2014**, *41*(12), 627-647. [http://dx.doi.org/10.1016/j.jgg.2014.10.004] [PMID: 25527104]
- [61] Garg, R.; Aggarwal, J.; Thakkar, B. Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. *Sci. Rep.*, **2016**, *6*, 28211. [http://dx.doi.org/10.1038/srep28211] [PMID: 27324275]
- [62] Qin, Y.; Hurley, L.H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **2008**, *90*(8), 1149-1171. [http://dx.doi.org/10.1016/j.biochi.2008.02.020] [PMID: 18355457]
- [63] Yang, D.; Hurley, L.H. Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides Nucleotides Nucleic Acids*, **2006**, *25*(8), 951-968. [http://dx.doi.org/10.1080/15257770600809913] [PMID: 16901825]
- [64] Cogoi, S.; Xodo, L.E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **2006**, *34*(9), 2536-2549. [http://dx.doi.org/10.1093/nar/gkl286] [PMID: 16687659]
- [65] Dexheimer, T.S.; Sun, D.; Hurley, L.H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **2006**, *128*(16), 5404-5415. [http://dx.doi.org/10.1021/ja0563861] [PMID: 16620112]
- [66] Dai, J.; Dexheimer, T.S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R.A.; Yang, D. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **2006**, *128*(4), 1096-1098. [http://dx.doi.org/10.1021/ja055636a] [PMID: 16433524]
- [67] Nambiar, M.; Goldsmith, G.; Moorthy, B.T.; Lieber, M.R.; Joshi, M.V.; Choudhary, B.; Hosur, R.V.; Raghavan, S.C. Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Res.*, **2011**, *39*(3), 936-948. [http://dx.doi.org/10.1093/nar/gkq824] [PMID: 20880994]
- [68] Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **2006**, *34*(14), 3887-3896. [http://dx.doi.org/10.1093/nar/gkl529] [PMID: 16914419]
- [69] Wei, C.; Wang, L.; Jia, G.; Zhou, J.; Han, G.; Li, C. The binding mode of porphyrins with cation side arms to (TG4T)4 G-quadruplex: spectroscopic evidence. *Biophys. Chem.*, **2009**, *143*(1-2), 79-84. [http://dx.doi.org/10.1016/j.bpc.2009.04.005] [PMID: 19411133]
- [70] Jia, G.; Feng, Z.; Wei, C.; Zhou, J.; Wang, X.; Li, C. Dynamic insight into the interaction between porphyrin and G-quadruplex DNAs: time-resolved fluorescence anisotropy study. *J. Phys. Chem. B*, **2009**, *113*(50), 16237-16245. [http://dx.doi.org/10.1021/jp906060d] [PMID: 19924868]
- [71] Chen, B.J.; Wu, Y.L.; Tanaka, Y.; Zhang, W. Small molecules targeting c-Myc oncogene: promising anti-cancer therapeutics. *Int. J. Biol. Sci.*, **2014**, *10*(10), 1084-1096. [http://dx.doi.org/10.7150/ijbs.10190] [PMID: 25332683]
- [72] Moruno-Manchon, J.F.; Koellhoffer, E.C.; Gopakumar, J.; Hambarde, S.; Kim, N.; McCullough, L.D.; Tsvetkov, A.S. The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Breal in neurons. *Aging (Albany NY)*, **2017**, *9*(9), 1957-1970. [http://dx.doi.org/10.18632/aging.101282] [PMID: 28904242]
- [73] Edwards, T.K.; Saleem, A.; Shaman, J.A.; Dennis, T.; Gerigk, C.; Oliveros, E.; Gartenberg, M.R.; Rubin, E.H. Role for nucleolin/Nsr1 in the cellular localization of topoisomerase I. *J. Biol. Chem.*, **2000**, *275*(46), 36181-36188. [http://dx.doi.org/10.1074/jbc.M006628200] [PMID: 10967121]
- [74] Bates, P.J.; Reyes-Reyes, E.M.; Malik, M.T.; Murphy, E.M.; O'Toole, M.G.; Trent, J.O. G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochim. Biophys. Acta*, **2016**. [PMID: 28007579]
- [75] Palmieri, D.; Richmond, T.; Piovan, C.; Sheetz, T.; Zanesi, N.; Troise, F.; James, C.; Wernicke, D.; Nyei, F.; Gordon,

- T.J.; Consiglio, J.; Salvatore, F.; Coppola, V.; Pichiorri, F.; De Lorenzo, C.; Croce, C.M. Human anti-nucleolin recombinant immunogen for cancer therapy. *Proc. Natl. Acad. Sci. USA*, **2015**, *112*(30), 9418-9423.  
