communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07815-4

Mechanism of expression regulation of head-to-head overlapping protein-coding genes *INO80E* and *HIRIP3*

Natalia Ryczek [1], Aneta Łyś[1], Elżbieta Wanowska [1], Joanna Kozłowska-Masłoń [1,2] & Izabela Makałowska [1]

Although the existence of overlapping protein-coding genes in eukaryotic genomes is known for decades, their role in regulating expression remains far from fully understood. Here, the mechanism regulating the expression of head-to-head overlapping genes, a pair of *INO80E* and *HIRIP3* genes is presented. Based on a series of experiments, we show that the expression of these genes is strongly dependent on sense/antisense interactions. The overlapping transcripts form an RNA:RNA duplex that has a stabilizing effect on the mRNAs involved, and this stabilization may be mediated by the ELAVL1 protein. We also show that the transcription factor RARG is important for the transcription of both genes studied. In addition, we demonstrate that the overlapping isoform of *INO80E* forms an R-loop that may positively regulate *HIRIP3* isoforms. We propose that both structures, dsRNA and R-loops, help to keep the DNA loop open to allow the transcription of the remaining variants of both genes. However, experiments suggest that RNA:RNA duplex formation plays a major role, while R-loops play only a complementary one. The absence of this dsRNA structure leads to the loss of a stable DNA opening and consequently to transcriptional interference.

Overlaps between protein-coding genes, i.e., partial or complete sharing of genomic location, have been identified in prokaryotes as well as in eukaryotes, including humans[1–4]. There are several types of possible gene overlaps: the 5' ends (head-to-head), nested genes, the 3' end (tail-to-tail), and genes overlapping with their promoter regions. Antisense transcription of head-to-head overlapping protein-coding (PC) genes can have several consequences, none of which have been comprehensively investigated and explained. One of these is the possible transcriptional interference, which can cause aberrations in the expression of overlapping genes[5]. Possible types of transcriptional interference include polymerase collision, the so-called 'sitting duck' interference, or occlusion[6]. It is also plausible that two genes that overlap head-to-head are competing for bidirectional promoters[7]. The results of some studies suggest that head-to-head gene overlap has a downregulatory effect on the expression of the genes involved[8,9]. However, ours and other studies did not show any negative effect of gene overlap. In contrast, a positive correlation of antisense transcripts was observed[10,11].

The transcripts of overlapping genes can regulate transcription at the level of RNA–DNA and RNA–RNA interactions. Examples of the first one are DNA methylation and demethylation[12]. Chromatin modification and silencing of the sense promoter have also been demonstrated[13]. Another mode of regulation is the formation of R-loops. The term R-loop refers to the DNA-RNA hybrid structure and a displaced single-stranded DNA (ssDNA)[14]. It has been shown that R-loop formation can have negative and positive effects on gene expression[15–24] and can also cause genome instability[24–26]. Studies have also shown that head-to-head transcription promotes the formation of an R-loop structure, which has been associated with positive expression regulation when two genes are in close proximity on opposite DNA strands[27,28]. However, the regulatory effects of the formation of R-loops were confirmed only for overlapping pairs of protein-coding genes and long noncoding RNAs (lncRNAs)[29].

The complementarity of two antisense transcripts can lead to RNA–RNA interactions by the formation of RNA:RNA duplexes[30]. Duplexes between protein-coding mRNAs and lncRNAs are well documented[31–33], and there are many reports explaining their regulatory function[31,32,34]. Such a duplex can physically hide access to splicing sites[35], affect transcript transport[36], contribute to endo-siRNA[37], or stabilize protein-coding sense transcripts by blocking[38] or by masking microRNA target sites[39]. It is also possible that the formation of duplexes allows the DNA strands to unwind, increasing the access of RNA Pol II to promoters, and thus increasing transcription. The phenomenon of RNA:RNA duplex formation from two complementary fragments of protein-coding gene transcripts has not yet been extensively studied, and the occurrence of

[1]Institute of Human Biology and Evolution, Faculty of Biology, Adam Mickiewicz University, ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland. [2]Laboratory of Cancer Genetics, Greater Poland Cancer Centre, ul. Garbary 15, 61-866 Poznań, Poland. e-mail: izabela.makalowska@amu.edu.pl

duplex formed by two mRNAs and its functions have so far been described only for genes *TP53* and *WRAP53*[40]. However, substantial levels of different transcript duplexes are not expected in the cytoplasm as they may trigger immune responses.

An overlap at the 5' ends between two PC genes can be either fixed or intermittently present. This depends mainly on the usage of alternative transcription start sites (TSS)[11,41,42], which dynamically changes under the influence of various intracellular and extracellular conditions[11,43,44]. TSSs switches in certain genes have been observed to have a translation regulatory function, for example, in cancers[45,46], during cerebellar development[47] and in neural differentiation[48].

Despite a number of studies, the functional implications of overlapping transcripts and their role in expression regulation are not fully understood, and the results of these studies do not provide clear answers[8,9,49]. Our previous studies demonstrated that genes that use TSSs that overlap with the gene located on the opposite DNA strand had, on average, higher expression levels than the same genes that use only non-overlapping TSSs[11]. To investigate this and to decipher the mechanism regulating the expression of head-to-head overlapping genes, we studied a pair of overlapping genes: *INO80E* and *HIRIP3*. The *INO80E* gene encodes a protein involved in chromatin remodeling processes as a part of the multi-subunit INO80-chromatin remodeling complex[50–52]. It is located on chromosome 16 and overlaps head-to-head with the *HIRIP3* gene encoding protein, which binds the H2B and H3 core histones and the HIRA protein to regulate chromatin and histone metabolism[53–55].

Results

INO80E and *HIRIP3* genomic organization and expression

According to the Ensembl[56] and NCBI[57] databases, the *INO80E* gene has fourteen mRNA isoforms, one of which is elongated at the 5' end (INO80E-201). The *HIRIP3* gene has seven mRNA isoforms, of which four overlap with the INO80E-201 isoform. To confirm these annotations, especially the presence of 5' extended transcripts, we analyzed 31 ENCODE long-read RNA-seq (LR-RNA-seq) libraries from the Pacific Biosciences sequencing platform[58] (Supplementary materials, Table S1) and found an additional isoform with even longer 5' UTR. Analysis has also revealed that the protein-coding HIRIP3-201 transcript has an alternative TSS, which is more downstream and has a longer 5' UTR (marked as HIRIP3-PB1) (Fig. 1A, B and Supplementary materials, Fig. S1). In addition, we checked the presence of TSSs of the *HIRIP3* and *INO80E* genes using Fantom5 Cap Analysis of Gene Expression (CAGE) data[59,60]. These results are consistent with isoforms' annotations complemented with transcripts identified in LR-RNA-seq data (Fig. 1B). Isoforms selected for further studies are presented in Fig. 1A. Their presence in HEK293T cells was validated by RT-PCR experiments (Fig. 1C–F) and sequencing. When isoforms differed only in the length of 5'UTR and specific starters could not be designed for both isoforms, their presence was confirmed by a comparison of the band's intensity using ImageJ[61] software.

In further studies, for simplification, we divided the isoforms of both genes into two categories: *INO80E*-O (isoforms overlapping with *HIRIP3*), *INO80E*-NO (not overlapping *INO80E* isoforms), *HIRIP3*-O (overlapping isoforms of *HIRIP3* including *HIRIP3-PB1*), and *HIRIP3*-S (HIRIP3-201) (Fig. 1A). The isoforms HIRIP3-203 and HIRIP3-205 did not appear to be significant in further experiments. Therefore, all the results regarding these isoforms are presented in the Supplementary Materials (Fig. S2).

We estimated and compared the abundance of overlapping and non-overlapping isoforms in HEK293T cells. The results of absolute quantification, based on the standard qPCR curve, showed that the *HIRIP3* gene is expressed at higher levels than the *INO80E* gene (Fig. 1G) and that the quantitative ratio of *INO80E*-O transcript to *INO80E*-NO transcripts is 1:766. This very low level of *INO80E*-O expression may indicate a regulatory rather than a protein-coding role. When compared to the expression of combined *HIRIP3*-S and *HIRIP3*-PB1 transcripts, the expression of *HIRIP3*-O transcripts was lower. The ratio was 1:4 (Fig. 1G) which indicated that the expression of these two isoforms is much higher than the expression of all overlapping isoforms.

It has been suggested that overlapping genes have a negative expression correlation due to possible transcriptional interference[8,9]. Using the results of our previous studies of 784 RNA-seq samples from different tissues and cell lines[62], we performed an analysis of the expression correlation of these genes in non-cancerous and cancerous tissues/cell line samples. In both cases, a strong and statistically significant positive correlation was observed (Fig. 1H). Similarly, a positive expression correlation was found when the data was divided according to the tumor type and organ system (Supplementary materials, Fig. S3A, B).

To investigate the mutual influence of *INO80E* and *HIRIP3* on their expression pattern, we performed knock-down experiments with antisense locked nucleic acids (LNA®) GapmeRs (Qiagen) followed by qRT-PCR. We designed three LNA® GapmeRs, which targeted all *HIRIP3* transcripts, all *INO80E* transcripts, and *INO80E*-O isoforms, respectively (Fig. 2A). After silencing all *HIRIP3* transcripts, the relative expression of the *INO80E*-O decreased to minimal levels, while the expression of the rest of the gene transcripts, *INO80E*-NO, increased (Fig. 2B). When the *INO80E* gene was knocked-down, the expression of overlapping *HIRIP3*-O isoforms decreased by almost 60%. However, there was a ten-fold increase in the expression of the *HIRIP3*-S + PB1 transcripts (Fig. 2C).

Silencing of 5'extended transcripts - *INO80E*-O resulted in an almost 70% decrease in the expression of the *HIRIP3*-O isoforms and ten times increased expression of the *HIRIP3*-S + PB1 transcript (Fig. 2D). Interestingly, the expression of all *INO80E*-NO isoforms decreased by almost 90%. These results show that the presence of the *INO80E*-O isoforms is required for a greater expression of not overlapping isoforms of *INO80E* as well as *HIRIP3*-O. All knock-down results, with tested control samples, are shown in the Supplementary Materials (Fig. S4A–D). Decreased expression of all *HIRIP3*-O transcripts and increased levels of *HIRIP3*-S + PB1 also suggest that after a knock-down of both total *INO80E* and *INO80E*-O, there was an increase in transcription of the *HIRIP3*-S isoform, but a decrease in the overlapping HIRIP3-PB1. This was confirmed by an RT-PCR experiment (Fig. 2E), which showed that after silencing of *INO80E*-O, there was a significant shift to more downstream TSS and an increased transcription of *HIRIP3*-S. Taking together, the silencing experiments suggest a regulatory role of *INO80E*-O.

