

R-loop biology: from a few gene cases to genome scale

Vladimir A. Kuznetsov*, Piroon Jenjaroenpun, Thidathip Wongsurawat

Bioinformatics Institute/A*STAR, Singapore

*Corresponding author

R-loops, which are triple-stranded RNA-DNA hybrid structures, can often occur in the human genome and play crucial roles in many normal biological processes. Such RNA-DNA hybrids could initiate mutations, DNA breaks, genome instability and diseases. However, until 2011 only a few cases of the R-loop formation have been experimentally documented, indicating the roles of R-looping in gene functions. The R-loops, involving in transcription through switch recombination regions at immunoglobulin heavy chain loci in a genome of mammalian B cells, were the well-studied examples. In 2011, we have developed our data-driven quantitative model of RLFS (QmRLFS), which easily demonstrated strong co-localization of predicted RLFS with most genic regions in the human genome and the genome regions associated with open chromatin, promoters and others gene expression control signals, transcript isoforms, splicing, triggering mutation and DNA break loci, fragile and critical disease regions (Wongsurawat et al, 2011). We found that many oncogenes, tumor suppressors and neurodegenerative diseases could be prone to significant R-loop formation. These predictions have been confirmed with several experimental systems and methods including DRIP-qPCR (Yeo et al, 2013, Ginno et al, 2012), DRIP-seq methods (Ginno et al, 2012, Ginno et al, 2013).

The accurate computational prediction (83-92%; Jenjaroenpun et al, 2015) and experimental genome mapping of RLFSs has opened up intriguing possibilities for the studies of RNA-DNA interactome complexity in vivo and R-loop's use targets for diagnostics and treatment of many diseases. Here we review the current knowledge about the mechanisms controlling R-loop formation, methods of experimental R-loop detection, and computational models of R-loop forming sequences at genic and genome-wide scales. Finally, we discuss the observed and putative relationships of R-loops with several basic biological mechanisms, evolution of RLFS motifs and medical conditions including that of cancer, autoimmune and neurodegenerative diseases.