HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2019 November 21.

Published in final edited form as:

Cell Rep. 2019 October 29; 29(5): 1369–1380.e5. doi:10.1016/j.celrep.2019.09.052.

Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach

Qingqing Yan[1,4,5], **Emily J. Shields**[2,3,4,5], **Roberto Bonasio**[2,4,*], **Kavitha Sarma**[1,4,6,*]

[1]Gene Expression and Regulation Program, The Wistar Institute, Philadelphia, PA 19104, USA

[2]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

[3]Graduate Group in Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Epigenetics Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

[5]These authors contributed equally

[6]Lead Contact

SUMMARY

R-loops are three-stranded DNA:RNA hybrids that are implicated in many nuclear processes. While R-loops may have physiological roles, the formation of stable, aberrant R-loops has been observed in neurological disorders and cancers. Current methods to assess their genome-wide distribution rely on affinity purification, which is plagued by large input requirements, high noise, and poor sensitivity for dynamic R-loops. Here, we present MapR, a method that utilizes RNase H to guide micrococcal nuclease to R-loops, which are subsequently cleaved, released, and identified by sequencing. MapR detects R-loops formed at promoters and active enhancers that are likely to form transient R-loops due to the low transcriptional output of these regulatory elements and the short-lived nature of enhancer RNAs. MapR is as specific as existing techniques and more sensitive, allowing for genomewide coverage with low input material in a fraction of the time.

In Brief

Yan et al. report a fast, easy, antibodyindependent strategy, MapR, to identify native R-loops *in vivo* without the need for generating stable cell lines. MapR uses the natural affinity and specificity of RNase H to detect R-loops. MapR identifies dynamic R-loops formed at enhancers with high sensitivity.

*Correspondence: roberto@bonasiolab.org (R.B.), kavitha@sarmalab.com (K.S.).

AUTHOR CONTRIBUTIONS

Q.Y. developed MapR and performed all the experiments. E.J.S. performed all bioinformatic analyses. Q.Y., E.J.S., R.B., and K.S. wrote the paper.

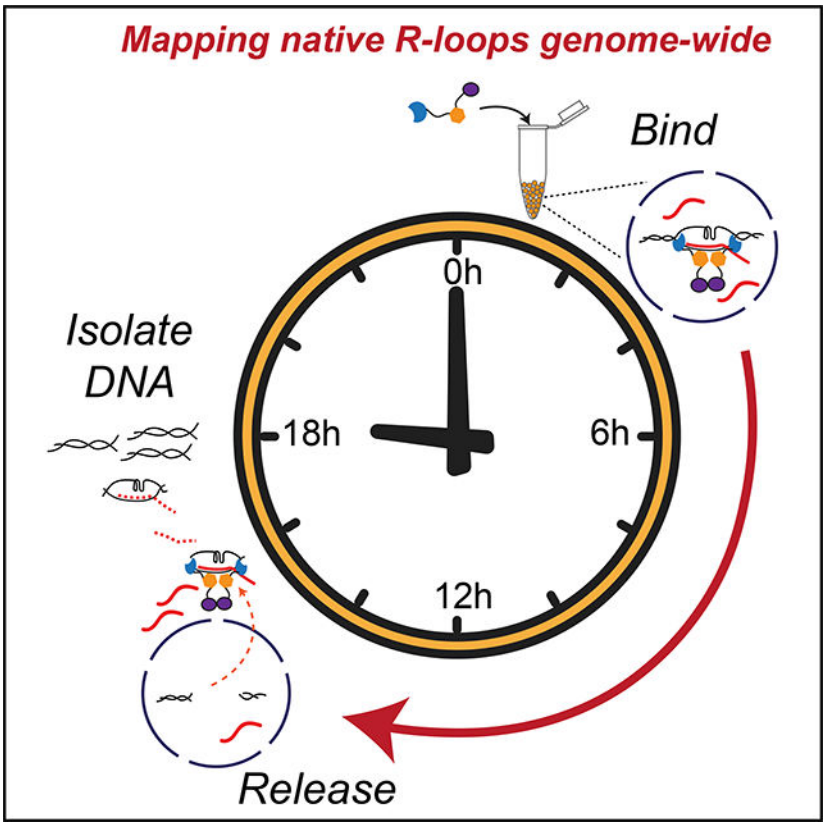
DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.09.052.

Graphical Abstract



INTRODUCTION

R-loops are three-stranded nucleic acid structures that contain a DNA:RNA hybrid and a displaced single strand of DNA (Thomas et al., 1976). R-loops are dynamic structures whose levels are tightly controlled across the genome (Chédin, 2016; Santos-Pereira and Aguilera, 2015; Skourti-Stathaki and Proudfoot, 2014). Alterations in nuclear R-loop levels are associated with disruption of transcription, DNA repair, and other key genomic processes (Cristini et al., 2018; Ribeiro de Almeida et al., 2018; Skourti-Stathaki et al., 2011; Song et al., 2017; Yasuhara et al., 2018). Identification of changes in R-loop abundance and distribution in different cell types could suggest mechanisms that lead to cell-type-specific pathology (Groh and Gromak, 2014; Hatchi et al., 2015; Perego et al., 2019; Richard and Manley, 2017; Sollier and Cimprich, 2015; Wang et al., 2015). However, efforts to study the regulatory functions of R-loops have been hindered because of the sub-optimal methods used to enrich for and recover these chromatin structures. Therefore, there is a critical need to develop new methods that will allow for enhanced and systematic discovery of R-loops.

Currently, two distinct strategies are used to map the distribution of R-loops. The predominant strategy relies on the immunoprecipitation of chromatin containing R-loops by using a monoclonal antibody, S9.6, specific for DNA:RNA hybrids (Boguslawski et al., 1986). DNA:RNA immunoprecipitation (DRIP) and all its variants (Dumelie and Jaffrey,

2017; Ginno et al., 2012; Nadel et al., 2015; Wahba et al., 2016) (bisulfite-DNA-RNA immunoprecipitation [bis-DRIP], S1 nuclease DRIP [S1-DRIP], and RNA:DNA immunoprecipitation [RDIP]) were foundational to the study of genome-wide R-loop localization but share similar disadvantages: (1) they prepare chromatin for immunoprecipitation by using harsh physical and biochemical treatments (high temperatures, strong detergents, sonication, and/or prolonged enzymatic digestion of chromatin) in the absence of fixation, which might disrupt less stable R-loops before they can be detected; and (2) they rely on the S9.6 antibody, whose strict specificity for DNA:RNA hybrids remains a subject of debate (e.g., it might also bind double-stranded RNA [dsRNA]) (Hartono et al., 2018). The second, more recent, strategy to map R-loops takes advantage of the natural affinity of RNase H for DNA:RNA hybrids. RNase H is an enzyme that degrades the RNA strand of DNA:RNA heteroduplexes. Two published methods, DNA:RNA *in vitro* enrichment (DRIVE) (Ginno et al., 2012) and R-loop chromatin immunoprecipitation (R-ChIP) (Chen et al., 2017), target R-loops by using a catalytic-deficient version of RNase H (RHΔ) that retains its affinity for DNA:RNA hybrids but does not cleave the RNA strand. In both cases the DNA:RNA hybrids bound by RHΔ are enriched by affinity purification. In DRIVE, RHΔ is fused to the maltosebinding protein (MBP), incubated with sheared chromatin *in vitro*, and bound R-loops are recovered by affinity purification on amylose resin (Ginno et al., 2012). In R-ChIP, V5-RHΔ is expressed *in vivo*, and the R-loops are recovered by immunoprecipitation using the V5 affinity tag (Chen et al., 2017). Although both DRIVE and R-ChIP take advantage of the exquisite specificity of RNase H for targeting DNA:RNA species, they still suffer from the limitations typical of affinity purifications, including high backgrounds, requirement for large amounts of starting material, and time-consuming protocols.

DESIGN

To overcome these limitations, we have developed an R-loop mapping strategy, termed “MapR.” MapR combines the specificity of RNase H for DNA:RNA hybrids with the sensitivity, speed, and convenience of the cleavage under targets and release using nuclease (CUT&RUN) approach (Skene et al., 2018; Skene and Henikoff, 2017), whereby targeted genomic regions are released from the nucleus by micrococcal nuclease (MNase) and sequenced directly, without the need for affinity purification. Specifically, in MapR, cells are immobilized (Figure 1, step 1) and permeabilized, and a fusion protein comprising a catalytically inactive RNase H and MNase (RHΔ-MNase) is allowed to diffuse into the nuclei in the absence of calcium ions, thus keeping the MNase enzyme inactive (Figure 1, step 2). After equilibration (Figure 1, step 3, top), calcium is added, and the nuclei are incubated for 30 min at 0°C before stopping the reaction with EGTA. This results in the release of chromatin fragments targeted by RHΔ and, therefore, that contain R-loops in their native state (Figure 1, step 4). As a control for MapR, we perform the same experiment using MNase lacking the RHΔ moiety (Figure 1 step 3, bottom). Finally, released chromatin fragments are purified and sequenced (Figure 1, step 5).

RESULTS

Generation of Native R-Loop Maps with CUT&RUN

As a first step toward developing MapR, we sought to determine whether a conventional, antibody-mediated CUT&RUN approach in a cell line expressing tagged RHΔ could release R-loop-containing fragments. For this, we expressed a FLAG-tagged version of RHΔ containing a nuclear localization signal (Figure S1A) in HEK293 cells and performed a standard CUT&RUN assay using an anti-FLAG antibody to reveal the chromatin distribution of RHΔ and, therefore, R-loops. (Figure 2A, left). FLAG CUT&RUN for transgenic RHΔ (RHΔC&R) identified 28,353 peaks compared to an IgG control. These presumptive R-loops mapped to the promoters of 12,653 genes, of which 5,842 overlapped with R-loop-containing genes, as identified by immunoprecipitation of RHΔ from crosslinked chromatin in the R-ChIP approach (Figures 2B and 2C). This overlap is highly significant ($p < 10^{-15}$, hypergeometric distribution), indicating that CUT&RUN correctly recovers a large portion of previously identified R-loops. The majority of nuclear R-loops are known to occur co-transcriptionally (Ginno et al., 2012; Sanz et al., 2016). In agreement with this and consistent with R-ChIP, the majority (79%) of peaks identified by RHΔC&R occurred in genic regions, with 41% localized at promoters and 38% within the gene body (Figure 2D). We conclude that R-loops can be targeted *in vivo* by RHΔ and their distribution can be revealed using a CUT&RUN approach.

