

# UC Irvine

## UC Irvine Previously Published Works

### Title

The role of G-density in switch region repeats for immunoglobulin class switch recombination.

### Permalink

<https://escholarship.org/uc/item/1w53c3zp>

### Journal

Nucleic Acids Research, 42(21)

### Authors

Zhang, Zheng  
Pannunzio, Nicholas  
Hsieh, Chih-Lin  
[et al.](#)

### Publication Date

2014-12-01

### DOI

10.1093/nar/gku1100

Peer reviewed

# The role of G-density in switch region repeats for immunoglobulin class switch recombination

Zheng Z. Zhang<sup>1</sup>, Nicholas R. Pannunzio<sup>1</sup>, Chih-Lin Hsieh<sup>1</sup>, Kefei Yu<sup>2,\*</sup> and Michael R. Lieber<sup>1,\*</sup>

<sup>1</sup>USC Norris Comprehensive Cancer Center, Molecular and Computational Biology Program, Departments of Biological Sciences; Pathology, Biochemistry & Molecular Biology; Molecular Microbiology & Immunology; Urology; University of Southern California Keck School of Medicine, 1441 Eastlake Ave., Rm. 5428, Los Angeles, CA 90089-9176, USA and <sup>2</sup>Department of Microbiology and Molecular Genetics, Michigan State University, 5175 Biomedical Physical Sciences, East Lansing, MI 48824, USA

Received September 30, 2014; Revised October 19, 2014; Accepted October 22, 2014

## ABSTRACT

**The boundaries of R-loops are well-documented at immunoglobulin heavy chain loci in mammalian B cells. Within primary B cells or B cell lines, the upstream boundaries of R-loops typically begin early in the repetitive portion of the switch regions. Most R-loops terminate within the switch repetitive zone, but the remainder can extend a few hundred base pairs further, where G-density on the non-template DNA strand gradually drops to the genome average. Whether the G-density determines how far the R-loops extend is an important question. We previously studied the role of G-clusters in initiating R-loop formation, but we did not examine the role of G-density in permitting the elongation of the R-loop, after it had initiated. Here, we vary the G-density of different portions of the switch region in a murine B cell line. We find that both class switch recombination (CSR) and R-loop formation decrease significantly when the overall G-density is reduced from 46% to 29%. Short 50 bp insertions with low G-density within switch regions do not appear to affect either CSR or R-loop elongation, whereas a longer (150 bp) insertion impairs both. These results demonstrate that G-density is an important determinant of the length over which mammalian genomic R-loops extend.**

## INTRODUCTION

The murine immunoglobulin heavy chain (IgH) locus consists of eight different constant (C) genes, most of which, except C $\delta$ , contain a cytokine-inducible promoter and a highly repetitive switch (S) region (1,2). Class switch recombination (CSR) replaces the default C $\mu$  exons with exons

for a downstream constant chain (C $\alpha$ , C $\epsilon$  or C $\gamma$ ). Thus, instead of IgM, another isotype (IgA, IgE or IgG, respectively) is produced. During this process, two double-strand breaks, one in the donor S $\mu$  and the other in the acceptor switch region, are created, and the intervening DNA region is deleted. Mammalian switch regions are all long, repetitive and G-rich on the non-template DNA strand, but there is no apparent homology among different switch regions, except for two short motifs (AGCT and GGGG) (3). These two motifs are enriched in mammalian switch regions relative to their neighboring regions, or elsewhere in the genome.

Transcription through switch regions is essential for CSR (3). The transcription of the IgH S $\mu$  region is constitutive, but transcription of downstream regions is only active when certain cytokines are present at the time of antigen stimulation. A major role of transcription during CSR is to induce R-loop formation in the switch regions, which provides stable single-stranded DNA substrates for deamination of C to U by activation-induced deaminase (AID) (2,4,5). The G-richness on the non-template DNA strand is critical for R-loop formation in the switch region sequence in cell-free biochemical systems. The distance between the promoter and the switch regions reduces R-loop formation as well, probably due to a lower frequency of productive collisions between the G-rich portion of the transcript and the corresponding template DNA strand (6–8). Cellular (*in vivo*) studies at murine  $\gamma$ 3 and  $\gamma$ 2b regions have shown that most R-loops terminate within the switch regions, but some extend to within 600 bp downstream of the switch regions (9). The downstream boundaries of R-loops coincide with the zone where G-density gradually falls to the level of the genome average. However, no direct cellular analysis has been performed to evaluate the role of high G-density in mammalian switch regions during R-loop formation and CSR. Recently, we studied the role of G-clusters in initiating

\*To whom correspondence should be addressed. Tel: +323 865 0568; Fax: +323 865 3019; Email: lieber@usc.edu  
Correspondence may also be addressed to Kefei Yu. Tel: +517 884 5354; Fax: +517 353 8957; Email: yuke@msu.edu

ing R-loop formation, but we did not evaluate the role of G-density in permitting the elongation of the R-loop once it had initiated (10).

In biochemical studies with purified DNA and prokaryotic RNA polymerases, G-clusters (2–4 directly adjacent G's) are important for the newly transcribed RNA to thread back onto the template DNA strand to form an R-loop (6–8). This is almost indistinguishable from the formation of R-loops at DNA replication origins in prokaryotes, where a G-cluster of a newly transcribed RNA must anneal to the template DNA strand to initiate DNA replication (11). Factors that favor annealing of the RNA to the DNA template, such as negative superhelicity, non-template DNA strand nicks and proximity of the G-cluster to the 5' end of the RNA all favor more efficient R-loop formation. For switch sequences, we term the G-cluster rich region responsible for initiating the R-loops the 'R-loop initiation zone' or RIZ. Once initiated in the RIZ, we find that the R-loop can extend downstream of the RIZ into a region with no G-clusters but in which the G-density is relatively high (35–50% G). We call this region the 'R-loop elongation zone' or REZ.

