



Article

# R-Loop Tracker: Web Access-Based Tool for R-Loop Detection and Analysis in Genomic DNA Sequences

Václav Brázda <sup>1</sup>, Jan Havlík <sup>2</sup>, Jan Kolomazník <sup>2</sup>, Oldřich Trenz <sup>2</sup> and Jiří Šťastný <sup>2,3,\*</sup>

<sup>1</sup> Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic; vaclav@ibp.cz

<sup>2</sup> Department of Informatics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic; xhavlik@mendelu.cz (J.H.); jan.kolomaznik@gmail.com (J.K.); oldrich.trenz@gmail.com (O.T.)

<sup>3</sup> Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic

\* Correspondence: stastny@fme.vutbr.cz

**Abstract:** R-loops are common non-B nucleic acid structures formed by a three-stranded nucleic acid composed of an RNA–DNA hybrid and a displaced single-stranded DNA (ssDNA) loop. Because the aberrant R-loop formation leads to increased mutagenesis, hyper-recombination, rearrangements, and transcription-replication collisions, it is regarded as important in human diseases. Therefore, its prevalence and distribution in genomes are studied intensively. However, *in silico* tools for R-loop prediction are limited, and therefore, we have developed the *R-loop tracker* tool, which was implemented as a part of the *DNA Analyser* web server. This new tool is focused upon (1) prediction of R-loops in genomic DNA without length and sequence limitations; (2) integration of *R-loop tracker* results with other tools for nucleic acids analyses, including Genome Browser; (3) internal cross-evaluation of *in silico* results with experimental data, where available; (4) easy export and correlation analyses with other genome features and markers; and (5) enhanced visualization outputs. Our new *R-loop tracker* tool is freely accessible on the web pages of *DNA Analyser* tools, and its implementation on the web-based server allows effective analyses not only for DNA segments but also for full chromosomes and genomes.

**Keywords:** sequence analysis; RNA–DNA hybrid; non-B structure



**Citation:** Brázda, V.; Havlík, J.; Kolomazník, J.; Trenz, O.; Šťastný, J. *R-Loop Tracker: Web Access-Based Tool for R-Loop Detection and Analysis in Genomic DNA Sequences*. *Int. J. Mol. Sci.* **2021**, *22*, 12857. <https://doi.org/10.3390/ijms222312857>

Academic Editor: Daniela Montesarchio

Received: 29 October 2021  
Accepted: 25 November 2021  
Published: 27 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In addition to the basic DNA structure first described by Watson and Crick in 1953 [1], new findings show many nucleic acid structures differing from this canonical B-DNA structure. The best-described local non-B nucleic structures are cruciform [2,3], left-handed Z-DNA [4,5], triplexes [6,7], and quadruplexes [8,9]. Contemporary research has demonstrated that these structures are present within the genomes of all organisms and play important roles in many basic biological functions [10–13]. R-loops are regular three-stranded non-B nucleic acid structures constituted from an RNA–DNA hybrid [14]. The formation of R-loops plays a role in DNA replication, mutations, and homologous recombination, but it also is suggested to be an inducer of trinucleotide repeat expansion associated with human neuromuscular degenerative diseases [14,15]. Furthermore, R-loop resolution defects have been found to be associated with an increasing number of human diseases [16,17]. Negative supercoiling of DNA during transcription facilitates the formation of local nucleic acid structures, including cruciform and R-loop [18,19]. R-loop structures have engendered immense interest due to their involvement in such important biological processes as transcription, mRNA splicing, DNA replication, recombination, and repair [20]. Although there are several tools to study inverted repeats that form cruciform, G-quadruplexes, and other non-B DNA structures in nucleic acids [11,21–24], *in silico* tools for the prediction of R-loops are limited [22]. The authors of the QmRLFS-finder tool used

a quantitative structural model for an R-loop-forming sequence written in Python-based code [25]. In our *R-loop tracker* implementation, we used a modern Java environment to enhance performance and optimize the algorithm for full genome analyses without length limitations. The *R-loop tracker* tool has been implemented as part of the *DNA Analyser* web server, and this completely new implementation enables server-based prediction of R-loops while integrating its results with other tools for nucleic acids analyses, internal cross-evaluation of results with experimental data, easy export and correlation analyses with other genome features and markers, and enhanced visualization outputs. The *R-loop tracker* tool is integrated into the web pages of *DNA Analyser* tools <http://bioinformatics.ibp.cz/> (accessed on 28 October 2021) and freely accessible.

