Journal of Visualized Experiments

Mapping R-loops and RNA:DNA hybrids with S9.6-based immunoprecipitation methods

--Manuscript Draft--

Article Type:	Invited Methods Collection - JoVE Produced Video
Manuscript Number:	JoVE62455R3
Full Title:	Mapping R-loops and RNA:DNA hybrids with S9.6-based immunoprecipitation methods
Corresponding Author:	Frederic Chedin UNITED STATES
Corresponding Author's Institution:	
Corresponding Author E-Mail:	flchedin@ucdavis.edu
Order of Authors:	Lionel Sanz
	Daisy Castillo-Guzman
	Frederic Chedin
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please specify the section of the submitted manuscript.	Genetics
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Davis, California, United States of America
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please provide any comments to the journal here.	
Please indicate whether this article will be Standard Access or Open Access.	Standard Access ($1400)

TITLE:

Mapping R-Loops and RNA:DNA Hybrids with S9.6-Based Immunoprecipitation Methods

AUTHORS AND AFFILIATIONS:

Lionel A. Sanz, Daisy Castillo-Guzman, Frédéric Chédin

Department of Molecular and Cellular Biology and Genome Center, University of California, Davis, California, USA.

Email addresses of co-authors:

Lionel A. Sanz (lsanz@ucdavis.edu)
Daisy Castillo-Guzman (dcastilloguzman@ucdavis.edu)
Frédéric Chédin (flchedin@ucdavis.edu)

Corresponding authors:

Frédéric Chédin (flchedin@ucdavis.edu)

KEYWORDS:

R-loops, DNA:RNA hybrids, S9.6, R-loop mapping, sonication, immunoprecipitation, strand-specific library

SUMMARY:

R-loops constitute a prevalent class of transcription-driven non-B DNA structures that occur in all genomes depending on both DNA sequence and topological favorability. In recent years, R-loops have been implicated in a variety of adaptive and maladaptive roles and have been linked to genomic instability in the context of human disorders. As a consequence, the accurate mapping of these structures in genomes is of high interest to many investigators. DRIP-seq (DNA:RNA Immunoprecipitation followed by high throughput sequencing) is described here. It is a robust and reproducible technique that permits accurate and semi-quantitative mapping of R-loops. A recent iteration of the method is also described in which fragmentation is accomplished using sonication (sDRIP-seq), which allows strand-specific and high-resolution mapping of R-loops. sDRIP-seq thus addresses some of the common limitations of the DRIP-seq method in terms of resolution and strandedness, making it a method of choice for R-loop mapping.

INTRODUCTION:

R-loops are three-stranded nucleic acid structures that form primarily during transcription upon hybridization of the nascent RNA transcript to the template DNA strand. This results in the formation of an RNA:DNA hybrid and causes the displacement of the non-template DNA strand in a single-stranded looped state. Biochemical reconstitution[1–4] and mathematical modeling[5], in combination with other biophysical measurements[6,7], have established that R-loops are more likely to occur over regions that exhibit specific favorable characteristics. For instance, regions that display strand asymmetry in the distribution of guanines (G) and cytosines (C) such that the RNA is G-rich, a property called positive GC skew, are favored to form R-loops when transcribed owing to the higher thermodynamic stability of the DNA:RNA hybrid compared to the DNA duplex[8]. Regions that have evolved positive GC skew, such as the early portions of many eukaryotic genes[4,9–11], are prone to forming R-loops *in vitro* and *in*

vivo[3,4,12]. Negative DNA superhelical stress also greatly favors structure formation[13,14] because R-loops efficiently absorb such topological stresses and return the surrounding DNA fiber to a favorable relaxed state[5,15].

Historically, R-loop structures were considered to result from rare, spontaneous, entanglements of RNA with DNA during transcription. However, the development of DNA:RNA immunoprecipitation (DRIP) coupled with high-throughput DNA sequencing (DRIP-seq) allowed the first genome-wide mapping of R-loops and revealed that those structures are far more prevalent than expected in human cells[4,16]. R-loops occur over tens of thousands of conserved, transcribed, genic hotspots in mammalian genomes, with a predilection for GC-skewed CpG islands overlapping the first intron of genes and the terminal regions of numerous genes[17]. Overall, R-loops collectively occupy 3%–5% of the genome in human cells, consistent with measurements in other organisms, including yeasts, plants, flies, and mice[18–22].

Analysis of R-loop forming hotspots in human cells revealed that such regions associate with specific chromatin signatures[23]. R-loops, in general, are found over regions with lower nucleosome occupancy and higher RNA polymerase density. At promoters, R-loops associate with increased recruitment of two co-transcriptionally deposited histone modifications, H3K4me1 and H3K36me3[17]. At gene termini, R-loops associate with closely arranged genes that undergo efficient transcription termination[17], consistent with prior observations[24]. R-loops were also shown to participate in the initiation of DNA replication at the replication origins of bacteriophage, plasmid, mitochondrial, and the yeast genomes[25–31]. In addition, 76% of R-loop-prone human CpG island promoters function as early, constitutive replication origins[32–35], further reinforcing the connections between R-loops and replication origins. Collectively, these studies suggest that R-loops represent a novel type of biological signal that can trigger specific biological outputs in a context-dependent manner[23].

Early on, R-loops were shown to form at class switch sequences during the process of immunoglobulin class switch recombination[3,36,37]. Such programmed R-loops are thought to initiate class switch recombination through the introduction of double-stranded DNA breaks[38]. Since then, harmful R-loop formation, generally understood to result from excessive R-loop formation, has been linked to genomic instability and processes such as hyper recombination, transcription-replication collisions, replication, and transcriptional stress (for review[39–43]). As a consequence, improved mapping of R-loop structures represents an exciting and essential challenge to better decipher the distribution and function of these structures in health and disease.

DNA:RNA immunoprecipitation (DRIP) relies on high affinity of the S9.6 monoclonal antibody for DNA:RNA hybrids[44]. DRIP-seq permits robust genome-wide profiling of R-loop formation[4,45]. While useful, this technique suffers from limited resolution due to the fact that restriction enzymes are used to achieve gentle DNA fragmentation. In addition, DRIP-seq does not provide information on the directionality of R-loop formation. Here, we report a variant of DRIP-seq that permits the mapping of R-loops at high resolution in a strand-specific manner. This method relies on sonication to fragment the genome prior to immunoprecipitation and the method is thus called sDRIP-seq (sonication DNA:RNA immunoprecipitation coupled with high throughput sequencing) (**Figure 1**). The use of sonication permits an increased resolution and limits restriction enzyme-linked fragmentation biases observed in DRIP-seq approaches[46].

sDRIP-seq produces R-loop maps that are in strong agreement with the results from both DRIP-seq and the previously described high-resolution DRIPc-seq method in which sequencing libraries are built from the RNA strands of immunoprecipitated R-loop structures[45].

Faced with a plethora of methods to choose from, users may wonder which particular DRIP-based approach is preferable for their needs. We offer the following advice. DRIP-seq, despite its limitations, is technically easiest and is the most robust (highest yields) of all three methods discussed here; it thus remains broadly useful. Numerous DRIP-seq datasets have been published, which provide a useful comparison point for new datasets. Finally, the bioinformatic analysis pipeline is simpler as the data is not stranded. It is recommended that new users begin honing their R-loop mapping skills with DRIP followed by quantitative polymerase chain reaction (qPCR) and DRIP-seq. sDRIP-seq represents a slightly higher degree of technical difficulty: the yields are slightly reduced due to sonication (discussed below) and the sequencing library process is slightly more complex. Yet, the gain of strandedness and higher resolution is invaluable. It is noted that sDRIP-seq will capture both two-stranded RNA:DNA hybrids and three-stranded R-loops. Due to the library construction steps, DRIP-seq will not capture two-stranded RNA:DNA hybrids. DRIPc-seq is the most technically demanding and requires higher amount of starting materials. In return, it offers the highest resolution and strandedness. Because sequencing libraries are built from the RNA moiety of R-loops or hybrids, DRIPc-seq may suffer from possible RNA contamination, especially since S9.6 possesses residual affinity for dsRNA[19,47,48]. sDRIP-seq permits strand-specific, high resolution mapping without worries about RNA contamination since sequencing libraries are derived from DNA strands. Overall, these three methods remain useful and present differing degrees of complexity and slightly different caveats. All three, however, produce highly congruent datasets[48] and are highly sensitive to RNase H pre-treatment, which represents an essential control to ensure signal specificity[45,49]. It is noted that given the size selection imposed on sequencing libraries, small hybrids (estimated <75 bp), such as those forming transiently around lagging strand DNA replication priming sites (Okazaki primers) will be excluded. Similarly, since all DRIP methods involve DNA fragmentation, unstable R-loops that require negative DNA supercoiling for their stability will be lost[5]. Thus, DRIP approaches may underestimate R-loop loads, especially for short, unstable R-loops that may be best captured using *in vivo* approaches[45,48]. It is noted that R-loops can also be profiled in an S9.6-independent manner at deep coverage, high-resolution, and in a strand-specific manner on single DNA molecules after sodium bisulfite treatment[12]. Additionally, strategies using a catalytically inactive RNase H1 enzyme have been employed to map native R-loops *in vivo*, highlighting short, unstable R-loops that form primarily at paused promoters[50–52].