Here, we vary the G-density in the elongation zone (REZ) of IgH switch regions in the genome of a murine B cell line, in contrast to our earlier genome study of G-clusters in the RIZ (10), to investigate its role in R-loop elongation. Consistent with our earlier *in vitro* studies (6–8), we find that a low G-density in the REZ dramatically decreases R-loop formation and CSR efficiency at the natural S $\alpha$  locus in murine B cells. Short insertions (50 bp) within the switch regions do not appear to affect either CSR or R-loop elongation, regardless of the G-richness of the insertions. However, both CSR and R-loop elongation are impaired when longer, low G-density sequences are added within the switch regions. These findings explain the importance of a high G-density at mammalian switch regions.

## MATERIALS AND METHODS

### Cell culture and CSR assay

CH12F3.2a and its derivative cells were cultured in 10% fetal calf serum (FCS) RPMI medium supplemented with 50  $\mu$ M  $\beta$ -mercaptoethanol. As for CSR assay, healthy cells in log phase were seeded at  $5 \times 10^4$  cells/ml in medium with or without cytokine stimulation (abbreviated CIT) as follows: 1  $\mu$ g/ml anti-CD40 (eBioscience #16-0404-86), 5 ng/ml IL-4 (R&D #404-ML-010) and 0.5 ng/ml TGF- $\beta$ 1 (R&D #240-B-002), and grown for 72 h. Cells were stained with fluorescein isothiocyanate-conjugated anti-mouse IgA antibody (BD #559354) and analyzed by flow cytometry. CSR efficiency was determined based on the percentage of IgA+ cells minus the percentage without CIT stimulation.

### Plasmid construction

The 80 nt oligonucleotides corresponding to each REZ repeat unit were 5'-phosphorylated and purified on a denaturing polyacrylamide gel (Supplementary Table S1). A pair of complementary oligonucleotides was annealed and self-ligated at a final concentration of 4  $\mu$ M. Ligation products were resolved on agarose gel, and DNA with expected

length was recovered, blunted with Klenow and cloned into the exchange vector backbone. The 50-bp insertions were 5'-phosphorylated, annealed and cloned into the exchange vector backbone. The 150-bp insertion was amplified from a bacteria plasmid pGG49, and cloned into the exchange vector. The sequence of the exchange vectors was confirmed by DNA sequencing.

Given the DNA sequence complexity of IgH class switch regions, there are well over  $10^{100}$  variations that one could consider constructing. Combined with our earlier study of G-clustering at the IgH switch regions (10), we have analyzed only slightly over 100 variations (each with 5–7 mammalian cell clones). Given the limited number of variations analyzed, we have tried to be cautious about extrapolating. Relevant to this, we do not endorse any of the current computer programs that predict R-loop formation based on DNA sequence (12).

### Cellular targeting and screening

Five micrograms exchange vector and one microgram Cre-expression vector were co-transfected into 1F7 cells by electroporation (Lonza). Transfected cells were serially diluted and seeded in 96-well plates. After 72 h, ganciclovir (Sigma-Aldrich #G2536-100MG) was added at a final concentration of 2  $\mu$ g/ml. At 7 days after transfection, single clones were tested for puromycin sensitivity at a final drug concentration of 1  $\mu$ g/ml. Puromycin-sensitive clones were screened by polymerase chain reaction (PCR) adjacent to the upstream and downstream boundaries of LoxP sites and across the entire switch region. A subset of clones were also examined by Southern blot. At least five clones were used for CSR assay.

### Native bisulfite sequencing

The principle of native (non-denaturing) bisulfite genomic sequencing is that all C's in duplex nucleic acid configuration (DNA:DNA or RNA:DNA) are resistant to bisulfite conversion from C to U (then remain to read as C during sequence analysis). But regions of single-stranded nucleic acid are subject to C to U (then read as T in the subsequent sequencing analysis) conversion (4,13).

For analyses, 1 million healthy cells at a density  $\sim 10^5$  cells/ml were supplemented with anti-CD40, IL-4 and TGF- $\beta$ 1 for 48 h, and genomic DNA was extracted. Five micrograms of genomic DNA was incubated with the bisulfite solution (Lightning Conversion Reagent by Zymo Research D5030) for 16 h at 37°C, and recovered as described by the manufacturer. The switch regions were amplified with one converted primer and one native primer by PCR (Supplementary Table S2), and subcloned into the pGEM-T Easy vector (Promega A1360). At least 12 clones from each switch region were randomly chosen and analyzed.

As discussed in our earlier descriptions of bisulfite genomic sequencing on native nucleic acid, there is some level of background C to U (read as a T after PCR) even when there is no R-loop (4,13). Therefore, we set 80 bp as the minimal length of C conversion before we regard this as indicating an R-loop. In this study, each synthetic S $\alpha$  repeat has 4 AGCT sites, and thus 4 C's. This means that four consecutive C conversions is the minimal length of an R-loop.

### S9.6 purification

ATCC HB-8730 hybridoma line (generously provided by Bradley Cairns) was cultured in a CELLLine 1000 bioreactor (Satorius Biotech, NY, USA) according to manufacturer's instructions. Harvested antibody (culture supernatant) was purified on a column packed with Protein G Sepharose 4 Fast Flow (GE Healthcare) equilibrated with  $1 \times$  phosphate buffered saline.

