RedChIP identifies noncoding RNAs associated with genomic sites occupied by Polycomb and CTCF proteins

Alexey A. Gavrilov⁶, Rinat I. Sultanov⁷, Mikhail D. Magnitov⁴, Aleksandra A. Galitsyna⁵, Erdem B. Dashinimaaev⁶, Erez Lieberman Aiden⁶, and Sergey V. Razin⁴,⁵,¹

*Institute of Gene Biology, Russian Academy of Sciences 119334 Moscow, Russia; ²Federal Research and Clinical Center of Physical-Chemical Medicine, Federal Medical Biological Agency 119435 Moscow, Russia; ³Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow 119997, Russia; ⁴The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; ⁵Center for Theoretical Biological Physics, Department of Computer Science, Rice University, Houston, TX 77030; and ⁶Faculty of Biology, Lomonosov Moscow State University, Moscow 119991, Russia

Edited by Alberto Kornblihtt, Departamento de Fisiologia, Biologia Molecular y Celular, IFByNE-UBA-CONICET, Universidad de Buenos Aires, Buenos Aires, Argentina; received September 15, 2021; accepted December 8, 2021

Nuclear noncoding RNAs (ncRNAs) are key regulators of gene expression and chromatin organization. The progress in studying nuclear ncRNAs depends on the ability to identify the genome-wide spectrum of contacts of ncRNAs with chromatin. To address this question, a panel of RNA–DNA proximity ligation techniques has been developed. However, neither of these techniques examines proteins involved in RNA–chromatin interactions. Here, we introduce RedChIP, a technique combining RNA–DNA proximity ligation and chromatin immunoprecipitation for identifying RNA–chromatin interactions mediated by a particular protein. Using antibodies against architectural protein CTCF and the EZH2 subunit of the Polycomb repressive complex 2, we identify a spectrum of cis- and trans-acting ncRNAs enriched at Polycomb- and CTCF-binding sites in human cells, which may be involved in Polycomb-mediated gene repression and CTCF-dependent chromatin looping. By providing a protein-centric view of RNA–DNA interactions, RedChIP represents an important tool for studies of nuclear ncRNAs.

Author contributions: A.A. Gavrilov, E.L.A., and S.V.R. designed research; A.A. Gavrilov and E.B.D. performed research; A.A. Gavrilov, R.I.S., M.D.M., and A.A. Galitsyna analyzed data; and A.A. Gavrilov and S.V.R. wrote the paper.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: sergey.v.razin@usa.net.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116222119/-/DCSupplemental.

Published December 30, 2021.

PNAS 2022 Vol. 119 No. 1 e2116222119
https://doi.org/10.1073/pnas.2116222119 | 1 of 3
gene) and compared the number of cis contacts between EZH2-precipitated and input fractions. We found that the degree of enrichment in the IP fraction correlated with the percentage of contacts detected in Polycomb-specific and poised promoter-specific chromatin types but not any other chromatin types in the area under study (Fig. 2A). We identified 10 long intergenic ncRNAs (lincRNAs) with a fold enrichment of >1.3 in both replicates (Fig. 2C). Among the identified RNAs is Kcnq1ot1 (fold change = 1.5), a well-known example of antisense lincRNA involved in the Polycomb-mediated silencing of several genes in the same locus (7). High fold enrichment was also observed for AC078778, antisense lincRNAs from the HOXC locus. Notably, antisense RNAs, on average, demonstrate higher fold enrichment than other lincRNAs and mRNAs of protein-coding genes (Fig. 2G), indicating the enrichment of the group of antisense RNAs with RNAs that may mediate Polycomb targeting.

At the final step of the analysis, we searched for trans-acting RNAs that could participate in Polycomb functioning genome-wide. We compared the number of trans contacts (contacts with nonparental chromosomes) for each RNA between EZH2-precipitated and input fractions and looked for RNAs showing an elevated number of contacts in the IP fraction. The highest enrichment was observed for antisense RNA KIF5C-AS1, snRNA RNU5B-1, and SNORD3a RNA (Fig. 2D). These RNAs are good candidates to act as global mediators of Polycomb activity.