PROTOCOL:

The following protocol is optimized for the human Ntera-2 cell line grown in culture, but it has been successfully adapted without modification to a range of other human cell lines (HEK293, K562, HeLa, U2OS), primary cells (fibroblasts, B-cells) as well as in other organisms with small modifications (mice, flies).

1. Cell harvest and lysis

1.1. Culture Ntera-2 cells to 75%–85% confluency. Ensure that the optimal cell count is 5 to 6 million cells with >90% viable counts to start any DRIP procedure.

1.2. Wash the cells once with 1x PBS, add 1.5 mL of Trypsin-EDTA 1x, and then incubate for 2 min at 37 °C until the cells dissociate from the dish.

1.3. Add 5 mL of warm media and pipette well to resuspend cells into a single cell suspension. Transfer the content into a new 15 mL tube and gently pellet the cells at 300 x *g* for 3 min.

1.4. Wash the cells once with 5 mL of 1x PBS and gently pellet the cells at 300 x *g* for 3 min.

1.5. Fully resuspend the cells in 1.6 mL of TE buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA pH 8.0). Add 5 µL of proteinase K (20 mg/mL stock solution) and 50 µL of SDS (20% stock solution) and gently invert the tubes five times until the solution become viscous. Do not try to pipette the solution, only mix by inversion.

1.6. Incubate the tubes overnight at 37 °C.

2. DNA extraction

2.1. Pour the DNA lysate into a pre-spun 15 mL high density phase lock gel tube and add 1 volume (1.6 mL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Gently invert five times and spin down at 1,500 x *g* for 5 min.

2.2. Add 1/10 volume of 3 M sodium acetate (NaOAc) (pH 5.2) and 2.5 volumes of 100% Ethanol to a new 15 mL tube. Pour in the top aqueous phase from the phase lock gel tube and gently invert until the DNA is fully precipitated (up to 10 min).

2.3. Spool the DNA threads using a wide bore 1,000 µL tip and transfer to a clean 2 mL tube while taking care not to carry over the residual supernatant.

2.4. Wash the DNA by adding 1.5 mL of 80% ethanol and gently invert the tube five times. Incubate for 10 min.

2.5. Repeat the previous step twice. Do not centrifuge during the wash steps. Carefully remove as much ethanol as possible by pipetting after the last wash while trying not to disturb the DNA.

2.6. Allow the DNA to air dry completely while inverting the tube. This step can take 30 min –1 h depending on the amount of DNA.

2.7. Add 125 µL of TE buffer directly on the DNA pellet to fragment the DNA through restriction enzyme digestion or 100 µL of TE buffer to shear the DNA through sonication. Keep on ice for 1 h and gently resuspend the DNA by pipetting a few times with a wide bore 200 µL tip. Leave on ice for 1 h before starting the fragmentation step.

3. DNA fragmentation

NOTE: For restriction enzyme-based DRIP-seq, follow step 3.1. For sonication-based DRIP-seq, skip to step 3.2.

3.1. Restriction enzyme (RE) fragmentation

3.1.1. Digest the resuspended genomic DNA (very viscous) using a cocktail of REs according to supplier's instructions.

3.1.1.1. Add 0.1 mM spermidine to the final reaction. Use a cocktail of 4–5 enzymes with 30 U of each enzyme in a total volume of 150 µL.

NOTE: The initial cocktail for DRIP-seq (HindIII, SspI, EcoRI, BsrGI, XbaI)[4] was developed to generate an average fragment length of 5 kilobases. Avoid any interference with CpG methylation and spare GC-rich regions of the genome. Other cocktails are also possible[16]). These cocktails are suitable for both the human and mouse genomes but can be adjusted as needed.

3.1.1.2. Incubate the reaction mixture overnight at 37 °C.

NOTE: The DNA mixture post digest should no longer be viscous. Any remaining viscosity at this step is indicative of an incomplete digestion.

3.1.1.3. If observed, add an additional 10 U of each enzyme and incubate for another 2–4 h at 37 °C.

NOTE: Users may not digest the entire pellet in the event they harvested more cells than recommended here.

3.1.2. Gently pipette the overnight digested DNA (150 µL) into a pre-spun 2 mL phase lock gel light tube. Add 100 µL of water and one volume (250 µL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Gently invert five times and spin down at 16,000 x *g* for 10 min.

3.1.3. Add 1.5 µL of glycogen, 1/10 volume of 3 M NaOAc (pH 5.2) and 2.5 volumes of 100% Ethanol to a new 1.5 mL tube. Pipette the DNA from the phase lock gel tube and mix by inverting five times. Incubate for 1 h at -20 °C.

3.1.4. Spin at 16,000 x *g* for 35 min at 4 °C. Wash the DNA with 200 µL of 80% ethanol and spin at 16,000 x *g* for 10 min at 4 °C.

3.1.5. Air dry the pellet and add 50 µL of TE buffer to the pellet. Leave the tube on ice for 30 min and gently resuspend the DNA.

3.1.6. Measure the concentration (OD_{260}) of the fragmented DNA using a spectrophotometer.

3.1.7. Optional but recommended: Load 1 µg of digested DNA on a 0.8% agarose gel alongside a size marker to verify that the digestion is complete. Run the gel for an hour at 100 V.

NOTE: If incomplete, additional enzyme can be added. Incomplete digestion can lead to the loss of resolution after immunoprecipitation.

3.1.8. After this step, treat 10 µg of digested DNA with 4 µL of ribonuclease H (RNase H) for 1–2 h at 37 °C to ensure that the signal retrieved upon immunoprecipitation is derived from DNA:RNA hybrids. Then, proceed to S9.6 immunoprecipitation (step 4).

NOTE: The digested DNAs can be kept frozen at -80 °C for up to one month without significant loss of yield.

3.2. Sonication

3.2.1. Sonicate all or a part of the extracted DNA in a 0.5 mL microcentrifuge tube in 100 µL total volume. Perform 15–20 cycles of 30 s ON / 30 s OFF on a sonicator (spin after 5, 10, and 15 cycles to ensure homogeneous sonication).

3.2.2. Measure the concentration (OD_{260}) of sonicated DNA on a spectrophotometer.

NOTE: At this step, the viscosity of the DNA should have disappeared.

3.2.3. Run an agarose gel to confirm the size distribution of the sonicated DNA (300–500 bp).

NOTE: Over-sonicating the DNA can lead to significant reduction in yield resulting from breakage and dissociation of R-loop structures.

3.2.4. After this step, treat 10 µg of sonicated DNA with 4 µL of RNase H for 1–2 h at 37 °C to ensure that the signal retrieved upon immunoprecipitation is derived from DNA:RNA hybrids. Then, proceed to S9.6 immunoprecipitation (step 4).

4. S9.6 immunoprecipitation

NOTE: The immunoprecipitation steps are similar regardless of whether DNA was fragmented through REs or sonication.

4.1. Prepare three tubes and aliquot 4.4 µg of fragmented DNA in a final volume of 500 µL of TE buffer per tube. Save 50 µL (1/10 of the volume) from each tube to use later as an input DNA.

4.2. Add 50 µL of 10x binding buffer (100 mM NaPO4 pH 7, 1.4 M NaCl, 0.5% Triton X-100) and 10 µL of S9.6 antibody (1 mg/mL) to the 450 µL of the diluted DNA.

4.3. Incubate overnight at 4 °C on a mini-tube rotator at 7–10 rpm.

4.4. For each tube, wash 50 µL of Protein A/G agarose bead slurry with 700 µL of 1x binding buffer by inverting the tubes on a mini-rotator at 7–10 rpm at room temperature for 10 min. Spin down the beads at 1,100 x *g* for 1 min and discard the supernatant. Repeat this step once.

4.5. Add the DNA from step 4.3 to the 50 µL of beads and incubate for 2 h at 4 °C while inverting at 7–10 rpm on a mini-rotator.