### S9.6 immunoprecipitation (IP)

Healthy cells in log phase were seeded at  $3 \times 10^5$  cells/ml in medium with anti-CD40, IL-4 and TGF- $\beta$ 1 and grown for 24 h. Genomic DNA was prepared by overnight proteinase K digestion, phenol-chloroform extraction and ethanol precipitation. Genomic DNA was digested with EcoRI in NEB buffer 2 [10 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 50 mM NaCl, 1 mM DTT pH 7.9]; importantly, RNase A, at a final concentration of 0.4 mg/ml, was added at this step to prevent S9.6 antibody binding to RNA species in subsequent steps (14). [In a separate methods study, we find that RNase A addition is needed to obtain adequate IP of RNA:DNA hybrids using S9.6. Without RNase A treatment, S9.6 yields a much lower (near background) IP (ZZZ, NP and MRL, submitted).] Five micrograms of fragmented genomic DNA was incubated with 5  $\mu$ g S9.6 antibody in 400  $\mu$ l IP buffer (10 mM sodium phosphate (pH 7.0), 140 mM NaCl, 0.1% Tween 20) for 2 h at 4°C. Ten microliter pre-blocked Dynabeads (Invitrogen 10004D) were added into the mixture and gently rotated at 4°C. After 2 h, beads were washed with IP buffer three times, and treated with proteinase K overnight. DNA bound to the beads was recovered by phenol-chloroform extraction, and quantified by quantitative (real-time) PCR (Supplementary Table S3).

### Germ-line transcript quantification

Two million healthy cells at a density of  $10^6$  cells/ml were supplemented with anti-CD40, IL-4 and TGF- $\beta$ 1 for 6 h, and total RNA was extracted with GenElute™ Mammalian Total RNA Kit (Sigma-Aldrich #RTN350). Note that 10% of RNA was reverse transcribed (RT) into cDNA with M-MuLV Reverse transcriptase (NEB M0253S), and 10% of the RT products were analyzed with quantitative PCR using Taqman probe/primers.  $\beta$ -Actin was used as an internal control. Each sample was done in duplicate, and at least three independent cellular clones were analyzed for each construct.

## RESULTS

### Experimental strategy

To examine the importance of a high G-density in mammalian switch regions, we used a system that allows us to exchange out the genomic switch region and replace it with any desired synthetic switch sequence in a mouse B-cell line, CH12F3.2a (15). Such an exchange permits us to study how switch sequence variations affect R-loop formation and cellular CSR *in vivo*. This cell line is able to specifically and efficiently switch to IgA upon cytokine stimula-

tion. Using this cellular assay system, the endogenous  $\alpha$  locus in CH12F3.2a was replaced with a positive-negative selection cassette (Puro $\Delta$ TK, provides puromycin resistance and ganciclovir sensitivity) flanked by two different LoxP sites (10,15). The resulting cell line is called 1F7. The sequence of interest is cloned into an exchange vector with the same LoxP sites as 1F7 cells. This exchange vector is co-transfected into 1F7 cells along with a Cre-expressing vector. Cre mediates recombination between the exchange vector and the 1F7 chromosome at the corresponding LoxP sites (Figure 1A). The successful replacement of the selection cassette by the sequence of interest gives rise to ganciclovir resistant clones, which are screened with a puromycin sensitivity test and PCR analysis for all clones. The CSR assay is described in the Materials and Methods. R-loop formation is determined using native bisulfite genomic sequencing and by IP using an RNA:DNA antibody called S9.6 (see Materials and Methods).

### Overall G-density is important for efficient CSR

Mammalian switch regions are extremely G-rich on the non-template DNA strand. G-clusters have been shown to be critical for CSR (10). Previous *in vitro* experiments indicate that a high G-density is important for CSR, but no direct *in vivo* evidence has addressed this issue (9,16).

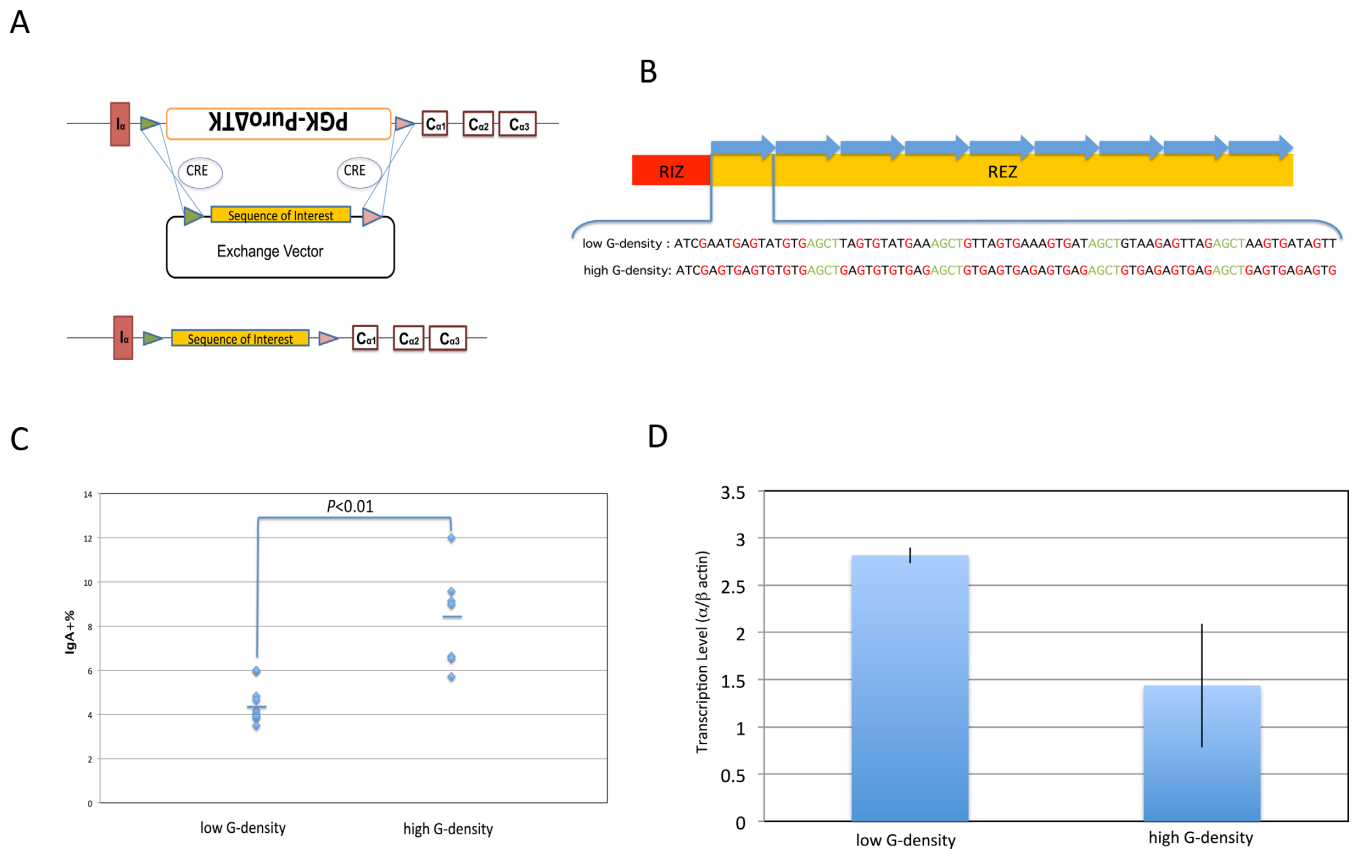
To evaluate the potential importance of high G-density for CSR, we constructed two cell lines with exactly the same length (9 repeats) and number of AGCT sites (4 within each repeat), but different G-densities on the non-template DNA strand (Figure 1B). One has an average G-density of 29% on the non-template DNA strand, whereas the other one has an average G-density of 46%. Using the same RIZ, we find that an REZ with higher G-density causes a significant increase in CSR (Figure 1C). The high G-density REZ leads to a decrease in transcription compared with a low G-density REZ (Figure 1D). The drop in transcription may be due to obstruction of the RNA polymerase by the R-loop, as we have seen in our biochemical studies (17). Regardless of the cause, this finding rules out increased transcription as the cause of the increased CSR, and clearly indicates that higher G-density in the REZ results in higher CSR.