## 2. Methods and Results

### 2.1. Features

*R-loop tracker* is part of the *DNA Analyser* web server, which provides multiple analyses and sequence operations at a single location. Our implementation utilizes the Python-based QmRLFS finder algorithm [25] and benefits from modern, rapid processing of Java-based code, a web-based interface with rich graphical results presentation, and server data management. The workflow and implementation were completely rewritten for a server-based platform, coded, and constructed so that all imported sequences and analyses are retained in the user database. The server architecture operates in batches and provides an API, which can be used for full genome sequence analyses.

### 2.2. Input and Analysis

Sequences can be imported individually or in larger numbers according to NCBI ID (RefSeq). Another option is to import a sequence via file either in a plain text format or in FASTA format. For quick operations and short sequence tests, there is an option to import a sequence directly from a computer's clipboard into a web application form. One can also add tags to each sequence to group them according to the project or planned analysis. Upon importing a sequence, it will remain linked to your account so that one can work with the same sequence later at any time. Individual sequence length is limited only by the maximum file upload size, which is currently 2048 MiB. Therefore, one can easily import and analyze whole chromosomes. Such an analysis takes longer to process, of course, and once such an operation is completed, the results will be stored on the server, where they can be found later and displayed from the "Stored results" tab.

Two detection models for the R-loop analysis are available in the "R-loop tracker" tab: The RIZ 3g-cluster is looking for three consecutive guanine clusters with cluster size of at least three guanines (equal to model *m1* in the original implementation), and the RIZ 2G-cluster is looking for two consecutive guanine clusters with cluster size of at least four guanines (equal to model *m2*). The first of these is the default model for an analysis. R-loops are detected on both DNA strands simultaneously. The *R-loop tracker* application creates the complementary strand and analyzes the parallel sequence as well. R-loops on the strand of the imported sequence are marked as +, R-loops from the complementary strand are marked as −.

