Integration of HiMoRNA and RNA-Chrom: Validation of the Functional Role of Long Non-coding RNAs in the Epigenetic Regulation of Human Genes Using RNA-Chromatin Interactome Data

I. S. Ilnitskiy[1,2,3†*], G. K. Ryabykh[1,2†], D. A. Marakulina[3,4], A. A. Mironov[1,2], Yu. A. Medvedeva[3,4]
[1]Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, 119234 Russia
[2]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991 Russia
[3]Skryabin Institute of Bioengineering, FRC Biotechnology, Russian Academy of Sciences, Moscow, 117312 Russia
[4]School of Biomedical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow, 141701 Russia
*E-mail: nfsus96@gmail.com
†Authors contributed equally to the work.
Received: October 21, 2024; in final form, March 31, 2025
DOI: 10.32607/actanaturae.27543

ABSTRACT **Long non-coding RNAs (lncRNAs) play a crucial role in the epigenetic regulation of gene expression by recruiting chromatin-modifying proteins to specific genomic loci. Two databases, previously developed by our groups, HiMoRNA and RNA-Chrom, provide valuable insights into this process. The former contains data on epigenetic modification regions (peaks) correlated with lncRNA expression, while the latter offers genome-wide RNA–chromatin interaction data for tens of thousands of RNAs. This study integrated the two resources to generate experimentally supported, interpretable hypotheses regarding lncRNA-mediated epigenetic gene expression regulation. We adapted the web interfaces of HiMoRNA and RNA-Chrom to enable the retrieval of chromatin contacts for each "lncRNA–epigenetic modification–associated gene" triad from HiMoRNA, either at specific genomic loci or genome-wide via RNA-Chrom. The integration analysis revealed that for the lncRNAs MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, and MEG3, most HiMoRNA peaks are located within 25 kb of their RNA-Chrom contacts. Further investigation confirmed the RNA–chromatin contacts of MIR31HG and PVT1 lncRNAs, with HiMoRNA peaks for H3K27ac and H3K27me3 marks in the loci of the genes *GLI2* and *LATS2*, respectively, which are known to be regulated by these RNAs. Thus, the integration of HiMoRNA and RNA-Chrom offers a powerful platform to elucidate the role of specific lncRNAs in the regulation of histone modifications at both individual loci and genome-wide levels. We expect this integration to help significantly advance the functional annotation of human lncRNAs.**
KEYWORDS **long non-coding RNA, histone modification, RNA-chromatin interaction.**
ABBREVIATIONS **ncRNA – non-coding RNA; lncRNA – long non-coding RNA.**

INTRODUCTION

Human cells transcribe a vast number of long non-coding RNAs (lncRNAs), with their quantity comparable to that of protein-coding genes [1, 2]. Functional annotation of lncRNAs is challenging due to their low expression levels, tissue specificity, and low sequence conservation [3–5]. Nevertheless, lncRNAs were observed to preserve certain characteristics, notably synteny with neighboring genes, secondary structure, and similarity in short sequence fragments [6]. In addition, the transcriptional regulation of lncRNA transcription exhibits intricacy comparable to that of protein-coding RNAs, facilitating their involvement in diverse molecular processes [7]. Suppression of lncRNAs has been shown to result in significant changes in the transcriptional profile of

cells [8]. These findings indicate a functional role for numerous lncRNAs. Most lncRNAs were shown to interact with chromatin and to be involved in the epigenetic regulation of genomic loci and the structural organization of chromosomes [9–12]. Accordingly, identifying the functional genomic targets of chromatin-associated lncRNAs is of high value.

Previously, we developed the HiMoRNA database [13], which catalogues over 5 million epigenetic "peaks," genomic regions exhibiting one of 10 histone modifications, with their modification level significantly correlating with lncRNA expression across at least 20 cell lines and tissues. Where possible, histone modification peaks in HiMoRNA are associated with genes, forming a triad: "lncRNA–epigenetic modification peak–associated gene." Within such a triad, the lncRNA is hypothesized to modulate the expression of the corresponding gene through the promotion or repression of histone modification within the gene's associated peak region. In the case of promotion, peaks positively correlate with lncRNA expression ("+" peaks), while in the case of repression, the correlation is negative ("–" peaks). These associations enable the formulation of hypotheses regarding the role of lncRNAs in the modulation of epigenetic modifications at specific genomic loci and the regulation of gene expression. However, to formulate and empirically test reasonable hypotheses, the 5-million-peak dataset has to be pre-processed to select the most reliable peaks. Experimental methods for detecting RNA-chromatin interactions can provide valuable data for this purpose. Several experimental methods exist to identify chromatin regions interacting with non-coding RNAs. These methods can be broadly classified into two categories: "one-to-all" [11, 14–18], which identify the contacts of a specific RNA with chromatin, and "all-to-all" [19–24], which capture all possible RNA–DNA contacts in a cell [25]. Notwithstanding their utility, these approaches are prone to high false positive rates. Additionally, "all-to-all" methods show low sensitivity to lowly expressed RNAs and bias toward nascent transcripts. Despite these challenges, genome-wide data on non-coding RNA (ncRNA) interactions with chromatin are crucial for elucidating their mechanisms of action. In this context, the RNA-Chrom database [26] was recently created. It contains experimental data on thousands of RNA-chromatin contacts and offers two analytical modes ("from RNA" and "from DNA") that can be used for research purposes.

To improve and streamline the functional annotation of lncRNAs, we have integrated the HiMoRNA and RNA-Chrom databases. The web interfaces of HiMoRNA and RNA-Chrom were modified to provide direct access to chromatin contacts for 4,124 out of the 4,145 lncRNAs from HiMoRNA in RNA-Chrom. This integration enables the generation of hypotheses regarding the mechanisms of epigenetic regulation of human gene expression by long non-coding RNAs, supported by experimental data on their interactions with chromatin. We anticipate that this unified resource will prove a valuable tool for identifying high-confidence "lncRNA–epigenetic modification–associated gene" triads for further experimental investigation of lncRNA mechanisms in gene regulation. The HiMoRNA database is available to users at https://himorna.fbras.ru (as of 20.10.2024).

EXPERIMENTAL PART

Integration of the HiMoRNA and RNA-Chrom databases

Due to differences in gene annotation sources (HiMoRNA uses GENCODE v31, while RNA-Chrom uses GENCODE v35), we established gene correspondence using three similarity metrics: (1) matching gene names ('gene_name', *Fig. 1A*); (2) matching gene identifiers ('gene_id', *Fig. 1A*); and (3) a Jaccard index (the ratio of the length of gene overlap to the length of their union) greater than 0.99 (Jaccard index > 0.99, *Fig. 1A*). Due to differences in naming conventions and genomic coordinates between annotation versions, gene identifiers and positions often do not align directly. To address this issue, we intersected 4,145 lncRNA genes from HiMoRNA with 60,619 genes from RNA-Chrom based on genomic coordinates using the "*intersect*" command from "*bedtools*". This yielded 6,778 gene pairs, exceeding the number of HiMoRNA entries, because some HiMoRNA genes intersected multiple times with RNA-Chrom genes. The two HiMoRNA genes (ENSG00000267034.1 and ENSG00000280076.1) did not intersect with any RNA-Chrom genes. Subsequently, the Jaccard index was calculated for each gene pair. The gene pairs were classified into six groups based on similarity metrics (*Fig. 1A*). By using a Jaccard index > 0.99 as the primary similarity metric, we identified four groups (groups 2, 4, 5, and 6) as having no overlapping genes (*Fig. 1B*), establishing 4,100 unambiguous gene correspondences. For the remaining 43 genes from HiMoRNA, 24 additional matches to RNA-Chrom genes were established using the 'gene_name' matching metric. In total, we identified 4,124 lncRNA genes common to both databases (see *Supplementary Table 1* for a full correspondence table that is available for download on the HiMoRNA web resource).