### High G-density is important to maintain R-loops

Given the identical RIZ and number of AGCT sites, the R-loop initiation and AID target number (WGCW site number) are both the same between the two constructs above. The only difference, G-density within the REZ, may affect R-loop elongation, based on our earlier cell-free biochemical observations (6–8).

We wondered if R-loops collapse more in the switch region with a low G-density on the non-template DNA strand. We performed bisulfite sequencing for the cellular constructs above. We find that 8 out of 12 of the molecules have long stretches of conversion when the G-density is high in the REZ (see legend to Figure 2A), and only 2 out of 12 when the G-density is low in the REZ (Figure 2B). Moreover, the ratio of converted C nucleotides in the cell line with the higher G-density on the non-template DNA strand is higher than that with the lower G-density (44% versus 31%,





**Figure 1.** Overall G-density is important for efficient CSR. (A) 1F7 cells, which are resistant to puromycin and sensitive to ganciclovir, are transfected with both an exchange vector containing the sequence to be integrated (sequence of interest) and a CRE-expression vector. Cre mediates recombination between the exchange vector and the 1F7 chromosome at the corresponding LoxP sites. Cells that have successfully replaced the selection cassette by the sequence of interest would survive during ganciclovir selection, which are further screened with a puromycin sensitivity test and PCR. (B) Nine repeats of 80 bp oligonucleotides (top strand of duplex shown below the RIZ/REZ diagram) with 4 AGCT sites (orange) and different G-densities were used for the REZ in the synthetic switch regions. The top panel has 29% G-density on the non-template DNA strand, whereas the bottom panel has 46% G-density. The RIZ contains 5 G-clusters. (C) Fluorescence-activated cell sorting (FACS) analysis of CSR is shown. Each data symbol represents an independent clone (after subtraction of the IgA<sup>+</sup> level for the same clone without CIT stimulation). The bars represent the mean of each group. (D) Germ-line transcription (GLT) Quantification. Healthy cells grown to  $10^6$ /ml were treated with 1  $\mu$ g/ml anti-CD40, 5 ng/ml IL-4 and 0.5 ng/ml TGF- $\beta$ 1 (CIT) for 6 h. Total RNA was extracted, reverse-transcribed into cDNA and quantified by real-time PCR.  $\beta$ -Actin was used as an internal control. In parallel, the same cellular clones without CIT treatment were used to measure the background of GLT. The error bar represents the SEM of 3–5 independent clones.

$P < 0.05$ ). No converted C nucleotides were observed on the template DNA strand.

To further evaluate the relationship between G-density and R-loop formation, we used an antibody, called S9.6, to detect R-loops (Figure 2C). This antibody has been documented to preferably bind RNA:DNA hybrids over DNA:DNA or RNA:RNA duplexes, if free RNA is removed using RNase A. We find that there is consistently more pull-down of R-loops at this location of the genome when an REZ with 46% G (non-template strand) is used rather than 29%.

The difference in S9.6 pull-down is not large, though it is quite reproducible (Figure 2C), likely because the short R-loops that are expected to arise in the RIZ are sufficient to permit S9.6 pull-down.