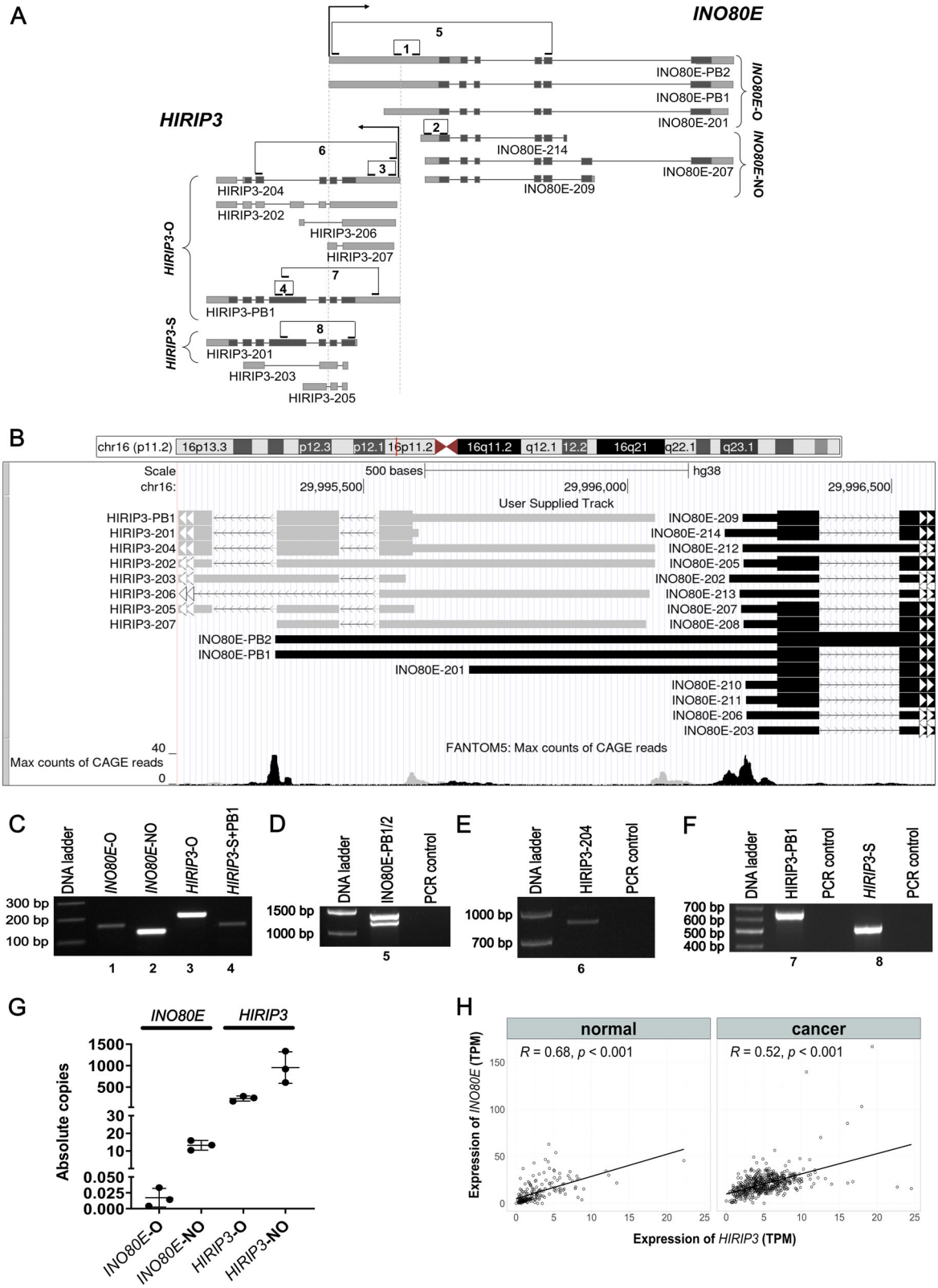
Considering the above results, we checked the expression correlation of *HIRIP3*-S and *INO80E*-O transcripts using the same RNA-seq data from normal and cancerous tissues. The correlation coefficient suggests that the expression of these two transcripts is not correlated in normal cells, which might seem contradictory to our result. However, Fig. 2F shows that the expression values of both transcripts are aligned along the axes. This means that a high expression of *INO80E*-O is mostly associated with a very low expression of *HIRIP3*-S and vice versa. It seems that any increase in the expression of one transcript is coupled with a minimized expression level of another. A different phenomenon is observed in cancer tissues, where there is a small but significant positive expression correlation.

The overlapping mRNA isoforms of *INO80E* gene are localized in the cell nucleus

The obtained results indicated a regulatory function of the overlapping mRNA isoforms. If so, these isoforms should be localized predominantly in the nucleus. To confirm this, their localization was verified by subcellular fractionation followed by qRT-PCR. *INO80E*-O was found to be located in the nucleus and associated with chromatin (Fig. 3A), confirming its putative regulatory function. All other transcripts of the *INO80E* gene are found in the greatest amount in the cytoplasm (Fig. 3A). *HIRIP3* isoforms are similarly abundant in the cytoplasm and the nucleus. Confirmation of the purity of the RNA fractions is presented in Supplementary Fig. S5A–C.

Overlapping isoforms of *INO80E* and *HIRIP3* form RNA:RNA duplexes

The complementary 5' UTR fragments of the overlapping transcripts may interact with each other and form RNA:RNA duplexes. This has already been demonstrated but mainly based on bioinformatic analyses of high-



throughput sequencing data[63–65]. The only verified example comes from the studies of *WRAP53* and *TP53* genes[40]. We tested whether *INO80E* and *HIRIP3* form the RNA:RNA duplex using an RNA antisense purification (RAP) assay (Fig. 3B), followed by PCR detection[66]. The 4'-aminomethyltrioxsalen (AMT)-fixed RNA:RNA duplexes were captured using a biotinylated oligonucleotide probe complementary to the mRNA isoforms of the *HIRIP3* gene (Fig. 3C and Supplementary Materials, Table S2). The pull-down effect was confirmed by RT-qPCR (Supplementary Materials, Fig. S6). The subsequent PCR experiment detected the *INO80E* gene, confirming the presence of RNA:RNA duplexes (Fig. 3D).

Complementary tethering of transcripts may, among other things, have a stabilizing effect and could delay mRNA degradation. To confirm

Fig. 1 | *INO80E* and *HIRIP3* genomic organization and expression. A Transcripts of the *HIRIP3* and selected transcripts of the *INO80E* gene pair expressed in HEK293T cells. The *HIRIP3*-O and *INO80E*-O transcripts overlap at the 5' ends. Light gray represents non-coding exon sequences, while dark gray – coding exon sequences. Numbers represent primers used for amplification of: 1 – *INO80E*-O transcripts; 2 – *INO80E*-NO transcripts; 3 – *HIRIP3*-O transcripts; 4 – HIRIP3-201/HIRIP3-PB1 transcripts; 5 – INO80E-PB1/PB2 transcripts; 6 – HIRIP3-204; 7 – HIRIP3-PB1; 8 – HIRIP3-201. **B** Fantom5 results of CAGE analysis showing TSSs of the *HIRIP3* (gray) and *INO80E* (black) genes. **C** *HIRIP3* and *INO80E* transcripts expression in HEK293T cells. (DNA ladder - GeneRuler 100 bp DNA Ladder (Thermo Scientific)). **D** Confirmation of INO80E-PB1/PB2 mRNA isoforms presence (DNA ladder - GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). **E** Confirmation of HIRIP3-204 mRNA isoform presence (DNA ladder - GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific)). **F** Confirmation of HIRIP3-201/PB1 mRNA isoforms presence (DNA ladder - GeneRuler 100 bp DNA Ladder (Thermo Scientific). **C**–**F** Numbers below the gel image represent the numbers of primer pairs used for amplification presented in Fig. 1A. **G** Absolute quantification of *HIRIP3* and *INO80E* transcripts in HEK293T cells (data are represented as mean ± SD; $n = 3$). **H** Correlation between *INO80E* gene and *HIRIP3* gene expression (TPM) in cancer and non-cancer cells.