[http://dx.doi.org/10.1073/pnas.1507087112] [PMID: 26170308]
- [76] Bates, P.J.; Laber, D.A.; Miller, D.M.; Thomas, S.D.; Trent, J.O. Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp. Mol. Pathol.*, **2009**, *86*(3), 151-164.  
[http://dx.doi.org/10.1016/j.yexmp.2009.01.004] [PMID: 19454272]
- [77] Teng, Y.; Girvan, A.C.; Casson, L.K.; Pierce, W.M., Jr; Qian, M.; Thomas, S.D.; Bates, P.J. AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. *Cancer Res.*, **2007**, *67*(21), 10491-10500.  
[http://dx.doi.org/10.1158/0008-5472.CAN-06-4206] [PMID: 17974993]
- [78] Cheng, Y.; Zhao, G.; Zhang, S.; Nigim, F.; Zhou, G.; Yu, Z.; Song, Y.; Chen, Y.; Li, Y. AS1411-Induced Growth Inhibition of Glioma Cells by Up-Regulation of p53 and Down-Regulation of Bcl-2 and Akt1 via Nucleolin. *PLoS One*, **2016**, *11*(12), e0167094.  
[http://dx.doi.org/10.1371/journal.pone.0167094] [PMID: 27907160]
- [79] Soundararajan, S.; Chen, W.; Spicer, E.K.; Courtenay-Luck, N.; Fernandes, D.J. The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer Res.*, **2008**, *68*(7), 2358-2365.  
[http://dx.doi.org/10.1158/0008-5472.CAN-07-5723] [PMID: 18381443]
- [80] Mirkin, E.V.; Mirkin, S.M. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.*, **2007**, *71*(1), 13-35.  
[http://dx.doi.org/10.1128/MMBR.00030-06] [PMID: 17347517]
- [81] Su, X.A.; Dion, V.; Gasser, S.M.; Freudenreich, C.H. Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.*, **2015**, *29*(10), 1006-1017.  
[http://dx.doi.org/10.1101/gad.256404.114] [PMID: 25940904]
- [82] Kim, H-M.; Narayanan, V.; Mieczkowski, P.A.; Petes, T.D.; Krasilnikova, M.M.; Mirkin, S.M.; Lobachev, K.S. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J.*, **2008**, *27*(21), 2896-2906.  
[http://dx.doi.org/10.1038/emboj.2008.205] [PMID: 18833189]
- [83] Tang, W.; Dominska, M.; Greenwell, P.W.; Harvanek, Z.; Lobachev, K.S.; Kim, H.M.; Narayanan, V.; Mirkin, S.M.; Petes, T.D. Friedreich's ataxia (GAA)<sub>n</sub>(TTC)<sub>n</sub> repeats strongly stimulate mitotic crossovers in *Saccharomyces cerevisiae*. *PLoS Genet.*, **2011**, *7*(1), e1001270.  
[http://dx.doi.org/10.1371/journal.pgen.1001270] [PMID: 21249181]
- [84] Piazza, A.; Serero, A.; Boulé, J.B.; Legoix-Né, P.; Lopes, J.; Nicolas, A. Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in *Saccharomyces cerevisiae* depends on G-quadruplexes or Cdc13. *PLoS Genet.*, **2012**, *8*(11), e1003033.  
[http://dx.doi.org/10.1371/journal.pgen.1003033] [PMID: 23133402]
- [85] Ribeyre, C.; Lopes, J.; Boulé, J.B.; Piazza, A.; Guédin, A.; Zakian, V.A.; Mergny, J.L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.*, **2009**, *5*(5), e1000475.  
