Cellular Type Is a Major Determinant of R-Loop Genomic Distribution

K. Yu. Oleynikova[1,2], N. A. Zhigalova[1], A. P. Hutchins[3], A. S. Ruzov[1*]
[1]Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, 117312 Russia
[2]Infochemistry Scientific Center, ITMO University, St. Petersburg, 197101 Russia
[3]Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, 518055 China
*E-mail: alexey.ruzov@gmail.com
Received September 26, 2025; in final form, January 27, 2026
DOI: 10.32607/actanaturae.27833
Copyright © 2026 National Research University Higher School of Economics. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT **R-loops that contain a DNA:RNA hybrid and unpaired single-stranded DNA are important determinants of normal cell physiology and of the pathogenesis of numerous diseases. Although several different approaches to R-loop mapping in the genome have been developed, these techniques can produce conflicting results. In order to assess their robustness, a recent study by Chedin et al. compared the R-loop genomic distribution assessed using different methods in normal and cancer cell lines. Importantly, that study assumed a high degree of similarity between R-loop genomic distributions across different cellular types. Here, we compared DRIP datasets produced using the same protocol in different cell lines to show that only 26% of R-loop peaks are shared between chronic myeloid leukemia-derived HAP1 cells and human pluripotent stem cells. Meanwhile, HAP1-derived double knockout cell lines are characterized by much higher fractions of R-loop peaks that are identical both to each other (most of them) and to the R-loop peaks of their parental line (71 and 55%). We conclude that cellular type represents a major determinant of R-loop genomic distribution and, therefore, that only a systematic comparison of a large array of various cell/tissue type-derived R-loop datasets may address the inconsistencies between different R-loop mapping techniques.**
KEYWORDS **R-loops, DNA:RNA hybrids, DNA-RNA Immunoprecipitation (DRIP), cellular differentiation, hPSCs, HAP1 cells.**
ABBREVIATIONS **DRIP – DNA–RNA immunoprecipitation; HNRNPA2B1 – heterogeneous nuclear ribonucleoprotein A2/B1; hPSC – human pluripotent stem cells.**

INTRODUCTION

R-loops that contain a DNA:RNA hybrid and unpaired single-stranded DNA are abundant in the genome and can be involved in the regulation of a broad range of biological processes, such as transcription termination, DNA repair, telomere homeostasis, and immunoglobulin class-switch recombination [1]. Meanwhile, irregular or pathological R-loops can disrupt transcription and replication, causing accumulation of DNA double-strand breaks and thus becoming a major source of genetic stress and instability in mammalian cells [1]. Taking into account the association between genomic instability and oncogenic transformation, the interest in the regulation of the R-loop distribution across various systems has led to the development of experimental methods for genome-wide R-loop mapping [2]. Some of these methods are based on antibodies specific to RNA:DNA hybrids (the S9.6 antibodies) [3]. These techniques include DNA:RNA immunoprecipitation (DRIP), with its numerous variants (DRIPc-seq; ssDRIP-seq, etc.), and R-loop cleavage under targets & tagmentation [4, 5]. Other methods are based on either mapping single-stranded DNA within R-loops using bisulfite [6] or the application of RNase H1, an enzyme responsible for the recognition and cleavage of RNA:DNA hybrids [7]. Importantly, R-loop sets obtained using different (and sometimes similar) methods on different cell lines can often be substantially different [8].

In order to resolve these contradictions, Chedin et al. compared the datasets on R-loop genomic distribution obtained for various cell lines using different methods and assessed their degree of similarity [8]. That study does not fully consider the cellular

types used to obtain the datasets. The implicit assumption was that there exists a fundamental similarity in R-loop distribution in housekeeping genes across different cellular types. Nonetheless, the validity of this assumption is unclear. Chedin et al. neither evaluated nor discussed the potential extent of differences in the R-loop genomic distribution in various cellular systems such as human pluripotent stem cells (hPSCs) and tumor cell lines (U2OS or HeLa). Instead, they compared the signal coverage profiles from corresponding DRIP experiments performed using different cellular types to identify the datasets considered "discordant" [8]. We believe that this approach, which implicitly presumes similarity in the R-loop genomic distribution across cell lines of different origins, has the potential to distort the interpretation of experimental data.