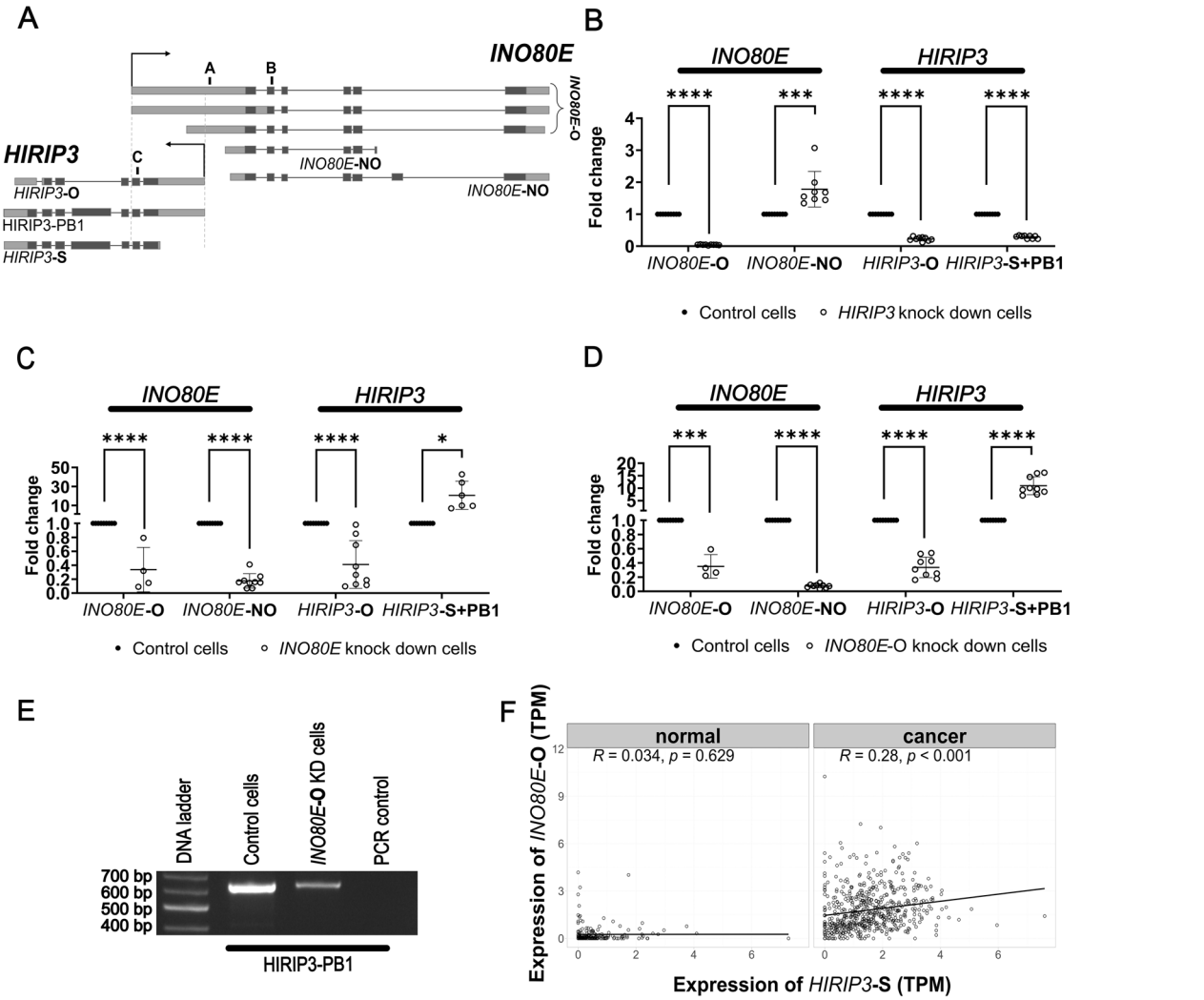


Fig. 2 | The expression relationship between the *INO80E* and *HIRIP3* genes. A Localization of three LNA® GapmeRs designed to target the transcripts of the *HIRIP3* (gapmeR C), *INO80E* (gapmeR B), and *INO80E*-O isoform (gapmeR A). **B** Fold change of *HIRIP3* and *INO80E* transcripts expression after silencing *HIRIP3* (data were represented as mean ± SD; ***$p < 0.001$; ****$p < 0.0001$; $N = 3$, $n = 3$). **C** Fold change of *HIRIP3* and *INO80E* transcripts expression after silencing *INO80E* (data are represented as mean ± SD; *$p < 0.05$; ****$p < 0.0001$; $N = 3$, $n = 3$). **D** Fold change of *HIRIP3* and *INO80E* transcripts expression after silencing *INO80E*-O (data are represented as mean ± SD; ***$p < 0.001$; ****$p < 0.0001$; $N = 3$, $n = 3$). **E** *HIRIP3*-PB1 expression in control cells and after *INO80E*-O knock-down cells (DNA ladder - GeneRuler 100 bp DNA Ladder (Thermo Scientific). **F** Correlation between *INO80E*-O and *HIRIP3*-S mRNA isoforms expression (TPM) in cancer and non-cancer tissues/cells.

this, two types of cells were treated with actinomycin D: control cells and cells transfected for 72 h with gapmeR targeting the *INO80E*-O isoform. After treatment, the expression of *HIRIP3* gene transcripts was checked at four time points (1, 2, 3 and 8 h) and compared to cells not treated with actinomycin D (ActD). This experiment showed that after the silencing of *INO80E*-O, the overlapping *HIRIP3*-O isoforms are degraded faster than under controlled conditions without the silencing of *INO80E*-O. The effect was seen 3 h after the addition of ActD (Fig. 3E).

RNA:RNA duplexes may also provide a scaffold for the proteins required for the correct expression and processing of overlapping genes. We used RAP coupled with mass spectrometry (MS) analysis to investigate which proteins bind to paired transcripts. HEK293T cells were UV-

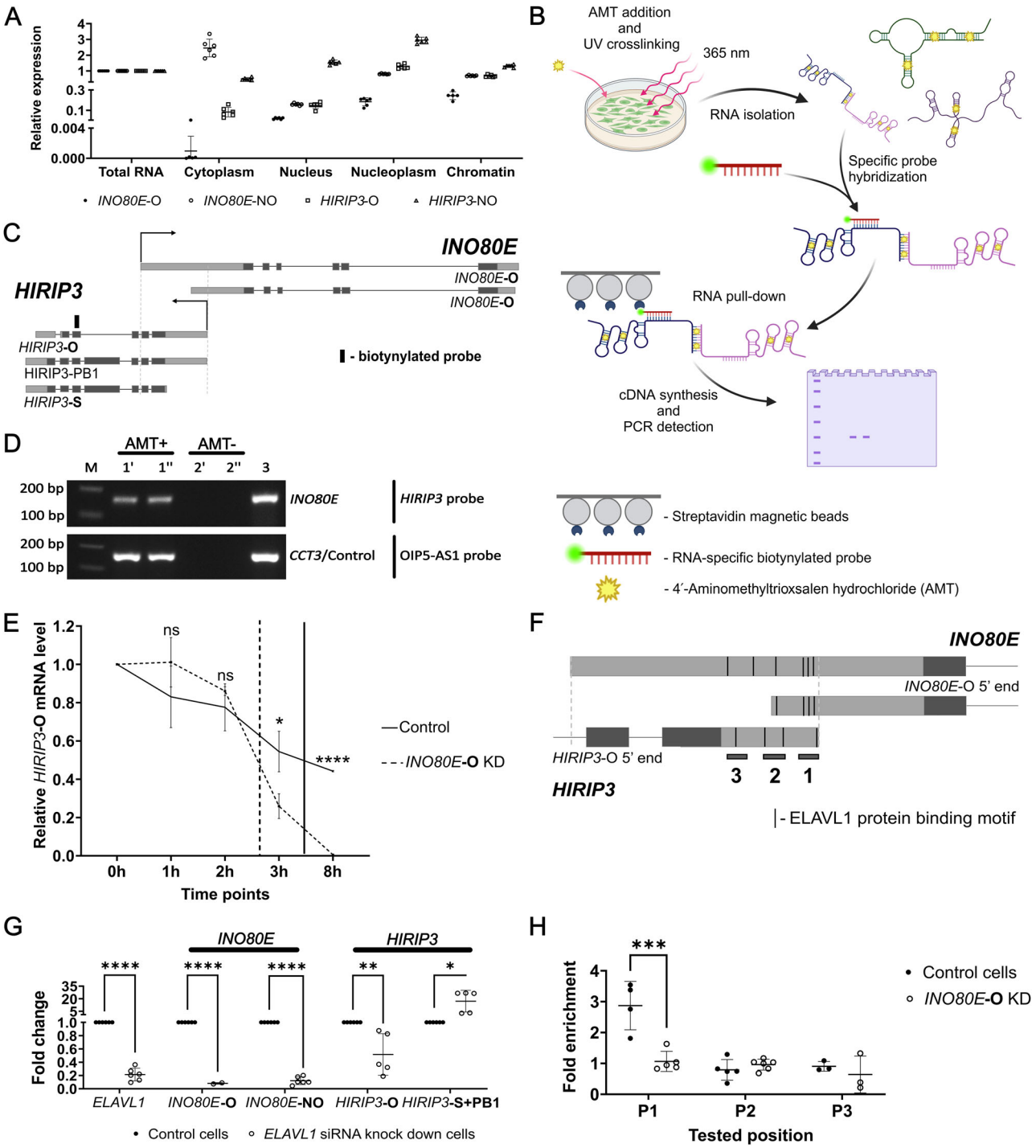


Fig. 3 | Overlapping *INO80E*-O and *HIRIP3*-O isoforms form RNA:RNA duplexes in the cell nucleus and are stabilized by ELAVL1 protein. A Location of *INO80E* and *HIRIP3* transcripts in control/standard conditions (data are represented as mean ± SD; $N = 2$, $n = 3$). **B** RAP assay workflow. Double-stranded RNA is crosslinked in HEK293T cells with UV 365 nm in the presence of AMT. Subsequent pull-down of specific RNAs is performed using biotinylated probes complementary to the target RNA and purified on magnetic streptavidin beads. The RNA interactions are then detected by PCR after transcription of RNA into cDNA. Created in BioRender. Wanowska, E. (2025) https://BioRender.com/s27t743. **C** Location of the biotinylated probe used in the RAP experiment. **D** The results of the RAP followed by PCR detection of captured transcripts. The 1,5% agarose gel electrophoresis results. PCR detection of the *INO80E* gene was performed to confirm duplexes between *INO80E*/*HIRIP3* after RAP and *CCT3* to confirm positive control duplexes between *CCT3*/*OIP5-AS1*. M - Thermo Scientific™ GeneRuler 100 bp Plus DNA Ladder (100–3000 bp), 1', 1'' – PCR products on cDNA from HEK293T treated with AMT crosslinker, two biological repetitions; 2', 2'' – PCR products on cDNA from HEK293T without crosslinker, two biological repetitions; 3 – control cDNA from total HEK293T RNA. **E** Degradation of *HIRIP3*-O isoforms in HEK293T cells in control/standard conditions and after silencing the *INO80E*-O isoform (data were represented as mean ± SD; ns not significant; $^{*}p < 0.05$; $N = 3$, $n = 3$, vertical lines mark the isoforms half-lives in two conditions). **F** ELAVL1 protein binding motifs predicted in 5' UTR sequences of the *HIRIP3*-O and *INO80E*-O isoforms (marked in black vertical lines). Numbers 1-3 mark the places in which CLIP confirmation of ELAVL1 binding was performed. **G** *ELAVL1* gene siRNA knock-down effect on the *INO80E* and *HIRIP3* genes (data were represented as mean ± SD; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{****}p < 0.0001$; $N = 2$; $n = 3$). **H** ELAVL1 protein binding to the *HIRIP3* 5'UTR region (tested positions are marked in Fig. 3F) tested in two conditions: control and *INO80E*-O knock-down cells (data were represented as mean ± SD; $^{***}p < 0.001$; $N = 2$; $n = 3$).