MapR, a Recombinant-Protein-Based Technology Identifies Similar R-Loops as CUT&RUN

Although the above strategy successfully retrieved native R-loops without affinity purification steps, it still required genetic manipulation of the cells to express a FLAG-tagged version of RHΔ. This presents obvious limitations when studying R-loops in cells that are difficult to transfect or sub-clone, such as patient-derived primary cells that do not divide *in vitro*. To overcome these limitations, we reasoned that the FLAG antibody step could be bypassed by fusing RHΔ directly to MNase and providing this recombinant protein exogenously after cell immobilization and permeabilization. Toward this end, we expressed and purified glutathione S-transferase (GST)-RHΔ-MNase (henceforth RHΔ-MNase) from *Escherichia coli* (Figure S1B). As a control, we used GST-MNase in our experiments to assess for non-specific cleavage across the genome (Figure 1; Figure S1B). To determine the substrate specificity of GST-RHΔ-MNase, we performed an electrophoretic mobility shift assay (EMSA) by using synthetic double-stranded DNA (dsDNA) or DNA:RNA hybrid substrates. GST-RHΔ-MNase specifically bound DNA:RNA hybrids (Figure S1C), whereas GST-MNase did not bind either substrate. Furthermore, to ascertain that the presence of the RHΔ moiety did not affect the enzymatic activity of MNase, we digested chromatin with equimolar amounts of MNase and RHΔ-MNase. We found that the two fusion proteins had comparable enzymatic activity because they produced similar patterns of nucleosomal ladders after 10 and 30 min (Figure S1D).

Next, we immobilized and permeabilized HEK293 cells and incubated them with either MNase (control) or RHΔ-MNase (MapR) (Figure 1). We activated the MNase moiety in both recombinant proteins by the addition of calcium at the same time and for the same duration. As in CUT&RUN, we constructed libraries from cleaved DNA fragments that diffused out

of the nucleus and sequenced them. Genome-wide profiles obtained by MapR (i.e., with the exogenous RHΔ-MNase fusion protein) were very similar to those obtained by expressing RHΔ *in vivo* and performing a conventional FLAG CUT&RUN (Figure 2E), whereas no discernible signal was obtained with MNase alone. MapR-enriched regions were predominantly genic (77%), with 41% of peaks mapping to promoters and 36% to gene bodies, consistent with the idea that this technology effectively identifies R-loops *in vivo* (Figure 2D).

Co-transcriptional R-loops are known to occur at the 5′ end of active genes immediately downstream of the promoter and, to a lesser extent, at the 3′ end of active genes (Ginno et al., 2013; Skourti-Stathaki et al., 2011, 2014). As an example, we inspected the *XIST* long non-coding RNA (lncRNA) gene. HEK293 are female cells and, therefore, one of the two X chromosomes is subject to X chromosome inactivation, a process that is dependent on the expression of the XIST lncRNA (da Rocha and Heard, 2017; Jégu et al., 2017). Both MapR and RHΔC&R signals are clearly higher than the respective controls at the 5′ end of the *XIST* gene (Figure 2E). *XIST* also contains an antisense gene, *TSIX* (Chao et al., 2002), that is expressed only in early development and is silent in HEK293. In contrast to the *XIST* locus, the *TSIX* gene showed no detectable signal from either MapR or RHΔC&R (Figure 2E).

MapR using RHΔ-MNase identified 14,769 peaks compared to an MNase-only control (Figure 2D). These peaks mapped to the promoters of 6,201 genes, of which 5,713 overlapped significantly ($p < 10^{-15}$) with promoter R-loop-containing genes, as identified by RHΔC&R (Figure 2F). Despite the ~7,000 genes where R-loop peaks were called by the peak-calling algorithm only in RHΔC&R, read densities from MapR and RHΔC&R over all the peaks were highly correlated (Figure 2G; Spearman, r = 0.76), demonstrating that the two approaches detected broadly comparable genomic regions as being occupied by R-loops. We analyzed the strength of MapR and RHΔC&R signals (Figure 2H) at all transcription start sites (TSS) and found that enriched regions from both datasets tracked closely with actively transcribed genes, as determined by global run-on sequencing (GRO-seq), by the presence of the activating chromatin mark, histone H3 lysine 27 acetylation (H3K27ac) (Frietze et al., 2012), and by the corresponding depletion of the repressive chromatin mark H3K27 trimethylation (H3K27me3) (Gao et al., 2012). No distinguishable MapR or RHΔC&R signal was observed at the TSS of inactive genes (Figure 2I). Consistent with a predominant localization of R-loops near and upstream of TSSs, metagene analyses for both MapR and RHΔC&R revealed an accumulation of signal starting 2 kb upstream and peaking at the TSS (Figure 2I). Metagene analyses for MapR and RHΔC&R at the 3′ end of genes corresponding to the transcription end site (TES) showed a lack of signal compared to the TSS, consistent with previous work (Chen et al., 2017). Thus, we conclude that genomic regions enriched by our MapR approach are specifically found at active genes and are broadly consistent with previously reported profiles for R-loops (Chen et al., 2017). Importantly, these analyses show that MapR, a technique that bypasses the need for transgenic cells, identifies the same genomic regions as FLAG CUT&RUN performed on RHΔ-expressing cells.

MapR Identifies Bona Fide R-Loops

Having demonstrated that regions identified by MapR have genomic features consistent with R-loops (i.e., they localized to the 5′ end of active genes), we next wished to determine if they also displayed known biochemical properties of R-loops. Because the majority of cellular R-loops are a consequence of active transcription, we reasoned that a general transcription inhibitor should cause decreased MapR signal (Figure 3A). Consistent with this, treating cells with actinomycin D, an inhibitor of transcription elongation, caused a decrease in MapR signal at specific genes (*GBAP1* and *IPP*, Figure 3B) and genome-wide (Figures 3C and 3G). We confirmed these findings in a second cell type, U87T cells (Figures S2A and S2B).

Bona fide R-loops are defined by the presence of a DNA:RNA heteroduplex, whose recognition by RNase H is the foundation for MapR. We expect that pre-treating immobilized and permeabilized cells with an enzymatically active RNase H would result in degradation of the RNA strand, restoration of dsDNA, and loss of MapR signal (Figure 3D). On the other hand, if our RHΔ-MNase fusion protein bound non-specifically to chromatin regions devoid of R-loops, these interactions should not be affected by pre-incubation with active RNase H. Indeed, MapR signal at the 5′end of the *RWDD1* and *ANP32E* genes was considerably reduced by pre-treatment with active RNase H (Figure 3E), an observation that held true throughout the genome (Figures 3F and 3G), demonstrating that most if not all peaks detected by MapR contained DNA:RNA heteroduplexes. A similar RNase H-dependent reduction in signal intensities was observed in MapR experiments performed in U87T cell lines (Figures S2C and S2D). The reduction in signal, as opposed to loss, is likely a result of incomplete digestion of R-loops given the short treatment time with RNase H, as required by the fact that MapR is carried out in unfixed cells. These results show that the genomic regions recovered by MapR contain DNA:RNA hybrids that are degraded by RNase H and whose formation is prevented by transcription inhibition. Therefore, MapR detects genomic features with the biochemical properties of R-loops *in vivo*.

MapR relies on the release of DNA fragments containing R-loops by MNase cleavage, an enzyme that digests regions of open chromatin. To confirm that all accessible chromatin regions were not non-specifically targeted by the MNase moiety in our fusion protein, we compared MapR signals with regions that show DNase I hypersensitivity, a feature of open chromatin. Although our results show a correlation between MapR and DNase I signals (Figure S2E), not all open chromatin regions contain R-loops (Figures 3H and 3I). Base-level overlap between MapR and DNase I datasets indicate that not all DNase I hypersensitive sites overlap with MapR peaks (Figure 3J), with only 10% of DNase I peaks containing an R-loop. Similarly, a comparison of R-loop signals detected by other techniques and DNase I shows that most DNase I peaks do not contain an R-loop. Thus, MapR specifically detects R-loops, not all regions of open chromatin.

Comparison of MapR with Existing Antibody and RNase H-Based Techniques

Next, we asked how MapR compared to existing R-loop detection strategies. We selected for comparison the two methods representing the two strategies outlined above: DRIP for methods that use the S9.6 antibody to purify DNA:RNA hybrids and R-ChIP for methods

that use RNase H. Importantly, datasets obtained with these techniques in the same cell type (HEK293) were publicly available (Chen et al., 2017; Manzo et al., 2018). We found that the MapR signal broadly resembled that of DRIP and R-ChIP (Figure 4A). We also compared MapR data with RDIP, a variation of DRIP where input material is digested with RNase I to remove free RNAs (i.e., not DNA:RNA hybrids) prior to S9.6-based enrichment (Nadel et al., 2015). Promoters that contained an R-loop according to MapR overlapped significantly ($p < 10^{-15}$) with genes identified by R-ChIP (Figure 4B). While there is a clear correlation between MapR and R-ChIP datasets (Figure S4A and S4B), the extent of overlap at the promoter level appears lower. This may be attributed to differences in peak numbers between the datasets and overexpression of RHΔ in R-ChIP versus the addition of RHΔ *ex vivo* in MapR. This significant overlap is also observed on comparison of MapR and DRIP datasets from 293 and K562 cells (Figure 4B; Figure S3A). Interestingly, MapR signals overlap to a considerably lower extent with RDIP in 293 cells (Figure S3B). Metagene analysis showed that MapR, RHΔC&R, and R-ChIP showed similar profiles where signals were significantly enriched at TSS (Figure 2I; Figure S3C). In contrast, RDIP and DRIP show a lower and broader enrichment pattern around TSS (Figure S3C). Our results also show that MapR detected thousands of additional genes compared to both previous technologies (Figure 4B), raising the question of whether these newly detected genes contained bona fide R-loops and were previously missed.

