doi 10.18699/bgrs2024-1.4-21

Roles of the transcriptional R-loop forming sequences (RLFS) and RLFS(+) R-loops in normal and cancer cell states. Critical review

Kapali O., Adala J., Kuznetsov V.A.
Department of Urology, Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, USA

Key words: R-loops; R-loop forming sequence; transcription regulation; cancer; regulatory sites

Motivation and Aim: Transcriptional R-loops are dynamical three-stranded guanine-rich RNA-DNA hybrid structures in which a nascent RNA is hybridized to a template DNA strand, with the other (non-template) DNA strand looped out. We have proposed that such single-stranded DNA sequences determine the initiation and stabilization of R-loops genome-wide and found that they are ubiquitous in eukaryotic genomes [1–4]. We called a family of these sequences the R-loop forming sequences (RLFS).

We have introduced the Quantitative models of R-Loop Forming Sequences (QmRLFS) tools [1, 2, 4]. As per QmRLFS model, RLFS is a non-template guanine-rich sequence consisting of three zones: R-loop initiation zone (RIZ, includes few G repeats), linker sequence, and R-loop elongation zone (REZ, high rich G sequence). QmRLFS identifies the sizes, *positions, and boundaries* of RLFS with high specificity and sensitivity. The model predicts the size and boundaries of the R-loop at an accuracy of 86–92 % in vitro [1, 2, 4, 5] providing objective controls and improved R-loop detection methods [6, 7]. We have discovered that in the human genome RLFSs are present in the vicinity of transcription regulatory sites (Transcription Start Site (TSS), Transcription Termination Site (TTS), Intron-Exon junctions, transcribed enhancers, etc.) of more than 75 % of protein-coding genes, about 25 % of non-coding transcribed loci and they are highly enriched with G-quadruplexes [3, 4]. We have classified RLFS regulatory functions genome-wide in humans and other eukaryotes [3–5]. About 90 % of the predicted RLFS(+) merged regions formed R-loops/RNA:DNA hybrids were supported by genome-wide data [2–4]. Our model and facts have been highlighted in recent reviews and used in databases and web tools [7–9]. Over 1,000 proteins have been reported to interact with RNA:DNA hybrids/R-loops, most of which are RLFS-positive, suggesting the existence of a robust RNA:DNA hybrid-protein interactome [10, 11]. These findings have suggested multiple structural and functional roles of the R-loops/RNA:DNA hybrids in cell physiology, stress response/adaptation, and pathobiology, including cancer [12]. In vitro and cell line model studies have shown that R-loops/RNA:DNAs can provide essential roles in genotoxic stress, genome instability, transcription alterations, pro-oncogenic mutagenesis, and links with several other cancer hallmarks. It has been proposed that R-loops may be relevant to oncogenesis and aggressive cancer phenotypes [1–4, 12]. Despite recent discoveries in R-loop biology, our current knowledge of the effects of R-loops and RLFS in the formation and stability of RNA:DNA hybrids/R-loops in cancer is poorly understood. Lack of

knowledge about R-loop initiation, formation, and regulation in cancer biology and limitations of detection methods prevent clinical studies and evaluation of R-loops and their subsets' roles in cancer.

The objective of our review is to consider the computationally predicted RLFSs and R-loop biology studies in the context of improvement of quality control, accuracy, reproducibility, and resolution experimental methods. We also consider the perspective of integrative computational biology, bioinformatics and functional genomics, single-cell and image analyses of R-loops, and mathematical modeling to move forward to cancer R-loop pathobiology and clinical oncology needs.

Consideration: Perspectives of RLFS, RNA-DNA hybrids and R-loop detection genome-wide technologies, big data collection and analyses. R-loop detection technologies, reproducibility, standards, quality control, specificity, and sensitivity, analytical tools and computational methods, and databases. Development of biological models, and the integrative databases of comprehensive clinical data study. Roles of RLFS-defined RNA:DNA hybrids in the pathogenesis, cells and clinical samples heterogeneity, diagnostic and prognosis of cancer using breast cancer data analysis as our analysis example.

Conclusion: The RLFS subset due to their sequence uniqueness and commonness in transcription regulatory sites, as well as the high accuracy of the computationally predicted RLFS and corresponding R-loop positions, may be defined as novel structural and functional determinants that are essential to cell physiology and genetic disease (including cancer). After adopting them for clinical practice and their detection, such structures may be useful targets in cancer diagnostics, personalized treatment, and predictors.

References

1. Wongsurawat T. et al. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.* 2012;40(2):e16
2. Jenjaroenpun P. et al. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res*. 2015;43(W1):W527-W534
3. Jenjaroenpun P. et al. R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. *Nucleic Acids Res*. 2017;45(D1):D119-D127
4. Kuznetsov V.A. et al. Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res*. 2018;46(15):7566-7585
5. Yeo A.J. et al. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS One*. 2014;9(3):e90219
6. Safari M. et al. R-loop-mediated ssDNA breaks accumulate following short-term exposure to the HDAC inhibitor romidepsin. *Mol Cancer Res*. 2021;19(8):1361-1374
7. Lin R. et al. R-loopBase: a knowledgebase for genome-wide R-loop formation and regulation. *Nucleic Acids Res*. 2022;50(D1):D303-D315.
8. Miller H.E. et al. Exploration and analysis of R-loop mapping data with RLBase. *Nucleic Acids Res.* 2023;51(D1):D1129-D1137
9. Shi X., Teng H., Sun Z. An updated overview of experimental and computational approaches to identify non-canonical DNA/RNA structures with emphasis on G-quadruplexes and R-loops. *Brief Bioinform*. 2022;23(6):bbac441. doi 10.1093/bib/bbac441
10. Wang I.X. et al. Human proteins that interact with RNA/DNA hybrids. *Genome Res.* 2018;28(9):1405-1414
11. Cristini A. et al., RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep*. 2018;23(6):1891-1905
12. Gaillard H., García-Muse T., Aguilera A. Replication stress and cancer. *Nat Rev Cancer*. 2015;15(5):276-89. doi 10.1038/nrc3916