To support database integration and streamline access to chromatin interaction data, we made sever-

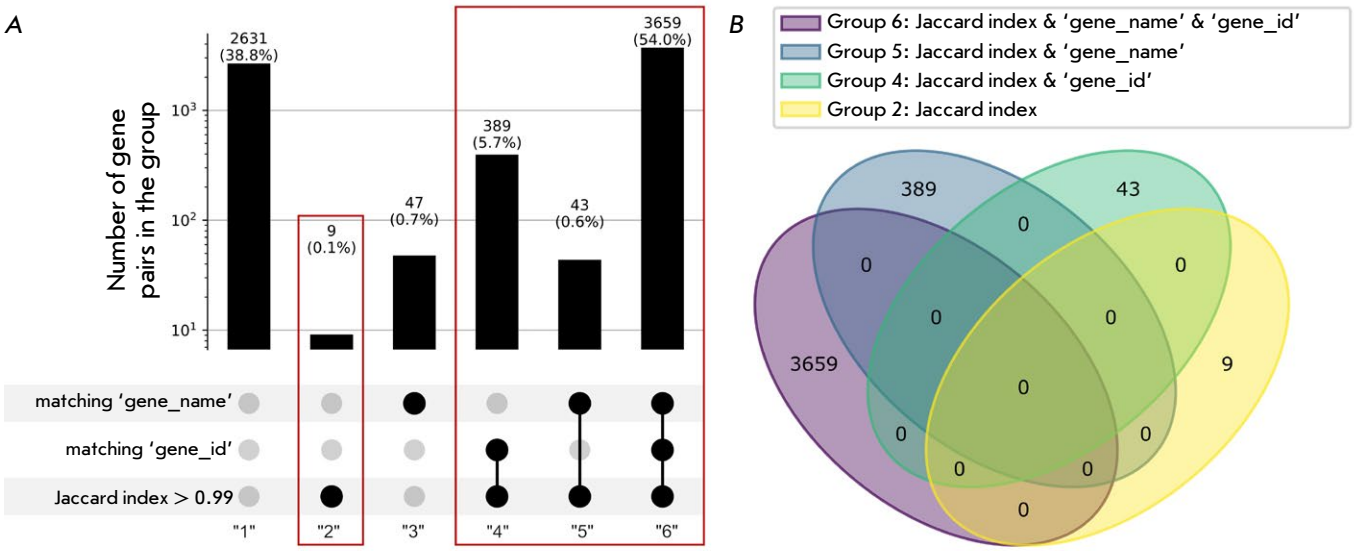

Fig. 1. Intersection of 4,145 genes from HiMoRNA with 60,619 genes from RNA-Chrom. (*A*) Division of gene pairs into six groups based on the similarity metrics they satisfy. Groups achieving unambiguous gene correspondence are highlighted with red rectangles. (*B*) Venn diagram showing the overlap between the gene groups 2, 4, 5, and 6 (the total number of gene pairs in these four groups is 4,100)

al enhancements to the RNA-Chrom and HiMoRNA interfaces. These include parameter processing (locus, RNA name, RNA-Chrom internal RNA identifier, organism) from a specific type of URL (e.g., https://rnachrom2.bioinf.fbb.msu.ru/basic_graphical_summary_dna_filter?locus=chrX:23456-24253566&name=XIST&rnaID=227896&organism=Homo+sapiens) and providing information on the chromatin contacts of the requested lncRNA across different experiment types on a new browser page.

Enhancements to the HiMoRNA interface include the incorporation of the gene correspondence table between RNA-Chrom and HiMoRNA to ensure correct URL generation.

A "Go to RNA-Chrom DB" button with a dropdown menu (*Fig. 2*) was added to the "Search Results Page," allowing the user to generate three types of URL links to navigate to the RNA-Chrom page:

a) Contacts of the given lncRNA in a specific genomic locus, extended by 1 / 5 / 10 / 25 / 50 / 100 kb;

b) All contacts of the given lncRNA;

c) All RNAs with contacts in a specific genomic locus.

One-sided Fisher's exact test

In most triads, histone modification peaks show both negative and positive correlations between lncRNA expression and the peak signal level (hereinafter denoted "–" and "+" peaks, respectively). The presence of a "+" peak suggests a role for the lncRNA in establishing the histone modification, whereas a "–" peak implies its role in removing the modification.

The alignment of predictions with published experimental results was evaluated by selecting lncRNAs and their corresponding histone peaks (positively and negatively correlated), extended by ±25 kb, and filtering for those where the proportion of peaks supported by contacts for at least one histone mark exceeded 40%. We subsequently conducted independent right-tailed and left-tailed Fisher's exact tests for each lncRNA-histone mark pair. For example, a representative contingency table for the "PVT1–H3K27ac" pair is provided in *Supplementary Table 2*.

Red-ChIP data

As a case study, we used the lncRNA PVT1 to validate integration between HiMoRNA and RNA-Chrom using independent experimental data. Specifically, we incorporated Red-ChIP data [27], available in the Gene Expression Omnibus under accession number GSE174474, samples GSM5315228 and GSM5315229 (hES cell line). The Red-ChIP method captures RNA-chromatin contacts mediated by the EZH2 protein, a component of the PRC2 complex, which establishes, among others, the H3K27me3 histone modification.