4.6. Spin down the beads for 1 min at 1,100 x *g* and discard the supernatant.

4.7. Wash the beads with 750 µL of 1x binding buffer by inverting at 7–10 rpm on a mini-rotator for 15 min. Spin down for 1 min at 1,100 x *g* and discard the supernatant. Repeat this step once.

4.8. Add 250 µL of the elution buffer (50 mM Tris-Cl pH 8, 10 mM EDTA pH 8, 0.5% SDS) and 7 µL of proteinase K (20 mg/mL stock) to the beads and incubate with rotation at 55 °C (12 rpm) for 45 min.

4.9. Spin down the beads for 1 min at 1,100 x *g*. Transfer the supernatant to a pre-spun 2 mL phase lock gel light tube and add one volume (250 µL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Invert the tubes five times and spin down for 10 min at 16,000 x *g* at room temperature.

4.10. Add 1.5 µL of glycogen, 1/10 volume 3M NaOAc (pH 5.2) and 2.5 volumes of 100% Ethanol to a new 1.5 mL tube. Pipette the DNA from the phase lock gel tube and mix by inverting five times. Incubate for 1 h at -20 °C.

4.11. Spin at 16,000 x *g* for 35 min at 4 °C. Wash the DNA with 200 µL of 80% ethanol and spin at 16,000 x *g* for 10 min at 4 °C.

4.12. Air dry the pellets and add 15 µL of 10 mM Tris-Cl (pH 8) in each tube. Leave the tubes on ice for 20 min and gently resuspend. Combine the contents of the three tubes into one tube (45 µL).

4.13. Check the DRIP efficiency by qPCR using 5 µL of the 45 µL resuspended DNA (see **Representative Results**). Dilute the 5 µL in 10 µL of water and use 2 µL per reaction.

5. Pre-library step for sonicated DNA only

NOTE: Sonication leads the displaced ssDNA strand of R-loops to break. Thus, three-stranded R-loop structures are converted into two-stranded DNA:RNA hybrids upon sonication. As a result, these DNA:RNA hybrids must be converted back to double-stranded DNA prior to library construction. Here, a second strand synthesis step is employed. An alternative approach that has been successfully used is to instead perform a single-stranded DNA ligation followed by a second strand synthesis[53].

5.1. To the 40 µL of DRIP'ed DNA from step 4.12, add 20 µL of 5x second strand buffer (200 mM Tris pH 7, 22 mM $MgCl_2$, 425 mM KCl), 10 mM dNTP mix (dATP, dCTP, dGTP, and dTTT or

dUTP if the user is planning to achieve strand-specific DRIP sequencing), 1 µL 16 mM NAD, and 32 µL water. Mix well and incubate for 5 min on ice.

5.2. Add 1 µL of DNA polymerase I (10 units), 0.3 µL of RNase H (1.6 units) and 0.5 µL of *E. coli* DNA ligase. Mix and incubate at 16 °C for 30 min.

5.3. Immediately clean up the reaction using paramagnetic beads with a ratio of 1.6x. Elute the DNA in 40 µL of 10 mM Tris-Cl (pH 8).

6. Pre-library sonication step for RE DNA only

NOTE: DRIP leads to the recovery of RE fragments that are often kilobases in length and thus not suited for immediate library construction.

6.1. To reduce the size of the material for library construction, sonicate the immunoprecipitated DNA in a 0.5 mL microcentrifuge tube. Perform 12 cycles of 15 s ON / 60 s OFF on a sonicator (spin after six cycles to ensure homogeneous sonication). Proceed to step 7.

NOTE: The immunoprecipitated material still carries the three-stranded R-loops which respond to sonication differently than the flanking double-stranded DNA.

6.2. Optional step: To even out DRIP profiles, treat the immunoprecipitated material with 1 µL of RNase H in 1x RNase H buffer for 1 h at 37 °C prior to sonication.

7. Library construction

7.1. Perform end repair by adding to the 40 µL from step 4.12 (RE fragmentation) or step 5.3 (sonication shearing) 5 µL of 10x end repair module buffer, 2.5 µL of 10 mM ATP and 2.5 µL of End repair module enzyme (50 µL total). Mix well and incubate for 30 min at room temperature. Include 1 µg of RE-digested and sonicated (DRIP) or sonicated (sDRIP) input DNA to create control sequencing libraries corresponding to the input DNA.

7.2. Clean up the reaction using paramagnetic beads (1.6x ratio) and elute in 34 µL of 10 mM Tris-Cl (pH 8).

7.3. Perform A-tailing by adding 5 µL of buffer 2, 10 µL of 1 mM dATP, and 1 µL of Klenow exo- (50 µL total). Mix well and incubate the mixture for 30 min at 37 °C.

7.4. Clean up the reaction using paramagnetic beads (1.6x ratio) and elute in 12 µL of 10 mM Tris-Cl (pH 8).

7.5. Ligate adapters by adding 15 µL of 2x quick ligation buffer, 1 µL of 15 µM adapters, and 2 µL of quick ligase (30 µL total). Mix well and incubate for 20 min at room temperature.

7.6. Clean up the reaction using paramagnetic beads (1x ratio) and elute in 20 µL of 10 mM Tris-Cl (pH 8).

7.7. If sonication shearing was performed and dUTP was used in step 5.1, add 1.5 µL (1.5 U) of Uracil N-glycosylase and incubate for 30 min at 37 °C to obtain a strand-specific DRIP.

7.8. PCR amplify 10 µL of the library from step 6.6 or 6.7. Add 1 µL of PCR primer 1.0 P5 (see **Table of Materials**), 1 µL of PCR primer 2.0 P7 (see **Table of Materials**), 15 µL of master mix, and 3 µL of water. Mix well.

7.9. In a thermo cycler, run the program as shown in **Table 1**.

7.10. Proceed to a two-step clean-up of the library using paramagnetic beads. First use a ratio of 0.65x to remove fragments over 500 bp. Keep the supernatant. Proceed to a 1x ratio on the supernatant to remove fragments under 200 bp. Elute in 12 µL of 10 mM Tris-HCl (pH 8).

8. Quality control

8.1. To check R-loop enrichments with qPCR on two negative and three positive loci using the Pfaffl method, use 1 µL of the clean-up library from step 6.10. Dilute 1 µL of the library in 10 µL of water and use 2 µL per locus.

8.2. Check the size distribution of the cleaned-up library from step 6.10 using a high sensitivity DNA kit.

REPRESENTATIVE RESULTS:

DRIP as well as sDRIP can be analyzed through qPCR (**Figure 2A**) and/or sequencing (**Figure 2B**). After the immunoprecipitation step, the quality of the experiment must be first confirmed by qPCR on positive and negative control loci, as well as with RNase H-treated controls. Primers corresponding to frequently used loci in multiple human cell lines are provided in **Table 2**. The results from qPCR should be displayed as a percentage of input, which corresponds to the percentage of cells carrying an R-loop at the time of the lysis for a given locus. In a successful DRIP experiment, the yield for negative loci should be less than 0.1% whereas positive loci can vary from 1% to over 10% for highly transcribed loci such as *RPL13A* (**Figure 2A**). For sDRIP, yields are typically lower (20%–50%) as judged by DRIP-qPCR but appear to affect recovery uniformly such that no particular subset of R-loops is affected more than another. As a result, maps derived from DRIP, sDRIP, and DRIPc are in good agreement (**Figure 2B**). qPCR data can also be displayed as fold enrichment of the percentage of input for positive loci over negative loci, thus assessing the specificity of the experiment. Fold enrichments typically range from a minimal of 10-fold to over 200-fold depending on the loci chosen for analysis. When precise quantification across multiple samples representing gene knockdowns, knockouts, or various pharmacological treatments, is required, the use of spike in controls to normalize inter-sample experimental variation is highly encouraged. Such spike-ins can correspond to synthetic hybrids[53] or genomes of unrelated species[54].

DRIP and sDRIP materials can be sequenced using single or paired-end sequencing strategies. Data can be extracted and analyzed similarly to most ChIP data using standard computational pipelines (see[45] for DRIP-relevant information). After adapter trimming and removal of PCR

duplicates, reads can be mapped to a reference genome and uploaded to a genome browser. A typical expected output of DRIP and sDRIP is shown in **Figure 2B**. The DRIP output is represented by the only green track as it does not allow strand specificity whereas sDRIP shows R-loop mapping to the positive and negative strands indicated respectively in red and blue. Control tracks corresponding to a sample pre-treated with RNase H show a clear reduction of signals, confirming the specificity of the technique for RNA:DNA hybrid-derived materials. The gains in resolution permitted by sDRIP are clearly illustrated when comparing the sizes of input DNA material (**Figure 2C**). The reproducibility of sDRIP-seq, along with the global impact of RNase H1 pre-treatment and the correlation between sDRIP-seq and DRIPc-seq are depicted by XY plots in **Figure 2D**.