EXPERIMENTAL

Cell lines and cell culture, DRIP, and high-throughput sequencing library preparation

HNRNPA2B1 KO and YTHDF2 KO cells, as well as isogenic parental wild-type HAP1 cells (Horizon Discovery, # HZGHC007378c010, # HZGHC006678c001, and # C631), were cultured in the DMEM/F12 medium (Gibco Life Technologies, USA # 11320033). DNA–RNA immunoprecipitation (DRIP) was conducted in compliance with the previously published procedure, including control over DRIP signal specificity by treating samples with RNase H [9]. The genomic libraries were constructed using a NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, USA # E7645) according to the manufacturer's protocol.

Bioinformatics analysis of the DRIP-seq data

Datasets on HAP1 (WT, HNRNPA2B1 KO and YTHDF2 KO) and the previously published dataset on hPSC (PRJNA474076) were analyzed simultaneously. The peaks were identified using a MACS2 peak caller. Broad peaks were detected using the following parameters: –format BAM -g hs –keep-dup all. The consensus peaks were identified using the bedtools intersect tool with the standard parameters to ensure minimal input peak intersection (-f 1E-9 -F 1E-9). A detailed description of bioinformatics analysis is available online at https://github.com/katerinaoleynikova/human_samples_paper_25.

RESULTS AND DISCUSSION

In order to assess the degree of similarity in the R-loop genomic distribution across cells of different origin, we compared DRIP datasets generated using our previously published protocol [9] in three different cellular models: human pluripotent stem cells (hPSCs), wild-type HAP1 (HAP1 WT) cells derived from the KBM-7 chronic myeloid leukemia cell line [10], and two isogenic HAP1 cell lines with *YTHDF2* or *HNRNPA2B1* gene knockout generated via gene editing (HAP1 YTHDF2 KO and HAP1 2B1 KO cell lines, respectively). The genomic regions enriched in R-loops in these cell lines were identified by DRIP-sequencing; the reads were mapped to the human genome and R-loop peaks were identified. Next, consensus peaks for each sample were identified by comparing two corresponding replicates. We demonstrated that, although a similar number of consensus peaks were identified for hPSCs and wild-type HAP1 cells (4,482 and 4,458, respectively), only 1,187 (~26%) peaks were common to both cell lines (*Fig. 1A*). Interestingly, while *YTHDF2* and *HNRNPA2B1* gene knockout substantially reduced the total number of peaks (1,496 and 950, respectively) compared to wild-type HAP1 cells, the HAP1 YTHDF2 KO and HAP1 2B1 KO datasets were similar to each other. The vast majority of peaks in HAP1 2B1 KO cells (873 out of 950) overlapped with those identified in HAP1 YTHDF2 KO cells (*Fig. 1B*). Furthermore, 71% (681) of the peaks identified in HAP1 2B1 KO cells and 55% (864) of the peaks in HAP1 YTHDF2 KO cells were identical to the peaks of R-loops detected in their parental HAP1 WT cell line (*Fig. 1B*). The genomic region surrounding the *RPL13A* housekeeping gene, which was used by Chedin et al. for dataset comparison and called "the gold-standard region," contained only peaks from wild-type HAP1 cells, but not the peaks from R-loops derived from hPSCs or the two other knockout cell lines tested in our study (*Fig. 1C*). We also identified other loci where R-loop peaks were present across all four tested cell lines (*Fig. 1D*, left panel). However, we believe that in order to refer to any genomic regions as "gold-standard" ones, datasets from additional cell lines of diverse origins need to be obtained, since the R-loop genomic distribution appears to be highly cellular-type specific. Hence, we demonstrated that R-loop sets obtained in cell lines of different origins using the same method differ substantially and that only 25% of R-loop peaks coincide between hPSCs and the HAP1 line derived from chronic myeloid leukemia cells. Meanwhile, the R-loop distribution in *YTHDF2* and *HNRNPA2B1* knockout HAP1 cells and in wild-type HAP1 cells is apparently more similar for these two cellular types compared to their distribution in hPSCs. Since the recently identified factors involved in R-loop regulation include the chromatin structure [10], nucleosome positioning [11], and RNA modifications [9], all varying significantly across cellular types, these findings are not particularly un-