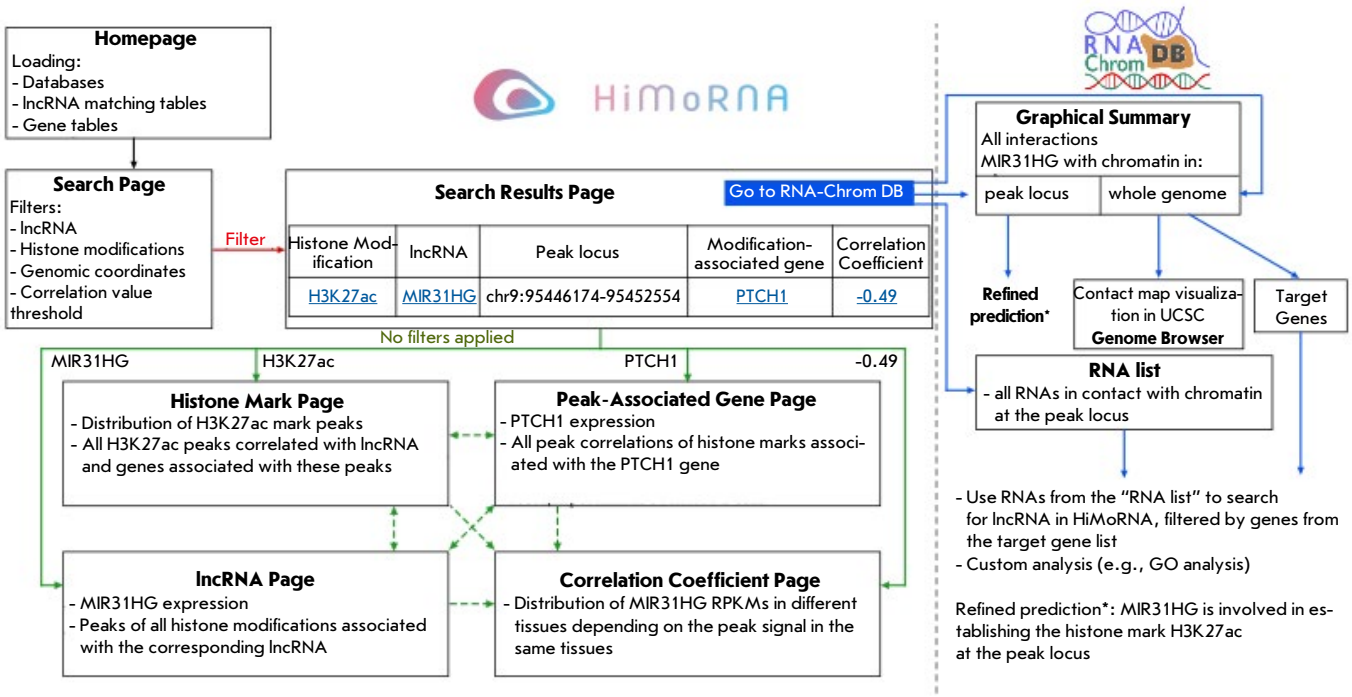

Fig. 2. Usage scenario of the HiMoRNA and RNA-Chrom databases after integration. Rectangles represent web pages, and arrows indicate movement between them

Primary data processing followed the established RNA-Chrom database protocol. Primary data processing followed the established RNA-Chrom database protocol. Subsequently, we identified genomic regions enriched with lncRNA PVT1 chromatin contacts using the BaRDIC program (--qval_type all; --qval_threshold 1) [28]. Consequently, 3,242 genomic regions exhibiting potential functionality and EZH2-mediated PVT1 binding were identified.

RESULTS

Integration of Databases

Given that HiMoRNA contains millions of epigenetic peaks, selecting the most reliable ones for further analysis is critical. For this task, we integrated HiMoRNA peak data with RNA-chromatin interactome data from the RNA-Chrom database. This integration was achieved by establishing one-to-one correspondences between genes in both databases and modifying their web interfaces (see "Experimental Section," subsection "Integration of HiMoRNA and RNA-Chrom Databases"). This strategy allows HiMoRNA to generate specific URL queries for 4,124 out of the 4,145 lncRNAs listed in RNA-Chrom, enabling, in particular, the identification of other chromatin loci with which the investigated RNA interacts. This approach significantly expands our understanding of the function of a specific RNA.

The general integration scheme is presented in *Fig. 2*. To utilize the integration, the user first needs to find the target lncRNA in the HiMoRNA database. From the HiMoRNA homepage, users can download the database itself, as well as the "Gene Table" and the "lncRNA Correspondence Table" added as part of the integration, to search for genes/lncRNAs of interest by genomic coordinates. This feature accommodates potential mismatches between user-specified Ensembl identifiers/lncRNA names and those present in the HiMoRNA database. On the search page, the user must configure filters according to their task, specifying the lncRNAs, histone modifications, genomic coordinates, and genes associated with the selected histone modifications.

On the search results page, users may conduct a more thorough examination of the retrieved predictions, including accessing the RNA-Chrom database. To do this, the user should select the desired

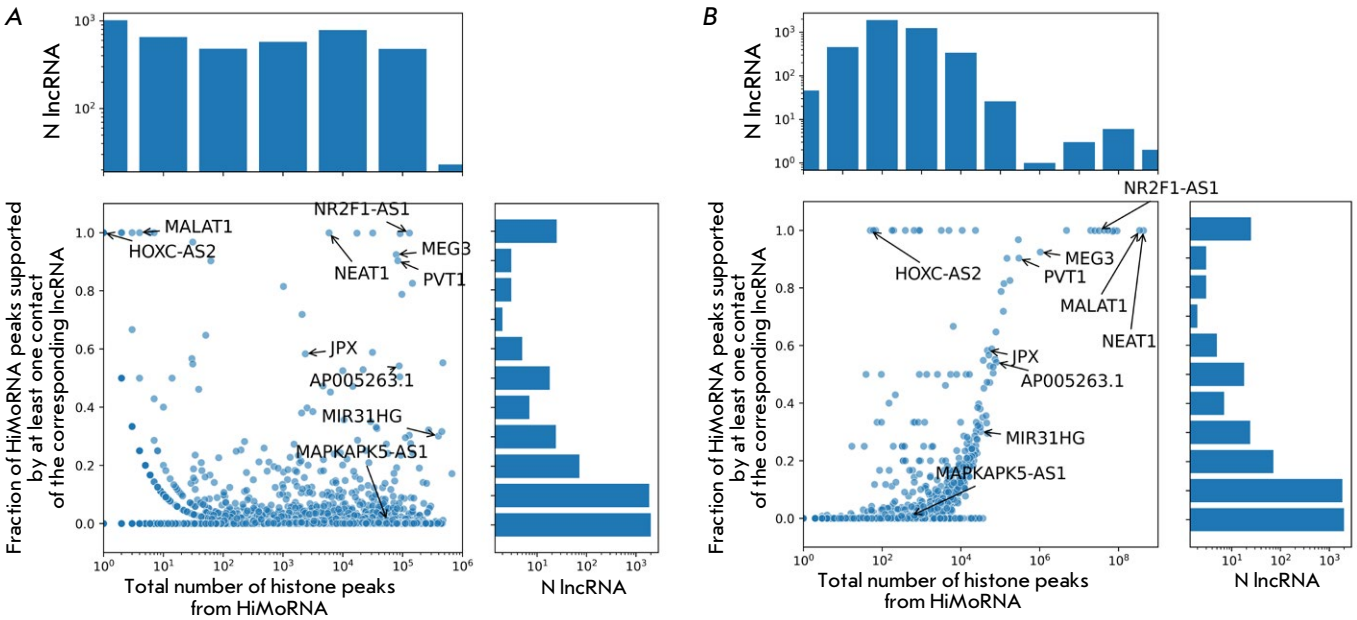

Fig. 3. Proportions of HiMoRNA peaks confirmed by at least one contact of the corresponding lncRNA from RNA-Chrom, relative to the total number of HiMoRNA peaks for the corresponding lncRNA (*A*) and the total number of contacts for the corresponding lncRNA from RNA-Chrom (*B*). Genomic coordinates of contacts are extended by ± 25 kb

"lncRNA–epigenetic modification peak–associated gene" triad from the interactive results table and then click the "Go to RNA-Chrom DB" button. In the dropdown menu, the user should select the appropriate option to navigate to a page displaying: 1) contacts of the given lncRNA in the region of a specific peak (with the option to choose how much to extend the peak coordinates when searching for contacts); 2) all contacts of the given lncRNA; or 3) all lncRNAs with contacts in the specified locus. The user will be automatically redirected to a graphic summary of the lncRNA-chromatin interactome on the RNA-Chrom database webpage. This summary allows the user to determine whether the functional relationship "lncRNA–epigenetic modification" from HiMoRNA is mediated by the physical presence of the lncRNA at the corresponding genomic locus, as well as to identify other lncRNAs potentially involved in the regulation of that locus. Visual analysis is enabled by utilizing the UCSC Genome Browser to load data from all relevant experiments by clicking "VIEW IN GENOME BROWSER." By selecting a single RNA-chromatin interactome experiment, the user can obtain a list of the genes located in the genomic region of interest, along with statistics on their interactions with the lncRNA, by clicking "ALL TARGET GENES." This list of genes can be downloaded for further research, such as performing a GO analysis. The "Use Cases" section offers a comprehensive discussion and exemplification of the HiMoRNA and RNA-Chrom database integration.