Table 1 | Proteins associated with *INO80E*:*HIRIP3* RNA:RNA duplex

Protein symbol	Protein full name	Score
TADA2B	Transcriptional adapter 2-beta	1479
CNOT9	CCR4-NOT transcription complex subunit 9	516
NPM1*	Nucleophosmin	409
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	407
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	267
LSM4	LSM4 homolog, U6 small nuclear RNA, and mRNA degradation associated	251
HNRNPH1	Heterogeneous nuclear ribonucleoprotein H1	214
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	189
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	98
SSBP1*	Single-stranded DNA binding protein	96
HNRNPH3	Heterogeneous nuclear ribonucleoprotein H3	93
PARP1*	Poly [ADP ribose] polymerase 1	88
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	79
KHSRP #	KH-type splicing regulatory protein	72
HNRNPF	Heterogeneous nuclear ribonucleoprotein F	71
ELAVL1	ELAV-like RNA-binding protein 1	66
SON	SON DNA and RNA-binding protein	57
HNRNPM #	Heterogeneous nuclear ribonucleoprotein M	56
TARDBP #	TAR DNA binding protein 43	44
DDX39A #	ATP dependent RNA helicase DDX39A (DExD box helicase 39A)	43
SNRPF	Small nuclear ribonucleoprotein F	42
HNRNPA3	Heterogeneous nuclear ribonucleoprotein A3	42
NONO #	Non-POU domain-containing octamer binding protein	41
SNRPD2	Small nuclear ribonucleoprotein Sm D2	41
PCBP2	Poly(rC)-binding protein 2	37

*R-loop related proteins; # proteins related to R-loop and involved in splicing and mRNA stability; without mark proteins involved in splicing and mRNA stability; the protein score is derived from the ions scores; $N = 2$).

crosslinked and subjected to a RAP procedure using biotinylated probes similar to the capture of RNA:RNA duplexes (Supplementary Materials, Table S2). Here, double cross-linking was used, one for RNA:RNA duplexes and another for cross-linking RNA with proteins. The captured and purified proteins were then analyzed by mass spectrometry. This experiment revealed two main groups of proteins: ones related to R-loops and ones involved in splicing and mRNA stability (Table 1).

The formation of an RNA:RNA duplex may block the formation of the secondary mRNA structure at the 5' ends, exposing the binding sites to proteins with a stabilizing function. One such protein, ELAV-like RNA-binding protein 1 - ELAVL1, was identified by RAP-MS. ELAVL1 binds selectively to AU-rich regions which signal mRNA degradation[67,68]. Using the Scan for Motifs bioinformatics tool[69] and RNA-Binding Protein DataBase (RBPDB)[70], we confirmed the presence of ELAVL1 protein binding motifs in the 5' UTR sequences of the *HIRIP3* and *INO80E* genes (Fig. 3F).

To test whether this protein affects the stability of the *HIRIP3* and *INO80E* transcripts, siRNA silencing of the *ELAVL1* gene was performed. This resulted in a decreased expression of the *INO80E*-O, *INO80E*-NO, and *HIRIP3*-O isoforms, suggesting that the RNA:RNA duplex may be stabilized by the ELAVL1 protein. However, the expression of the *HIRIP3*-S + PB1 isoforms increased significantly, which is consistent with *INO80E*-O and *HIRIP3* silencing experiments (Fig. 3G and Supplementary materials - Fig. S7). To confirm the binding of the ELAVL1 protein in the 5'UTR region of the *HIRIP3*-O isoforms, a cross-linking and immunoprecipitation (CLIP) assay was performed. The results showed that there is a significant enrichment of ELAVL1 protein at the very 5' end of the *HIRIP3*-O transcript in the presence of an RNA:RNA duplex (control) compared to the cell lines with *INO80E*-O silencing and, therefore, without an RNA:RNA duplex (*INO80E*-O knock-down) (Fig. 3H). This indicates that ELAVL1 protein-mediated stabilization occurs in the presence of RNA:RNA duplexes.

Overlapping mRNA isoforms of the *INO80E* gene are involved in R-loop formation

The presence of proteins involved in the regulation of R-loops in RAP-MS led to the hypothesis that R-loops may participate in the regulation of the expression of the genes studied. A search of the R-loopBase database[71] revealed the presence of an R-loop in the 5' UTR region of the *INO80E*-O isoforms. The presence of these structures was then confirmed by a native bisulfite sequencing method (Fig. 4A and Supplementary materials - Fig. S8)[72].

The presence of an R-loop within the 5' UTR of the overlapping transcript may cause RNA polymerase II (Pol II) to pause and block *INO80E*-O transcription. It may also promote transcription of the remaining *INO80E*-NO and *HIRIP3* isoforms without the risk of transcriptional interference. To test this hypothesis, we performed an experiment using human RNase H1 overexpression. We transfected HEK293T cells with a pEGFP-RNASEH1 plasmid (Addgene) and checked the effect using a microscope with fluorescence detection (Supplementary materials, Fig. S9). Cells expressing a green fluorescent protein (GFP) were collected and the effectiveness of RNase H1 overexpression was checked by qRT-PCR. The level of RNase H1 increased more than 5,500 times compared to control cells (Fig. 4B). Overexpression of RNase H1 resulted in a significantly increased expression of the *INO80E*-O transcript involved in R-loop formation and overlapping *HIRIP3*-O isoforms. No significant changes in expression were observed for the non-overlapping *INO80E*-NO and the *HIRIP3*-S + PB1 isoforms (Fig. 4C). This suggests that the formation of an R-loop blocks the transcription of the *INO80E*-O isoform, which in turn leads to faster degradation of the overlapping *HIRIP3*-O isoforms.

An observed increase in the expression of the *INO80E*-O isoform can be expected since the R-loops block the binding of the polymerase to the promoter. A forming R-loop may also prevent the formation of RNA:RNA duplexes and, thus, a faster degradation of overlapping *HIRIP3*-O isoforms. However, the lack of an effect on *HIRIP3*-S + PB1 transcripts was somewhat surprising, as many reports indicate a positive effect of R-loops on genes located on the opposite DNA strand[27,73].

Chromatin accessibility in the absence of overlapping *INO80E* gene isoforms

Based on the results obtained so far, we hypothesize that RNA:RNA duplexes help to keep the chromatin more accessible and reduce the possibility of transcriptional interference. We performed a polymerase activity study using a chromatin immunoprecipitation (ChIP) experiment with an RNA polymerase II antibody in wild-type cells and cells 72 h after silencing of the *INO80E*-O isoforms (Supplementary materials - Fig. S10). The results of the experiment showed a fold enrichment of RNA polymerase II in the promoter region of the *HIRIP3* and *INO80E* genes after silencing the *INO80E*-O isoforms. A significant decrease in RNA polymerase II was observed in the body of *INO80E* (Fig. 4D, E). However, an enrichment of polymerase II RNA was detected in the exon specific for the *HIRIP3*-