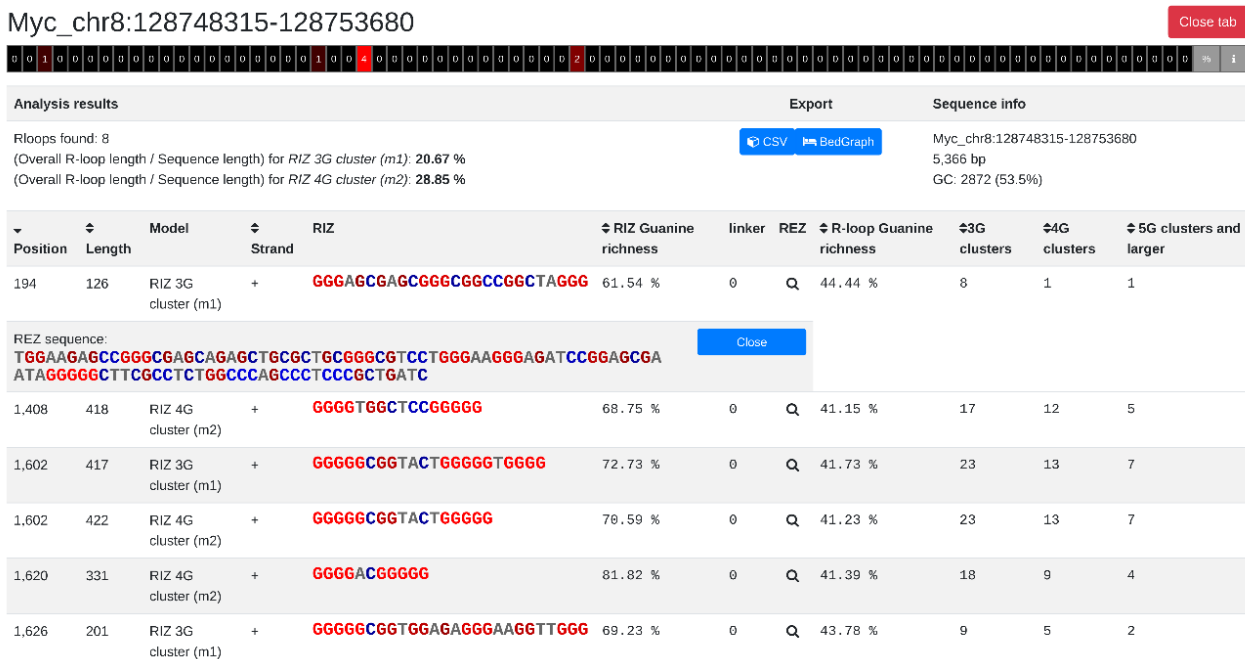
### 2.3. R-Loop Detection

The detection algorithm is based on a quantitative structural model for an R-loop-forming sequence [25]. This algorithm has been verified by comparison to in vitro experiments [26] and provides satisfactory results for R-loop detection in selected human/mouse genes.

### 2.4. R-Loop Tracker Web Application Output

All results are displayed directly in the browser using AJAX technology. It is possible to run multiple analyses at once using batches and asynchronous events. This ensures that the web server is still able to receive requests while processing analyses in the background. Each analysis is displayed in a separate tab providing detailed information about the

detected R-loops (Figure 1). The top part of the tab displays the name of the analyzed sequence and a heat map of R-loop localization in the sequence (showing the number of R-loops found in the location or % of sequence covered by R-loops). Below the heat map, the general sequence information is shown, which includes the number of R-loops found, the R-loop rate throughout the sequence, the sequence length, and the CG content. There are also buttons that provide data export options.



**Figure 1.** Example of an *R-loop tracker* result, analysis of *Myc* gene location on human chromosome 8.

The main windows show *R-loop tracker* results in the form of a table, where every row represents a detected R-loop. The columns provide information about the position of the R-loop in the sequence, R-loop length, model, location of the R-loop according to DNA strand, R-loop initiation zone (RIZ) sequence, guanine density in RIZ sequence, length of the linker, R-loop elongation zone (REZ, which is opened by clicking the magnifier icon), guanine density in REZ sequence, and number of G-clusters based on the G-cluster length. To improve readability and show the important features of R-loop sequences, we formatted the RIZ and REZ sequences based on the numbers of Gs and Cs. The more Gs there are in the cluster, the brighter the intensity of the cluster displayed in red (blue color is used for C clusters). Columns can be sorted by clicking on the arrows in the first row, and therefore, it is very easy to find R-loops with parameters of interest.

### 2.5. Output Formats

In addition to the graphical representation described above, there is also the possibility to export the analysis data in two different file formats. The first choice is the widely used CSV format, which contains the same data as seen in the web browser and is useful for machine processing. The second option is the *bedGraph* format, which can be used for visualization in *Genome Browser* [27]. A *bedGraph* file can be added directly as a custom track into the *Genome Browser* engine. Be aware that this feature requires a specific sequence name for correct integration (see *R-loop tracker* help available on the tool's web pages). A *bedGraph* file contains a header with information about the analysis and R-loop records consisting of sequence name, start and end position of the R-loop in input sequence, and R-loop score.

The R-loop score reflects the guanine density and cluster occurrences in an R-loop according to the following formula

$$\text{score} = (g_3 + 2 \cdot g_4 + 3 \cdot g_5) \cdot \text{RIZ}_{G\%} \cdot \text{RLOOP}_{G\%}, \quad (1)$$

where  $g_3$ ,  $g_4$ , and  $g_5$  represent a guanine cluster count of specific size, and these weighted attributes are multiplied by R-loop initiation zone guanine density ( $\text{RIZ}_{G\%}$ ) and R-loop guanine density ( $\text{RLOOP}_{G\%}$ ).

### 2.6. API Usage

*DNA Analyser* provides an application programming interface to integrate one's scripts or web server with *R-loop tracker*. The API documentation is available at the *DNA Analyser* web site.