Consistency of HiMoRNA and RNA-Chrom Results

To assess the completeness of the integration, we analyzed the frequency of confirmation of histone peaks from HiMoRNA, correlated with lncRNA expression, using data on the corresponding lncRNA chromatin contacts from RNA-Chrom. Out of the 4,145 lncRNAs present in HiMoRNA, 4,011 (96.8%) were found to have at least one contact in the RNA-Chrom database, with 29 RNAs not matching between the databases and 105 (2.5%) having no contacts in RNA-Chrom. Among the 4,011 lncRNAs of interest, only 35.5% had at least one peak supported by the contacts of the corresponding lncRNA. However, due to the design of experimental protocols, actual lncRNA-chromatin interactions may occur at a distance from the experimentally detected contact. To address this issue, we extended the contact coordinates for a more accurate assessment of the correspondence between HiMoRNA predicted peaks and RNA-Chrom data. Extending contact ranges by ±1, ±5, ±10, ±25, and ±50 kb resulted in a respective increase in the per-

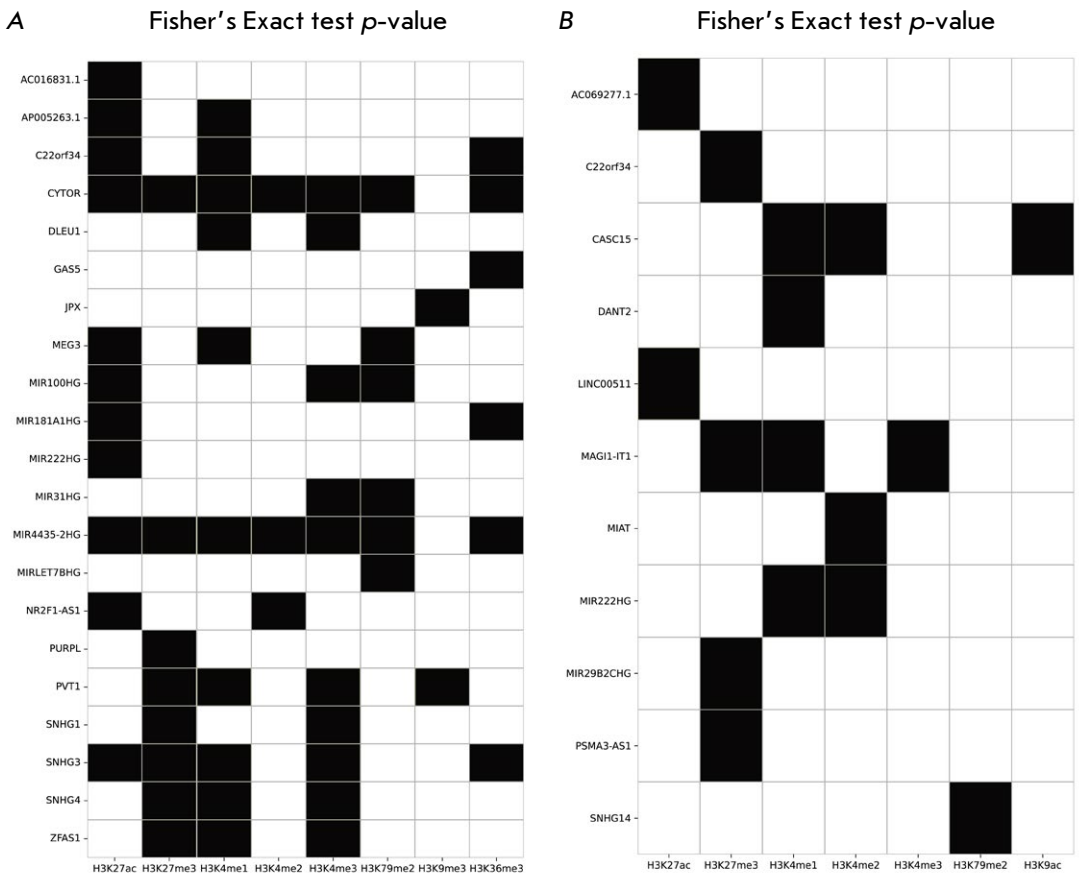

Fig. 4. Heatmap showing the results of Fisher's exact test for pairs of "lncRNA – histone mark peaks extended by ±25 kb". Black indicates that the proportion of "–" or "+" histone peaks supported by contacts of the corresponding lncRNA is greater than 0.4 and the p-value of the Fisher's exact test is less than 10^{-3}; otherwise, it is white. (*A*) Right-tailed Fisher's test: "+" peaks of corresponding histone marks are better supported by RNA-Chrom contacts than "–" peaks. (*B*) Left-tailed Fisher's test: "–" peaks of corresponding histone marks are better supported by RNA-Chrom contacts than "+" peaks

centage of RNAs with HiMoRNA peaks confirmed by at least one contact to 38.5%, 42.7%, 45.7%, 50.1%, and 53%. Some lncRNAs (e.g., MALAT1, HOXC-AS2, NEAT1, NR2F1-AS1, PVT1, MEG3) had nearly all HiMoRNA peaks confirmed with ±25 kb extension (*Fig. 3*). However, lncRNAs with a significantly lower proportion of peaks extended by ±25 kb and confirmed by contacts are more common (e.g., JPX, AP005263.1, MIR31HG) or have a proportion approaching 0 (e.g., MAPKAPK5-AS1). This discrepancy likely arises from the incomplete datasets of lncRNAs in HiMoRNA and RNA-Chrom, resulting from the stringent criteria used for prediction filtering and the limitations inherent in available experimental RNA-chromatin interaction data. For example, in RNA-Chrom, half of the lncRNAs considered in this study have fewer than 200 contacts (*Fig. 3B*), as most lncRNAs rely on "all-to-all" experimental data, which insufficiently captures contacts of lowly expressed RNAs.

HiMoRNA triads display either negative or positive correlations between lncRNA expression and the signal of epigenetic peaks ("–" and "+" peaks, respectively). To evaluate biological consistency with known data, we selected 30 lncRNAs and their corresponding histone peaks, for which "+" or "–" peaks of at least one histone mark are statistically significantly predominant (one-sided Fisher's exact test, p-value < 0.001), confirmed by RNA-Chrom contacts extended by ±25 kb (see "Experimental Section," subsection

"One-Sided Fisher's Exact Test," *Fig. 4*). After filtering the results by a p-value < 0.001, we obtained the following "lncRNA–histone mark" pairs.

Twenty-one lncRNAs with "+" peaks of corresponding histone marks were better supported by RNA-Chrom contacts than "–" peaks (right-tailed Fisher's exact test, p-value < 0.001).

Eleven lncRNAs with "–" peaks of corresponding histone marks were better supported by RNA-Chrom contacts than "+" peaks (left-tailed Fisher's exact test, p-value < 0.001).

Previous studies have shown the potential involvement of many identified lncRNAs in epigenetic regulation via histone modifications. Let us consider cases where "+" peaks are statistically significantly better supported by RNA-Chrom contacts than "–" peaks. For example, MIR4435-2HG is involved in establishing the activator mark H3K27ac in the enhancer region of the RPTOR locus [29]. Our data suggest that MIR4435-2HG, in addition to H3K27ac, likely targets other epigenetic modifications, such as H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, and H3K79me2 (*Fig. 4A*). Similarly, based on the data for MIR31HG [30], SNHG1, PVT1 [31–33], and the mouse lncRNA lnc-Nr2f1 (presumed to have functional conservation with human NR2F1-AS1) [34], we identified consistent histone modifications: NR2F1-AS1 – H3K27ac, MIR31HG – H3K4me3, SNHG1 – H3K27me3, PVT1 – H3K27me3. Additionally, we uncovered functional associations of these lncRNAs with other epigenetic marks: NR2F1-AS1 – H3K4me2, MIR31HG – H3K79me2, SNHG1 – H3K4me3, PVT1 – H3K4me1, H3K4me3, H3K9me3 (*Fig. 4A*).