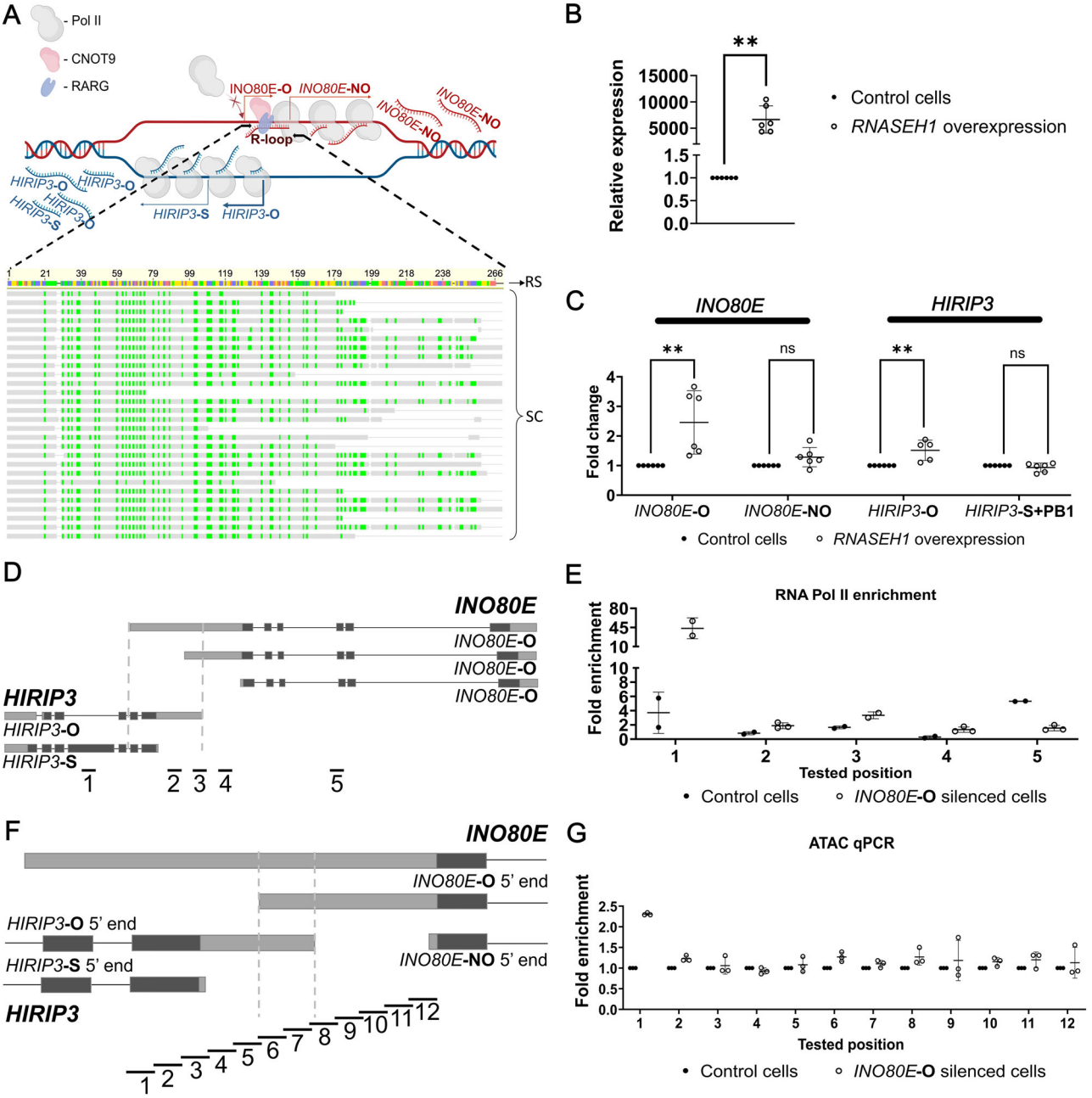


Fig. 4 | *INO80E*-O isoform is involved in R-loop formation, and the absence of *INO80E*-O isoform leads to transcriptional interference. A R-loop formation within the *INO80E* gene isoform *INO80E*-O results from the native bisulfite sequencing R-loop method. The sequenced clones (SC) are compared to the *INO80E*-O isoform encoding sequence (RS – reference sequence). The SC results were obtained through molecular cloning of PCR product into pGEM-T easy plasmid and Sanger sequencing. CNOT9 protein is light pink color, RARG is purple, and Pol II is gray. *INO80E* isoforms are red and *HIRIP3* isoforms are blue. C to T nucleotide changes are shown in green in the sequenced clones. The diagram is based on one-quarter of the figure created in BioRender. Wanowska, E. (2025) https://BioRender.com/j29t536. **B** *RNaseH1* gene overexpression confirmation in HEK239T cells after FACS) ($N = 2$; $n = 3$). **C** RNase H1 overexpression influence on *INO80E* and *HIRIP3* genes expression (data were represented as mean ± SD; ns not significant; $^{**}p < 0.01$; $N = 2$; $n = 3$). **D** Location of the RNA Polymerase II binding detection primers for chromatin immunoprecipitation with an RNA Polymerase II antibody. **E** Chromatin immunoprecipitation with an RNA Polymerase II antibody (ab10338, Abcam) (data are represented as mean ± SD; $n = 3$). The numbers indicate the location of the RNA Polymerase II binding detection primers. **F** Location of the detection primers for an assay for transposase-accessible chromatin. **G** Fold enrichment of the qPCR detection after an assay for transposase-accessible chromatin (data are represented as mean ± SD; $N = 3$; $n = 3$). The numbers indicate the location of the detection primers.

S + PB1 isoform, supporting previous results showing increased expression of *HIRIP3*-S upon silencing of the *INO80E*-O isoforms.

We also performed an assay for transposase-accessible chromatin (ATAC) followed by qPCR detection in order to check for changes in chromatin accessibility after *INO80E-O* silencing. The results showed a significant difference in chromatin accessibility around the first exon of the *HIRIP3*-S isoform. This perfectly corresponds with our previous results, which indicated an increase in *HIRIP3*-S isoform expression under such conditions. No significant differences were observed elsewhere in the studied region (Fig. 4F, G). This could be somewhat expected since although the ATAC technique is a valuable tool for identifying open chromatin regions, the variability in transposition reaction efficiency can compromise the reliability of the results when attempting to determine differentially accessible areas[74,75].

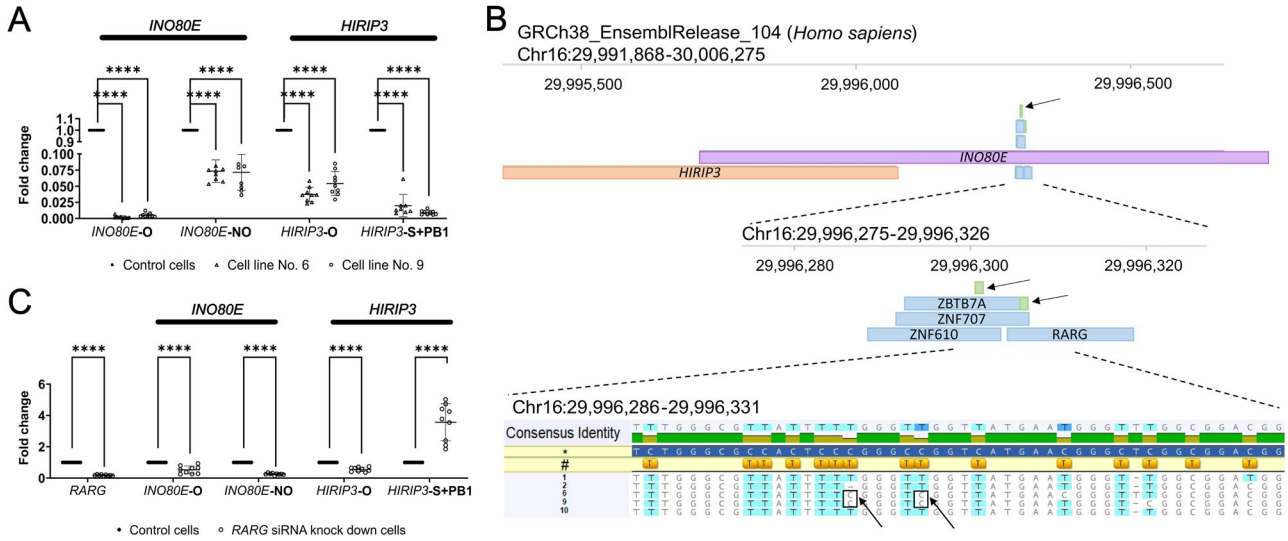


Fig. 5 | *INO80E* and *HIRIP3* promoters de novo methylation reveals the RARG transcription factor influence on *HIRIP3* and *INO80E* gene expression. A The fold change of *INO80E* and *HIRIP3* isoforms expression after promoter methylation in two selected HEK239T cell lines No.6 and No.9 (data were represented as mean ± SD; $^{****}p < 0.0001$; $N = 3$; $n = 3$). **B** Genomic localization of site-specific methylations identified in cell lines 6 and 9. 1, 2, 6, 9, 10 — numbers of tested cell lines, RARG, ZNF610, ZNF707, and ZBTB7A — target sites for these transcription factors (in blue boxes) (∗ control sequence; # Variants: nucleotide alignment). *INO80E* gene 5' end is marked with purple box and *HIRIP3* 5' end in orange box. Green boxes or black rectangles marked with arrows indicate the specific C methylations for cell lines 6 and 9. **C** *RARG* gene knock-down effect on the *INO80E* and *HIRIP3* genes (data were represented as mean ± SD; $^{****}p < 0.0001$; $N = 3$; $n = 3$).

De novo methylation of the *INO80E* promoter reveals the influence of the RARG transcription factor on the expression of the *HIRIP3* and *INO80E* genes

In the final step, we altered the promoter regions of both genes. De novo cytosine methylations were introduced into the sequence between non-overlapping *INO80E*-NO TSS and *HIRIP3*-O using dCas9-DNMT3A (Supplementary materials - Fig. S11A, B). For further analysis, we selected cell lines No. 6 and 9 with the greatest effects of methylation (Fig. 5A). Significant decreases in expression were observed for all isoforms, including *HIRIP3*-S expression. This could indicate that some common regulatory elements were methylated.

Binding sites for four transcription factors: RARG, ZNF610, ZNF707, and ZBTB7A (Fig. 5B) were identified in the location of modification sites specific to selected cell lines. We selected RARG for further investigation, since the RAP-MS experiment identified an interaction with the RARG - CNOT9 protein. The *RARG* gene was knocked down using siRNA and the effect was observed for *INO80E*-O, *HIRIP3*-O, and *INO80E*-NO, where the total expression of these isoforms decreased significantly (Fig. 5C and Supplementary materials - Fig. S12). An increased expression was again observed for *HIRIP3*-S. The results obtained indicate that the RARG protein could be involved in the regulation of the transcription of the *INO80E* and *HIRIP3* genes and are consistent with our previous results.

Discussion

Overlapping protein-coding genes are found in most organisms, including humans. The effects of such gene arrangements are seen at the DNA and RNA levels. One of the proposed consequences is a reduction in gene expression levels due to transcriptional interference. The study performed by Bendtsen et al. shows that RNA polymerase (RNAP), by binding to one promoter, can repress the activity of the overlapping promoter located on the opposite DNA strand[76]. However, not all studies support this[41,77,78]. Interference such as collision of two polymerase complexes is well demonstrated in yeast[79] and bacteria[80], but is questioned when it comes to human and other mammalian genomes[81], where genes are transcribed in short burst[82]. Johnnson et al. demonstrated that lncRNAs have much longer intervals between bursts than protein-coding genes and that genes that are divergently transcribed, including pairs of protein-coding genes, had, on average, higher burst frequencies. They also showed that lncRNAs can modulate bursts of nearby genes, but it remains to be determined if antisense transcription of lncRNA and mRNA or two mRNAs is associated with a similar bursting effect[81].

Here, we present the results of studies on *INO80E* and *HIRIP3* genes. Our experiments indicate that the overlapping *INO80E*-O transcript is mostly located in the nucleus, suggesting a regulatory function. Isoforms of the *HIRIP3* gene and non-overlapping isoforms of *INO80E* are equally abundant in all fractions. Using RAP-PCR, we have also shown that the overlapping isoforms of the *INO80E* and *HIRIP3* form an RNA:RNA duplex.

Several of our experiments showed that decreased expression of the *INO80E*-O isoforms is associated with decreased expression of the remaining *INO80E* isoforms, as well as overlapping isoforms of *HIRIP3* – *HIRIP3*-O. At the same time, the expression level of the *HIRIP3*-S isoform increased. Subsequently, the results of the *HIRIP3* gene knock-down showed that the overlapping *HIRIP3*-O isoforms may also stabilize the *INO80E*-O isoform, as its expression decreased significantly. However, the expression of the remaining *INO80E*-NO isoforms increases upon silencing of the *HIRIP3* gene.

The results of silencing experiments strongly suggest that the expression of the genes studied is dependent on sense/antisense interactions. Similar results were obtained by Mahmoudi and colleagues[40], who studied two protein-coding genes: the *TP53* gene and the antisense *WRAP53*. siRNA knock-down of *WRAP53* gene expression resulted in a significant decrease in *TP53* mRNA. Conversely, overexpression of *WRAP53* increased *TP53* mRNA and p53 protein levels. A blocking of potential *WRAP53/TP53* RNA hybrids also resulted in reduced *TP53* level. We propose that the resulting dsRNA structure acts as a stabilizer of the open DNA loop, allowing the transcription of the remaining variants of both genes. The possible stabilizing effect of the RNA:RNA duplex was confirmed by transcription termination in HEK293T cells using actinomycin D. The results of this experiment showed that the *HIRIP3*-O isoforms were less abundant in cells with a knock-down *INO80E*-O isoform, which may indicate their faster degradation.