## 3. Discussion

R-loops are three-stranded DNA–RNA hybrid molecules that play important roles in a wide range of biological processes, including processes initiating molecular events regulating gene transcription and chromatin modifications [28]. R-loops have also been suggested as hot-spots for double-stranded breaks leading to DNA damage and human diseases [29,30]. These structures are very dynamic, and a subset of R-loops maintained after differentiation has been shown to be associated with repressive chromatin marks on silent pluripotency genes [31]. Meanwhile, several methods for their profiling have been developed in vivo [32–34]. Although many tools exist for analyzing other non-B DNA structures in silico, the options for R-loop analyses are limited. Therefore, we integrated a tool for R-loops into our *DNA Analyser* server. This new tool, *R-loop tracker*, is a web application capable of processing multiple analyses simultaneously while having the capability to process large input sequences, including full-length chromosomes. In addition to this advantage, the results are presented in a web browser with enhanced visualization and are sortable, thereby enabling effective information mining from *R-loop tracker* analyses. The web server runs on up-to-date technologies and provides means of integration with other tools and servers. Contemporary DRIP-seq analysis can provide experimental data of R-loop presence in various organisms, in various cell lines, and under diverse conditions [35–37]. A comparison of *R-loop tracker* results with DRIP-seq data showed that *R-loop tracker* can help find target sequences in silico. Interestingly, R-loops, and especially the RIZ parts, are also similar to sequences predicted as G-quadruplexes by the G4Hunter algorithm [23]. These results suggest an important role of other non-B DNA structures for effective R-loop formation. On the other hand, the overlap between various R-loop data sets is still poor in some cases. Therefore, future training of the *R-loop tracker* algorithm on an experimental data set can be beneficial. Fortunately, the *R-loop tracker* tool can be easily updated to improve the results and detection algorithms when more validated data sets are available.

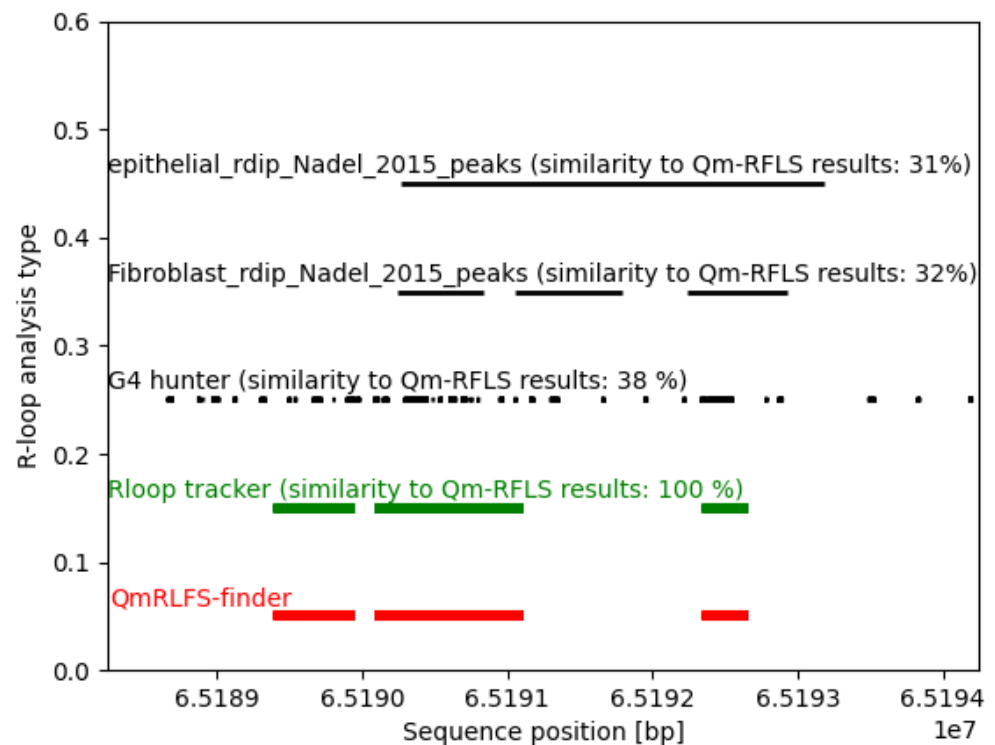
## 4. Materials and Methods

### 4.1. Algorithm Validation

To test our implementation of the *R-loop tracker* tool, we created a command line utility to obtain and compare results of different analyses. This utility uses data downloaded from *Genome Browser* which have to be manually transformed into input files. The utility only loads files in a selected folder, proceeds with comparison, and provides a graphical representation. An algorithm output can also be compared with in vitro data sets present in *Genome Browser*. The data set is available in the utility repository: <https://github.com/janhavlik/genome-comparator> (accessed on 28 October 2021), and the repository also contains README to manually recreate the process with a different input sequence if needed.