[http://dx.doi.org/10.1371/journal.pgen.1000475] [PMID: 19424434]
- [86] McLuckie, K.I.; Di Antonio, M.; Zecchini, H.; Xian, J.; Caldas, C.; Krippendorff, B.F.; Tannahill, D.; Lowe, C.; Balasubramanian, S. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.*, **2013**, *135*(26), 9640-9643.  
[http://dx.doi.org/10.1021/ja404868t] [PMID: 23782415]
- [87] Piazza, A.; Boulé, J.B.; Lopes, J.; Mingo, K.; Largy, E.; Teulade-Fichou, M.P.; Nicolas, A. Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **2010**, *38*(13), 4337-4348.  
[http://dx.doi.org/10.1093/nar/gkq136] [PMID: 20223771]
- [88] Rodriguez, R.; Miller, K.M.; Forment, J.V.; Bradshaw, C.R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S.P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **2012**, *8*(3), 301-310.  
[http://dx.doi.org/10.1038/nchembio.780] [PMID: 22306580]
- [89] Kim, N.; Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.*, **2012**, *13*(3), 204-214.  
[http://dx.doi.org/10.1038/nrg3152] [PMID: 22330764]
- [90] Aguilera, A.; Gaillard, H. Transcription and recombination: when RNA meets DNA. *Cold Spring Harb. Perspect. Biol.*, **2014**, *6*(8), 6.  
[http://dx.doi.org/10.1101/cshperspect.a016543] [PMID: 25085910]
- [91] Azvolinsky, A.; Giresi, P.G.; Lieb, J.D.; Zakian, V.A. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell*, **2009**, *34*(6), 722-734.  
[http://dx.doi.org/10.1016/j.molcel.2009.05.022] [PMID: 19560424]
- [92] Freedman, J.A.; Jinks-Robertson, S. Genetic requirements for spontaneous and transcription-stimulated mitotic recombination in *Saccharomyces cerevisiae*. *Genetics*, **2002**, *162*(1), 15-27.  
[PMID: 12242220]
- [93] Kim, N.; Abdulovic, A.L.; Gealy, R.; Lippert, M.J.; Jinks-Robertson, S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair (Amst.)*, **2007**, *6*(9), 1285-1296.  
[http://dx.doi.org/10.1016/j.dnarep.2007.02.023] [PMID: 17398168]
- [94] Sun, D.; Hurley, L.H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **2009**, *52*(9), 2863-2874.  
[http://dx.doi.org/10.1021/jm900055s] [PMID: 19385599]
- [95] Duquette, M.L.; Handa, P.; Vincent, J.A.; Taylor, A.F.; Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **2004**, *18*(13), 1618-1629.  
[http://dx.doi.org/10.1101/gad.1200804] [PMID: 15231739]
- [96] Cheung, I.; Schertzer, M.; Rose, A.; Lansdorp, P.M. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.*, **2002**, *31*(4), 405-409.  
[http://dx.doi.org/10.1038/ng928] [PMID: 12101400]
- [97] Bacolla, A.; Tainer, J.A.; Vasquez, K.M.; Cooper, D.N. Translocation and deletion breakpoints in cancer genomes

- are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res.*, **2016**, *44*(12), 5673-5688. [http://dx.doi.org/10.1093/nar/gkw261] [PMID: 27084947]
- [98] Katapadi, V.K.; Nambiar, M.; Raghavan, S.C. Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. *Genomics*, **2012**, *100*(2), 72-80. [http://dx.doi.org/10.1016/j.ygeno.2012.05.008] [PMID: 22659239]
- [99] Nambiar, M.; Srivastava, M.; Gopalakrishnan, V.; Sankaran, S.K.; Raghavan, S.C. G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t(10;14) translocation in T-cell leukemia. *Mol. Cell Biol.*, **2013**, *33*(21), 4266-4281. [http://dx.doi.org/10.1128/MCB.00540-13] [PMID: 24001773]
- [100] Kim, N.; Jinks-Robertson, S. Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast *DNA Repair (Amst)*, **2011**.
- [101] Pommier, Y.; Sun, Y.; Huang, S.N.; Nitiss, J.L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.*, **2016**, *17*(11), 703-721. [http://dx.doi.org/10.1038/nrm.2016.111] [PMID: 27649880]
- [102] Kim, N.; Jinks-Robertson, S. The Top1 paradox: Friend and foe of the eukaryotic genome *DNA Repair (Amst)*, **2017**.