The formation of dsRNA may additionally stabilize overlapping isoforms of the *HIRIP3* gene by suppressing mRNA folding at the 5' end, which

in turn may affect the exposure of protein binding sites. Using the Scan for Motifs[69] and RBPDB[70] databases, we found motifs characteristic of ELAVL1. We identified this protein by RAP-MS as being associated with the *INO80E:HIRIP3* duplex. These motifs are located in the 5' UTR of overlapping isoforms of both genes, downstream of the sequences involved in the RNA:RNA duplex formation and in the duplex region. At this stage, it is difficult to say whether ELAVL1 binding is necessary for duplex formation by preventing the formation of secondary structures at the 5' ends of transcripts, or whether duplex formation uncovers ELAVL1 binding sites that would otherwise be masked by hairpins. Nevertheless, the knockdown of the *ELAVL1* gene resulted in a decreased expression of all transcripts, with the exception of *HIRIP3*-S. Although the ELAVL1 protein is most often described in terms of its role in stabilizing transcripts after interacting with 3'UTR, there are also reports demonstrating a similar function of this protein by interacting with 5' UTR[83,84].

The results of the *HIRIP3* silencing experiment showed that the overlapping *HIRIP3*-O isoforms may also stabilize the *INO80E*-O isoform, as its expression significantly decreased. However, the expression of *INO80E* non-overlapping isoforms increased upon the silencing of the *HIRIP3* gene. This may indicate that silencing of the gene on one strand of DNA makes a promoter region more available for polymerases to transcribe the gene on the opposite strand. With a significantly reduced possibility of transcriptional interference, the presence of an RNA:RNA duplex may not be required for a higher expression. A possible transcriptional interference was confirmed by ChIP using an antibody directed against RNA polymerase II. After the silencing of *INO80E-O*, a decrease in RNA polymerase II was observed along the *INO80E* gene body. On the other hand, there was an enrichment of RNA polymerase in the region of the *HIRIP3*-S isoform. This is consistent with the observation of an increased expression following *INO80E*-O silencing. It is plausible that the absence of a stabilizing duplex facilitates the degradation of overlapping *HIRIP3*-O isoforms, making the downstream TSS of *HIRIP3*-S more accessible to polymerase. The ATAC assay confirmed increased accessibility of chromatin after the silencing of *INO80E*-O. Such a switch in TSS in response to a changing environment has previously been described by us based on a STAT6 silencing experiment[11].

We found that the *INO80E*-O isoforms form an R-loop and could, therefore, positively regulate the antisense *HIRIP3* gene isoforms. This type of mechanism has already been observed and described for other genes[27,28,73]. R-loop formation at the 5' end by the *INO80E*-O isoform was confirmed by native bisulfite conversion sequencing and was supported by RAP-MS results which identified proteins responsible for R-loop regulation, including the PARP1 protein[71,85–89]. However, RNase H1 overexpression experiments showed that in the absence of an R-loop, the expression of all overlapping isoforms increased. This result may indicate that the R-loop does not play a major role in the regulation of the *HIRIP3* gene and that a higher dose of *INO80E*-O leads to a higher amount of RNA:RNA duplexes, thus to a higher expression of *HIRIP3*.

We also introduced de novo methylation in the promoter region of the *HIRIP3* and *INO80E* genes using the CRISPR/Cas system. As a result, reduced expression of all the isoforms of the *HIRIP3* and *INO80E* genes was observed in the modified cell lines. Methylated cytosines were found in the binding site of four transcription factors, including RARG[90]. After the knock-down of *RARG* gene expression, a decreased expression of all *INO80E* and *HIRIP3-O* gene isoforms was observed. This suggests that RARG is involved in the regulation of the expression of both genes.

Based on our experiments, we propose that the *INO80E*-O isoforms regulate the expression of both genes by two alternative mechanisms: R-loops formation and RNA:RNA duplex formation. Both mechanisms may help chromatin to be more accessible and consequently prevent transcriptional interference. The first few hundred nucleotides at the 5' end of *INO80E*-O are complementary to the *HIRIP3*-O isoforms and can form an RNA:RNA duplex. A large part of this region is also involved in R-loops formation. It is very probable that these two structures have the same function, i.e., to maintain chromatin more accessible.

Upon the binding of a RARG transcription factor, the *INO80E*-O isoform is transcribed and an R-loop is formed, which blocks the transcription of subsequent *INO80E*-O isoforms but keeps chromatin open. This allows the polymerases to transcribe all *HIRIP3* isoforms and non-overlapping transcripts of *INO80E*-NO (Fig. 6A). When an R-loop is not formed or resolved, the 5' UTR end of *INO80E*-O is available for interaction with *HIRIP3*-O isoforms and forms an RNA:RNA duplex. The created duplex can stabilize the isoforms involved and, similar to the R-loop, keep the chromatin more accessible. In addition, the mRNA stabilizing protein ELAVL1 binds to *HIRIP3*-O or *INO80E*-O isoforms near or in the RNA:RNA duplex (Fig. 6B). ELAVL1 could have a stabilizing role, but also its binding could prevent the 5' ends of the overlapping isoforms from forming a secondary structure, which would block duplex formation (Supplementary Materials - Fig. S13).

Alternatively, both structures may be required simultaneously. It is known that R-loops stimulate transcription on the opposite DNA strand[27,28,73]. Therefore, an R-loop formed by the *INO80E*-O isoform may induce the transcription of overlapping *HIRIP3*-O isoforms. In turn, *HIRIP3*-O isoforms form a duplex with *INO80E*-O, either with its terminus (Fig. 6C) or with the longer fragment after R-loop resolution (Fig. 6B). The structure involving an R-loop and a dsRNA has already been described by the Proudfoot group[28]. However, the overexpression of RNAase H1 had no effect on the expression of the *INO80E*-NO and *HIRIP3*-S isoforms. This suggests that duplexes formation may be more important and sufficient for keeping chromatin open and for normal gene transcription. The formation of R-loops seems to play only a complementary role.

Keeping chromatin wide open is essential for undisturbed transcription and could be critical in the case of 5' overlapping genes, where transcriptional interference is very likely to occur. A lack of expression of *INO80E*-O corresponds to a lack of RNA:RNA duplexes or R-loops. Consequently, chromatin is less accessible, and polymerase cannot bind to or move freely in the promoter region (Fig. 6D). At the same time, the reduced transcription of the overlapping *HIRIP3*-O isoforms makes the downstream TSS available to the polymerase, resulting in an increased transcription of the *HIRIP3*-S isoform.

Altogether, our results indicate that the formation of an RNA:RNA duplex is necessary for a stable simultaneous expression of both genes. The absence of this dsRNA structure leads to changes in the DNA accessibility and consequently to transcriptional interference, such as described in the proposed model. These results also explain our previous finding that the use of TSSs, which lead to the overlap, is associated with higher gene expression[11]. We have also shown that transcriptional interference can occur in the case of overlapping genes. However, not when the overlapping transcripts are expressed, as would be expected, but on the contrary, when only TSSs that are not in the overlapping region are used.

The regulatory function of RNA:RNA duplexes formed by natural antisense transcripts (NATs) has been demonstrated in a number of different studies, although mostly in the context of lncRNA-mRNA antisense pairs[10,33]. However, we cannot exclude other factors and mechanisms that may be involved in the regulation of the expression of these two genes. One of these is bidirectional promoters[33], which are quite common in the humans[7]. The promoter from which transcripts are produced in different directions acts as a regulatory unit that couples the expression of two genes. Simultaneous activation of two genes that overlap at their 5' ends can, as mentioned before, lead to a collision of polymerase complexes on two different DNA strands. It is convincible that bidirectional promoters may prevent this by alternating transcription bursts, for example.

Another important factor is enhancers, which are typically bound by multiple transcription factors (TFs) and act in concert to influence transcription from a target gene promoter[91,92]. Enhancers are bound by multiple TFs, and it has been shown that the binding of one TF molecule to DNA affects the binding of another TF molecule[93]. It is, therefore, possible that the silencing of *INO80E*-O led to changes in chromatin accessibility and altered the binding of some TFs. As a result, regulation by the enhancers may also be altered.