Numerous H3K27me3 and H3K4me3 "+" peaks, validated through a chromatin interaction analysis, were identified for several lncRNAs (ZFAS1, SNHG4, SNHG1, SNHG3, PVT1, MIR4435-HG, and CYTOR). These peaks exhibit significantly stronger support from RNA-chromatin contacts than "–" peaks (*Fig. 4A*), representing opposing chromatin states. By analogy with known lncRNAs that establish both marks depending on their association with different effector proteins (e.g., ncRNA SRA [35], ANRIL [36]), it can be hypothesized that these lncRNAs also exhibit more complex mechanisms of chromatin activity regulation.

Cases where "–" peaks are statistically significantly better supported by RNA-Chrom contacts than "+" peaks can be explained by the corresponding lncRNAs regulating the removal of histone marks by recruiting demethylases and deacetylases to specific genomic loci (*Fig. 4B*). A lack of corroborating experimental data precludes a quality assessment of our predictions for these lncRNAs. We suggest that the "lncRNA–histone mark" pairs reported in this section (*Fig. 4*) are suitable candidates for further study.

Usage examples

The objective of integrating HiMoRNA and RNA-Chrom is to refine the functional relationship within the "lncRNA–epigenetic modification peak–associated gene" triads using data on the localization of the corresponding lncRNA in the genomic region near peaks of specific histone modifications. Below, we provide examples of user studies of several lncRNAs with known mechanisms of action.

lncRNA MIR31HG

The long non-coding RNA MIR31HG is a known regulator of the histone marks H3K4me1, H3K4me3, and H3K27ac. Previous studies have reported a reduction in the H3K4me1 and H3K27ac levels in the enhancer region of the *GLI2* gene and H3K4me3 and H3K27ac in the promoter region of the *FABP4* gene following MIR31HG knockdown [30, 37]. This observation can be validated using our integration of HiMoRNA and RNA-Chrom.

To this end, we created a query in HiMoRNA: lncRNA MIR31HG, histone marks H3K4me1 and H3K27ac, with the coordinates of the two selected genes specified with an extended promoter region of 10 kb in the genomic coordinates field (*Fig. 5A*). As a result, the HiMoRNA web resource generated a table with H3K27ac and H3K4me1 peaks correlated with MIR31HG expression across various tissues (*Fig. 5B*). We then selected a triad with an H3K27ac peak and navigated to the RNA-Chrom page displaying experimentally detected MIR31HG chromatin contacts in the region of the selected peak (by clicking "Go to RNA-Chrom DB," *Fig. 5C*). By selecting an RNA-chromatin experiment from the top table and clicking "All target genes" (*Fig. 6A*), we obtained a table that, in particular, reflected the interaction of MIR31HG with the *GLI2* gene (*Fig. 6B*). A step-by-step analysis is presented in *Supplementary Table 3*.

To explore the potential for novel biological insights into the lncRNA function, we hypothesized that integrating the HiMoRNA and RNA-Chrom datasets would reveal that MIR31HG regulates additional components of the Sonic hedgehog signaling pathway (KEGG:04340) besides *GLI2*. For this purpose, we used the KEGG Pathway database [38] to identify relevant genes. Next, we formulated a new HiMoRNA query consisting of lncRNA MIR31HG, histone modifications H3K4me1 and H3K27ac, and the 56 genes associated with the Hedgehog signaling pathway (*Supplementary Table 4*). The outcome was a table of 162 triads, which can be validated with the RNA-Chrom resource. For

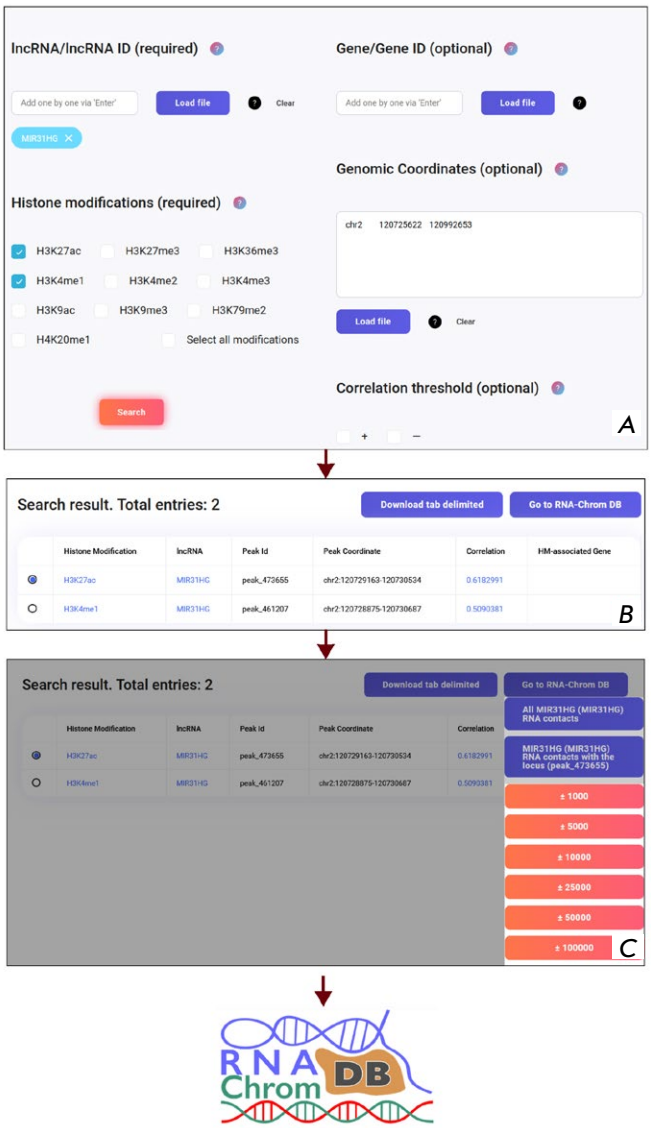

Fig. 5. Use case of the integration of the HiMoRNA and RNA-Chrom databases using the example of lncRNA MIR31HG. (*A*) Creating a query in HiMoRNA for MIR31HG, histone modifications H3K4me1 and H3K27ac, and genes *GLI2* and FABP4. (*B*) Table with search results. (*C*) Navigation to RNA-Chrom