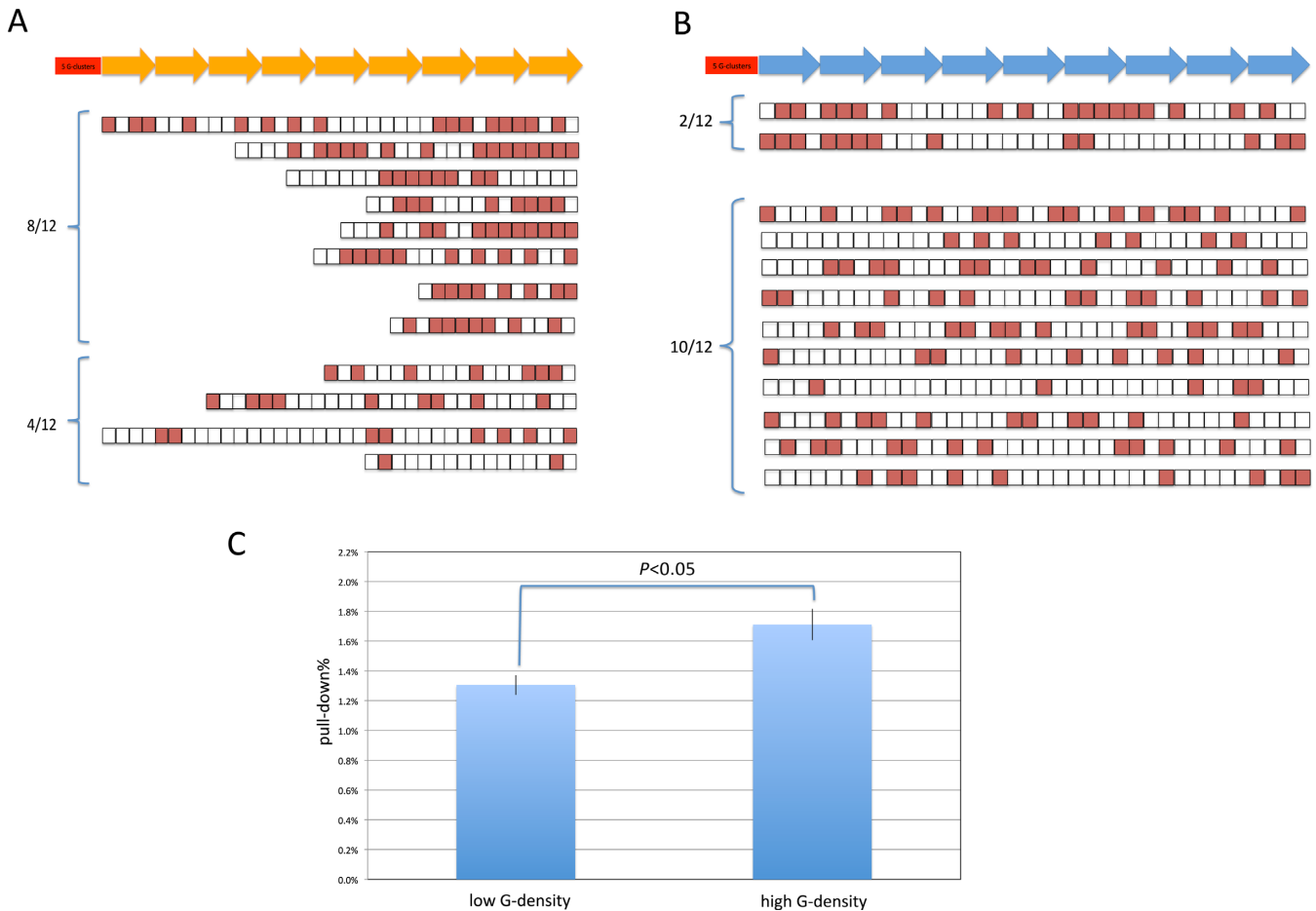
Therefore, a high G-density in the REZ supports more frequent R-loop elongation within the cells.

### R-loops can proceed across a short region with low G-density

Although the distance of DNA between switch regions and their upstream promoters (typically 200–600 bp in length) are not G-rich (on the non-template strand) at mammalian IgH loci, a few G-clusters (e.g. three G-clusters of 3 or 4 nt each at  $S\alpha$ ) typically exist within these regions. The R-loops initiated from those G-clusters may extend into the switch regions, where the R-loop extension may proceed efficiently. Thus, we wondered how regions of low G-density may affect R-loop stability between regions of high G-density.

To explore whether R-loops can tolerate short regions of low G-density, we inserted two pieces of DNA with different sequences between the RIZ and the REZ (Figure 3A). If R-loops are sensitive to G-density and collapse once the G-density is low, then the R-loops that begin in the RIZ would not reach the REZ portion, and AID/UNG/APE1 would not have an opportunity to create DNA breaks. In this case, CSR efficiency would drop dramatically.

We find that CSR efficiency is not affected by the 50-bp insertion between the RIZ and REZ (Figure 3B). How-



**Figure 2.** Longer and more frequent R-loops are detected using bisulfite sequencing when the REZ has a high G-density. Bisulfite sequencing analysis was done on the cellular constructs. Detailed R-loop information is shown with high (A) or low (B) overall G-density on the non-template DNA strand, respectively. Solid boxes represent C nucleotides that have been converted to U (and then T) nucleotides due to bisulfite. The results are grouped according to alleles that contain at least the equivalent of one 80 bp repeat unit of C conversion. This corresponds to four or more consecutive C conversions (top). Those with fewer than four consecutive C conversions (less than one 80 bp repeat unit of conversion) are shown in the lower group and designated as short regions of conversion. (C) IP with the S9.6 antibody was performed on the cellular constructs with different overall G-density in Figure 1. Background signals from mock samples with no antibody were subtracted. Values were normalized to the total input DNA to calculate the pull-down percentage. Three independent IP experiments were performed for each cell line.

ever, when the size of the insert between the RIZ and REZ is increased to 150 bp with a low G-density sequence on the non-template DNA strand, the CSR does drop substantially (Figure 3C), even though transcription is unaffected (Figure 3D).