FIGURE AND TABLE LEGENDS:

Figure 1: Overview of the DRIP-seq and sDRIP-seq procedures. Both approaches start by the same DNA extraction steps developed to preserve R-loops (RNA strands within R-loops are represented by squiggly lines). For DRIP-seq, the genome is fragmented using restriction enzymes, often resulting in kilobase-size fragments within which shorter R-loops are embedded. For sDRIP-seq, the genome is fragmented via sonication, which results in smaller fragments and the shearing and loss of the displaced single-strand of R-loops (indicated by dashed lines). Following immunoprecipitation with the S9.6 antibody, DRIP leads to the recovery of three-stranded R-loops embedded within restriction fragments, while sDRIP recovers two-stranded RNA:DNA hybrids with little flanking DNA, ensuring higher resolution. For sDRIP, a library construction step must be included to convert RNA:DNA hybrids back to duplex DNA. As shown here, this is an opportunity to build strand-specific libraries. As detailed in the protocol itself, exogenous treatment with RNase H represents a key control for the specificity of both procedures; they are not shown here.

Figure 2: Result of R-loop mapping strategies. (**A**) qPCR results from successful immunoprecipitations using the DRIP and sDRIP method (corresponding to qPCR check step 4.13). Results are from two independent experiments from human Ntera-2 cells at a negative locus and two positive loci, including the highly R-loop-prone *RPL13A* locus and the moderately R-loop-prone *TFPT* locus. The y-axis indicates the yield of the immunoprecipitation as a percentage of the input DNA. Note that the recovery is slightly more robust for DRIP than sDRIP. (**B**) The results of R-loop mapping conducted in human Ntera-2 cells are shown over a region centered around the *CCND1* and neighboring *ORAOV1* genes. The first two tracks correspond to DRIP-seq results, without and with RNase H treatment, respectively. The position of the restriction enzymes used to fragment the genome are shown at the top. The next six tracks represent the results of strand-specific sDRIP-seq, broken down between (+) and (-) strands (two replicates each) and pre-treated with RNase H, or not, as indicated. The last four tracks represent the results of R-loop mapping via the high-resolution strand-specific DRIPc-seq method (Sanz et al., 2016; Sanz and Chedin, 2019), where libraries are built from the RNA strands of R-loops. As can be clearly seen, the *CCND1* and *ORAOV1* genes lead to R-loop formation on the (+) and (-) strands, respectively, consistent with their directionality. RNase H treatment abolishes the signal, as expected. (**C**) Input DNA materials after restriction enzyme fragmentation (left) and sonication (right) are shown after the materials were separated by agarose gel electrophoresis. The DNA ladder corresponds to a 100 bp ladder and the 500 bp band is highlighted by an asterisk. (**D**) XY signal correlation plots are shown to illustrate the reproducibility of sDRIP-seq (left), the overall sensitivity of sDRIP-seq to RNase

H1 pre-treatment (middle), and the global correlation between sDRIP-seq and DRIPc-seq (right). All data are from Ntera-2 human cells.

Table 1: PCR program settings. The duration and temperature settings for the PCR cycles are listed.

Table 2: Primers used for qPCR validation in human cell lines. All sequences are listed in the 5' to 3' direction.

DISCUSSION:

Described here are two protocols to map R-loop structures in potentially any organism using the S9.6 antibody. DRIP-seq represents the first genome-wide R-loop mapping technique developed. It is an easy, robust, and reproducible technique that allows one to map the distribution of R-loops along any genome. The second technique, termed sDRIP-seq, is also robust and reproducible but achieves higher resolution and strand-specificity owing to the inclusion of a sonication step and a stranded sequencing library construction protocol. Both techniques are highly sensitive to RNase H treatment prior to immunoprecipitation, confirming that the signal is principally derived from genuine RNA:DNA hybrids. Finally, when comparing immunoprecipitation yields between R-loop positive and R-loop negative loci, both techniques offer up to a 100-fold difference in several human cell lines, providing high specificity mapping with low background.

When considering which method to implement, it is useful to consider their respective strengths and limitations. As previously noted, DRIP-seq produces maps with a lower resolution and does not give information on the strandedness of R-loop formation. The lower resolution is mainly a product of the use of REs to fragment the genome. This gentle method is best at preserving R-loops, thereby allowing unsurpassed recovery of such structures, and making DRIP-seq very robust. To circumvent the issue of limited resolution while preserving high recovery, RE cocktails can be adapted and/or maps resulting from different RE cocktails can be combined to improve resolution[16]. A technique using 4 bp cutters has been developed to improve the resolution of DRIP-seq and may achieve strand-specific mapping[22,55], although the resulting datasets have not yet been systematically compared to other human datasets. It is important to note that in RE-based approaches, larger fragments tend to be recovered more efficiently because they can carry multiple R-loop forming regions. This bias must be taken into account when analyzing DRIP-seq datasets. Similarly, peak calling for DRIP-seq data must be ultimately translated into R-loop-positive RE fragments, since it is these fragments that are immunoprecipitated and the position of R-loops within these fragments can't be inferred. In general, it is recommended that users first adopt RE-based DRIP-seq to learn the method and build their confidence in achieving the yields documented in **Figure 2A**. sDRIP-seq typically results in lower yields, which could result in maps with lower signal-to-noise ratios in untrained hands. The use of sonication as a means of fragmenting the genome offers in return a great improvement in resolution since the non-R-looped portions that typically constitute the majority of RE fragments will be broken off, allowing S9.6 to principally retrieve the R-looped portions (**Figure 1**). It is worth noting that sonication causes the displaced ssDNA strand of R-loops to break. It is therefore essential to add a second strand synthesis after immunoprecipitating sonicated DNA:RNA hybrids, which will convert these hybrids back to dsDNA, prior to building sequencing libraries. Without this step, the only fragments that can

be ligated to dsDNA adapters will be background dsDNA fragments; thus, the resulting maps will be devoid of any signal. Strand-specificity provides numerous further benefits to the understanding of R-loop formation mechanisms, making sDRIP-seq a method of choice for the study of R-loops.

Importantly, maps obtained via DRIP-seq and sDRIP-seq represent the average distribution of R-loops through a cell population; thus, the length and position of individual R-loops cannot be addressed with those techniques. For this, an independent and complementary method termed single-molecule R-loop footprinting (SMRF-seq)[12] can be leveraged to reveal individual R-loops at high-resolution in a strand-specific manner. Assessment of R-loop formation using SMRF-seq over 20 different loci, including independently of S9.6, revealed a strong agreement between collection of individual R-loop footprints and the population average distribution gather by DRIP-based approaches[12], lending strong support to DRIP-based approaches. It is also important to consider that R-loop mapping data only provides a snapshot of R-loop genomic distribution and does not provide information on the dynamics of R-loop formation, stability, and resolution. DRIP approaches, combined with specific drug treatments and an evaluation of R-loop distributions through time series, can nonetheless be deployed to address these parameters[17,53]. The limitations of R-loop profiling methodologies are particularly important to keep in mind when the goal is to characterize altered R-loop distributions in response to genetic, environmental, or pharmacological perturbations. In addition to those already described above, it is key to consider any possible change to nascent transcription since these will inherently cause R-loop changes owing to the co-transcriptional nature of these structures. These issues and guidelines for developing rigorous R-loop mapping approaches have been extensively discussed[48,56] and readers are encouraged to refer to these studies.

ACKNOWLEDGMENTS:

Work in the Chedin lab is supported by a grant from the National Institutes of Health (R01 GM120607).

DISCLOSURE:

The authors declare no conflicts of interest.