- [103] Wu, H.Y.; Shyy, S.H.; Wang, J.C.; Liu, L.F. Transcription generates positively and negatively supercoiled domains in the template. *Cell*, **1988**, *53*(3), 433-440. [http://dx.doi.org/10.1016/0092-8674(88)90163-8] [PMID: 2835168]
- [104] Mirkin, S.M. Discovery of alternative DNA structures: A heroic decade (1979-1989), *Front. Biosci.*, **2008**, *13*, 1064-1071.
- [105] Conover, H.N.; Lujan, S.A.; Chapman, M.J.; Cornelio, D.A.; Sharif, R.; Williams, J.S.; Clark, A.B.; Camilo, F.; Kunkel, T.A.; Argueso, J.L. Stimulation of Chromosomal Rearrangements by Ribonucleotides. *Genetics*, **2015**, *201*(3), 951-961. [http://dx.doi.org/10.1534/genetics.115.181149] [PMID: 26400612]
- [106] Andersen, S.L.; Sloan, R.S.; Petes, T.D.; Jinks-Robertson, S. Genome-destabilizing effects associated with top1 loss or accumulation of top1 cleavage complexes in yeast. *PLoS Genet.*, **2015**, *11*(4), e1005098. [http://dx.doi.org/10.1371/journal.pgen.1005098] [PMID: 25830313]
- [107] Allen-Soltero, S.; Martinez, S.L.; Putnam, C.D.; Kolodner, R.D. A saccharomyces cerevisiae RNase H2 interaction network functions to suppress genome instability. *Mol. Cell Biol.*, **2014**, *34*(8), 1521-1534. [http://dx.doi.org/10.1128/MCB.00960-13] [PMID: 24550002]
- [108] Huertas, P.; Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell*, **2003**, *12*(3), 711-721. [http://dx.doi.org/10.1016/j.molcel.2003.08.010] [PMID: 14527416]
- [109] Li, X.; Manley, J.L. Cotranscriptional processes and their influence on genome stability. *Genes Dev.*, **2006**, *20*(14), 1838-1847. [http://dx.doi.org/10.1101/gad.1438306] [PMID: 16847344]
- [110] Yadav, P.; Harcy, V.; Argueso, J.L.; Dominska, M.; Jinks-Robertson, S.; Kim, N. Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed G-quadruplex-forming sequence. *PLoS Genet.*, **2014**, *10*(12), e1004839. [http://dx.doi.org/10.1371/journal.pgen.1004839] [PMID: 25473964]
- [111] Yadav, P.; Owiti, N.; Kim, N. 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.*, **2016**, *44*(2), 718-729. [http://dx.doi.org/10.1093/nar/gkv1152] [PMID: 26527723]
- [112] Lopez, C.R.; Singh, S.; Hambarde, S.; Griffin, W.C.; Gao, J.; Chib, S.; Yu, Y.; Ira, G.; Raney, K.D.; Kim, N. Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA. *Nucleic Acids Res.*, **2017**, *45*(10), 5850-5862. [http://dx.doi.org/10.1093/nar/gkx201] [PMID: 28369605]
- [113] Shuai, L.; Deng, M.; Zhang, D.; Zhou, Y.; Zhou, X. Quadruplex-duplex motifs as new topoisomerase I inhibitors. *Nucleosides Nucleotides Nucleic Acids*, **2010**, *29*(11), 841-853. [http://dx.doi.org/10.1080/15257770.2010.530635] [PMID: 21128171]
- [114] Arimondo, P.B.; Riou, J.F.; Mergny, J.L.; Tazi, J.; Sun, J.S.; Garestier, T.; Hélène, C. Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.*, **2000**, *28*(24), 4832-4838. [http://dx.doi.org/10.1093/nar/28.24.4832] [PMID: 11121473]
- [115] Marchand, C.; Pourquier, P.; Laco, G.S.; Jing, N.; Pommier, Y. Interaction of human nuclear topoisomerase I with guanosine quartet-forming and guanosine-rich single-stranded DNA and RNA oligonucleotides. *J. Biol. Chem.*, **2002**, *277*(11), 8906-8911. [http://dx.doi.org/10.1074/jbc.M106372200] [PMID: 11756434]