To evaluate whether the signals obtained exclusively from MapR experiments represented genuine R-loops, we first confirmed that these regions correspond to sites of active transcription, as ascertained by the presence of a GRO-seq signal (Figures 2H and 4C; Figure S3D). Next, we compared the effect of actinomycin D treatment (see Figures 3A–3C and 3G) on presumptive R-loops detected by MapR and other techniques. Actinomycin D treatment resulted in a similar or increased reduction of MapR signal in MapR genes as well as genes identified by R-ChIP and DRIP (Figure 4D). We observed that the MapR signals from regions identified by RDIP showed the least reduction in actinomycin D treatment. The MapR signal over a control gene set that according to all techniques did not contain R-loops did not show any appreciable change upon actinomycin D addition, confirming treatment specificity. Finally, we analyzed the distribution of sequences predicted to give rise to G-quadruplex structures, which is a common feature of the displaced DNA strand in R-loops (Chen et al., 2017; Duquette et al., 2004; Ginno et al., 2012; Roy et al., 2008; Yu et al., 2003). The frequency of G-quadruplexes in the promoter regions of genes with promoter R-loops by MapR were comparable to those measured in genes called using our RHΔC&R as well as the other R-loop detection strategies, whereas non-R-loop genes have a lower frequency of G-quadruplexes in their promoters (Figure 4E). Consistent with these results, enriched regions identified by MapR and the other R-loop detection strategies contained a higher GC content compared to regions without R-loops (Figure S3E). Thus, MapR identifies bona fide native R-loops.

We further analyzed the published data and observed that R-ChIP peaks had a genomic distribution similar to MapR, with a majority of peaks (79%) mapping to genes and a small number (21%) to intergenic sites. Similarly, 85% of DRIP peaks mapped to genes and 15% mapped to intergenic sites. In comparison, 49% of RDIP peaks occurred at intergenic sites and only 51% within genes (Figure S3F). Of the genic peaks, only 8% mapped to promoter

regions in the RDIP dataset. As an orthogonal method to test concordance between MapR and other R-loop detection technologies, we looked at promoters that had peaks from both MapR and other methods and measured the distance between the peaks. We found that most RHΔC&R and R-ChIP peaks showed minimal separation from the center of MapR peaks (Figure 4F). In contrast, a smaller number of DRIP and RDIP peaks resided in close proximity to MapR peaks and many were distributed over 2 kb away from the closest MapR peak. This is consistent with the similarity in the location of MapR, RHΔC&R, and R-ChIP signal accumulation at promoters and the relative dissimilarity between the location of MapR and RDIP/DRIP (Figure 2I; Figure S3C and S3F).

To thoroughly investigate similarities and differences between R-loop detection methods, we tested the extent of correlation between MapR, RHΔC&R, R-ChIP, DRIP, and RDIP datasets, for which data are available in 293 cells. Relying on peak calling to compute similarity between techniques has limitations, including the necessity of setting arbitrary thresholds and relying on peak-finding software to detect true peaks from noisy data. To avoid this bias, we compared read densities obtained with these four methods at all promoters. Our two CUT&RUN-based methods, MapR and RHΔC&R, showed a high degree of correlation (Spearman, r = 0.82; Figure S4A), consistent with the similarity of their signals at peaks genome-wide as shown above (Figure 2G). The signal from R-ChIP, which, similar to MapR, also uses a modified RNase H to detect R-loops, correlated better to MapR (Spearman, r = 0.59) than the signal from S9.6-antibody-based RDIP and DRIP (Spearman, r = 0.44 and r = 0.50, respectively; Figures S4A and S4B). The dissimilarity between the signal detected by RDIP and the other techniques at promoters may be attributed to the tendency of RDIP signals to accumulate at STRs, in contrast to the RNase-H-based methods (Figure 5B; Figure S4C). This difference between methods that use RNase H and the S9.6 antibody has been noted before (Crossley et al., 2019; Vanoosthuyse, 2018).

We next examined why peak calling identified R-loops in certain genes with MapR but not R-ChIP and vice versa. We analyzed read densities at the promoters of genes containing an R-loop peak according to MapR alone (“MapR only”), R-ChIP alone (“R-ChIP only”), or both techniques (“common”; Figure S4C). We detected strong R-ChIP signal at R-ChIP only and common genes, consistent with the fact that R-loop peaks had been called in those regions (Figure S4C, left and middle panels). However, we also found robust, if reduced, R-ChIP signal at promoters that were identified only by MapR as containing an R-loop (Figure S4C, right panel). As expected of bona fide R-loops, R-ChIP performed with an RNase H mutant unable to bind RNA-DNA hybrids (Chen et al., 2017) did not yield any signal at any of these R-loop containing genes, including those that were not identified by peak-calling in the R-ChIP dataset. This observation further supports our conclusion that signal from MapR and R-ChIP correlate strongly and suggests that the incomplete overlap of genes identified by the two techniques (e.g., Figure 4B) is mostly due to the vagaries of peak calling. The converse of this analysis also shows evidence of similarity between MapR and R-ChIP; MapR signal is present consistently at R-ChIP-only genes, although reduced compared to MapR and common genes, and is diminished by actinomycin D treatment (Figure S4D), indicating genuine R-loops. In contrast, R-loops identified by RDIP alone had minimal MapR signal that was largely unaffected upon actinomycin D treatment (Figure S4E). Also, in agreement with the significant overlap between MapR and DRIP datasets (Figure 4B),

peaks identified by DRIP alone showed MapR signal that was reduced upon actinomycin D treatment (Figure S4F).

Variable R-Loop Detection at Simple Tandem Repeats

Intriguingly, we found that many sites of exclusive RDIP enrichment overlapped with simple tandem repeats (STRs) (Figure 4G). The established link between R-loops and tandem repeats in the genome and their relevance to several neurodegenerative diseases (Groh and Gromak, 2014; Perego et al., 2019; Richard and Manley, 2017) prompted us to further investigate the enrichment of STRs by RDIP. In HEK293 cells, we found that 64% of RDIP peaks contained STRs, whereas only 7% of MapR peaks contained STRs (Figure 4H). To determine whether this STR enrichment was observed in other technologies, we analyzed RHΔC&R, DRIP, and R-ChIP from HEK293. We also analyzed published RDIP and DRIP datasets from IMR90 and K562 cells, respectively, to exclude experimental- and cell-type-specific bias (Nadel et al., 2015; Sanz et al., 2016). We found that MapR and RHΔC&R, which rely on cleavage and release of nucleic acid followed by direct sequencing (as opposed to the enrichment strategies used in RDIP, DRIP, and R-ChIP) showed lower overlap with STRs (Figure 4H). Interestingly, peaks called by RDIP showed a frequency of STR overlap that correlated with the strength of peak enrichment (Figure S3G). Such a correlation was absent in the MapR, RHΔC&R, R-ChIP, and DRIP datasets. To further validate this observation, we measured the distance between peaks from different methods to the closest STR and found that RDIP peaks from two independent datasets showed minimal distance from STRs compared to all the other methods that showed a median distance of ≈1 kb from the closest STR (Figure 4I). When we analyzed the transcription output at STRs that overlap with R-loops, we found that although MapR, R-ChIP, and DRIP showed clear signal at many STRs, only a small fraction of STRs detected by RDIP showed any GRO-seq signals (Figure 4J).

MapR Identifies R-Loops Formed at Active Enhancers

Next, we asked whether MapR is able to detect dynamic R-loops with higher efficiency than R-ChIP, DRIP, and RDIP. For this, we focused on enhancer elements. Enhancer RNAs (eRNAs) are known to be less abundant than other cellular RNAs and short-lived (Rabani et al., 2014; Schwalb et al., 2016) and are, therefore, likely to form transient R-loops. We used a 10-state chromHMM model (Ernst and Kellis, 2012) to identify genomic regions corresponding to active enhancers (H3K27ac+ and H3K4me1+, state 5) or poised enhancers (H3K27ac− and H3K4me1+, state 4,) in HEK293 cells (Figure S5A). Metaplot analyses of these regions confirmed an enrichment for the relevant chromatin marks (H3K27ac and H3K4me1 at active enhancers, H3K4me1 alone at poised enhancers) (Figure S5B). Consistently, active enhancers contained more GRO-seq signal compared to poised enhancers (Figure S5C), and MapR signal from both active and poised enhancers were reduced upon treatment with either RNase H or actinomycin D, indicating that these regions contain bona fide R-loops (Figure S5D). Furthermore, active enhancers with an R-loop showed significant overlap ($p < 10^{-15}$) between MapR and RHΔC&R datasets (Figure S5E). The intergenic R-loops identified by MapR, RHΔC&R, R-ChIP, DRIP, and RDIP contained H3K27ac and H3K4me1 active enhancer signatures (Figure 5A). Our analyses also showed

that intergenic R-loops that were detected exclusively by MapR and RHΔC&R also correspond to active enhancer regions (Figure 5B).

Having established that MapR can detect R-loops that form at enhancers, we compared its sensitivity at enhancers genome-wide with that of R-ChIP, RDIP, and DRIP. We focused on the intergenic peaks in MapR, RHΔC&R, R-ChIP, RDIP, and DRIP datasets to avoid confusion with co-transcriptional R-loops formed within genes and tested their degree of overlap with enhancers. We found that both MapR and RHΔC&R intergenic peaks showed a higher overlap (32.2% and 53.3%, respectively) with enhancer regions compared to R-ChIP, RDIP, and DRIP (4.99%, 3.04%, and 9.18% respectively) intergenic peaks (Figure 5C). Genomic regions classified as active enhancers by chromHMM (H3K4me+ and H3K27ac+) were much more likely to contain R-loops detected by MapR and RHΔC&R than genomic regions classified as poised enhancers, consistent with the conclusion that these enhancer R-loops are formed by eRNAs (Figure 5D). Our data indicate that MapR and RHΔC&R can identify R-loops at enhancers with higher sensitivity than other techniques.

R-Loops Can Be Detected with Low Input Material

We wished to probe the detection limits of our technology. Our MapR experiments above were performed with five million cells, which is within the range used in most ChIP and DRIP experiments; however, a main advantage of CUT&RUN over immunoprecipitation methods to map chromatin marks (ChIP) is that the lack of an affinity purification step decreases the amount of input material required (Skene et al., 2018; Skene and Henikoff, 2017). Thus, we tested whether MapR could identify R-loops starting from 50-fold fewer cells. The MapR profiles obtained from 10^5 cells closely resembled those obtained with 5 million cells (Figure S5F), with a similar genome-wide enrichment profile at and upstream of active TSSs (Figures S5G and S5H). Thus, we conclude that MapR offers the ability to discover R-loops with high sensitivity and is robust even when cell numbers are limiting. Notably, these improvements on sensitivity and specificity are accompanied by a greatly streamlined experimental protocol that can be completed in 1 day, which is ~4X less than the fastest alternative.