### Comparison Method

Every interval defined with an R-loop start position and an R-loop end position was expressed as a discrete interval (set) of numbers. These numbers were merged into one set of indices representing every nucleic base occurring in the detected R-loops. We then compared those positions across all methods and we present them here in graphic form. This method of comparison does not take into account the R-loop direction, which may cause irregularities when comparing *in vitro* results with *in silico* results. Comparison of analysis results was made using the first 6000 bp in the sequence of the human gene NEAT1 (Figure 2). Within this 6000-bp sequence, the R-loop tracker detected 20 R-loops in three main regions. Comparison with available data sets from the R-loop DB [25] and RDIP (RNA:DNA immunoprecipitation) [38] data sets downloaded from Genome Browser showed that the results of the R-loop tracker were identical to those of the QmRLFS mapper, thus confirming the correctness of the algorithm. The overlay with the Fibroblast RDIP-seq data set was 32% and was located in similar regions of this sequence. Interestingly, the comparison with the G4Hunter results (G4Hunter analysis parameters—windows 25, G4Hunter score > 1) showed a 38% overlap of R-loops with G-prone sequences.



**Figure 2.** Example of an *R-loop tracker* comparison with QmRLFS finder algorithm on human chromosome 8.

#### 4.2. Validation

To validate our implementation, we decided to compare the output of an algorithmic detection with the experimental sequencing method DRIPc [35]. The choice was made due to the availability of the source data in *Genome Browser*, which can be accessed via the API. To validate the effectiveness of the tool, we defined the following metrics for result comparison.

- **TP** (true positive)—at least one R-loop was detected both by DRIPc sequencing and the R-loop tracker algorithm.
- **TN** (true negative)—not a single R-loop was detected with the experimental method nor with the R-loop tracker algorithm.
- **FP** (false positive)—DRIPc sequencing did not detect any R-loop in a given area, but the R-loop tracker found at least one R-loop in a given area.

- **FN** (false negative)—at least one R-loop was detected by DRIPc sequencing but none was found by the R-loop tracker in a given area.

The choice of gene sequences for the comparison was based on previously published fourteen verified genes containing R-loops [25]. The source data and the comparison script are available on GitHub. We measured the following metrics to validate the data for the R-loop tracker:

1. Accuracy
2. Sensitivity
3. Specificity
4. Precision
5. Matthews Correlation Coefficient

We have compared the results for both positive and negative strands, and the results differed by the false positive/false negative ratio (Table 1).

**Table 1.** Standard performance measure for both strands.

	Positive Strand	Negative Strand
<i>Accuracy [%]</i>	78.57	64.29
<i>Sensitivity [%]</i>	25	40
<i>Specificity [%]</i>	100	77.78
<i>Precision [%]</i>	76.92	70
<i>Matthews Correlation Coefficient</i>	0.44	0.19

R-loop tracker showed 71.4% in accuracy, 88.9% in sensitivity, and 75% positive Matthews Correlation Coefficient. The validation dataset is available in Supplementary Material online at [https://github.com/jan-havlik/validation\\_dataset](https://github.com/jan-havlik/validation_dataset) (accessed on 28 October 2021).

#### 4.3. R-Loop Tracker Effectivity

The effectivity of the R-loop tracker tool was measured by comparing the processing speed of each analysis (Figure 3). The input sequence was split into smaller batches of the following size:

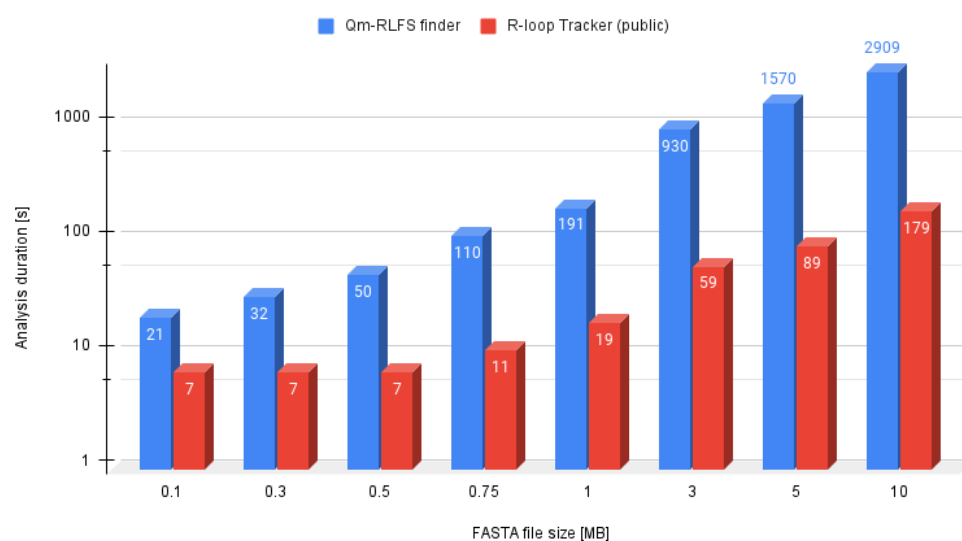
- 100 kB
- 300 kB
- 500 kB
- 750 kB
- 1 MB
- 3 MB
- 5 MB
- 10 MB

Because the QmRLFS finder website only allows files of maximum size of 300 kB, we were using the available command line tool written in Python 2.7.

As we can see in the figure, R-loop tracker was faster compared to QmRLFS finder. This was also caused by greater computational resources of the standalone server which is publicly accessible to everyone. The time difference was taken using two different approaches:

- Linux *time* utility measuring the script run time
- Time difference calculation from web server logfile

The data set for comparison is available at <https://github.com/jan-havlik/comparison> (accessed on 28 October 2021).



**Figure 3.** R-loop tracker web server effectivity compared to that of the QmRLFS finder tool.

**Supplementary Materials:** The validation dataset is available online at [https://github.com/jan-havlik/validation\\_dataset](https://github.com/jan-havlik/validation_dataset).

**Author Contributions:** V.B. was responsible for conceptualization, formal analyses, and project administration. J.H. was responsible for methodology, back-end development, data curation, and visualization. J.K. developed the front end. V.B., O.T. and J.Š. acquired the funding and provided overall supervision. J.H. and V.B. wrote the initial draft. All coauthors participated in writing, reviewing, and editing the text. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a project of Brno University of Technology FME-S-20-6538.

**Data Availability Statement:** The tool is available at: <https://bioinformatics.ibp.cz/> (accessed on 28 October 2021), the data are available at: <https://github.com/jan-havlik/comparison> (accessed on 28 October 2021).

**Acknowledgments:** We thank Jean-Louis Mergny for a consultation and providing motivation to develop this tool.

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Watson, J.D.; Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)]
2. Brázda, V.; Laister, R.C.; Jagelská, E.B.; Arrowsmith, C. Cruciform Structures Are a Common DNA Feature Important for Regulating Biological Processes. *BMC Mol. Biol.* **2011**, *12*, 33. [[CrossRef](#)] [[PubMed](#)]
3. Gentry, M.; Hennig, L. A Structural Bisulfite Assay to Identify DNA Cruciforms. *Mol. Plant* **2016**, *9*, 1328–1336. [[CrossRef](#)] [[PubMed](#)]
4. Rich, A.; Zhang, S. Timeline: Z-DNA: The Long Road to Biological Function. *Nat. Rev. Genet.* **2003**, *4*, 566–572. [[CrossRef](#)]
5. Li, H.; Xiao, J.; Li, J.; Lu, L.; Feng, S.; Droge, P. Human Genomic Z-DNA Segments Probed by the Z Domain of ADAR1. *Nucleic Acids Res.* **2009**, *37*, 2737–2746. [[CrossRef](#)] [[PubMed](#)]
6. Jain, A.; Rajeswari, M.R.; Ahmed, F. Formation and Thermodynamic Stability of Intermolecular (R\*R Center Dot Y) DNA Triplex in GAA/TTC Repeats Associated with Friedreich's Ataxia. *J. Biomol. Struct. Dyn.* **2002**, *19*, 691–699. [[CrossRef](#)] [[PubMed](#)]
7. Lee, H.-T.; Khutsishvili, I.; Marky, L.A. DNA Complexes Containing Joined Triplex and Duplex Motifs: Melting Behavior of Intramolecular and Bimolecular Complexes with Similar Sequences. *J. Phys. Chem. B* **2010**, *114*, 541–548. [[CrossRef](#)] [[PubMed](#)]
8. Huppert, J.L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [[CrossRef](#)]
9. Lam, E.Y.N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat. Commun.* **2013**, *4*, 1796. [[CrossRef](#)]
10. Kamura, T.; Katsuda, Y.; Kitamura, Y.; Ihara, T. G-Quadruplexes in MRNA: A Key Structure for Biological Function. *Biochem. Biophys. Res. Commun.* **2020**, *526*, 261–266. [[CrossRef](#)]

11. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [[CrossRef](#)] [[PubMed](#)]
12. Brázda, V.; Coufal, J. Recognition of Local DNA Structures by P53 Protein. *Int J Mol Sci* **2017**, *18*, 375. [[CrossRef](#)] [[PubMed](#)]
13. Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24*, 1711. [[CrossRef](#)]
14. Pan, X.; Jiang, N.; Chen, X.; Zhou, X.; Ding, L.; Duan, F. R-Loop Structure: The Formation and the Effects on Genomic Stability. *Yi Chuan Hered.* **2014**, *36*, 1185–1194. [[CrossRef](#)]
15. Groh, M.; Lufino, M.M.P.; Wade-Martins, R.; Gromak, N. R-Loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genet.* **2014**, *10*, e1004318. [[CrossRef](#)]
16. Richard, P.; Manley, J.L. R Loops and Links to Human Disease. *J. Mol. Biol.* **2017**, *429*, 3168–3180. [[CrossRef](#)]
17. Cristini, A.; Gromak, N.; Sordet, O. Transcription-Dependent DNA Double-Strand Breaks and Human Disease. *Mol. Cell. Oncol.* **2020**, *7*, 1691905. [[CrossRef](#)] [[PubMed](#)]
18. Chasovskikh, S.; Dimtchev, A.; Smulson, M.; Dritschilo, A. DNA Transitions Induced by Binding of PARP-1 to Cruciform Structures in Supercoiled Plasmids. *Cytom. Part J. Int. Soc. Anal. Cytol.* **2005**, *68*, 21–27. [[CrossRef](#)] [[PubMed](#)]
19. Shen, X.; Mizuguchi, G.; Hamiche, A.; Wu, C. A Chromatin Remodelling Complex Involved in Transcription and DNA Processing. *Nature* **2000**, *406*, 541–544. [[CrossRef](#)] [[PubMed](#)]
20. Chakraborty, P. New Insight into the Biology of R-Loops. *Mutat. Res.* **2020**, *821*, 111711. [[CrossRef](#)]
21. Cer, R.; Bruce, K.; Donohue, D.; Temiz, N.; Mudunuri, U.; Yi, M.; Volfovsky, N.; Bacolla, A.; Luke, B.; Collins, J.R.; et al. Searching for Non-B DNA-Forming Motifs Using NBMST (Non-B DNA Motif Search Tool). *Curr. Protoc. Hum. Genet.* **2012**, *73*, 18.7.1–18.7.22. [[CrossRef](#)] [[PubMed](#)]
22. Brázda, V.; Kolomazník, J.; Lýsek, J.; Hároníková, L.; Coufal, J.; Šťastný, J. Palindrome Analyser—A New Web-Based Server for Predicting and Evaluating Inverted Repeats in Nucleotide Sequences. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 1739–1745. [[CrossRef](#)] [[PubMed](#)]
23. Brázda, V.; Kolomazník, J.; Lýsek, J.; Bartas, M.; Fojta, M.; Šťastný, J.; Mergny, J.-L. G4Hunter Web Application: A Web Server for G-Quadruplex Prediction. *Bioinformatics* **2019**, *35*, 3493–3495. [[CrossRef](#)]
24. Puig Lombardi, E.; Londoño-Vallejo, A. A Guide to Computational Methods for G-Quadruplex Prediction. *Nucleic Acids Res.* **2019**, *48*, 1603. [[CrossRef](#)]
25. Jenjaroenpun, P.; Wongsurawat, T.; Yenamandra, S.P.; Kuznetsov, V.A. QmRLFS-Finder: A Model, Web Server and Stand-Alone Tool for Prediction and Analysis of R-Loop Forming Sequences. *Nucleic Acids Res.* **2015**, *43*, W527–W534. [[CrossRef](#)] [[PubMed](#)]