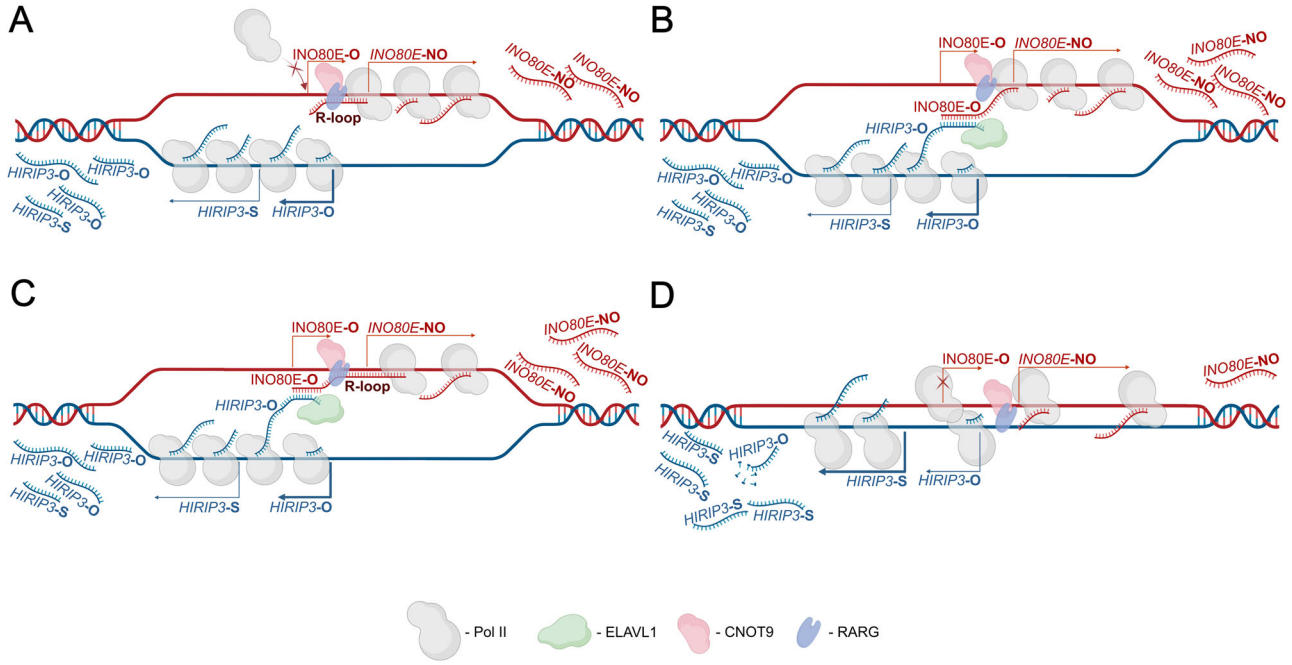


Fig. 6 | Mechanism of overlapping *INO80E* and *HIRIP3* expression. A The *INO80E* and *HIRIP3* expression when only R-loop is formed by *INO80E*-O isoform. The expression of the next *INO80E*-O is blocked, but the chromatin is open, which gives a space for the transcription of *INO80E*-NO and *HIRIP3*. **B** The *INO80E* and *HIRIP3* expression when only RNA:RNA duplex is formed by *INO80E*-O isoform and *HIRIP3*-O isoforms. The presence of the duplex helps to maintain the chromatin open, and both genes are expressed at a high level. **C** The *INO80E* and *HIRIP3* expression when RNA:RNA duplex is formed by *INO80E*-O isoform and *HIRIP3*-O isoforms along with the R-loop by *INO80E*-O isoform. The presence of the duplex helps to maintain the chromatin open; the transcription of the next *INO80E*-O is blocked, but there is a high expression of *INO80E*-NO and *HIRIP3*. **D** The *INO80E* and *HIRIP3* expression when there is a lack of expression of *INO80E*-O and, in consequence, a lack of RNA:RNA duplexes or/and R-loop. Chromatin is not accessible enough, and polymerase molecules cannot pass freely and block each other in the promoter region – transcriptional interference; *HIRIP3*-O are transcribed at very limited amounts, which makes downstream TSS more open for polymerase and the transcription of *HIRIP3*-S significantly increases. Throughout Fig. 6, ELAVL1 is light green in color, CNOT9 protein is light pink color, RARG is purple, and Pol II is gray. *INO80E* isoforms are red and *HIRIP3* isoforms are blue. Created in BioRender. Wanowska, E. (2025) https://BioRender.com/j29t536.

It should also be noted that our model assumes that the RNA polymerase slides along the DNA independently on each strand. However, many studies provide strong evidence for stationary polymerases and transcription factors where the DNA template is threaded through the polymerase to initiate transcription[94]. At present, it is difficult to explain how a stationary polymerase could transcribe genes on both the sense and the antisense strands of the same DNA region. In such a scenario, the DNA would be pulled in opposite directions by the polymerase, which could prevent any movement. One possibility is that the DNA fragment is pulled back and forth, but studies of transcription factors have not yet provided evidence for such a mechanism.

Methods

Bioinformatics analysis

In silico analyses were performed using the results of expression estimation studies performed in our laboratory[62]. The analysis of LR-RNA-seq was performed on data from the Pacific Biosciences sequencing platform[58]. The following calculations were performed on 784 ENCODE libraries (listed in the Supplementary Data: ENCODE_libraries file) obtained from normal and cancer tissue/cell line samples. The data was obtained from an openly available database gathering RNA-seq data - ENCODE database and who's downloading does not require additional consent or ethical approvals. Using the Scan for Motifs bioinformatics tool[69] and the RNA-Binding Protein DataBase (RBPDB)[70], we confirmed the presence of ELAVL1 protein-binding motifs in the 5' UTR sequence of the *HIRIP3* gene. The possibility of R-loop formation within the studied gene pair, *INO80E* and *HIRIP3*, was checked in the R-loopBase database[71]. Prediction of DNA binding proteins/transcription factors was performed using the ReMap Atlas of Regulatory Regions database at UCSC Genome Browser[95] and the JASPAR database[96]. De novo methylation and R-loop native bisulfite-converted DNA were visualized using Geneious Prime 2023.1.2. In addition, the RNAfold online tool[97] was used to predict the secondary structure of the overlapping end of the *HIRIP3* gene.

AMT cross-linking

HEK293T cells were cultured to a ~ 80% confluence (~25 million cells) on two separate 15-cm plates, one for crosslinked (+4'-aminomethyl trioxsalen (AMT)) and one for mock-crosslinked (+PBS) conditions. Trypsinized and centrifuged cells were washed once with room temperature PBS, then centrifuged again, and finally resuspended in 4 mL of ice-cold 0.5 mg/ml AMT solution in PBS (+AMT sample) or in an ice-cold PBS alone (-AMT control sample). Both samples were incubated on ice for 15 min and then transferred to a pre-chilled 10-cm tissue culture dishes. The cross-linking was performed on ice, 3–4 cm away from the light source, under a long-wave UV lamp (350 nm) in a UV Stratalinker 2400 (Stratagene) for 7 min, with mixing pauses every 2 min. After this procedure, the cells were centrifuged at 330×*g* for 4 min and the cell pellets were used directly in the standard RNA isolation procedure.

RNA isolation, reverse transcription, and PCR

RNA was extracted using TRI Reagent® (Molecular Research Center) according to a standard protocol. RNA yield was measured using a DeNovix DS-11 series spectrophotometer (DeNovix). RNA was reverse transcribed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Standard PCRs were performed using StartWarm HS-PCR Mix (A&A Biotechnology) or EconoTaq PLUS2X Master Mix (Lucigen). All primers used in the work were designed with Primer Blast and ordered from Merck. PCR products were analyzed on a 1.5% agarose gel containing GelRed (Biotium) in a 1X TAE buffer, and subsequently, the gel images

were captured using G:Box EF2 (Syngene) with GeneSys image analysis software (Syngene).

RNA antisense purification (RAP)

Biotinylated probes and RAP were performed according to the protocol described[66]. Briefly, samples were incubated with denaturized 15 pmol of biotinylated ssDNA probes for two hours. The sequence of the designed and ordered probe (Merck) can be found in Supplementary Materials - Table S2. Samples were then incubated for 30 min with freshly prepared Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (Invitrogen™, Thermo Fisher Scientific). Magnetic beads with the captured probe:target transcript complexes were purified by washing: three times with a low stringency wash buffer (1× SSPE, 0.1% SDS, 1% NP-40, 4 M urea), three times with High Stringency Wash Buffer (0.1× SSPE, 0.1% SDS, 1% NP-40, 4 M urea), and twice with an RNase H Elution Buffer (50 mM Tris-HCl pH 7.5, 75 mM NaCl, 3 mM $MgCl_2$, 0.125% *N*-lauroylsarcosine, 0.025% sodium deoxycholate, 2.5 mM TCEP). To remove the ssDNA probe bound to the RNA, samples were incubated with RNase H (Thermo Fisher Scientific) at 37 °C for 30 min. Eluted RNA was purified using the TRI Reagent® (Molecular Research Center). As a positive control, we used a probe against OIP5-AS1 that forms a duplex with *CCT3* transcripts.

RAP-MS

RAP-MS was performed according to the described protocol[66]. Concisely, cells were UV-crosslinked at a wavelength of 254 nm at 0.8 J/cm^2. The cell pellet was then resuspended in 1 ml of Cell Lysis Buffer - Nuclear I (10 mM HEPES pH 7.4, 20 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM TCEP, and 0.5 mM PMSF). To pull out duplexes created by transcripts of the *INO80E* and *HIRIP3* genes and associated proteins, a biotinylated probe complementary to the *HIRIP3* gene isoforms was designed. The sequence of the designed and ordered probe (Merck) can be found in Supplementary Materials - Table S2. The MS analysis was performed in the Mass Spectrometry Laboratory, IBB PAS, Warsaw. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[98] partner repository with the dataset identifier PXD060587.

Cell culture

Experiments were performed using HEK239T cells (CRL-1573™, ATCC). Cells were maintained under standard conditions at 37 °C in a humidified incubator with 5% CO_2 in DMEM (Genos) supplemented with a 10% fetal bovine serum – FBS (Genos) and 1% penicillin-streptomycin (10,000 U/mL) (Gibco™, Thermo Fisher Scientific). Cells were trypsinized with trypsin (0.25%) (Gibco™, Thermo Fisher Scientific).

RNA knock-down protocol

Silencing of the selected transcripts was performed using Antisense LNA® GapmeRs (Qiagen) or Silencer Select siRNAs (Thermo Fisher Scientific). Antisense LNA® GapmeRs are short 16-nucleotide single-stranded antisense oligonucleotides that catalyze RNase H-dependent degradation of complementary RNA targets and function in both the cytoplasm and the cell nucleus. For Antisense LNA® GapmeRs (Qiagen) silencing, the 30 nM of GapmeRs were used, and for Silencer Select siRNAs silencing the 75 nM of siRNA. The specificity of the LNA gapmeRs was tested using positive and negative controls. A gapmeR against *MALAT1* was used as a positive control, and a scrambled gapmeR, which should not cleave any RNA, was used as a negative control. After transfection, cells were cultured for 72 h, then harvested and used for RNA isolation.

Cell transfection

Genetic material was delivered to cells using Lipofectamine™ 3000 Transfection Reagent (Invitrogen™, Thermo Fisher Scientific). Prior to transfection, cells were grown to a ~70% confluence in 12-well dishes. Lipofectamine™ 3000 in an amount of 3 µl was diluted in 50 µl of Opti-MEM™ I Reduced Serum Medium, no phenol red (Gibco™, Thermo Fisher Scientific) per one well. Transfection material was diluted in 50 µl of Opti-MEM™ I Reduced Serum Medium, no phenol red (Gibco™, Thermo Fisher Scientific), and 2 µl of P3000™ Reagent was added per well. Diluted Lipofectamine™ 3000 and a transfection material were then mixed and incubated for 15 min at room temperature. After incubation, the lipid complexes were gently added to the cells.

Subcellular fractionation

Isolation of RNA from cytoplasmic, total nuclear, nucleoplasmic, and chromatin fractions were performed from cells cultured to a ~90% confluence on 10 cm plates according to the described protocol[99]. Briefly, harvested cells were resuspended in 380 µl of an ice-cold hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% NP-40, and 10% glycerol) supplemented with 100 U of a RiboLock RNase inhibitor (Thermo Fisher Scientific) and were incubated on ice for 10 min. After centrifugation (200×*g* at 4 °C for 2 min), supernatants were mixed with 1 ml of RNA precipitation solution (5% 3 M sodium acetate (pH 5.5) in 99,6% ethanol) and stored at −20 °C for at least one hour. Pellets were washed three times with 1 ml of ice-cold hypotonic lysis buffer. Total nuclear RNA samples were suspended in TRI Reagent® (Molecular Research Center) for direct RNA extraction. Pellets for the isolation of nucleoplasmic and chromatin fractions were suspended in 380 µl of modified Wuarin-Schibler buffer (10 mM Tris-HCl pH 7.0, 4 mM EDTA, 0.3 M NaCl, 1 M urea, and 1% NP-40) supplemented with 100 U of RiboLock RNase inhibitor (Thermo Fisher Scientific) and incubated on ice for 10 min. After centrifugation (1000×*g* at 4 °C for 3 min), the supernatants were mixed with 1 ml of RNA precipitation solution and stored as described above. Pellets were washed three times with 1 ml of ice-cold modified Wuarin-Schibler buffer. Chromatin samples were suspended in TRI Reagent® (Molecular Research Center) for direct RNA extraction. Samples incubated in RNA precipitation solution were centrifuged at 18,000×*g* at 4 °C for 15 min and washed in ice-cold 70% ethanol. Semi-dry pellets were suspended in TRI Reagent® (Molecular Research Center) for direct RNA extraction.