DISCUSSION

MapR is an efficient, convenient, and fast method to generate genome-wide maps of R-loops. MapR uses an antibody-independent strategy that can be used in any cell type without the need to generate stable transgenic lines. MapR also identifies transient R-loops that are formed at active enhancers. Importantly, MapR can identify R-loops in small cell numbers, and this can facilitate its future application to study aberrant R-loops formed in diseases by using patient-derived material.

The S9.6 antibody and RNase H are thought to recognize distinct molecular features of R-loops (Crossley et al., 2019; Vanoosthuyse, 2018). Therefore, methods based on these strategies likely enrich for different R-loops. Our results indicate that R-loops identified by S9.6-based approaches (RDIP and DRIP) have a lower GC content compared to R-loops detected by RNase H methods (Figure S3E). This is in agreement with the *in vitro* preference of S9.6 for R-loops with lower GC content (König et al., 2017). Therefore, our

comprehensive comparison of the different R-loop detection strategies reveals that one feature that may be distinguished by S9.6 and RNase H is the GC content within R-loops. Improvement of RNase-H-based methods, like MapR, and their use in combination with S9.6 antibody-based approaches will allow for a more detailed picture of R-loop deregulation in both neurodegenerative diseases and cancers.

Limitations

Although MapR is a robust technique for the identification of R-loops, in its current form it has a few drawbacks. MapR isolates the DNA component of R-loops and does not provide information on the displaced strand. Modification of MapR to sequence both DNA and RNA in parallel will provide valuable strand information and perhaps allow for better resolution. Prior to DNA isolation, MapR chromatin fragments are treated with RNase A to remove RNAs. RNase A specificity for single-stranded RNA (ssRNA) and its inability to digest RNA that is paired to DNA would theoretically leave the RNA component of R-loops intact. Therefore, with a few careful experimental adjustments in enzyme concentration, length of digestion, and/or use of other nucleases with different specificities, MapR may also be adapted to identify the RNA components within R-loops.

STAR★METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

All unique reagents generated in this study are available from the Lead Contact, Kavitha Sarma (kavitha@sarmalab.com) with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HEK293 and U87T cell lines were grown in DMEM supplemented with 10% serum. HEK293 cells are female and U87T are male. Cell lines have not been authenticated.

METHOD DETAILS

Plasmid Construction—RNaseHdcat was amplified from pICE-RNaseHI-D10R-E48R-NLS-mCherry (Addgene plasmid: 60367) and sub-cloned into pGEX-6p-1-MNase and pLT3GEPIR (Fellmann et al., 2013). Primer sequences can be found in STAR Methods.

Generation of Stable Cell Lines—Stable cell lines were generated by transfection with Lipofectamine 2000 (Invitrogen) and selection with puromycin (1 μg/ml). Protein expression was induced by addition of doxycycline (1μg/ml final concentration) and analysed by western blot with antibodies as indicated.

Protein Expression and Purification—GST-MNase and GST-RHΔMNase were cloned into pGEX plasmid and transformed into BL21 (DE3) (ThermoFisher C601003) for expression. The transformed BL21 were grown in 500 mL LB medium containing 100 μg/mL ampicillin in a 37°C shaker until they reached 0.4–0.6 OD at 600nm. Protein expression was induced by addition of 0.5mM IPTG (Fisher scientific BP17755–10) and grown for an additional 3 hours in a 37°C shaker. Bacterial cell pellets were lysed in cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, and 1.8 mM KH2PO4, pH 7.4), and

sonicated (BRANSON Sonifier 450 at power setting 7) 3 times for 10 s each time. Lysates were centrifuged in a SORVALL LYNX 6000 centrifuge at 10,000 rpm for 20 min at 4°C and supernatant was transferred to a fresh tube. 150 μL GST-agarose beads (Affymetrix 1602/002) was added to lysates and the mixture was rotated overnight at 4°C. Beads were collected by centrifugation at 2,000 rpm for 10 min, washed 3 times with 1 mL cold PBS. Bound proteins were eluted using 100 μL GST elution buffer (125mM Tris-HCl, 150mM NaCl and 10mM glutathione, pH 8.0) and incubating on ice for 30 minutes with intermittent mixing. 3 elutions were performed and eluates analyzed for purity by SDS-PAGE. Purified proteins were dialyzed against 2L BC100 buffer (25 mM Tris-HCl pH7.6, 0.2 mM EDTA, 100mM KCl, 20% glycerol, and 1 mM β-mercaptoethanol), with two changes of dialysis buffer, at 4°C for 2 hours. Purified proteins were aliquoted and stored at −80°C.

MNase Activity Assay—3×10^6 HEK293 cells were resuspended in Buffer A (10 mM MES pH 6.5, 0.25M Sucrose, 60 mM KCl,15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5% Triton X-100, 0.5 mM PMSF) and incubated on ice for 20 min. Cells were centrifuged at 1000 g for 10 min and resuspended in 160 μL Buffer B (10 mM PIPES pH 6.8, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM PMSF) and divided into two tubes. 1.5 μM of GST-MNase and GST-RHΔ-MNase proteins was added and chromatin digestion performed at 37°C. 25 μL of the digestion reaction was transferred at different time points (0, 10, 30 min) to a tube containing 1 μL 0.5 M EDTA, 15 μL 10% SDS, 10 μL 5 M NaCl and 40 μL H_2O. DNA was extracted using phenol-chloroform and resolved on 2% agarose gels.

EMSA—RNA and DNA substrates were synthesized (IDT)-

RNA: 5′ GAAAUAUGGCGAGGAAAACUGAAAAAGGUGGAAAA 3′,

DNA_template: 5′ AF488:TTTTCCACCTTTTTCAGTTTTCCTCGCCATATTTC 3′,

DNA_non-template: 5′ GAAATATGGCGAGGAAAACTGAAAAAGGTGGAAAA 3′.

DNA:RNA hybrid and dsDNA were formed using 10 μM of each oligonucleotide, denatured at 95°C for 5 min, then cooled gradually to 21°C (Bio-Rad T100 Thermal Cycler). Reactions were assembled in 20 μL volume with purified protein and dsDNA or DNA:RNA hybrid in binding buffer containing 50 mM Tris pH 8.0, 100mM NaCl, 10 μg/ml BSA, 1 mM DTT, 0.1 mM EDTA, 5% Glycerol, and 2 μg yeast tRNA (Ambion cat# AM7119). Binding reactions were incubated at 30°C for 30 minutes and resolved on an 8% native polyacrylamide gel at 120V for 2 hours in 0.5 × TBE (44.5 mM Tris, 44.5 mM Boric Acid, 1 mM EDTA) at 4°C. Nucleic acids were visualized using the Amersham Typhoon Gel and Blot Imaging Systems (GE).

MapR and CUT&RUN—CUT&RUN was performed exactly as described in Skene et al. (2018) using 5 μg of FLAG M2 antibody or mouse IgG. MapR buffer volumes and incubation times follow the standard CUT&RUN protocol unless otherwise specified. 5×10^6 cells were washed twice with 1.5 mL room temperature wash buffer (20 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM Spermidine, 1 mM protease inhibitors) and immobilized on Cocanavalin A-coated beads. Immobilized cells were divided equally into two tubes and

resuspended in 50 μL wash buffer (20 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM Spermidine, 1 mM protease inhibitor) containing 0.02% Digitonin. GST-MNase and GST-RHΔ-MNase proteins were added to a final concentration of 1 μM and incubated overnight at 4°C with rotation. The beads were washed three times, resuspended in 100 μL Dig-wash buffer and place on ice. 2 μL 0.1 M CaCl2 added to activate MNase and digestion was carried out for 30 minutes. Reaction was stopped by adding equal volume of 2x STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% Digitonin, 5 μg RNaseA, 5 μg linear acrylamide and 2 pg/ml heterologous spike-in DNA). The samples were incubated at 37°C for 10 minutes to release the protein-DNA fragments and spun down at 16000 g for 5 minutes at 4°C. Supernatants were transferred to fresh tubes, 2 μL 10% SDS and 5 μg proteinase K was added and reactions were incubated at 70°C for 10 minutes. DNA was extracted using phenol-chloroform.

For RNase H treatment, after immobilization of cells to beads, 150 U RNase H in 50 μL Dig-wash buffer was added and incubated at room temperature for 1 hr before proceeding with MapR. HEK293 and U87T cells were treated with Actinomycin D (5 μg/ml) for 8 hr and processed for MapR.

Library Preparation and Sequencing—DNA was end-repaired using End-It Repair Kit, tailed with an A using Klenow exo minus, and ligated to custom adapters with T4 DNA ligase. Fragments >150 bp were size-selected with SPRI and subjected to ligation-mediated PCR amplification (LM-PCR) with custom barcoded adapters for Illumina sequencing using Q5 DNA polymerase. All enzymes except Q5 (NEB) were from Enzymatics (a QIAGEN company). Sequencing was performed on a NextSeq 500 (Illumina).

Step-by-Step MapR Protocol

1. 2–5 million cells are used per MapR experiment. Cells for 4 MapR experiments (8–20 million cells) can be processed together until Step 9.
2. Harvest cells and centrifuge at 600 × g for 3 min at room temperature and remove media.
3. Gently resuspend cells in 1 mL room temperature wash buffer (20 mM HEPES pH 7.5, 150mM NaCl, 0.5 mM Spermidine, 1 mM protease inhibitors (Roche, Cat#11836170001).
4. Centrifuge cells at 600 × g for 3 min at room temperature.
5. Repeat steps 3 and 4.
6. *Wash Concanavalin A-coated beads (Polysciences, Cat # 86057–3) 2 times with 1ml binding buffer (20mM HEPES -KOH pH 7.9, 10mM KCl, 1mM $CaCl_2$, 1mM $MnCl_2$) for each wash.

*NOTE- Use 10 μL Concanavalin A bead slurry per MapR experiment. Beads for up to 4 experiments can be processed together.