REFERENCES:

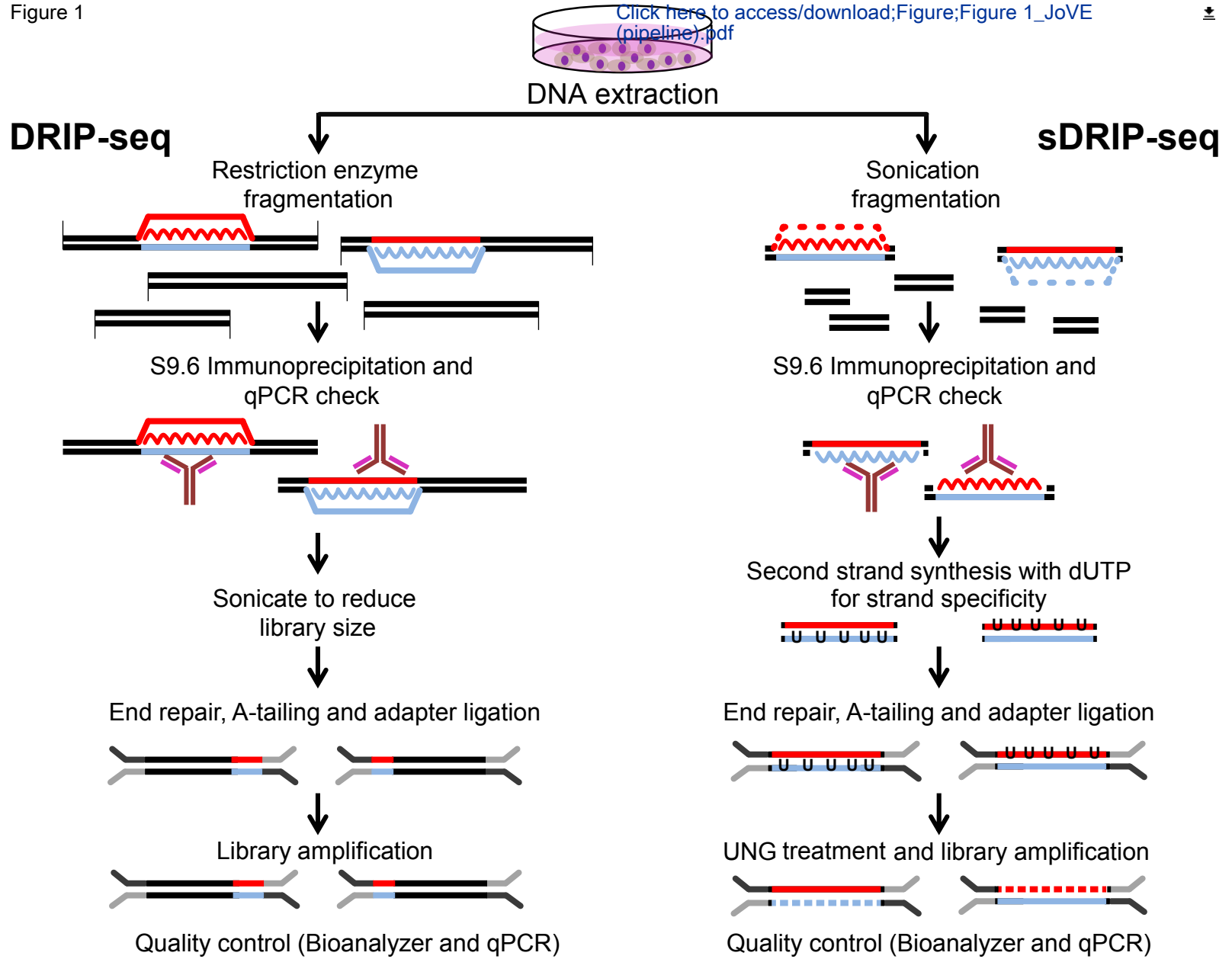
1. Reaban, M. E., Lebowitz, J., Griffin, J. A. Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. *The Journal of Biological Chemistry*. **269** (34), 21850–21857 (1994).
2. Daniels, G. A., Lieber, M. R. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Research*. **23** (24), 5006–5011 (1995).
3. 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. *Nature Immunology*. **4** (5), 442–451 (2003).
4. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell*. **45** (6), 814–825 (2012).

5. Stolz, R. et al. Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proceedings of the National Academy of Sciences of the United States of America*. **116** (13), 6260–6269 (2019).
6. 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 & Development*. **18** (13), 1618–1629 (2004).
7. Carrasco-Salas, Y. et al. The extruded non-template strand determines the architecture of R-loops. *Nucleic Acids Research*. **47** (13), 6783–6795 (2019).
8. Huppert, J. L. Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. *Molecular Biosystems*. **4** (6), 686–691 (2008).
9. Hartono, S. R., Korf, I. F., Chedin, F. GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Research*. **43** (20), 9729–9741 (2015).
10. Green, P., Ewing, B., Miller, W., Thomas, P. J., Green, E. D. Transcription-associated mutational asymmetry in mammalian evolution. *Nature Genetics*. **33** (4), 514–517 (2003).
11. Polak, P., Arndt, P. F. Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Research*. **18** (8), 1216–1223 (2008).
12. Malig, M., Hartono, S. R., Giafaglione, J. M., Sanz, L. A., Chedin, F. Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. *Journal of Moleclar Biology*. **432** (7), 2271–2288 (2020).
13. Masse, E., Phoenix, P., Drolet, M. DNA topoisomerases regulate R-loop formation during transcription of the rrnB operon in Escherichia coli. *The Journal of Biological Chemistry*. **272** (19), 12816–12823 (1997).
14. Drolet, M. et al. The problem of hypernegative supercoiling and R-loop formation in transcription. *Frontiers in Bioscience: A Journal and Virtual Library*. **8**, d210–d221 (2003).
15. Chedin, F., Benham, C. J. Emerging roles for R-loop structures in the management of topological stress. *The Journal of Biological Chemistry*. **295** (14), 4684–4695 (2020).
16. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., Chedin, F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research*. **23** (10), 1590–1600 (2013).
17. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-Loop structures associate with specific epigenomic signatures in mammals. *Molecular Cell*. **63** (1), 167–178 (2016).
18. El Hage, A., Webb, S., Kerr, A., Tollervey, D. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genetics*. **10** (10), e1004716 (2014).
19. Hartono, S. R. et al. The affinity of the S9.6 Antibody for Double-Stranded RNAs impacts the accurate mapping of R-loops in fission yeast. *Journal of Molecular Biology*. **430** (3), 272–284 (2018).
20. Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & Development*. **30** (11), 1327–1338 (2016).
21. Alecki, C. et al. RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. *Nature Communications*. **11** (1), 1781 (2020).
22. Xu, W. et al. The R-loop is a common chromatin feature of the Arabidopsis genome. *Nature Plants*. **3** (9), 704–714 (2017).
23. Chedin F. Nascent connections: R-Loops and chromatin patterning. *Trends in Genetics: TIG*. **32** (12), 828–838 (2016).

609 24. Skourti-Stathaki, K., Proudfoot, N. J., Gromak, N. Human senataxin resolves RNA/DNA
610 hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination.
611 *Molecular Cell*. **42** (6), 794–805 (2011).
612 25. Kreuzer, K. N., Brister, J. R. Initiation of bacteriophage T4 DNA replication and
613 replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its
614 relatives. *Virology Journal*. **7**, 358 (2010).
615 26. Carles-Kinch, K., Kreuzer, K. N. RNA-DNA hybrid formation at a bacteriophage T4
616 replication origin. *Journal of Molecular Biology*. **266** (5), 915–926 (1997).
617 27. Masukata, H., Tomizawa, J. A mechanism of formation of a persistent hybrid between
618 elongating RNA and template DNA. *Cell*. **62** (2), 331–338 (1990).
619 28. Itoh, T., Tomizawa, J. Formation of an RNA primer for initiation of replication of ColE1
620 DNA by ribonuclease H. *Proceedings of the National Academy of Sciences of the United States*
621 *of America*. **77** (5), 2450–2454 (1980).
622 29. Stuckey, R., Garcia-Rodriguez, N., Aguilera, A., Wellinger, R. E. Role for RNA:DNA
623 hybrids in origin-independent replication priming in a eukaryotic system. *Proceedings of the*
624 *National Academy of Sciences of the United States of America*. **112** (18), 5779–5784 (2015).
625 30. Lee, D. Y., Clayton, D. A. Initiation of mitochondrial DNA replication by transcription
626 and R-loop processing. *The Journal of Biological Chemistry*. **273** (46), 30614–30621 (1998).
627 31. Xu, B., Clayton, D. A. A persistent RNA-DNA hybrid is formed during transcription at a
628 phylogenetically conserved mitochondrial DNA sequence. *Molecular and Cellular Biology*. **15**
629 (1), 580–589 (1995).
630 32. Cadoret, J. C. et al. Genome-wide studies highlight indirect links between human
631 replication origins and gene regulation. *Proceedings of the National Academy of Sciences of*
632 *the United States of America*. **105** (41), 15837–15842 (2008).
633 33. Sequeira-Mendes, J. et al. Transcription initiation activity sets replication origin
634 efficiency in mammalian cells. *PLoS Genetics*. **5** (4), e1000446 (2009).
635 34. Picard, F. et al. The spatiotemporal program of DNA replication is associated with
636 specific combinations of chromatin marks in human cells. *PLoS Genetics*. **10** (5), e1004282
637 (2014).
638 35. Mukhopadhyay, R. et al. Allele-specific genome-wide profiling in human primary
639 erythroblasts reveal replication program organization. *PLoS Genetics*. **10** (5), e1004319 (2014).
640 36. Huang, F. T., Yu, K., Hsieh, C. L., Lieber, M. R. Downstream boundary of chromosomal
641 R-loops at murine switch regions: implications for the mechanism of class switch
642 recombination. *Proceedings of the National Academy of Sciences of the United States of*
643 *America*. **103** (13), 5030–5035 (2006).
644 37. Huang, F. T. et al. Sequence dependence of chromosomal R-loops at the
645 immunoglobulin heavy-chain Smu class switch region. *Molecular and Cellular Biology*. **27** (16),
646 5921–5932 (2007).
647 38. Yu, K., Lieber, M. R. Current insights into the mechanism of mammalian
648 immunoglobulin class switch recombination. *Critical Reviews in Biochemistry and Molecular*
649 *Biology*. **54** (4), 333–351 (2019).
650 39. Crossley, M. P., Bocek, M., Cimprich, K. A. R-Loops as cellular regulators and genomic
651 threats. Molecular *Cell*. **73** (3), 398–411 (2019).
652 40. Santos-Pereira, J. M., Aguilera, A. R loops: new modulators of genome dynamics and
653 function. *Nature Reviews. Genetics*. **16** (10), 583–597 (2015).
654 41. Garcia-Muse, T., Aguilera, A. R Loops: From physiological to pathological roles. *Cell*. **179**
655 (3), 604–618 (2019).