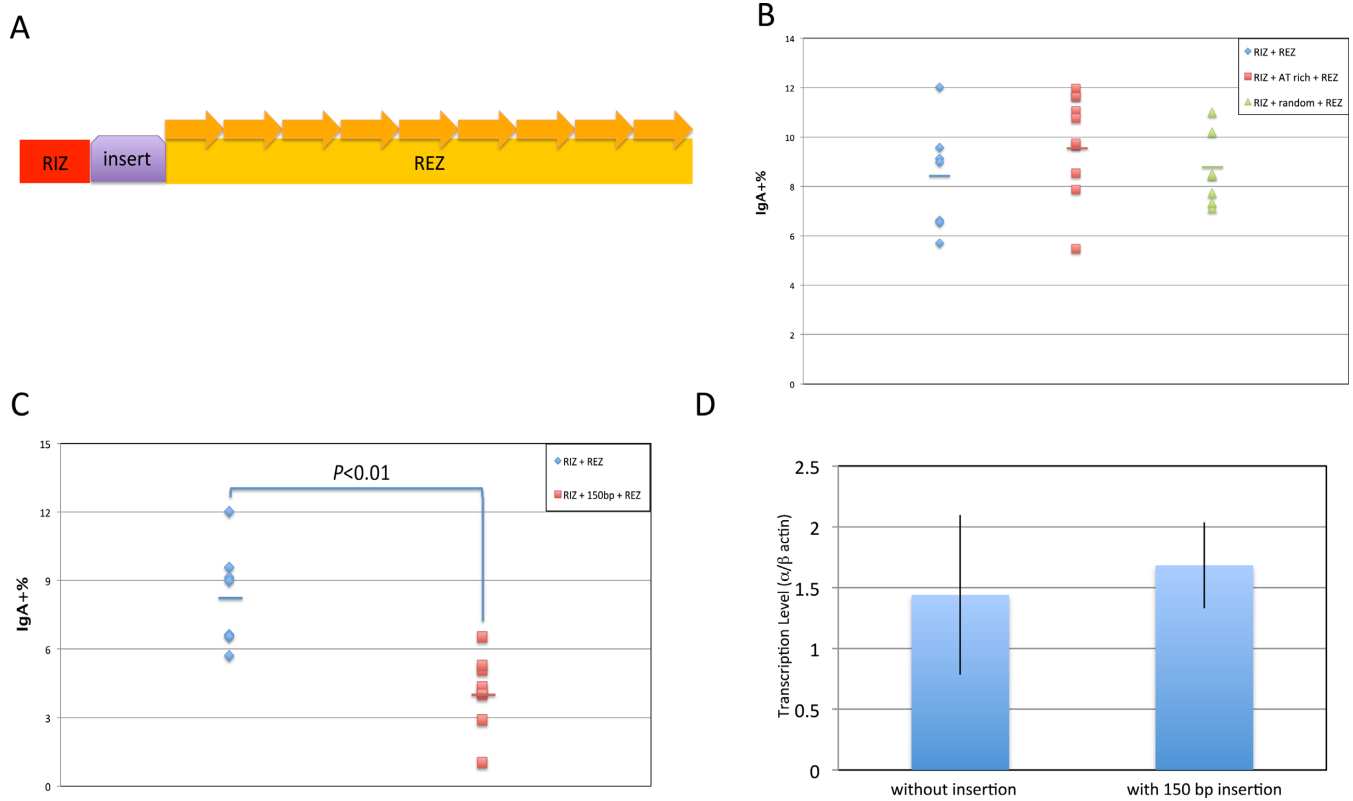
To evaluate whether there is a correlation between the CSR efficiency and the R-loop status, we performed bisulfite sequencing for the cellular clones above. We found that with the short insertion (50 bp), regardless of the G-richness, R-loop status is not affected (Table 1), which agrees with the fact that CSR efficiency of these cellular constructs is not affected. However, fewer R-loops are observed with a longer (150 bp) insertion of low G-density (Table 1). We conclude that R-loops can tolerate short interruptions by a sequence of low G-density.

We also tested whether insertion of 50 bp upstream of the entire RIZ/REZ region would affect CSR within cells (Figure 4A). Our cell-free biochemical work showed that a greater distance between the transcriptional promoter and the RIZ/REZ region reduces R-loop formation (8). Consis-

tent with our cell-free biochemical studies, CSR is reduced when the distance between the promoter and the RIZ/REZ switch sequences is increased (Figure 4B).

## DISCUSSION

Based on the much higher ratio of [switched isotype (e.g. IgG in mammals or IgX in amphibians) divided by unswitched isotype (IgM)] in mammals versus more distantly evolved vertebrates, we have speculated that mammalian Ig CSR may be more efficient than other species that carry out CSR. If so, some of the greater efficiency may be due to the distinctively G-rich nature of the switch sequence on the non-template strand. Upon transcription through switch regions, G-richness on the non-template DNA strand favors R-loop formation. R-loop formation provides stable single-stranded DNA, which is required as a substrate for AID. The *in vivo* boundaries of R-loops have been documented (9,16), and G-clusters have been demonstrated to be crucial for R-loop initiation *in vitro* and *in vivo*



**Figure 3.** R-loops can proceed across a short region of DNA that has a low G-density. (A) DNA with different lengths (purple) were inserted between RIZ (containing 5 G-clusters) and REZ (containing 9 repeats with high G-density and 4 AGCT sites within each repeat). (B) A 50-bp segment of DNA with either AT-rich sequence or random sequence was inserted between RIZ and REZ. FACS analysis is shown. Each data symbol represents an independent clone. The bars represent the mean of each group. (C) A 150-bp segment of DNA with AT-rich sequence was inserted between RIZ and REZ. FACS analysis is shown. Each data symbol represents an independent clone. The bars represent the mean of each group. The Student's *t*-test was used to analyze the effect of the insert. (D) Transcription assay for the clones in panel C. The same method was used as described in Figure 1D and in the Materials and Methods. We observe no significant difference in transcription here in panel D.