7. Gently resuspend cells in 1 mL room temperature wash buffer and add Concanavalin A-coated beads from Step 6. Cells for up to 4 experiments are immobilized together on beads in the same tube.
8. Rotate for 1 hr at room temperature.
9. Divide immobilized cells into separate tubes such that each tube contains 2–5 million cells.
10. Place tubes on a magnetic stand and remove the liquid.
11. Gently resuspend cells in 50 µL of wash buffer containing digitonin (Dig-wash buffer- 20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1 mM protease inhibitor, 0.02% digitonin).
12. Add purified GST-RHΔ-MNase and GST-MNase proteins to two separate tubes of immobilized cells to a final concentration of 1 µM. Pipette gently to mix and transfer to a new tube. This step ensures that beads with immobilized cells are contained within a small volume because of surface tension and do not get dispersed along the walls of the tube. Rotate at 4°C overnight.
13. Spin down with a quick pulse on a micro-centrifuge in case bead solution is displaced during overnight rotation.
14. Place tubes on a magnetic stand and remove all the liquid.
15. Wash cells with 1ml Dig-wash buffer.
16. Place tubes on a magnetic stand and remove the liquid.
17. Repeat Steps 15–16 two times (total of 3 washes).
18. Resuspend the cells in 100 µL of Dig-wash buffer by gently pipetting.
19. Incubate the tubes in wet ice for 2 minutes, to chill down to 0°C.
20. Add 2 µL of 100 mM $CaCl_2$ into the samples and put back immediately on wet ice.
21. Incubate the samples at 0°C for 30 min.
22. Stop reaction by addition of 100 µL 2XSTOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% Digitonin, 5 0µg RNaseA, 5 µg linear acrylamide and 2 pg/ml heterologous spike-in yeast DNA).
23. Place on a 37°C heat block for 10 min to release MapR chromatin fragments.
24. Centrifuge at 16,000 × g for 5 min at 4°C and place on a magnetic stand.
25. Transfer the supernatant to fresh tubes and add 2 µL 10% SDS and 2.5 µL Proteinase K (20 mg/ml).
26. Place on a 70°C heat block for 10 min.
27. Spin down with a quick pulse on a micro-centrifuge and add 300 µL phenol chloroform isoamyl alcohol (Sigma, Cat # P3803).

28. Transfer samples to phase-lock gel heavy tubes (Quanta bio, Cat # 2302830), and centrifuge at 16,000 × g for 5 min at room temperature.
29. Transfer upper aqueous phase to a fresh tube containing 0.5x volume of 7.5M NH_4OAc and 4 μL of linear acrylamide (VWR, Cat # 97063–560).
30. Add 2.5X volume of cold 100% ethanol and store at −80°C for 1 hr.
31. Centrifuge at 16,000 × g for 30 minutes at 4°C. Remove supernatant.
32. Wash the pellets with 1 mL 75% ethanol, and spin down at 16,000 × g for 10 min at 4°C.
33. Remove the supernatant and air-dry pellets for 2 minutes.
34. Dissolve the pellets in 20 μL buffer containing 1 mM Tris-HCl pH8 and 0.1 mM EDTA.
35. Use 7 μL DNA to make libraries as described in STAR Methods.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-wide Data Analysis—Raw MapR reads, RChIP reads (downloaded from GSE97072), RDIP reads (downloaded from GSE68948), or DRIP (downloaded from GSE70189) were mapped to the human genome (hg19) with Bowtie2 (Langmead and Salzberg, 2012) with default parameters. Normalized genome-wide read densities were computed using deeptools (Ramirez et al., 2016).

GRO-seq raw data were downloaded from GSE97072 (293 cells) and GSE92375 (U87T cells), and mapped to the human genome (hg19) with Bowtie2 (Langmead and Salzberg, 2012) with default parameters. Peaks were called using the HOMER tool findPeaks (Heinz et al., 2010) with default parameters. Any gene with overlapping GRO-seq peaks was considered active, while genes without GRO-seq peaks were considered inactive.

Peaks were called for each sample (with associated control as background, if possible) using MACS 2.1.1 (Zhang et al., 2008) using the parameters:–broad–broad-cutoff 0.1. Peak locations were computed by identifying R-loops in promoter regions (−2kb/+2kb of the TSS), gene bodies (the entirety of the gene including introns, but excluding the promoter region), and intergenic regions. Gene level overlaps were calculated by identifying genes in the hg19 NCBI RefSeq gene set with an R-loop at the promoter for each technology and reporting common genes.

Heatmaps and metagene plots were generated from signal surrounding the transcription start site (Figures 2, 3, S2, S3, S4, and S5H), middle of STR (Figures 4J and S5B–S5D). Duplicated reads (PCR and optical duplicates) were removed in processing. Reads per million were calculated in evenly spaced bins across the window by calculating the number of overlapping reads using Genomic Ranges countOverlaps, and normalizing by library size and bin size. Heatmaps were visualized using pheatmap.

Reduction of signal with ActD treatment was calculated using the total occupancy across the window from −2kb to +2 kb of each unique TSS. Read densities were computed (bedtools

coverage) over the merged peak co-ordinates from MapR and RHΔC&R and normalized to total mapped reads for each dataset.

Comparisons of read density (Figures 2G, S2E, S4A, and S4D) were calculated by determining the read density of each technology in either windows across the genome (Figure 2G), at the union of peaks (Figure S2E), or in a −2kb/+2kb window around the TSS (Figures S4A and S4D).

STRs—Short tandem repeats in hg19 were downloaded from UCSC. For percent of peaks overlapping with STRs, a peak was considered to contain an STR if it had at least a 1 bp overlap with an annotated STR. For the peak strength analysis, peaks were binned by q-value into 100 bins of even size. For each bin, the % of bp of all peaks in the bin that overlapped with an STR was reported.

G-Quadruplexes—G-quadruplexes were detected in promoters (−2kb/+2kb of TSS) of genes with an R-loop in the promoter region, as well as genes with no promoter R-loop, for all technologies using pqsfinder (Hon et al., 2017) and the number of G-quadruplexes per promoter was reported.

Active and Poised Enhancers—Active and poised enhancer regions were defined by H3K4me1 occupancy and H3K27ac presence (active enhancers) or absence (poised enhancers). Chromatin states for 293 cells were determined on hg19 by a 10-state ChromHMM model (Ernst & Kellis, 2012), run with default parameters.

Published Data—RDIP data (293, K562, IMR90) were downloaded from GEO: GSE68948. R-ChIP and Gro-seq data were downloaded from GEO: GSE97072. H3K27me3 ChIP-seq data were downloaded from GEO: GSM855015. H3K27ac (ENCSR000FCH), H3K4me1 (ENCSR000FCG), H3K4me3 (ENCSR000DTU), H3K9me3 (ENCSR000FCJ), DNase-seq (ENCSR000EJR), and input (ENCSR000EVA) ChIP-seq data were downloaded from ENCODE. U87 Gro-seq data were downloaded from GSE92375. DRIP data for K562 cells were downloaded from GEO: GSE70189. Read densities for R-ChIP were computed from analyzing raw R-ChIP sequencing data, but published peak files were used from R-ChIP and RDIP studies to call R-loop containing genes.

DATA AND CODE AVAILABILITY

Sequencing data generated for this study have been deposited in the NCBI GEO as GEO: GSE120637.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Guohong Li (IBP, Chinese Academy of Sciences) for the pGEX-6p-1MNase plasmid, Johannes Zuber (IMP) for the LT3GEPIR plasmid, and Hongwu Zheng (Cornell) for U87T cells. This work was supported by the NIH New Innovator Award DP2-NS105576 to K.S. R.B. acknowledges financial support from the NIH

(R01GM127408) and the Searle Scholars Program (15-SSP-102). E.J.S. acknowledges financial support from the NIH (T32HG000046).

REFERENCES

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, and Carrico RJ (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J. Immunol. Methods 89, 123–130. [PubMed: 2422282]

Chao W, Huynh KD, Spencer RJ, Davidow LS, and Lee JT (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295, 345–347. [PubMed: 11743158]

Chédin F (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet. 32, 828–838. [PubMed: 27793359]

Chen L, Chen JY, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell 68, 745–757.e745. [PubMed: 29104020]

Cristini A, Groh M, Kristiansen MS, and Gromak N (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep. 23, 1891–1905. [PubMed: 29742442]

Crossley MP, Bocek M, and Cimprich KA (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol. Cell 73, 398–411. [PubMed: 30735654]

da Rocha ST, and Heard E (2017). Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. Nat. Struct. Mol. Biol 24, 197–204. [PubMed: 28257137]

Dumelie JG, and Jaffrey SR (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. eLife 6, e28306. [PubMed: 29072160]

Duquette ML, Handa P, Vincent JA, Taylor AF, and Maizels N (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. 18, 1618–1629. [PubMed: 15231739]

Ernst J, and Kellis M (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216. [PubMed: 22373907]

Fellmann C, Hoffmann T, Sridhar V, Hopfgartner B, Muhar M, Roth M, Lai DY, Barbosa IA, Kwon JS, Guan Y, et al. (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell Rep. 5, 1704–1713. [PubMed: 24332856]

Frietze S, Wang R, Yao L, Tak YG, Ye Z, Gaddis M, Witt H, Farnham PJ, and Jin VX (2012). Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. Genome Biol. 13, R52. [PubMed: 22951069]

Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, and Reinberg D (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell 45, 344–356. [PubMed: 22325352]

Ginno PA, Lott PL, Christensen HC, Korf I, and Chédin F (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell 45, 814–825. [PubMed: 22387027]

Ginno PA, Lim YW, Lott PL, Korf I, and Chédin 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. [PubMed: 23868195]

Groh M, and Gromak N (2014). Out of balance: R-loops in human disease. PLoS Genet. 10, e1004630. [PubMed: 25233079]

Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, and Vanoosthuyse V (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J. Mol. Biol 430, 272–284. [PubMed: 29289567]

Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, et al. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol. Cell 57, 636–647. [PubMed: 25699710]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineagedetermining transcription factors prime cis-regulatory

elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

Hon J, Martínek T, Zendulka J, and Lexa M (2017). pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. Bioinformatics 33, 3373–3379. [PubMed: 29077807]

Jégu T, Aeby E, and Lee JT (2017). The X chromosome in space. Nat. Rev. Genet 18, 377–389. [PubMed: 28479596]

König F, Schubert T, and Längst G (2017). The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. PLoS One 12, e0178875. [PubMed: 28594954]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, and Carey VJ (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol 9, e1003118. [PubMed: 23950696]