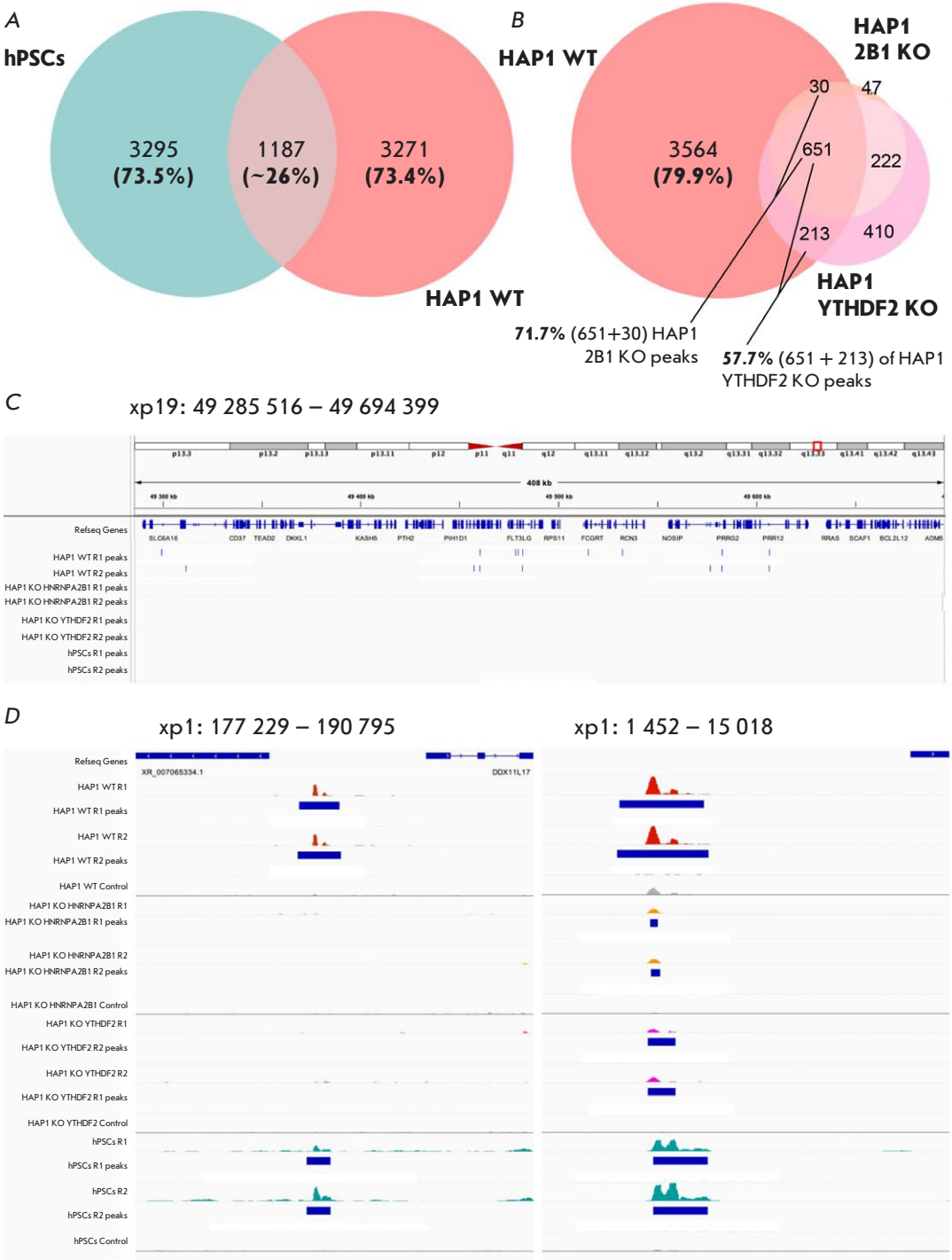


Fig. 1. Cellular type is a major determinant of the R-loop genomic distribution. (*A*) Venn diagrams showing the overlaps between the R-loop peak datasets obtained using DRIP experiments in hPSC and wild-type HAP1 (HAP1 WT) cells, as well as (*B*) wild-type HAP1cells and HAP1 cells with genetic knockouts of the *YTHDF2* (HAP1 YTHDF2 KO) and *HNRNPA2B1* (HAP1 2B1 KO) genes. The numbers of R-loop peaks in each category are indicated on the diagrams. (*C*) Genome browser view of the distribution of R-loop peaks (shown as blue vertical dashes) in the datasets generated in our DRIP experiments on the aforementioned cell lines over the region centered around the *RPL13A* housekeeping gene used by Chedin et al. as the "gold standard" for dataset comparison. (*D*) Genome browser views of the signal profiles of our R-loop DRIP datasets alongside the control input samples at two representative genomic loci. The locations of the R-loop peaks are shown with blue rectangles