The above types of analysis were then performed for the data from the experiment with CTCF antibodies. In the analysis of cis-acting RNAs, we observed a correlation of RNA enrichment in the CTCF-precipitated fraction with the percentage of contacts detected in promoter-, enhancer-, and insulator-specific chromatin type (Fig. 2B). We identified seven lincRNAs with a fold enrichment of >1.3 in both replicates (Fig. 2E). These lincRNAs might participate in loading CTCF to its DNA sites and organization of promoter-enhancer specific and other chromatin loops within genomic loci from where lincRNAs are produced. In the analysis of trans-acting RNAs, the highest fold enrichment was observed for snRNA RNU12 (Fig. 2F). Notably, U12 RNA is the second top by the total number of contacts among all RNAs in K562 cells (1.1% of all contacts). The potential involvement of RNU12 RNA in the functions of CTCF RNA in the functions of CTCF requires further experimental evidence.

Remarkably, the set of RNAs enriched in RedChIP significantly intersects the set of RNAs enriched in RNA IP (formaldehyde RNA IP-sequencing, fRIP-seq) experiments (Fig. 2H). Importantly, 18 of 22 ncRNAs overrepresented in CTCF- and EZH2-RedChIP samples are fRIP-positive, indicating these ncRNAs indeed interact with the studied proteins.

Collectively, the present study results demonstrate the utility of the RedChIP protocol for identifying RNAs that may target nonhistone proteins to various locations on chromosomes or mediate interactions of these proteins with DNA. The identification of RNAs that are known to target Polycomb complexes to repressed genomic domains strongly supports the validity of the experimental approach, whereas identifying a set of RNAs possessing similar characteristics will stimulate studies of their possible role in Polycomb and CTCF functioning. The RedChIP technique can be used for identifying RNAs associated with genomic regions occupied by any protein of interest.
Materials and Methods

RedChIP Procedure. Cells are fixed with formaldehyde, DNA is fragmented with NlaIII restriction enzyme, and the ends are blunted and A-tailed. RNA 3’ ends are ligated to a biotinylated bridge adapter followed by ligation of the opposite ends of the bridges with DNA ends in spatial proximity. Ligated complexes are solubilized by sonication, immunoprecipitated, and washed. RNA–DNA chimeras are purified, and DNA is digested with MmeI restriction enzyme. After biotin pull-down, reverse transcription is initiated from the bridge with template switching at the RNA 5’ end, allowing for the incorporation of an Illumina adapter. Another Illumina adapter is ligated to DNA ends, and the chimeras are amplified and paired-end sequenced. The detailed protocol and sequencing data processing are described in SI Appendix. For sample processing statistics, refer to Dataset S1. For read processing statistics, refer to Dataset S2.

Data Availability. Raw fastq reads and processed TSV files with contacts are available at Gene Expression Omnibus (accession no. GSE174474). The code for read processing is available as RedClib on GitHub: https://github.com/agalitsyna/RedClib.

ACKNOWLEDGMENTS. This work was supported by the Russian Science Foundation (21-64-00001) and by the Russian Ministry of Science and Higher Education (075-15-2021-1062).

1. S. V. Razin, A. A. Gavrilov, Non-coding RNAs in chromatin folding and nuclear organization. Cell. Mol. Life Sci. 78, 5489–5504 (2021).
2. J. M. Engertz, N. Ollikainen, M. Gutman, Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. Nat. Rev. Mol. Cell Biol. 17, 756–770 (2016).
3. Y. Long et al., RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. Nat. Genet. 52, 931–938 (2020).
4. R. Saldana-Meyer et al., RNA interactions are essential for CTCF-mediated genome organization. Mol. Cell 76, 412–422.e5 (2019).
5. A. A. Gavrilov et al., Studying RNA–DNA interactome by Red-C identifies noncoding RNAs associated with various chromatin types and reveals transcription dynamics. Nucleic Acids Res. 48, 6699–6714 (2020).
6. M. R. Mumbach et al., HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods 13, 919–922 (2016).
7. R. R. Pandey et al., Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 32, 232–246 (2008).