Manzo SG, Hartono SR, Sanz LA, Marinello J, De Biasi S, Cossarizza A, Capranico G, and Chedin F (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. Genome Biol. 19, 100. [PubMed: 30060749]

Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, ÓBroin P, Sato H, Zhang Z, Jeddeloh J, Montagna C, Golden A, et al. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics Chromatin 8, 46. [PubMed: 26579211]

Perego MGL, Taiana M, Bresolin N, Comi GP, and Corti S (2019). R-Loops in Motor Neuron Diseases. Mol. Neurobiol 56, 2579–2589. [PubMed: 30047099]

Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, Hacohen N, Schier AF, Blackshear PJ, Friedman N, et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell 159, 1698–1710. [PubMed: 25497548]

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165. [PubMed: 27079975]

Ribeiro de Almeida C, Dhir S, Dhir A, Moghaddam AE, Sattentau Q, Meinhart A, and Proudfoot NJ (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. Mol. Cell 70, 650–662.e658. [PubMed: 29731414]

Richard P, and Manley JL (2017). R Loops and Links to Human Disease. J. Mol. Biol 429, 3168–3180. [PubMed: 27600412]

Roy D, Yu K, and Lieber MR (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol. Cell. Biol 28, 50–60. [PubMed: 17954560]

Santos-Pereira JM, and Aguilera A (2015). R loops: new modulators of genome dynamics and function. Nat. Rev. Genet 16, 583–597. [PubMed: 26370899]

Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, and Chédin F (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol. Cell 63, 167–178. [PubMed: 27373332]

Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, and Cramer P (2016). TT-seq maps the human transient transcriptome. Science 352, 1225–1228. [PubMed: 27257258]

Skene PJ, and Henikoff S (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6, e21856. [PubMed: 28079019]

Skene PJ, Henikoff JG, and Henikoff S (2018). Targeted in situ genomewide profiling with high efficiency for low cell numbers. Nat. Protoc 13, 1006–1019. [PubMed: 29651053]

Skourti-Stathaki K, and Proudfoot NJ (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396. [PubMed: 24990962]

Skourti-Stathaki K, Proudfoot NJ, 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. [PubMed: 21700224]

Skourti-Stathaki K, Kamieniarz-Gdula K, and Proudfoot NJ (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516, 436–439. [PubMed: 25296254]

Sollier J, and Cimprich KA (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. 25, 514–522. [PubMed: 26045257]

Song C, Hotz-Wagenblatt A, Voit R, and Grummt I (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev. 31, 1370–1381. [PubMed: 28790157]

Thomas M, White RL, and Davis RW (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. USA 73, 2294–2298. [PubMed: 781674]

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. (2012). The accessible chromatin landscape of the human genome. Nature 489, 75–82. [PubMed: 22955617]

Vanoosthuyse V (2018). Strengths and Weaknesses of the Current Strategies to Map and Characterize R-Loops. Noncoding RNA 4, E9. [PubMed: 29657305]

Wahba L, Costantino L, Tan FJ, Zimmer A, and Koshland D (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. 30, 1327–1338. [PubMed: 27298336]

Wang J, Haeusler AR, and Simko EA (2015). Emerging role of RNA•DNA hybrids in C9orf72-linked neurodegeneration. Cell Cycle 14, 526–532. [PubMed: 25590632]

Yasuhara T, Kato R, Hagiwara Y, Shiotani B, Yamauchi M, Nakada S, Shibata A, and Miyagawa K (2018). Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. Cell 175, 558–570.e511. [PubMed: 30245011]

Yu K, Chedin F, Hsieh CL, Wilson TE, and Lieber MR (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol 4, 442–451. [PubMed: 12679812]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Highlights

- MapR is a fast, antibody-independent R-loop profiling strategy
- MapR merges DNA-RNA hybrid recognition by RNase H and MNase cleavage to detect R-loops
- MapR does not require the generation of stable transgenic cell lines
- Dynamic R-loops formed at enhancers are specifically and efficiently recovered by MapR

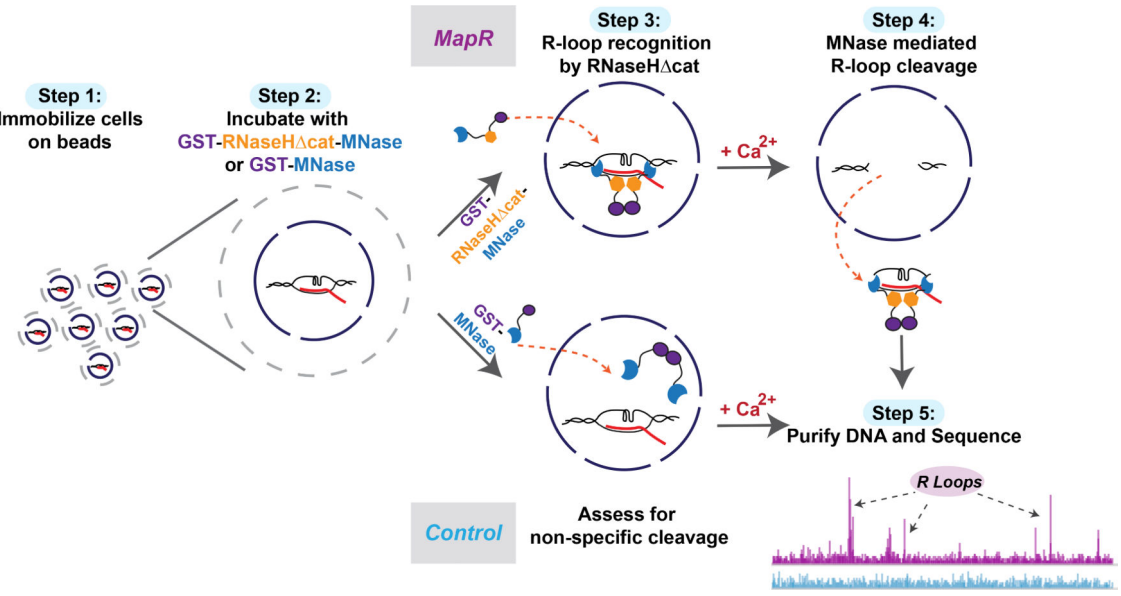


Figure 1. MapR, a Native and Antibody-Independent R-Loop Detection Strategy

R-loop recognition and recovery by MapR. Step 1: cells are immobilized on concanavalin A beads and permeabilized. Step 2: equimolar amounts of a catalytic deficient mutant of RNase H fused to micrococcal nuclease (GST-RHΔ-MNase) or GST-MNase is added to immobilized cells. Step 3: the RHΔ module recognizes and binds R-loops on chromatin. Step 4: controlled activation of the MNase moiety by addition of calcium results in cleavage of DNA fragments in proximity to R-loops. Step 5: Released R-loops diffuse out of the cell; the DNA is recovered and sequenced.

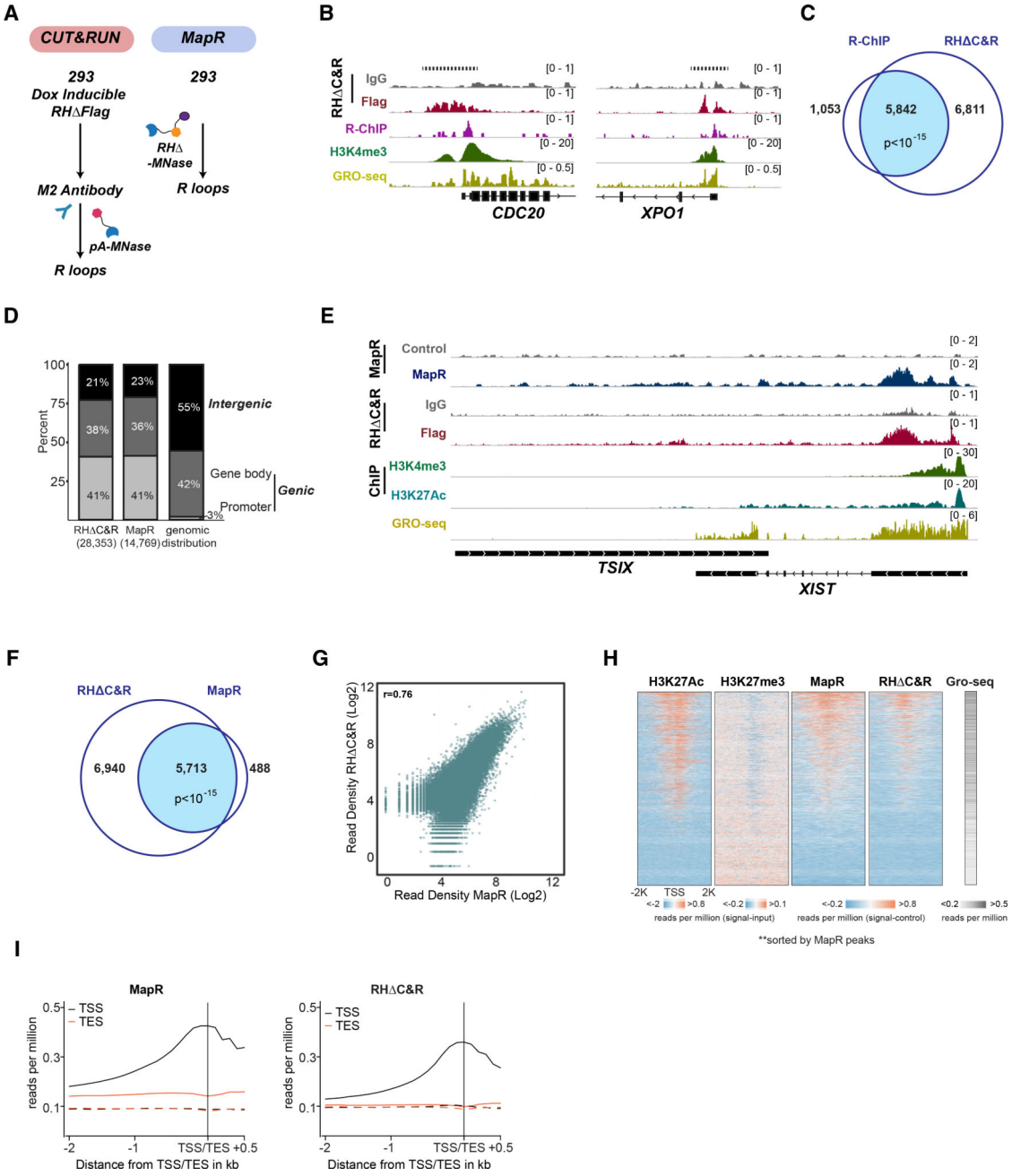


Figure 2. MapR and RHΔ CUT&RUN Signals Are Enriched at Similar Regions Genome-wide

(A) Schematic of RHΔC&R using FLAG M2 antibody (left) and MapR using GST-RHΔ-MNase (right) in HEK293.