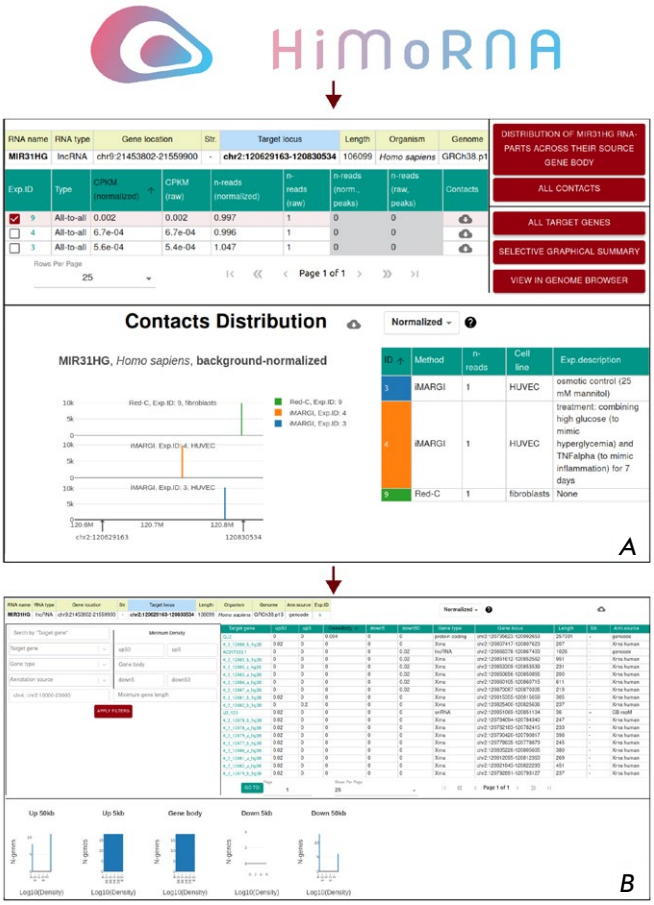

Fig. 6. Use case of the integration of the HiMoRNA and RNA-Chrom databases using the example of lncRNA MIR31HG. (*A*) RNA-Chrom page showing MIR31HG contacts with chromatin in the region of the extended HiMoRNA peak. (*B*) Table listing all genes in the region of the extended peak, indicating whether they interact with MIR31HG or not (Experiment ID: 9)

example, in the locus of the H3K27ac_963553 peak (chr9:95446174-95452554), MIR31HG interacts with the *PTCH1* gene, which encodes the "Sonic hedgehog" receptor. To determine if the gene set associated with the H3K27ac and H3K4me1 peaks, correlated with MIR31HG expression, exhibits significant enrichment of "Hedgehog signaling pathway" genes, a KEGG pathway enrichment analysis was conducted using the g:Profiler web resource [39]. The query included genes selected for MIR31HG and H3K27ac/H3K4me1, with all other genes associated with HiMoRNA peaks used as the background. The analysis revealed genes belonging to the "Hedgehog signaling pathway" to be enriched with H3K27ac peaks (p-value = 2.090×10^{-2}) but not with H3K4me1 peaks. This observation suggests the involvement of MIR31HG in regulating the "Hedgehog signaling pathway" through the establishment of the H3K27ac histone modification in the corresponding genomic loci.

lncRNA PVT1

The long non-coding RNA PVT1 is known to inhibit the expression of the *LATS2* gene in non-small cell lung cancer cells by recruiting EZH2 (a subu-

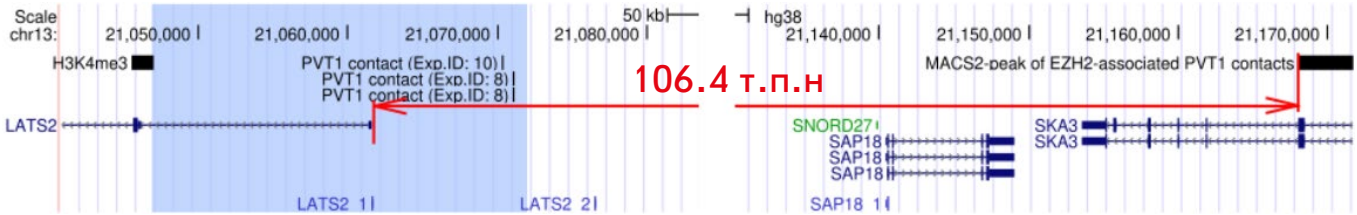

Fig. 7. Representation in the UCSC Genome Browser of the region encompassing the *LATS2* gene and its promoter vicinity, showing an H3K4me3 peak correlated with lncRNA PVT1 expression, lncRNA PVT1 contacts from two experiments (RNA-Chrom Exp. IDs: 8, 10), and an EZH2-mediated PVT1 contact peak. The blue region reflects the extension of the H3K4me3 peak coordinates by 25 kb, within which RNA-Chrom contacts were selected

nit of the PRC2 complex) to the corresponding promoter [40]. We performed a search for triads in HiMoRNA: lncRNA PVT1, all histone modifications, and the *LATS2* gene. Our findings revealed solely H3K4me3 activation peaks, exhibiting a negative correlation with PVT1 expression levels. This observation is consistent with previously published findings [39], wherein PVT1's recruitment of EZH2 contributes to the establishment of the repressive H3K27me3 mark. In RNA-Chrom, we observed contacts around one of the H3K4me3 peaks (peak_169403, chr13:21045571-21046978) in two experiments (K562 and MDA-MB-231 cell lines). Visualization of PVT1 contacts in the Genome Browser [41] confirms the presence of this peak in the promoter region of the *LATS2* gene (*Fig.* 7, step-by-step analysis presented in *Supplementary Table 5*). Additional confirmation of *LATS2* regulation by lncRNA PVT1 was obtained using Red-ChIP data (see "Experimental Section," subsection "Red-ChIP Data"). A peak of EZH2-mediated PVT1 contacts (chr13:21168000-21224000, q-value = 0.09) was identified 106.4 kb from the 5'-end of the *LATS2* gene (*Fig.* 7).

The absence of "lncRNA PVT1–H3K27me3 peak–*LATS2* gene" triads with a positive correlation in HiMoRNA is likely due to overly stringent filtering of H3K27me3 peaks during the database creation. The preceding examples from the "Use Cases" section illustrate the successful application of this integration in generating testable hypotheses regarding lncRNA involvement in the epigenetic regulation of specific genes.

DISCUSSION

Prospects for Use and Limitations of the Approach

The HiMoRNA database comprises genomic loci exhibiting a significant correlation between histone modification signals and lncRNA expression across diverse cell types and tissues. Currently, it contains over 5 million correlations for 10 types of histone modifications and 4,145 lncRNAs. We hypothesize that some of these correlations may represent false positives or indirect regulatory relationships, thus requiring further validation using external data. The RNA-Chrom database houses genome-wide RNA-chromatin interaction data. These data unfortunately lack sufficient representation of contacts from lowly expressed ncRNAs and are heavily biased towards contacts from nascent transcripts. Furthermore, these data are insufficient to formulate hypotheses regarding the functional roles of these interactions. With these limitations in mind, integrating the data from HiMoRNA and RNA-Chrom is a reasonable approach to characterize the impact of lncRNAs on epigenetic modifications and gene expression.

Despite the advantages of integration, some inconsistencies remain. No H3K27ac or H3K4me1 peaks were found in the *FABP4* gene locus in the HiMoRNA database. This observation contradicts experimental data and the RNA-Chrom database. There are other negative examples in both HiMoRNA and RNA-Chrom for well-known lncRNAs involved in epigenetic regulation and chromatin structure maintenance. For instance, no H3K27me3 peaks correlating with MEG3 were observed in HiMoRNA, despite evidence suggesting that MEG3 regulates the PRC2 complex and contributes to maintaining H3K27me3, particularly in the promoter regions of the *SMAD2*, *TGFB2*, and *TGFBR1* genes [11]. The absence of such peaks is likely attributable to the cell-specific expression of most lncRNAs, with the mentioned mechanism observed in a cell type not represented in HiMoRNA. Even when ChIP-seq data are available, the standard peak-calling procedure may be too stringent, potentially filtering out biologically significant interactions.

We observed instances where HiMoRNA predictions aligned with published experimental data, but the corresponding RNA-chromatin contacts were absent from the relevant genomic locus in the RNA-Chrom dataset. For example, for lncRNA MAPKAPK5-AS1, most correlated peaks from HiMoRNA are not supported by the expression of this RNA, resulting in a low number of observed chromatin contacts in "all-to-all" experiments. We posit that these instances may stem from variations in cell types across the two databases as a result of insufficient data.