- [116] Kato, L.; Stanlie, A.; Begum, N.A.; Kobayashi, M.; Aida, M.; Honjo, T. An evolutionary view of the mechanism for immune and genome diversity. *J. Immunol.*, **2012**, *188*(8), 3559-3566. [http://dx.doi.org/10.4049/jimmunol.1102397] [PMID: 22492685]
- [117] Chaudhuri, J.; Alt, F.W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.*, **2004**, *4*(7), 541-552. [http://dx.doi.org/10.1038/nri1395] [PMID: 15229473]
- [118] Revy, P.; Muto, T.; Levy, Y.; Geissmann, F.; Plebani, A.; Sanal, O.; Catalan, N.; Forveille, M.; Dufourcq-Labelouse, R.; Gennery, A.; Tezcan, I.; Ersoy, F.; Kayserili, H.; Ugazio, A.G.; Brousse, N.; Muramatsu, M.; Notarangelo, L.D.; Kinoshita, K.; Honjo, T.; Fischer, A.; Durandy, A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, **2000**, *102*(5), 565-575. [http://dx.doi.org/10.1016/S0092-8674(00)00079-9] [PMID: 11007475]
- [119] Muramatsu, M.; Kinoshita, K.; Fagarasan, S.; Yamada, S.; Shinkai, Y.; Honjo, T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **2000**, *102*(5), 553-563. [http://dx.doi.org/10.1016/S0092-8674(00)00078-7] [PMID: 11007474]
- [120] Bergsagel, P.L.; Chesi, M.; Nardini, E.; Brents, L.A.; Kirby, S.L.; Kuehl, W.M. Promiscuous translocations into immunoglobulin heavy chain switch regions in multiple myeloma. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*(24), 13931-13936.

- [121] Küppers, R.; Dalla-Favera, R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*, **2001**, *20*(40), 5580-5594. [http://dx.doi.org/10.1038/sj.onc.1204640] [PMID: 11607811]
- [122] Stavnezer, J.; Amemiya, C.T. Evolution of isotype switching. *Semin. Immunol.*, **2004**, *16*(4), 257-275. [http://dx.doi.org/10.1016/j.smim.2004.08.005] [PMID: 15522624]
- [123] Bottaro, A.; Lansford, R.; Xu, L.; Zhang, J.; Rothman, P.; Alt, F.W. S region transcription per se promotes basal IgE class switch recombination but additional factors regulate the efficiency of the process. *EMBO J.*, **1994**, *13*(3), 665-674. [http://dx.doi.org/10.1002/j.1460-2075.1994.tb06305.x] [PMID: 8313911]
- [124] Reaban, M.E.; Lebowitz, J.; Griffin, J.A. Transcription induces the formation of a stable RNA:DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.*, **1994**, *269*(34), 21850-21857. [PMID: 8063829]
- [125] Yu, K.; Chedin, F.; Hsieh, C.L.; Wilson, T.E.; Lieber, M.R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.*, **2003**, *4*(5), 442-451. [http://dx.doi.org/10.1038/ni919] [PMID: 12679812]
- [126] Roy, D.; Yu, K.; Lieber, M.R. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell Biol.*, **2008**, *28*(1), 50-60. [http://dx.doi.org/10.1128/MCB.01251-07] [PMID: 17954560]
- [127] Maizels, N.; Gray, L.T. The G4 genome. *PLoS Genet.*, **2013**, *9*(4), e1003468. [http://dx.doi.org/10.1371/journal.pgen.1003468] [PMID: 23637633]
- [128] Kobayashi, M.; Aida, M.; Nagaoka, H.; Begum, N.A.; Kitawaki, Y.; Nakata, M.; Stanlie, A.; Doi, T.; Kato, L.; Okazaki, I.M.; Shinkura, R.; Muramatsu, M.; Kinoshita, K.; Honjo, T. AID-induced decrease in topoisomerase I induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*(52), 22375-22380. [http://dx.doi.org/10.1073/pnas.0911879106] [PMID: 20018730]