(B) Enriched regions identified by RHΔC&R and R-ChIP in HEK293. GRO-seq and H3K4me3 tracks indicate active gene transcription.

(C) Venn diagram of gene-level overlap between RHΔC&R and R-ChIP. Total number of unique genes with an R-loop at the promoter region (−2kb/+2kb from the TSS) and their overlap are shown. $p < 10^{-15}$, hypergeometric distribution.

(D) Peak distribution of MapR and RHΔC&R showing percent of peaks mapping to promoter regions (−2kb/+2kb from the TSS), gene bodies (entirety of gene including introns, excluding promoter region), or intergenic regions. Total peak numbers are shown in parentheses. Background genomic distribution is shown for comparison.
(E) MapR and RHΔC&R signals at the *XIST* and *TSIX* genes. GST-MNase and IgG controls are shown for MapR and RHΔC&R, respectively. H3K4me3 (Thurman et al., 2012) and H3K27Ac (Frietze et al., 2012) chromatin immunoprecipitation sequencing (ChIP-seq) and GRO-seq (Chen et al., 2017) tracks are shown as proxies for transcriptional activity.
(F) Venn diagram of gene-level overlap between RHΔC&R and MapR. Total number of unique genes with an R-loop at the promoter region (−2kb/+2kb from the TSS) and their overlap are shown. $p < 10^{-15}$, hypergeometric distribution.
(G) Correlation scatterplot showing read densities for the union of peaks from MapR and RHΔC&R (log2 scale). r = 0.76, Spearman correlation coefficient.
(H) Heatmaps of H3K27Ac, H3K27me3, MapR, and RHΔC&R signal intensity across all TSS sorted by MapR signal. GRO-seq signals were summed and collapsed into a box per gene.
(I) Metagene plots of MapR (left) and RHΔC&R (right) signals at all TSSs (black) and TESs (red).

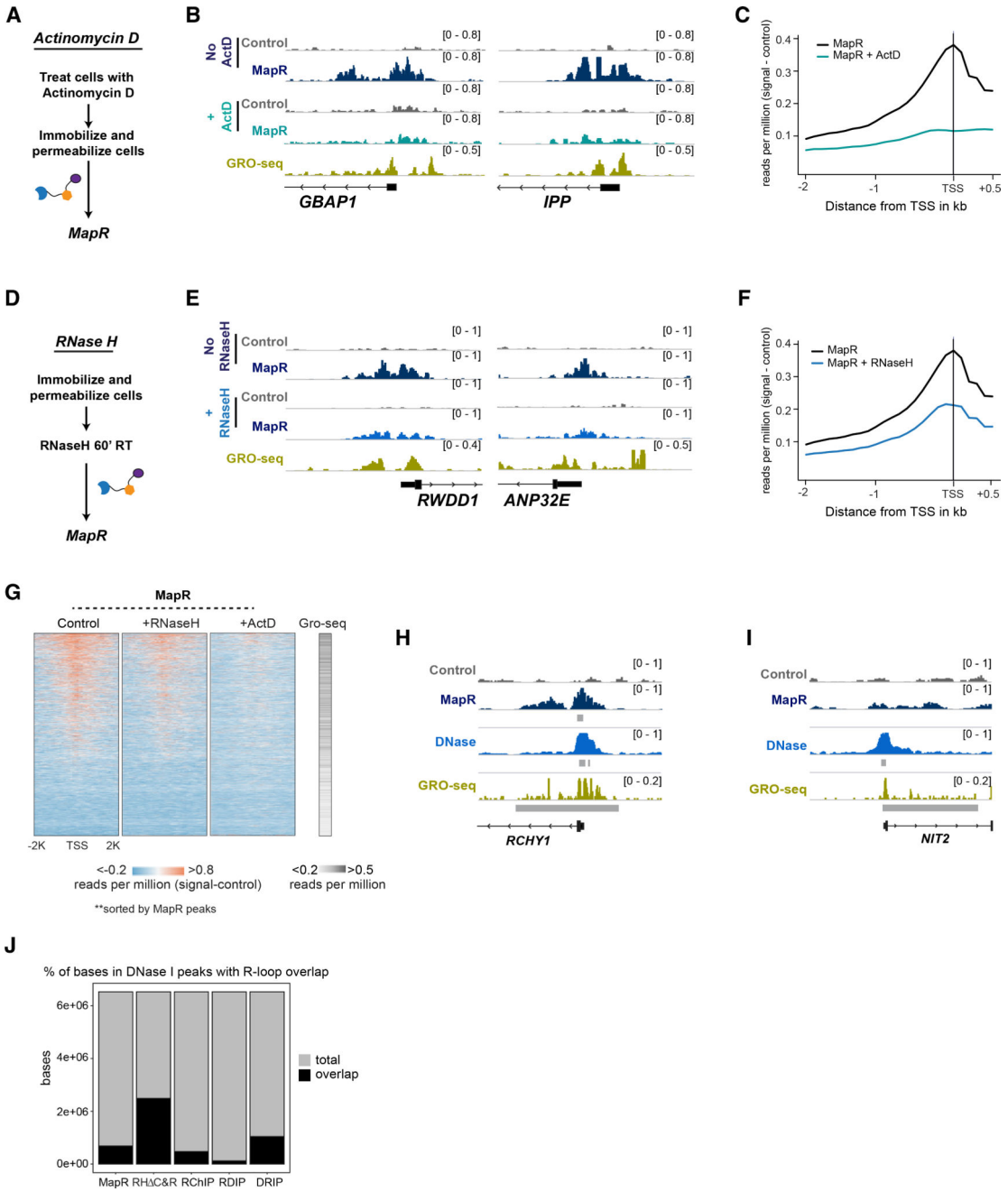


Figure 3. Characterization of R-Loops Obtained by MapR

(A) Schematic of actinomycin D (ActD) treatment followed by MapR identification of R-loops.

(B) Genome browser views of *GBAP1* and *IPP* genes showing MapR signals with and without ActD treatment. GRO-seq tracks show active transcription.

(C) Metagene plots of MapR signals at TSS of all genes with and without ActD treatment.

(D) Schematic of RNase H treatment followed by MapR identification of R-loops.

(E) Genome browser views *RWDD1* and *ANP32E* genes showing MapR signals with and without RNase H treatment. GRO-seq tracks show transcription at these genes.

(F) Metagene plots of MapR signals at TSS of all genes with and without RNase H treatment.
(G) Heatmaps of MapR signals across all TSS in control, RNase-H-, and ActD-treated HEK293 cells, sorted by MapR signal. GRO-seq signals from untreatedHEK293 were summed and collapsed into a box per gene.
(H) Genome browser views of *RCHY1* gene showing overlapping MapR and DNase I hypersensitivity signals. GRO-seq tracks show active transcription.
(I) Genome browser views of *NIT2* gene showing DNase I hypersensitivity signals without appreciable MapR signal. GRO-seq tracks show active transcription.
(J) Base-level overlap between DNase I signal and MapR, RHΔC&R, RChIP, RDIP, and DRIP signal.

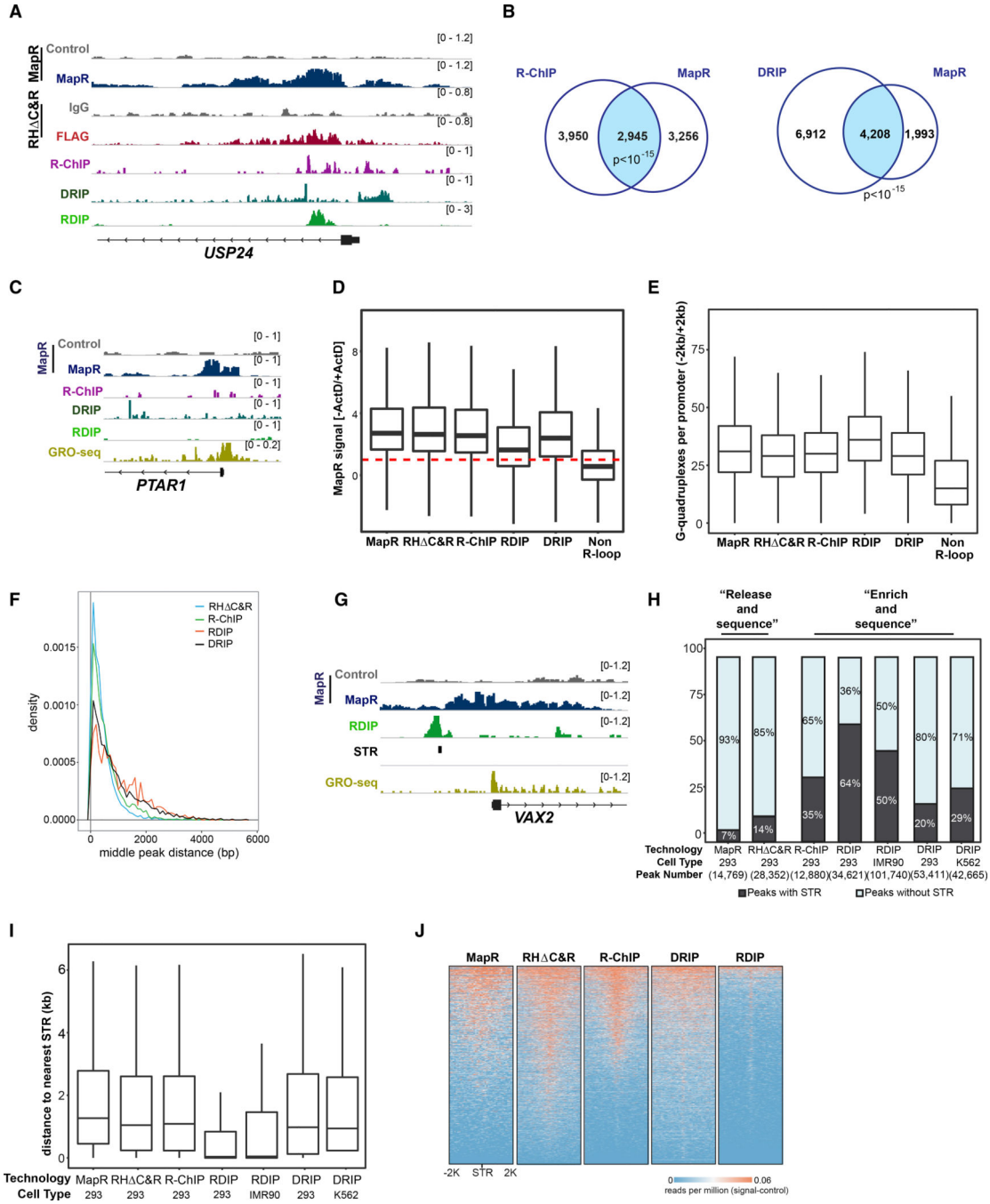


Figure 4. Similarities and Differences between MapR and Other R-Loop Detection Methods

(A) Genome browser view of the *USP24* gene showing MapR, RHΔC&R, R-ChIP, DRIP, and RDIP signals. The scale for the y axis is in RPM.