42. Skourti-Stathaki, K., Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & Development*. **28** (13), 1384–1396 (2014).
43. Costantino, L., Koshland, D. The Yin and Yang of R-loop biology. *Current Opinion in Cell Biology*. **34**, 39–45 (2015).
44. Boguslawski, S. J. et al. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of Immunological Methods*. **89** (1), 123–130 (1986).
45. Sanz, L. A., Chedin, F. High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nature Protocols*. **14** (6), 1734–1755 (2019).
46. Halasz, L. et al. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Research*. **27** (6), 1063–1073 (2017).
47. Phillips, D. D. et al. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *Journal of Molecular Recognition*. **26** (8), 376–381 (2013).
48. Chedin, F., Hartono, S. R., Sanz, L. A., Vanoosthuyse, V. Best practices for the visualization, mapping, and manipulation of R-loops. *The EMBO Journal*. **40** (4), e106394 (2021).
49. Smolka, J. A., Sanz, L. A., Hartono, S. R., Chedin, F. Recognition of RNAs by the S9.6 antibody creates pervasive artefacts when imaging RNA:DNA hybrids. *Journal of Cell Biology*. **220** (6), e202004079 (2021).
50. Chen, J. Y., Zhang, X., Fu, X. D., Chen, L. R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. *Nature Protocols*. **14** (5), 1661–1685 (2019).
51. Yan, Q., Sarma, K. MapR: A Method for Identifying native R-loops genome wide. *Current Protocols in Molecular Biology*. **130** (1), e113 (2020).
52. Wang, K. et al. Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Science Advances*. **7** (8), eabe3516 (2021).
53. Crossley, M. P., Bocek, M. J., Hamperl, S., Swigut, T., Cimprich, K. A. qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide. *Nucleic Acids Research*. **48** (14), e84 (2020).
54. Svikovic, S. et al. R-loop formation during S phase is restricted by PrimPol-mediated repriming. *The EMBO Journal*. **38** (3), e99793 (2019).
55. Yang, X. et al. m(6)A promotes R-loop formation to facilitate transcription termination. *Cell Research*. **29** (12), 1035–1038 (2019).
56. Vanoosthuyse, V. Strengths and weaknesses of the current strategies to map and characterize R-Loops. *Non-coding RNA*. **4** (2), 9 (2018).

Figure 1



DRIP-seq		sDRIP-seq
Low	**Resolution**	High
No	**Strand-specificity**	Yes
Highest	**Robustness**	High

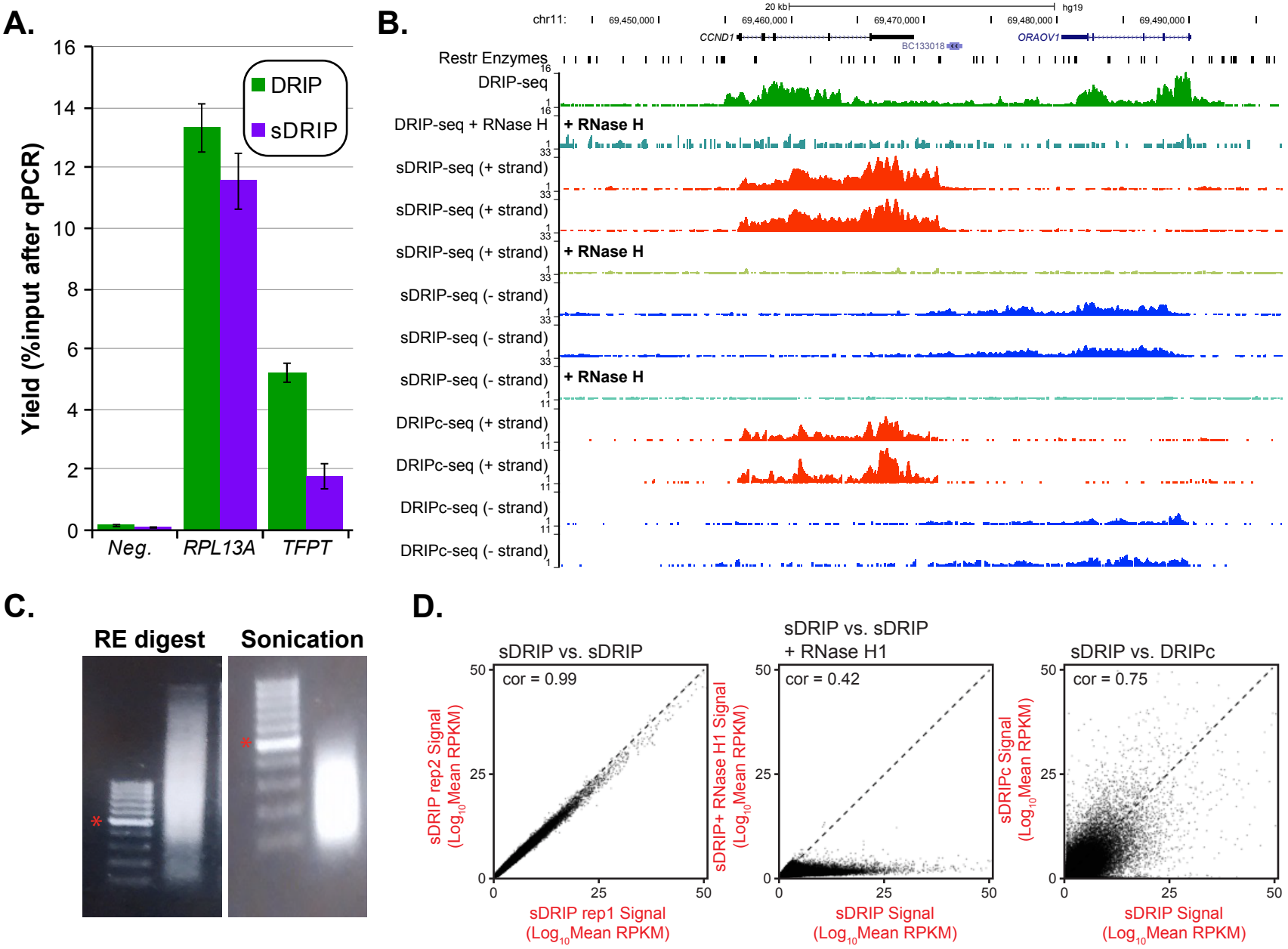


Table 1

Cycle number	Duration	Temperature
1	30 s	98 °C
2–15	10 s	98 °C
	30 s	60 °C
	30 s	72 °C
16	5 min	72 °C
17	Hold	12 °C

Table 2

Primers used for qPCR validation in human cell lines. All sequences are listed in the 5' to 3' direction.