expected [12]. Importantly, the interpretation of our results and the conclusions derived from them are confined to the selected models. Therefore, datasets from additional differentiated and tumor cell lines of different histogeneses need to be analyzed to ensure a more robust generalization.

CONCLUSIONS

Hence, we infer that the R-loop genomic distribution is cellular-type specific. Therefore, only the systematic and standardized comparison of a substantial array of R-loop genomic distribution datasets derived from different cellular types can resolve the discrepancies between the experimental results obtained using different R-loop mapping techniques. ●

This work was supported by the Russian Science Foundation (grant No. 22-65-00022).

The work conducted by K.Yu. Oleynikova involving the analysis of hPSC dataset was supported by the Ministry of Science and Higher Education of the Russian Federation. The authors have no conflicts of interest to declare.

REFERENCES

1. Aguilera P, Aguilera A. R-loop homeostasis in genome dynamics, gene expression and development. *Curr Opin Genet Dev.* 2025;92:102325. doi: 10.1016/j.gde.2025.102325
2. Yadav C, Yadav R, Nanda S, Ranga S, Ahuja P. The hidden architects of the genome: a comprehensive review of R-loops. *Mol Biol Rep.* 2024;51(1):1095. doi: 10.1007/s11033-024-10025-6
3. Boguslawski SJ, Smith DE, Michalak MA, et al. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods.* 1986;89(1):123-130. doi: 10.1016/0022-1759(86)90040-2
4. García-Rubio M, Soler-Oliva ME, Aguilera A. Genome-Wide Analysis of DNA–RNA Hybrids in Yeast by DRIPc-Seq and DRIP-Seq. *Methods Mol Biol.* 2022;2528:429-443. doi: 10.1007/978-1-0716-2477-7_28
5. Wang H, Li C, Liang K. Genome-wide native R-loop profiling by R-loop cleavage under targets and tagmentation (R-Loop CUT&Tag). *Methods Mol Biol.* 2022;2528:345-357. doi: 10.1007/978-1-0716-2477-7_23
6. Malig M, Hartono SR, Giafaglione JM, Sanz LA, Chedin F. Ultra-deep coverage single-molecule R-loop footprinting reveals principles of R-loop formation. *J Mol Biol.* 2020;432(7):2271-2288. doi: 10.1016/j.jmb.2020.02.014
7. Cerritelli SM, Sakhuja K, Crouch RJ. RNase H1, the gold standard for R-loop detection. *Methods Mol Biol.* 2022;2528:91-114. doi: 10.1007/978-1-0716-2477-7_7
8. Chédin F, Hartono SR, Sanz LA, Vanoosthuyse V. Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J.* 2021;40(4):e106394. doi: 10.15252/embj.2020106394
9. Abakir A, Giles TC, Cristini A, et al. N6-methyladenosine regulates the stability of RNA: DNA hybrids in human cells. *Nat Genet.* 2020;52(1):48-55. doi: 10.1038/s41588-019-0549-x
10. Bayona-Feliu A, Barroso S, Muñoz S, Aguilera A. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription–replication conflicts. *Nat Genet.* 2021;53(7):1050-1063. doi: 10.1038/s41588-021-00867-2
11. Werner M, Trauner M, Schauer T, et al. Transcription-replication conflicts drive R-loop-dependent nucleosome eviction and require DOT1L activity for transcription recovery. *Nucleic Acids Res.* 2025;53(4):gkaf109. doi: 10.1093/nar/gkaf109
12. Carette JE, Guimaraes CP, Varadarajan M, et al. Haploid genetic screens in human cells identify host factors used by pathogens. *Science.* 2009;326(5957):1231-1235. doi: 10.1126/science.1178955