Due to insufficient experimental data, neither database contains exhaustive information. Therefore, some documented biological examples might have been overlooked in the integration process. Nevertheless, their integration offers complementary strengths, such as mitigating various systematic errors that are due to the multi-omics nature of the combined data and expanding the generation of interpretable hypotheses about the mechanisms of epigenetic regulation of gene expression by long non-coding RNAs.

FURTHER DEVELOPMENT

The implemented integration could be significantly improved through the incorporation of supplementary genome-wide data and annotations. The dataset may contain information regarding the three-dimensional chromatin structure, gene expression and co-expression patterns (including long non-coding RNA expression, the target genes of the triad, and the genes associated with histone modification), in addition to the localization of DNA-binding and chromatin-modifying proteins. Given the current scarcity of such experimental data, it is worth considering the use of bioinformatics predictions. One potential direction could involve incorporating results from predictions of the type of lncRNA-chromatin interactions (for a comparison of programs determining the mechanisms of lncRNA interactions with other molecules; see, for example, [42]). The combination of predicted lncRNA–target interactions and multi-omics experimental data has facilitated effective hypothesis generation concerning the roles of specific lncRNAs [43–45]. Another important direction would be to include data on gene expression changes following artificial alterations in the concentration of specific lncRNAs in cells [8], as well as experimentally validated information on the involvement of specific lncRNAs in regulating particular histone modifications [46]. This would provide an additional layer of validation for the results of the HiMoRNA and RNA-Chrom integration. Furthermore, from a practical perspective, it would be useful to enhance the integration with an assessment of the statistical significance of the co-localization of HiMoRNA peaks and RNA-Chrom contacts for a specific lncRNA using specialized software tools such as Genometricorr [47], StereoGene [48], and RegioneR [49].

The field of lncRNA research is rapidly evolving. We will maintain support for the integration of HiMoRNA and RNA-Chrom as both databases expand their taxonomic scope and incorporate updated data. Upon the emergence of new, experimentally validated data, we intend to construct multiple predictive models for "lncRNA–histone epigenetic modifications–associated gene" interactions. We are confident that the continued collaborative expansion of the HiMoRNA and RNA-Chrom web resources will contribute to a deeper understanding of the functional role of lncRNAs in the epigenetic regulation of genes. ●

The authors express their gratitude to A. Nikolskaya for providing processed Red-ChIP data and to anonymous reviewers for their valuable suggestions.

This work was supported by the Russian Science Foundation grant No. 23-14-00371 (PI: Yu.A. Medvedeva).

Author Contributions

Improvement of the HiMoRNA web interface and database, software development, I.I.; adaptation of the RNA-Chrom web resource for integration with HiMoRNA, gene name correspondence, project administration, G.R.; use cases, G.R. and D.M.; general supervision, A.M. and Yu.M.; manuscript preparation, I.I., G.R., D.M., A.M., and Yu.M. All authors have read and agreed to the published version of the manuscript.

Supplementaries are available on the website https://doi.org/10.32607/actanaturae.27543.

REFERENCES

1. Carninci P., Sandelin A., Lenhard B., Katayama S., Shimokawa K., Ponjavic J., Semple C.A.M., Salmena L., Nishida M., Hayashizaki Y., et al. // Science. 2005. V. 309. № 5740. P. 1559–1563. doi: 10.1126/science.1112014.
2. Hon C.-C., Ramilowski J.A., Harshbarger J., Bertin N., Rackham O.J.L., Garmire L.X., Forrest A.R.R., Carninci P., Kawaji H., Hayashizaki Y., et al. // Nature. 2017. V. 543. № 7644. P. 199–204. doi: 10.1038/nature21374.
3. Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega

B., Regev A., Rinn J.L. // Genes Dev. 2011. V. 25. № 18. P. 1915–1927. doi: 10.1101/gad.17446611.
4. Andersson R., Gebhard C., Miguel-Escalada I., Hoof I., Bornholdt J., Boyd M., Chen Y., Zhao X., Schmidl C., Suzuki T., et al. // Nat. Commun. 2014. V. 5. № 1. P. 5336. doi: 10.1038/ncomms6336.
5. Ulitsky I., Shkumatava A., Jan C.H., Sive H., Bartel D.P. // Cell. 2011. V. 147. № 7. P. 1537–1550. doi: 10.1016/j.cell.2011.11.055.
6. Quinn J.J., Ilik I.A., Qu K., Georgiev P., Chu C., Akhtar A., Chang H.Y. // Genes Dev. 2016. V. 30. № 2. P. 191–207. doi: 10.1101/gad.272187.115.
7. Alam T., Medvedeva Y.A., Jia H., Brown J.B., Lipovich L., Baillie J.K. // PLoS One. 2014. V. 9. № 10. P. e109443. doi: 10.1371/journal.pone.0109443.
8. Ramilowski J.A., Yip C.W., Agrawal S., Chang J.-C., Ciani Y., Kulakovskiy I.V., Mendez M., Ooi J.L.C., Ouyang J.F., Nguyen A., et al. // Genome Res. 2020. V. 30. № 7. P. 1060–1072. doi: 10.1101/gr.254516.119.
9. Khalil A.M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D., Thomas K., Presser A., Bernstein B.E., van Oudenaarden A., et al. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 28. P. 11667–11672. doi: 10.1073/pnas.0904715106.
10. Grote P., Herrmann B.G. // RNA Biol. 2013. V. 10. № 10. P. 1579–1585. doi: 10.4161/rna.26165.
11. Mondal T., Subhash S., Vaid R., Enroth S., Uday S., Reinius B., Mitra S., Mohammed A., James A.R., Hoberg E., et al. // Nat. Commun. 2015. V. 6. № 1. P. 7743. doi: 10.1038/ncomms8743.
12. Goff L.A., Rinn J.L. // Genome Res. 2015. V. 25. № 10. P. 1456–1465. doi: 10.1101/gr.191122.115.
13. Mazurov E., Sizykh A., Medvedeva Y.A. // Non-Coding RNA. 2022. V. 8. № 1. P. 18. doi: 10.3390/ncrna8010018.
14. Engreitz J.M., Pandya-Jones A., McDonel P., Shishkin A., Sirokman K., Surka C., Kadri S., Xing J., Goren A., Lander E.S., et al. // Science. 2013. V. 341. № 6147. P. 1237973. doi: 10.1126/science.1237973.
15. Simon M.D., Wang C.I., Kharchenko P.V., West J.A., Chapman B.A., Alekseyenko A.A., Borowsky M.L., Kuroda M.I., Kingston R.E. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 51. P. 20497–20502. doi: 10.1073/pnas.1113536108.
16. Chu C., Qu K., Zhong F.L., Artandi S.E., Chang H.Y. // Mol. Cell. 2011. V. 44. № 4. P. 667–678. doi: 10.1016/j.molcel.2011.08.027.
17. Quinn J.J., Qu K., Chang H.Y. // Nat. Biotechnol. 2014. V. 32. № 9. P. 933–940. doi: 10.1038/nbt.2943.
18. Chu H.-P., Cifuentes-Rojas C., Kesner B., Aeby E., Lee H.- G., Wei C., Oh H.J., Boukhali M., Haas W., Lee J.T., et al. // Cell. 2017. V. 170. № 1. P. 86–101. doi: 10.1016/j.cell.2017.06.017.
19. Sridhar B., Rivas-Astroza M., Nguyen T.C., Chen W., Yan Z., Cao X., Hebert L., Zhong S. // Curr. Biol. 2017. V. 27. № 4. P. 602–609. doi: 10.1016/j.cub.2017.01.011.
20. Li X., Zhou B., Chen L., Gou L.-T., Li H., Fu X.-D. // Nat. Biotechnol. 2017. V. 35. № 10. P. 940–950. doi: 10.1038/nbt.3968.
21. Bell J.C., Jukam D., Teran N.A., Risca V.I., Smith O.K., Johnson W.L., Skotheim J.M., Greenleaf W.J., Straight A.F. // eLife. 2018. V. 7. P. e27024. doi: 10.7554/eLife.27024.
22. Yan Z., Huang N., Wu W., Chen W., Jiang Y., Zhang J., Zhang X., Zhang Y., Zhang Q., Zhang L., et al. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 8. P. 3328–3337. doi: 10.1073/pnas.1819788116.
23. Bonetti A., Agostini F., Suzuki A.M., Hashimoto K., Pascarella G., Gimenez J., Roos L., Nash A.J., Ghazanfar S., Carninci P., et al. // Nat. Commun. 2020. V. 11. № 1. P. 1018. doi: 10.1038/s41467-020-14337-6.
24. Gavrilov A.A., Zharikova A.A., Galitsyna A.A., Luzhin A.V., Rubanova N.M., Golov A.K., Petrova N.V., Kantidze O.L., Ulianov S.V., Misteli T., et al. // Nucl. Acids Res. 2020. V. 48. № 12. P. 6699–6714. doi: 10.1093/nar/gkaa456.
25. Ryabykh G.K., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // Mol. Biol. 2022. V. 56. № 2. P. 210–228. doi: 10.1134/S002689332202017X.
26. Ryabykh G.K., Dikstein N., Zharikova A.A., Galitsyna A.A., Ulianov S.V., Razin S.V., Gavrilov A.A. // Database J. Biol. Databases Curation. 2023. V. 2023. P. baad025. doi: 10.1093/database/baad025.
27. Gavrilov A.A., Golov A.K., Luzhin A.V., Zharikova A.A., Galitsyna A.A., Rubanova N.M., Kantidze O.L., Ulianov S.V., Razin S.V., Misteli T., et al. // Proc. Natl. Acad. Sci. USA. 2022. V. 119. № 1. P. e2116222119. doi: 10.1073/pnas.2116222119.
28. Mylarshchikov D.E., Baumgart S.J., Mazurov E., Saveliev A., Medvedeva Y.A. // NAR Genomics Bioinforma. 2024. V. 6. № 2. P. lqae054. doi: 10.1093/nargab/lqae054.
29. Hartana C.A., Rassadkina Y., Gao C., Yashiro-Ohtani Y., Mercier F., Deng S., Li X., Lin 7. R.H., Das S., Lian C., et al. // J. Clin. Invest. 2021. V. 131. № 9. P. e146136. doi: 10.1172/JCI146136.
30. Chen W., Zhang J., Xu H., Dai X., Zhang X., Wang W., Wu Y., Li Y., Wang X., Zhang Y., et al. // Oncogene. 2024. V. 43. № 18. P. 1328–1340. doi: 10.1038/s41388-024-02986-2.
31. Li B., Jiang Q., Liu X., Yang X., Li Z., Li J., Zhang Y., Wu Y., Li X., Zhang Y., et al. // Cell Death Dis. 2020. V. 11. № 10. P. 823. doi: 10.1038/s41419-020-03031-6.
32. Li Z., Guo X., Wu S. // Stem Cell Res. Ther. 2020. V. 11. № 1. P. 435. doi: 10.1186/s13287-020-01953-8.
33. Nylund P., Gimenez G., Laine I., Massinen S., Gordon S., Kuusela M., Kontturi S., Rusanen A., Vettenranta K., Lohi O., et al. // Haematologica. 2024. V. 109. № 2. P. 567–577. doi: 10.3324/haematol.2022.282375.
34. Ang C.E., Trevino A.E., Chang S., Soeung M., Ma L., Chidambaram S., Young C., Wernig M., Sudhof T.C., Chang H.Y., et al. // eLife. 2019. V. 8. P. e41770. doi: 10.7554/eLife.41770.
35. Wongtrakoongate P., Riddick G., Hashem O., Harris R., Jones S., Ramjaun A., Bunjobpol W., Felsenfeld G., Turner B.M., Schwabe J.W.R., et al. // PLoS Genet. 2015. V. 11. № 10. P. e1005615. doi: 10.1371/journal.pgen.1005615.
36. Alfeghaly C., Bozec D., Forterre M., Muller Q., Bernard D., Munch M., Schulte M.L., Nollet M., Paul N., Regnault B., et al. // Nucl. Acids Res. 2021. V. 49. № 9. P. 4954–4970. doi: 10.1093/nar/gkab267.
37. Huang Y., Zheng Y., Jia L., Li W. // Sci. Rep. 2017. V. 7. № 1. P. 8080. doi: 10.1038/s41598-017-08131-6.
38. Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. // Nucl. Acids Res. 2016. V. 44. № D1. P. D457–D462. doi: 10.1093/nar/gkv1070.
39. Kolberg L., Raudvere U., Kuzmin I., Vilo J., Peterson H. // Nucl. Acids Res. 2023. V. 51. № W1. P. W207–W212. doi: 10.1093/nar/gkad347.
40. Wan L., Zhang L., Fan K., Cheng Z.X., Sun Q.C., Wang J.J. // Mol. Cancer Ther. 2016. V. 15. № 5. P. 1082–1094. doi: 10.1158/1535-7163.MCT-15-0509.
41. Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler D. // Genome Res. 2002. V. 12. № 6. P. 996–1006. doi: 10.1101/gr.229102.
42. Antonov I.V., Mazurov E., Borodovsky M., Medvedeva

Y.A. // Brief. Bioinform. 2019. V. 20. № 2. P. 551–564. doi: 10.1093/bib/bbx132.
43. Antonov I., Medvedeva Y.A. // Genes. 2020. V. 11. № 12. P. E1483. doi: 10.3390/genes11121483.
44. Matveishina E., Antonov I., Medvedeva Y.A. // Int. J. Mol. Sci. 2020. V. 21. № 3. P. 830. doi: 10.3390/ijms21030830.
45. Ogunleye A.J., Romanova E., Medvedeva Y.A. // F1000Research. 2021. V. 10. P. 204. doi: 10.12688/f1000research.51844.1.
46. Marakulina D., Vorontsova Y., Mazurov E., Anisenko A., Korchagina A., Medvedeva Y.A., Kolmykov S., Fishman V., Kulakovskiy I.V., Meshcheryakov R., et al. // Nucl. Acids Res. 2023. V. 51. № D1. P. D564–D570. doi: 10.1093/nar/gkac809.
47. Favorov A., Mularoni L., Cope L.M., Medvedeva Y.A., Mironov A.A., Makeev V.J., Wheelan S.J. // PLoS Comput. Biol. 2012. V. 8. № 5. P. e1002529. doi: 10.1371/journal.pcbi.1002529.
48. Stavrovskaya E.D., Tepliuk N., Mironov A.A., Panchenko A.R., Favorov A.V., Makeev V.J. // Bioinforma. Oxf. Engl. 2017. V. 33. № 20. P. 3158–3165. doi: 10.1093/bioinformatics/btx379.
49. Gel B., Díez-Villanueva A., Serra E., Buschbeck M., Peinado M.A., Malinverni R. // Bioinforma. Oxf. Engl. 2016. V. 32. № 2. P. 289–291. doi: 10.1093/bioinformatics/btv562.