Species	Type	Name	Sequence
human	positive control locus	RPL13A F	AGGTGCCTTGCTCACAGAGT
human	positive control locus	RPL13A R	GGTTGCATTGCCCTCATTAC
human	positive control locus	TFPT F	TCTGGGAGTCCAAGCAGACT
human	positive control locus	TFPT R	AAGGAGCCACTGAAGGGTTT
human	negative control locus	EGR1neg F	GAACGTTCAGCCTCGTTCTC
human	negative control locus	EGR1neg R	GGAAGGTGGAAGGAAACACA
human	negative control locus	SNRPNneg F	GCCAAATGAGTGAGGATGGT
human	negative control locus	SNRPNneg R	TCCTCTCTGCCTGACTCCAT

Name of Material/ Equipment	Company	Catalog Number
15 mL tube High density Maxtract phase lock ge	Qiagen	129065
2 mL tube phase lock gel light	VWR	10847-800
Agarose A/G beads	ThermoFisher Scientific	20421
Agencourt AMPure XP beads	Beckman Coulter	A63881
AmpErase Uracil N-glycosylase	ThermoFisher Scientific	N8080096
Index adapters	Illumina	
Klenow fragment (3’ to 5’ exo-)	New England BioLabs	M0212S
NEBNext End repair module	New England BioLabs	E6050
PCR primers for library amplification		
PCR primers for library amplification		
Phenol/Chloroform Isoamyl alcohol 25:24:1	Affymetrix	75831-400ML
Phusion Flash High-Fidelity PCR master mix	ThermoFisher Scientific	F548S
Quick Ligation Kit	New England BioLabs	M2200S
Ribonuclease H	New England BioLabs	M0297S
S9.6 Antibody	Kerafast	ENH001
S9.6 Antibody	Millipore/Sigma	MABE1095
S9.6 Antibody	Abcam	ab234957

Comments/Description

Corresponds to the TrueSeq Single indexes

primer 1.0 P5 (5’ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA 3’)
PCR primer 2.0 P7 (5’ CAAGCAGAAGACGGCATACGAGAT 3’)

These three sources are equivalent

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

This paper describes two protocols to map RNA-DNA hybrid structures using the S9.6 antibody. Beyond the classic DRIP-seq, a variant method is also described (dubbed sDRIP-seq) that relies on sonication rather than restriction enzyme digestion to fragment the genome prior to S9.6 immunoprecipitation. However, sonication is routinely applied in DRIP protocols to fragment the genome, so the sDRIP variant is by no means new in this regard. The positive effect of ultrasound treatment on the resolution of DRIP-seq is thoroughly discussed in e.g. https://pubmed.ncbi.nlm.nih.gov/28341774/, which should be cited in the manuscript. In addition, differences between published DRIP protocols relying on sonication and their sDRIP-seq protocol should be clearly highlighted and elaborated in more detail.

The reviewer is correct that previous approaches have used sonication. We've added a reference to the paper mentioned above to justify the utility of sonication. We note however, that quite a few sonication-based datasets published so far appear rather discordant from the majority of other DRIP-based datasets (see PMID: 33411340). This raises questions about these prior protocols and lessens the utility of including a detailed side-by-side comparison of protocols. We note that the sDRIP-seq protocol described here is very close to the qDRIP-seq protocol published by the Cimprich lab (Crossley et al., 2020), which we cite in this work. sDRIP-seq and qDRIP-seq generate datasets that are highly congruent with each other and also in strong agreement with many other DRIP-based datasets (see PMID: 33411340).

Concerns:

1. Is there an optimal RE cocktail for different species (having different genomes in term of size, CG content, repetitive elements, DNA methylation levels)? Could the authors suggest suitable RE cocktails for the most common model organisms?

We mention (step 3.1.1) that the "classic" RE cocktail (BsRGI, EcoRI, HindIII, SspI, XbaI) is suitable for human and mouse cells, generating an average fragment length of 5 kb, in a manner insensitive to DNA methylation. We also refer to Ginno et al., (2013), where a second cocktail was developed to improve resolution. Beyond this we do not have direct experience with RE cocktails for other organisms. We advise users that these cocktails can be adjusted.

2. R-loop formation is prevalent in most organisms studied so far. How does the presence of R-loops affect the efficiency of RE fragmentation? (Most restriction enzymes do not cut R-loops.) Sonication seems to be more unbiased in this respect.

The reviewer is correct that R-loops are often not cut by restriction enzymes, which may lower the resolution for those instances where an R-loop covers an RE site. We agree that this is yet one more argument in favor of sonication.

3. In the classical DRIP protocol, RE fragmentation is followed by sonication to reduced fragment size. This step is needed to make the sample compatible with NGS library preparation but it will not improve the low resolution of the method. This should be clearly articulated in the manuscript.

We agree and believe that Step 6 clearly states that this sonication step is used "To reduce the size of the material for library construction". This is also clearly mentioned in Figure 1.

4. Please show the fragment size distribution of RE-fragmented and sonicated nucleic acid samples related to the DRIP protocols.

We added this information to revised Figure 2C. Thanks for the suggestion.

5. It is argued that "Sonication leads the displaced ssDNA strand of R-loops to break. Thus, three-stranded R-loop structures are converted into two-stranded DNA:RNA hybrids upon sonication." Is there any experimental proof or reference justifying this statement?

The experimental “proof” is that we systematically failed to build proper sequencing libraries until we incorporated a second strand synthesis step, which converts RNA:DNA hybrids to dsDNA. In the absence of this step, all enrichments were lost because the only fragments that could be ligated to our dsDNA adapters were background dsDNA fragments. Once hybrids were converted back to dsDNA, we were able to pick them up and the libraries displayed the expected enrichments for R-loop-positive loci versus negative controls.

Similar observations were derived from the Cimprich lab using qDRIP-seq. The qDRIP protocol uses identical conditions for virtually every step (which makes sense as it was adapted from our sDRIP protocol), except that at the end, libraries are built using single-strand ligation step following denaturation of IP-ed material. If sonication followed by IP resulted in three-stranded R-loops, then both DNA strands would have been captured by the ligation step and the resulting maps would not be stranded. However, qDRIP maps are clearly strand-specific. This strongly argues that the displaced strand of R-loops does not survive sonication and that two-stranded hybrids are immunoprecipitated.

We are not sure if these detailed considerations are useful to include in a step by step protocol and look forward to guidance on this point.

Reviewer #2:

Manuscript Summary:

In this manuscript, Sanz. L. et al describe a novel technique (sDRIP) designed to achieve highly sensitive and strand-specific mapping of R loops and compare it to DRIP-seq, the most used technique to study R loops currently. First, current knowledge on R loops filed is nicely introduced and then DRIP and sDRIP-seq techniques are described at high technical level. Importantly, specific details on DRIP-seq technique not previously reported (such as the optional RNH treatment of samples prior to library construction) are clarified which further increase the relevance of the protocols described. Finally, they discuss the benefits of using each technique and also compare them to other reported methodologies to map R loops genome-wide. Overall, in my opinion, both techniques are thoroughly described in the text and should be of knowledge to any scientist planning to map such structures genome-wide. However, I have several concerns/suggestions that they should address prior to publication (see below, please).

We thank the reviewer for his/her general positive comments.

Major Concerns:

1) The authors describe sDRIP-seq as a highly sensitive and stranded procedure to map R loops and compare it to DRIP-seq, a widely used technique to map R loops. In the last years, DRIPc-seq has also emerged as a technique allowing high sensitivity and stranded sequencing

of R loops (Sanz, L. et al. Mol Cell 2016). However, only little mention to DRIPc-seq methodology is present in the manuscript. Actually, there is only one reference to say DRIPc-seq may suffer from RNA contamination (lines 109-112). But, what are the efficiencies in R loop mapping in DRIPc-seq and sDRIP-seq? How much of the signal is RNH-sensitive and correlate with DRIP-seq in each case? Given that both techniques offer stranded mapping of R loops, the authors should offer convincing data to demonstrate that sDRIP-seq permits higher sensitivity and reproducibility in R loop mapping than DRIPc-seq. They should offer comparative results between both techniques to justify why sDRIP-seq is better than DRIPc-seq.

We thank the reviewer for raising this point and have revised the manuscript to provide better context as users consider which method to use. We added a new paragraph at the end of the Introduction that presents users with a guide on choosing the most appropriate R-loop mapping method, including a discussion of their caveats, limitations, and degree of technical difficulty.

In our minds, DRIPc-seq remains superior in terms of resolution and sensitivity but it is technically more challenging and requires significantly more starting material. Both sDRIP-seq and DRIPc-seq are highly reproducible and in strong agreement which each other overall (See Figure 2). This is now illustrated by adding a correlation plot and a screenshot illustrating mapping results. sDRIP-seq does not necessarily “permit higher sensitivity … in R-loop mapping” but the fact that it is technically much less demanding and that it is insensitive to any RNA contamination issues are significant advantages.