26. Roy, D.; Lieber, M.R. G Clustering Is Important for the Initiation of Transcription-Induced R-Loops in Vitro, Whereas High G Density without Clustering Is Sufficient Thereafter. *Mol. Cell. Biol.* **2009**, *29*, 3124–3133. [[CrossRef](#)]
27. Haeussler, M.; Zweig, A.S.; Tyner, C.; Speir, M.L.; Rosenbloom, K.R.; Raney, B.J.; Lee, C.M.; Lee, B.T.; Hinrichs, A.S.; Gonzalez, J.N.; et al. The UCSC Genome Browser Database: 2019 Update. *Nucleic Acids Res.* **2019**, *47*, D853–D858. [[CrossRef](#)]
28. Kim, A.; Wang, G.G. R-Loop and Its Functions at the Regulatory Interfaces between Transcription and (Epi)Genome. *Biochim. Biophys. Acta BBA Gene Regul. Mech.* **2021**, *1864*, 194750. [[CrossRef](#)]
29. Brambati, A.; Zardoni, L.; Nardini, E.; Pellicoli, A.; Liberi, G. The Dark Side of RNA:DNA Hybrids. *Mutat. Res.* **2020**, *784*, 108300. [[CrossRef](#)] [[PubMed](#)]
30. Ui, A.; Chiba, N.; Yasui, A. Relationship among DNA Double-Strand Break (DSB), DSB Repair, and Transcription Prevents Genome Instability and Cancer. *Cancer Sci.* **2020**, *111*, 1443–1451. [[CrossRef](#)] [[PubMed](#)]
31. Yan, P.; Liu, Z.; Song, M.; Wu, Z.; Xu, W.; Li, K.; Ji, Q.; Wang, S.; Liu, X.; Yan, K.; et al. Genome-Wide R-Loop Landscapes during Cell Differentiation and Reprogramming. *Cell Rep.* **2020**, *32*, 107870. [[CrossRef](#)]
32. Wang, K.; Wang, H.; Li, C.; Yin, Z.; Xiao, R.; Li, Q.; Xiang, Y.; Wang, W.; Huang, J.; Chen, L.; et al. Genomic Profiling of Native R Loops with a DNA-RNA Hybrid Recognition Sensor. *Sci. Adv.* **2017**, *7*, eabe3516. [[CrossRef](#)] [[PubMed](#)]
33. Sanz, L.A.; Castillo-Guzman, D.; Chédin, F. Mapping R-Loops and RNA:DNA Hybrids with S9.6-Based Immunoprecipitation Methods. *JoVE J. Vis. Exp.* **2021**, *174*, e62455. [[CrossRef](#)] [[PubMed](#)]
34. Guo, M.S.; Kawamura, R.; Littlehale, M.L.; Marko, J.F.; Laub, M.T. High-Resolution, Genome-Wide Mapping of Positive Supercoiling in Chromosomes. *eLife* **2021**, *10*, e67236. [[CrossRef](#)] [[PubMed](#)]
35. Sanz, L.A.; Chédin, F. High-Resolution, Strand-Specific R-Loop Mapping via S9.6-Based DNA-RNA Immunoprecipitation and High-Throughput Sequencing. *Nat. Protoc.* **2019**, *14*, 1734–1755. [[CrossRef](#)] [[PubMed](#)]
36. Russo, M.; De Lucca, B.; Flati, T.; Gioiosa, S.; Chillemi, G.; Capranico, G. DROPA: DRIP-Seq Optimized Peak Annotator. *BMC Bioinformatics* **2019**, *20*, 414. [[CrossRef](#)]
37. Zhang, P.; Feng, Y.; Wei, H.; Zhang, W. R-Loop Identification and Profiling in Plants. *Trends Plant Sci.* **2019**, *24*, 971–972. [[CrossRef](#)]
38. Nadel, J.; Athanasiadou, R.; Lemetre, C.; Wijetunga, N.A.; Broin, Ó.P.; Sato, H.; Zhang, Z.; Jeddeloh, J.; Montagna, C.; Golden, A.; et al. RNA:DNA Hybrids in the Human Genome Have Distinctive Nucleotide Characteristics, Chromatin Composition, and Transcriptional Relationships. *Epigenet. Chromatin* **2015**, *8*, 46. [[CrossRef](#)] [[PubMed](#)]