- [129] Kobayashi, M.; Sabouri, Z.; Sabouri, S.; Kitawaki, Y.; Pommier, Y.; Abe, T.; Kiyonari, H.; Honjo, T. Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc. Natl. Acad. Sci. USA*, **2011**, *108*(48), 19305-19310. [http://dx.doi.org/10.1073/pnas.1114522108] [PMID: 22080610]
- [130] Larson, E.D.; Duquette, M.L.; Cummings, W.J.; Streiff, R.J.; Maizels, N. MutSalpα binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Curr. Biol.*, **2005**, *15*(5), 470-474. [http://dx.doi.org/10.1016/j.cub.2004.12.077] [PMID: 15753043]
- [131] Cahoon, L.A.; Seifert, H.S. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science*, **2009**, *325*(5941), 764-767. [http://dx.doi.org/10.1126/science.1175653] [PMID: 19661435]
- [132] Cahoon, L.A.; Seifert, H.S. Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in *Neisseria gonorrhoeae*. *PLoS Pathog.*, **2013**, *9*(1), e1003074. [http://dx.doi.org/10.1371/journal.ppat.1003074] [PMID: 23349628]
- [133] Norris, S.J. vls Antigenic variation systems of lyme disease borrelia: eluding host immunity through both random, segmental gene conversion and framework heterogeneity. *Microbiol. Spectr.*, **2014**, *2*(6), 2. [PMID: 26104445]
- [134] Zhang, J.R.; Norris, S.J. Genetic variation of the *Borrelia burgdorferi* gene vlsE involves cassette-specific, segmental gene conversion. *Infect. Immun.*, **1998**, *66*(8), 3698-3704. [PMID: 9673251]
- [135] Lin, T.; Gao, L.; Edmondson, D.G.; Jacobs, M.B.; Philipp, M.T.; Norris, S.J. Central role of the Holliday junction helicase RuvAB in vlsE recombination and infectivity of *Borrelia burgdorferi*. *PLoS Pathog.*, **2009**, *5*(12), e1000679. [http://dx.doi.org/10.1371/journal.ppat.1000679] [PMID: 19997622]
- [136] Walia, R.; Chaconas, G. Suggested role for G4 DNA in recombinational switching at the antigenic variation locus of the Lyme disease spirochete. *PLoS One*, **2013**, *8*(2), e57792. [http://dx.doi.org/10.1371/journal.pone.0057792] [PMID: 23469068]
- [137] Lin, Y.; Hubert, L., Jr; Wilson, J.H. Transcription destabilizes triplet repeats. *Mol. Carcinog.*, **2009**, *48*(4), 350-361. [http://dx.doi.org/10.1002/mc.20488] [PMID: 18973172]
- [138] Lin, Y.; Dent, S.Y.; Wilson, J.H.; Wells, R.D.; Napierala, M. R loops stimulate genetic instability of CTG:CAG repeats. *Proc. Natl. Acad. Sci. USA*, **2010**, *107*(2), 692-697. [http://dx.doi.org/10.1073/pnas.0909740107] [PMID: 20080737]
- [139] Nolin, S.L.; Brown, W.T.; Glicksman, A.; Houck, G.E., Jr; Gargano, A.D.; Sullivan, A.; Biancalana, V.; Brøndum-Nielsen, K.; Hjalgrim, H.; Holinski-Feder, E.; Kooy, F.; Longshore, J.; Macpherson, J.; Mandel, J.L.; Matthijs, G.; Rousseau, F.; Steinbach, P.; Väisänen, M.L.; von Koskull, H.; Sherman, S.L. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet.*, **2003**, *72*(2), 454-464. [http://dx.doi.org/10.1086/367713] [PMID: 12529854]
- [140] Haeusler, A.R.; Donnelly, C.J.; Periz, G.; Simko, E.A.; Shaw, P.G.; Kim, M.S.; Maragakis, N.J.; Troncoso, J.C.; Pandey, A.; Sattler, R.; Rothstein, J.D.; Wang, J. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **2014**, *507*(7491), 195-200. [http://dx.doi.org/10.1038/nature13124] [PMID: 24598541]