2) Related to the first point. The authors claim that DRIPc-seq suffers from RNA contamination due to the residual affinity of S9.6 for dsRNA, but certainly part of such "contamination" may be present as well in sDRIP-seq. As described recently (Chedin, F. et al. EMBO J 2021), most of such signal may come from transposable elements such as retrotransposable Alu elements. As far as I understood from the description of sDRIP-seq technique, this type of contamination will be present as well in sDRIP-seq as retrotransposons may exist as DNA:RNA hybrids while retrotranposing and such structures will also be immunoprecipitated by S9.6 antibody and amplified according to the protocol. In fact, RNH will digest RNA strand from retrotransposable DNA:RNA hybrids and DNA polymerase I perform second strand synthesis in step from the protocol. Again, comparison between both techniques is missing.

To clarify, the issue of Alu-delineated signal in DRIPc-seq described in our EMBO J review doesn’t correspond to actively transposing retroelements being captured through an RNA:cDNA replication intermediate. These Alu-delineated signals instead result from contamination from unspliced gene transcripts strictly at the RNA level. This conclusion is supported by three main observations. One, these Alu-delineated signals map to transcribed genes only. If they came from independently transposing Alu elements, a significant portion of them would be expected to map outside of genes. Two, the strandedness of these Alu-delineated signals is always that of the gene in which these Alu repeats lie. If these Alu elements were engaged in independent transposition events, we would expect the strandedness to vary according to the orientation of each Alu in the genome (Alu possess their own RNA Pol III promoters). Three, if these signals were derived from RNA:DNA hybrids, they would be sensitive to RNase H pre-treatment. While no RNase H pre-treated dataset could be found for the Nojima et al, (2018) study, the Perez-Calero et al., (2020) study did include such a control and the signals are clearly RNase H resistant (please follow the link provided in our EMBO J review to access the datasets). Finally, we note that the Alu-delineated signals are most evident in datasets that already also show clear signs of contamination by spliced transcripts, creating exon-delimited signals (see EMBO J review).

Regardless of these considerations, the reviewer is correct that if there was true retroelement transposition, characterized by the formation of RNA:cDNA hybrids, sDRIP would indeed capture this (and the signals would be RNase H sensitive).

3) According to the described protocol, sDRIP-seq technique will allow sequencing on genomic places where displaced single strand of DNA will be preferentially broken thru sonication while the other strand will be much more protected from breaking thru hybridization with the RNA. But what happens with similar structures found in the genome that might also be IPed with S9.6 antibody. Will Okazaki fragments be sequenced also? What is the real extent of RNH-sensitive sDRIP-seq signal? This may complicate interpretation of the results.

Currently, given the size selection imposed on sequencing libraries, small hybrids (estimated < 75 bp), such as those found transiently at Okazaki primers will be excluded. We added a note for this in the revised version so that it is clearer to those users interested in addressing short hybrids. We thank the reviewer for this comment.

sDRIP signal is very sensitive to RNase H pre-treatment, as expected. We now cite a new publication in press in which this is illustrated.

4) What is used as input in sDRIP-seq? Simply sonicated gDNA or second strand synthesis on sonicated gDNA skipping S9.6 IP? Are inputs also sequenced in sDRIP-seq? Is peak calling performed against the input in this case also? This information is missing.

Input DNA is indeed sonicated DNA, which we recommend to sequence and to use in peak calling. We thank the reviewer for pointing out this omission and revised the manuscript accordingly.

Minor Concerns:
1) In Figure 2, authors show an example of a genomic screenshot from DRIP-seq and sDRIP-seq data. According to the text, sDRIP-seq is much more specific as sequencing will concentrate on DNA:RNA hybrid itself rather than the restriction fragment carrying the hybrid. Then, why sDRIP-seq profile at gene CCND1 is wider than DRIP-seq? It should be the other way around, no?

As it often happens, the *CCND1* gene and its neighbor are covered by R-loops almost through their entire gene bodies so the higher resolution is hard to detect going from DRIP to sDRIP (although detectable at the start of these genes if you look closely). We chose this region because it nicely illustrates the strandedness of the method, the agreement with DRIPc-seq, and the overall great RNase H sensitivity.

2) It would be nice to show which strands are DNA and which ones are RNA in Figure 1. I think it would increase understanding of the described methodology, especially at second strand synthesis step in sDRIP-seq technique.

We added in the Figure legend that “RNA strands within R-loops are represented by squiggly lines”.

Reviewer #3:
Manuscript Summary:
In this manuscript, Sanz et al describe DRIP-seq methods for genome-wide mapping of R-loops

using the anti RNA/DNA hybrid S9.6 antibody. They provide protocols for the classical DRIP-seq and an iteration of this technique, which they called sDRIP-seq (for sonication DRIP-seq). sDRIP-seq relies on sonication to fragment the genome prior to immunoprecipitation with the S9.6 antibody while DRIP-seq uses restriction enzymes. They report that the two main advantages of sDRIP-seq over DRIP-seq are (i) higher resolution of the R-loop mapping, and (ii) the possibility to map R-loops in a strand-specific manner without building the library from the RNA of RNA/DNA hybrids as in the case of DRIPc-seq. The protocols are very well detailed and comprehensive, and suitable examples of DRIP-qPCR and DRIP-seq profiles are provided. Given the growing interest in R-loop structures, these protocols will be very useful for the scientific community.

We thank the reviewer for his/her positive comments.

Major Concerns:
None

Minor Concerns:
It would be useful that the authors mention and discuss in the manuscript the S9.6-independent techniques to map R-loops.

We added references to various methods including bisulfite-based and RNase H1-based.

Reviewer #4:
Manuscript Summary:
In "Mapping R-loops and RNA:DNA hybrids with S9.6-based immunoprecipitation methods", the authors describe DRIP-Sequencing and sDRIP-Sequencing, two techniques which they developed for mapping R-loop locations across the genome. In their introduction, the authors describe the recent research demonstrating the characteristics of R-loops and their important roles in pathology and physiology. These findings indicate the importance of R-loop mapping experiments which seek to uncover R-loop locations across the genome. Overall, the introduction was well written, easy to follow and covered current understanding of the varieties of R-loop biologies.
The written protocol presented in the manuscript is easy to follow and highly detailed. The representative results clearly demonstrate the expected outcomes from DRIP-Seq and sDRIP-Seq experiments.
Aside from the minor critiques noted below, this reviewer finds this to be an excellent resource and an important contribution to the R-loop field.

We thank the reviewer for his/her positive comments.

Major Concerns:
One shortfall of the manuscript, though beyond its scope and focus, is the lack of attention is paid to the bioinformatics analysis, following sequencing. As the authors have already pointed out in a recent publication, the analysis step is a major contributor to the quality of the resulting R-loop map (https://www.embopress.org/doi/abs/10.15252/embj.2020106394). This reviewer would, therefore, expect a note on appropriate bioinformatics approaches used in order to analyze these data with an indication of where to find such protocols, particularly when considering any differences in data types (eg. strandedness) between sDRIP-Seq and DRIP-Seq.

We agree with the reviewer that the bioinformatic treatment of data is of strong importance. We are in the midst of preparing a MS dedicated to this topic, offering a second set of “best practices” and a uniform pipeline for this purpose. The format of this JoVE manuscript, however, doesn’t lend itself to such an extensive discussion and we note than many genomic-focused JoVE articles do not include any bioinformatics information. To nonetheless help guide users, we added a reference to our previous Nature Protocols manuscript in which the bioinformatics treatment of both unstranded and stranded data was specified.

An additional concern is a lack of information on the extent of reduction in R-loops caused by sonication (about 40-50%?). That may be important depending on the question being asked. The authors mention that sDRIP typically results in lower yield, so it would be valuable to have a comparison between DRIP and sDRIP and whether sonication affects any one class/subset of R-loops (eg. TSS vs TTS Rloops) and the authors should provide a measure for the amount of correlation between the two methods.

We added a comment that while yields appear lower in sDRIP compared to DRIP as judged after qPCR, the reduction does not appear to affect any particular subset of R-loops as best we can tell. Thus, maps derived from all three DRIP-based approaches discussed here are in good agreement. We thank the reviewer for bringing this point.

To further show the correlation between methods, we added three XY plots to revised Figure 2D that display both self-correlation (two independent sDRIP replicates), the impact of RNase H1 pre-treatment on sDRIP, and a correlation between sDRIP-seq and DRIPc-seq. We thank the reviewer for the suggestion.

It may also be fair to the reader if the authors also cautions on when to prefer RE based methods over sonication (say if probing for Rloops in a "normal" cell line that has fewer Rloops than a cancer cell line thus making retrieving sufficient material to see strong signal over background more of a concern).

We thank the reviewer for this comment and added a paragraph in the Introduction that addresses this topic, which we hope will be useful to users.

Minor Concerns:
Minor typographical error:
Line 130: 20 mg/ml should be mL
This was corrected.