(B) Gene-level overlap between MapR and R-ChIP datasets (left) and MapR and DRIP datasets (right). Total number of unique genes with an R-loop at the promoter region (−2kb/+2kb from the TSS) and their overlap are shown. $p < 10^{-15}$, hypergeometric distribution.

(C) Genome browser view of *PTAR1* that shows MapR but no R-ChIP, DRIP, or RDIP signals. GRO-seq tracks indicate active transcriptional status.

(D) Ratio in the −2kb/+2 kb window around the TSS in MapR signal in untreated cells and cells treated with actinomycin D to inhibit transcription. MapR signal from genes identified by each R-loop detection method and genes that did not contain R-loops are shown.
(E) Genes called from MapR, RHΔC&R, R-ChIP, DRIP, and RDIP in 293 cells show similar frequency of predicted G quadruplex structures at their promoter, indicative of R-loop presence, whereas non-R-loop genes have a lower frequency of G quadruplexes.
(F) Distance between MapR peaks and peaks detected by RHΔC&R, RChIP, RDIP, and DRIP at genes with R-loop detected at promoter.
(G) Genome browser view of *VAX2*, that shows MapR signals that do not overlap with RDIP peaks in close proximity. Simple tandem repeat (STR) and GRO-seq tracks are shown. The scale for the y axis is in RPM.
(H) Percent of peaks that contain STRs in MapR, RHΔC&R, RDIP, and R-ChIP experiments in HEK293. Results from published RDIP (IMR90)(Nadel et al., 2015) and DRIP (293 and K562) (Manzo et al., 2018; Sanz et al., 2016) datasets are also shown.
(I) Distance (kb) from peaks detected by each technology to the nearest STR.
(J) Heatmaps of GRO-seq signal intensity across STRs identified by each method in HEK293. Each heatmap is sorted by intensity.

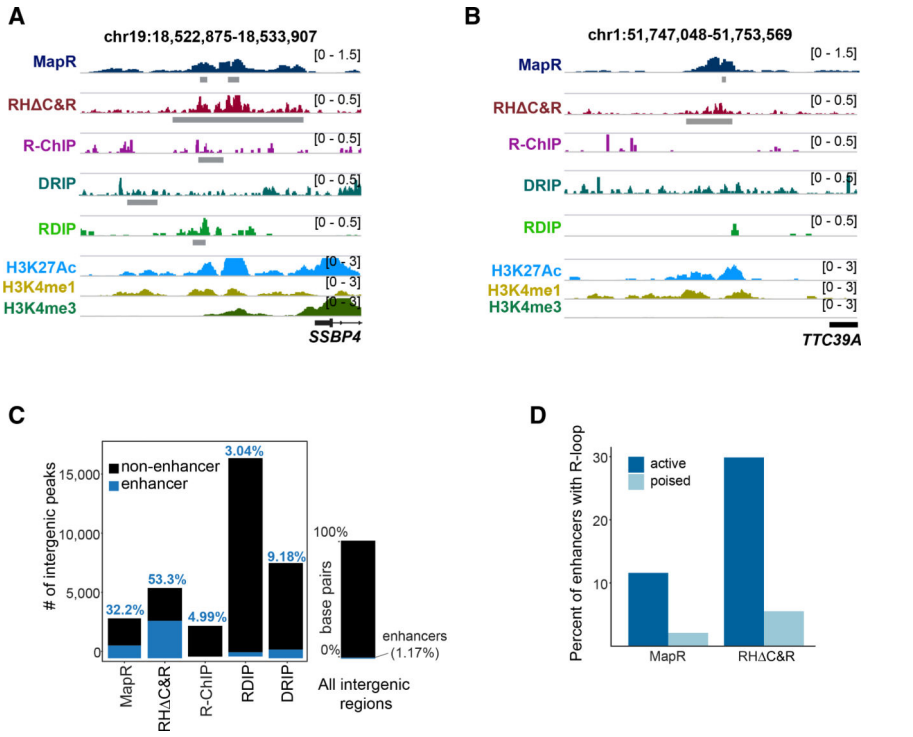


Figure 5. MapR Identifies R-Loops Formed at Enhancers with Higher Sensitivity

(A) Genome browser view of an intergenic enhancer with an R-loop detected by MapR, RHΔC&R, DRIP, RDIP, and R-ChIP. Chromosome and genome coordinates are shown above the panels. H3K27ac, H3K4me1, and H3K4me3 tracks are also shown to verify that the region has a signature characteristic of an active enhancer.

(B) Genome browser view of an intergenic enhancer with an R-loop detected by MapR and RHΔC&R but not by DRIP, RDIP, or R-ChIP. Chromosome and genome coordinates are shown above the panels. H3K27ac, H3K4me1, and H3K4me3 tracks are also shown to verify that the region has a signature characteristic of an active enhancer.

(C) Percentage of intergenic peaks that overlap with an enhancer for each technology (left), and the percentage of intergenic sequence that is characterized as an enhancer by chromHMM (right).

(D) Percentage of active and poised enhancers that contain a R-loop detected by MapR or RHΔC&R.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal ANTI-FLAG M2 antibody	Sigma	A2220
Rabbit IgG	Sigma	I5006–10MG
Bacterial and Virus Strains		
E.coli BL21 (DE3)	Thermo Fisher	C601003
E.coli Top10	Thermo Fisher	C404003
Chemicals, Peptides, and Recombinant Proteins		
Dulbecco's Modification of Eagle's Medium (DMEM)	Corning	10013CV
SuperCalf Serum	Gemini Bio-Products	100–510
Lipofectamine 2000	Invitrogen	11668–019
Doxycycline	Sigma	D9891–25 g
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Fisher Scientific	BP1755–10
Phenylmethane sulfonyl fluoride (PMSF)	RPI	329986
DL-Dithiothreitol (DTT)	Sigma	D0632
Digitonin	Sigma	D141
RNaseA	Roche	10109169001
Linear Acrylamide	VWR	97063–560
Proteinase K	Sigma	P2308–100MG
RNaseH	NEB	M0297L
Actinomycin D	Cell Signaling Technology	15021S
End-Repair Mix	Enzymatics	Y9140-LC-L
Klenow HC (3′ → 5′ exo-)	Enzymatics	P7010-HC-L
T4 DNA ligase	Enzymatics	L6030-HC-L
Agencourt AMPure XP	Beckman Coulter	A63881
Q5 HotStart High-Fidelity DNA Polymerase	NEB	M0493L
Thermolabile UDG	Enzymatics	G5020L
NEBNext Library Quant Kit for Illumina	NEB	E7630L
KAPA Library Quant Kit	Roche	07960255001
cOmplete EDTA-free protease inhibitor cocktail tablets	Roche	11836170001
Phenol: Chloroform: Isoamyl Alcohol 25:24:1	Sigma	P3803
Yeast tRNA	Ambion	AM7119
2-Mercaptoethanol	Sigma	M6250
Deposited Data		
Raw and Analyzed data	This study	GEO: GSE120637
Experimental Models: Cell Lines		
HEK293	Laboratory of Jeannie Lee (MGH)	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
U87T	Laboratory of Hongwu Zheng (Cornell)	N/A
Oligonucleotides		
RNA: 5′ GAAAUAUGGCGAGGAAAACUGAAAAAGGUGGAAAA 3′	IDT	N/A
DNA_template: 5′ AF488:TTTTCCACCTTTTTCAGTTTTCCTCGCCATATTTC 3′	IDT	N/A
DNA_non-template: 5′ GAAATATGGCGAGGAAAACTGAAAAAGGTGGAAAA 3′	IDT	N/A
RNaseH-NEBulider-F 5′ GTAAAGTCGAGCTTGCGTTGATGCTGAAACAGGTGGAAATCTTC 3′	IDT	N/A
RNaseH-NEBulider-R 5′ GATCCTGGCTGAATTCGGTGGCGACCGG 3′	IDT	N/A
RNaseH-BamHI-F 5′ TGCGGATCCATGCTGAAACAGGTGGAAATCT 3′	IDT	N/A
RNaseH-ECORI-R 5′ CCAGAATTCCACTTCCACCTGGTAGCCGGTA 3′	IDT	N/A
Recombinant DNA		
pICE-RNaseHI-D10R-E48R-NLS-mCherry	Addgene	60367
LT3GEPIR	Fellmann et al., 2013	N/A
pGEX-6p-1-MNase	Guohong Li's lab	N/A
pGEX-6p-1-GST-MNase	This paper	N/A
pGEX-6p-1-GST-ΔRNH-MNase	This paper	N/A
Software and Algorithms		
bowtie2	Langmead and Salzberg, 2012	N/A
deepTools	Ramirez et al., 2016	N/A
HOMER	Heinz et al., 2010	N/A
MACS 2.1.1	Zhang et al., 2008	N/A
GenomicRanges (R package)	Lawrence et al., 2013	N/A
pheatmap (R package)		N/A
Other		
PrepEase Protein Purification Glutathione Agarose 4B	Affymetrix	78820
BioMag Plus Concanavalin A (Cocanavalin A-coated magnetic beads)	Polysciences	86057–3