**Table 1.** R-loop analysis using bisulfite sequencing on different cellular constructs

		Mean% IgA <sup>+</sup>	% R-loop (actual #)
1	RIZ + REZ	8.4	67 (8/12)
2	RIZ + 50bp AT-rich + REZ	9.8	73 (16/22)
3	RIZ + 50bp random + REZ	8.6	60 (12/20)
4	RIZ + 150bp AT-rich + REZ	4.2	31 (11/36)

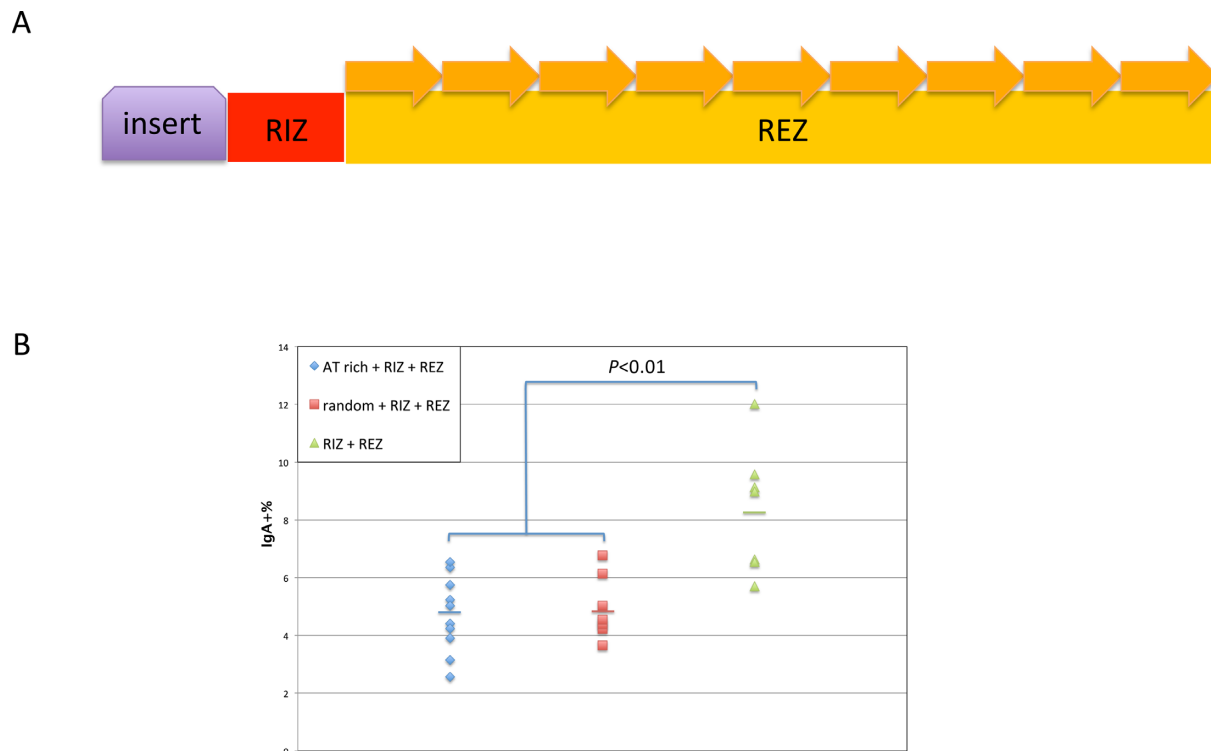
(6–8,10). Decline in G-density on the non-template DNA strand is thought to account for the collapse of R-loops, but no direct evidence has been reported. Here, we have shown that decreasing G-density at mammalian IgH switch regions impairs both R-loop formation and CSR (Figures 1 and 2).

We find that the distance between switch regions and their respective promoters affects CSR efficiency (Figure 4). This agrees with the biochemical studies on R-loop formation (8). Upstream portions of the RNA transcript that are not G-rich may reduce annealing of the G-rich portion of the RNA to the C-rich template DNA. Mammalian switch regions are located several hundred base pairs downstream of their respective promoters. Though these sequences between switch regions and promoters are not G-rich, ample G-clusters are found within these regions. Our finding that 50 bp insertions between RIZ and REZ do not affect CSR indicates that R-loops can extend across short regions with

low G-density. In this case, R-loops initiated from the G-clusters upstream of switch regions may extend and reach switch regions, or they may increase the frequency of collisions between RNA and DNA around switch regions, where longer and more stable R-loops can form.

The length of low G-density sequence through which R-loops can persist likely accounts for the wide spectrum of downstream boundaries of R-loops at mammalian switch regions. The downstream boundaries do not have a distinct point of R-loop termination at the end of the zone of high G-density (9). Rather, they are distributed up to 600 bp downstream, as the G-density gradually declines below 35%. The persistence of R-loops through regions of low G-density (50 bp here) illustrates that R-loops do not terminate acutely at a boundary of high to low G-density.

Overall, we see very good correlation between our cell-free biochemical studies of R-loop formation (6–8) and



**Figure 4.** Distance between G-clusters and promoter influences CSR. (A) The same 50 bp DNA segment as in Figure 3B was inserted immediately upstream of the RIZ (the RIZ contains 5 G-clusters). (B) FACS analysis is shown. Each data symbol represents an independent cellular clone. The bars represent the mean of each group. The Student's *t*-test was used to analyze the effect of the insert.

*in vivo* cellular R-loop formation and CSR at the mammalian chromosomal level. The study here focused on the G-density in the REZ. In our earlier study of how G-clusters affect CSR *in vivo*, we included some constructs where G-clusters were embedded in the REZ repeats (10). The results from those constructs were consistent with the findings here in indicating that higher G-density in the REZ increases CSR. The results here are more directly interpretable because only the G-density is varied, and not the G-clustering.

We have not done a formal test of how R-loops interfere with RNA polymerase II progression. We do note that the total number of transcripts produced *in vitro* when purified prokaryotic RNA polymerases are used to transcribe through switch sequences on purified DNA templates is lower than if random sequences are used (17). Perhaps RNA polymerase moves more slowly through GC-rich sequences. It is known that RNA polymerase II accumulates in switch regions (18,19). It is possible that such accumulation of RNA polymerase contributes to the stability of R-loops once formed.

There have been numerous recent reports of R-loops at other locations in the eukaryotic genome (14,20–32), outside of the IgH locus. Nearly all of these studies rely heavily on the S9.6 antibody for immunostaining or IP of RNA:DNA duplexes. As mentioned in the Materials and Methods, we find important differences in the IP of the IgH switch regions, depending on whether the harvested genomic nucleic acid is treated with RNase A prior to IP (ZZ, NP and MRL, submitted). In light of our experience and in light of studies demonstrating S9.6 binding to RNA:RNA

(14,33), it may be useful to confirm the existence of R-loops at other genomic locations using the same types of functional and biochemical studies that have been applied to the IgH switch regions over the last 25 years (2–8,13,17,34–38). If these other locations are confirmed, then the rules for R-loop formation defined here will help understand the broader contribution of R-loops to physiologic and pathologic events in the nucleus.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Funding for open access charge: NIH.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Stavnezer, J., Guikema, J.E. and Schrader, C.E. (2008) Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.*, **26**, 261–292.
2. Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A. *et al.* (2007) Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.*, **94**, 157–214.
3. Dunnick, W.A., Hertz, G.Z., Scappino, L. and Gritzmacher, C. (1993) DNA sequence at immunoglobulin switch region recombination sites. *Nucleic Acid Res.*, **21**, 365–372.
4. Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E. and Lieber, M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.*, **4**, 442–451.



5. Yu, K. and Lieber, M.R. (2003) Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair*, **2**, 1163–1174.
6. Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell. Biol.*, **29**, 3124–3133.
7. Roy, D., Yu, K. and Lieber, M.R. (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell. Biol.*, **28**, 50–60.
8. Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L. and Lieber, M.R. (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol. Cell. Biol.*, **30**, 146–159.
9. Huang, F.-T., Yu, K., Hsieh, C.-L. and Lieber, M.R. (2006) The downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. *Proc. Natl. Acad. Sci.*, **103**, 5030–5035.
10. Zhang, Z.Z., Pannunzio, N.R., Han, L., Hsieh, C.L., Yu, K. and Lieber, M.R. (2014) The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. *Cell Rep.*, **8**, 557–569.
11. Masukata, H. and Tomizawa, J. (1990) A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell*, **62**, 331–338.
12. Wongsurawat, T., Jenjaroenpun, P., Kwok, C.K. and Kuznetsov, V. (2012) Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.*, **40**, e16.
13. Yu, K., Roy, D., Huang, F.T. and Lieber, M.R. (2006) Detection and structural analysis of R-loops. *Methods Enzymol.*, **409**, 316–329.
14. Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H. and Leysath, C.E. (2013) The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J. Mol. Recognit.*, **26**, 376–381.
15. Han, L., Masani, S. and Yu, K. (2011) Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11584–11589.
16. Huang, F.-T., Yu, K., Balter, B.B., Selsing, E., Oruc, Z., Khamlichi, A.A., Hsieh, C.-L. and Lieber, M.R. (2007) Sequence-dependence of chromosomal R-loops at the immunoglobulin heavy chain Smu class switch region. *Mol. Cell. Biol.*, **27**, 5921–5932.
17. Daniels, G.A. and Lieber, M.R. (1995) RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.*, **23**, 5006–5011.
18. Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R. and Gearhart, P.J. (2009) Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J. Exp. Med.*, **206**, 1237–1244.
19. Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A. and Kenter, A.L. (2009) S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.*, **206**, 1817–1830.
20. Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S. and Hieter, P. (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.*, **26**, 163–175.
21. Skourti-Stathaki, K., Proudfoot, N.J. and Gromak, N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell*, **42**, 794–805.
22. Wahba, L., Amon, J.D., Koshland, D. and Vuica-Ross, M. (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell*, **44**, 978–988.
23. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J. and Dean, C. (2013) R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science*, **340**, 619–621.
24. El Hage, A., French, S.L., Beyer, A.L. and Tollervey, D. (2010) Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.*, **24**, 1546–1558.
25. Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A. and Proudfoot, N.J. (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell*, **41**, 21–32.
26. Yang, Y., McBride, K.M., Hensley, S., Lu, Y., Chedin, F. and Bedford, M.T. (2014) Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R Loop Accumulation. *Mol. Cell*, **53**, 484–497.
27. Pfeiffer, V., Crittin, J., Grolimund, L. and Lingner, J. (2013) The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J.*, **32**, 2861–2871.
28. Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K. and Luke, B. (2013) Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.*, **20**, 1199–1205.
29. Nakama, M., Kawakami, K., Kajitani, T., Urano, T. and Murakami, Y. (2012) DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells*, **17**, 218–233.
30. Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I. and Chedin, F. (2013) GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.*, **23**, 1590–1600.
31. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. and Chedin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell*, **45**, 814–825.
32. Skourti-Stathaki, K., Kamieniarz-Gdula, K. and Proudfoot, N.J. (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, doi:10.1038/nature13787.
33. Hu, Z., Zhang, A., Storz, G., Gottesman, S. and Leppla, S.H. (2006) An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res.*, **34**, e52.
34. Davis, M., Kim, S.K. and Hood, L.E. (1980) DNA sequences mediating class switching in a-immunoglobulins. *Science*, **209**, 1360–1365.
35. Shinkura, R., Tian, M., Khuong, C., Chua, K., Pinaud, E. and Alt, F.W. (2003) The influence of transcriptional orientation on endogenous switch region function. *Nat. Immunol.*, **4**, 435–441.
36. Reaban, M.E. and Griffin, J.A. (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature*, **348**, 342–344.
37. Reaban, M.E., Lebowitz, J. and Griffin, J.A. (1994) Transcription induces the formation of a stable RNA:DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.*, **269**, 21850–21857.
38. Yu, K., Roy, D., Bayramyan, M., Haworth, I.S. and Lieber, M.R. (2005) Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol. Cell. Biol.*, **25**, 1730–1736.