Quantitative PCR

Extracted total, cytoplasmic, total nuclear, nucleoplasmic, and chromatin RNA was used for cDNA synthesis using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's instructions. Quantitative PCR was performed using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems™, Thermo Fisher Scientific) on a QuantStudio™ 7 Flex Real-Time PCR System platform (Thermo Fisher Scientific). All experiments were performed in triplicate technical repeats for three biological replicates. For expression normalization, the expression of *GAPDH* was measured as an endogenous control. The data was then analyzed using the $2^{-\Delta\Delta Ct}$ method. Statistical significance of the results was performed using GraphPad Prism 9 (GraphPad Software).

CRISPR/Cas9 methylation

De novo methylation was introduced using pdCas9-DNMT3A-PuroR_v2, which was a gift from Vlatka Zoldoš (Addgene plasmid #74407; http://n2t.net/addgene:74407; RRID: Addgene_74407). sgRNA oligonucleotides were designed on the Benchling platform and ordered from Merck. Oligonucleotides were annealed and phosphorylated with 5U of T4 Polynucleotide Kinase (New England Biolabs). Methylation constructs were obtained using the Golden Gate assembly cloning strategy based on the *Bbs*I restriction enzyme. The resulting constructs were transformed into DH5α *E.coli* cells. After a control colony PCR, positive clones were cultured for isolation using the Plasmid Mini Kit (A&A Biotechnology). Correctly cloned constructs were verified using Sanger sequencing.

Bisulfite conversion

Bisulfite conversion was performed using the EpiJET Bisulfite Conversion Kit (Thermo Scientific) according to the manufacturer's protocol. Briefly, samples containing 500 ng of DNA were mixed with 120 µl of freshly

prepared modification reagent and incubated at 98 °C for 10 min and 60 °C for 150 min. Converted DNA samples were purified and used for PCR reactions.

RNase H1 overexpression

The pEGFP-RNASEH1 vector was used for RNase H1 overexpression (Addgene plasmid #108699; http://n2t.net/addgene:108699; RRID:Addgene_108699 as a gift from Andrew Jackson & Martin Reijns). Plasmids were introduced to the HEK293T cells with the Lipofectamine™ 3000 Transfection Reagent (see: Cell transfection). Transfection success was checked after 48 h using a ZOE Fluorescent Cell Imager (Bio-Rad). Fluorescence-activated cell sorting (FACS) was then used to separate GFP-positive cells that had taken up the pEGFP-RNASEH1 plasmid.

Chromatin immunoprecipitation

Proteins were crosslinked to DNA with 1% formaldehyde. After ten minutes of incubation, the cells were quenched with 125 mM glycine. The cells were then scraped and incubated for five minutes at 4 °C with two lysis buffers (L1: 50 mM TRIS pH 8, 1 mM EDTA pH 8, 0.1% NP-40, 10% glycerol, protease Inhibitors; L2: 1% SDS, 50 mM TRIS pH 8, 10 mM EDTA pH 8, protease inhibitors). The samples were sonicated for ten minutes (30 s ON – 30 s OFF intervals) using Bioruptor® (Diagenode Diagnostics). The chromatins were examined by gel electrophoresis. Chromatin immunoprecipitation was performed using the use of DynaBeads protein A/G (Thermo Scientific), histone H3 antibody - (ab1791) anti-histone H3 antibody (Abcam) – 5 µg, and RNA polymerase II antibody - (ab10338) anti-RPB2 antibody (Abcam) – 5 µg.

Assay for transposase-accessible chromatin (ATAC)

The ATAC was performed according to the protocol described[100]. Briefly, cells were collected and counted. For one ATAC reaction, 50,000 cells were collected. Then the cell lysis and Tn5 transposition were performed. Next, the samples were purified with DNA Clean-up and Concentration kit (Zymo Research). Quantification was performed with qPCR.

Cross-linking and immunoprecipitation (CLIP)

The CLIP assay was performed according to the protocol desscribed[101]. In brief, cells were irradiated with 150 mJ/cm^2 of UVA (365 nm) in a Stratalinker device. Then the cell lysis and RNase T1 digestion were performed. Next, the immunoprecipitation using normal mouse IgG: sc-2025 (Santa Cruz Biotechnology) and HuR/ELAV1 Antibody (G-8): sc-365816 (Santa Cruz Biotechnology) was performed. The amount of antibodies for the experiment was used as described in the published protocol[101]. Quantification was performed with RT-qPCR. In each case, the fold change after RT-qPCR was counted against IgG within a given condition and sample.

R-loop native bisulfite sequencing—chromatin preparation

The experiment was based on a protocol published by Boque-Sastre and colleques[72]. Cells at ~80% confluence were scraped from a six-well plate and centrifuged. Cell pellets were incubated in 500 µl of lysis buffer (10 mM TRIS pH 8, 1 mM EDTA, 0.5% SDS with 55 µg Proteinase K) at 37 °C overnight. Cell lysates were mixed with 250 µl of 5 M NaCl and centrifuged at maximum speed for 15 min. Precipitated chromatin was used for bisulfite conversion.

pGEM-T cloning of PCR products

The pGEM®-T Easy Vector System (Promega) was used according to the manufacturer's protocol. The standard ligation reaction mixtures were incubated overnight at 4 °C. The resulting vectors were then mixed with insert in 1:3 ratio and transformed into DH5α *E.coli* cells. Transformants were selected on LB/ampicillin/IPTG/X-Gal plates. Positive white-colored clones were cultured for isolation using the Plasmid Mini Kit (A&A Biotechnology). Vectors were verified by Sanger sequencing.

Statistics and reproducibility

The expression values for correlation analysis were examined using GraphPad Prism 9 (GraphPad, San Diego, CA, USA, www.graphpad.com) and R[102] with the packages "ggsignif"[103], "ggplot2"[104], "ggh4x"[105], "smplot2"[106], "tidyverse"[107], and "ggpubr"[108]. The Shapiro–Wilk normality test was first used to determine whether the data was normally distributed. The Mann–Whitney *U*-test was then conducted in order to compare differences between expression levels of different overlapping genes and transcripts. The relationship between *INO80E* and *HIRIP3* expression levels was examined using the Spearman correlation test. In all analyses, $p < 0.05$ was considered statistically significant. Statistical analysis (*T*-tests) of fold changes and relative expressions of tested transcripts after qRT-PCR was performed in GraphPad Prism 9 (GraphPad, San Diego, CA, USA, www.graphpad.com). In almost all experiments, three biological replicates and three technical replicates within each biological were performed. In some experiments, two biological replicates were performed due to high reproducibility. The number of replicates performed is described under the result of each experiment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated or analyzed during this study is included in this published article (and its supplementary information files). Uncropped and unedited gel images are presented in Supplementary Materials – Fig. S14. All source data were included in the Supplementary Data files. The mass spectrometry data were available via ProteomeXchange with identifier PXD060587.

Received: 1 August 2024; Accepted: 25 February 2025;
Published online: 08 March 2025

References

1. Bøvre, K. & Szybalski, W. Patterns of convergent and overlapping transcription within the b2 region of coliphage λ. *Virology* **38**, 614–626 (1969).
2. Henikoff, S., Keene, M. A., Fechtel, K. & Fristrom, J. W. Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. *Cell* **44**, 33–42 (1986).
3. Makałowska, I., Lin, C. F. & Hernandez, K. Birth and death of gene overlaps in vertebrates. *BMC Evol. Biol*. **7**, 193 (2007).
4. Wang, X.-J., Gaasterland, T. & Chua, N.-H. Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* **6**, 1–11 (2005). *2005 64*.
5. Rosikiewicz, W. & Makałowska, I. Biological functions of natural antisense transcripts. *Acta Biochim. Pol.* **63**, 665–673 (2016).
6. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference – a crash course. *Trends Genet.* **21**, 339–345 (2005).
7. Trinklein, N. D. et al. An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–66 (2004).
8. Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G. & Rowley, J. D. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends Genet.* **21**, 326–329 (2005).
9. Henz, S. R. et al. Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* **144**, 1247–1255 (2007).
10. Das, S., Zea Rojas, M. P. & Tran, E. J. Novel insights on the positive correlation between sense and antisense pairs on gene expression. *Wiley Interdiscip. Rev. RNA* **15**, e1864 (2024).
11. Rosikiewicz, W., Sikora, J., Skrzypczak, T., Kubiak, M. R. & Makałowska, I. Promoter switching in response to changing environment and elevated expression of protein-coding genes overlapping at their 5' ends. *Sci. Rep.* **11**, 1–14 (2021).

12. Cebrat, M. et al. Mechanism of lymphocyte-specific inactivation of RAG-2 intragenic promoter of NWC: implications for epigenetic control of RAG locus. *Mol. Immunol.* **45**, 2297–2306 (2008).
13. Yu, W. et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–206 (2008).
14. Aguilera, A. & García-Muse, T. R Loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
15. Castillo-Guzman, D. & Chédin, F. Defining R-loop classes and their contributions to genome instability. *DNA Repair* **106**, 103182 (2021).
16. Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
17. Hegazy, Y. A., Fernando, C. M. & Tran, E. J. The balancing act of R-loop biology: the good, the bad, and the ugly. *J. Biol. Chem.* **295**, 905–913 (2020).
18. Kaneko, S., Chu, C., Shatkin, A. J. & Manley, J. L. Human capping enzyme promotes formation of transcriptional R loops in vitro. *Proc. Natl Acad. Sci. USA* **104**, 17620–17625 (2007).
19. Stirling, P. C. et al. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* **26**, 163–175 (2012).
20. Wimberly, H. et al. R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat. Commun.* **4**, 1–11 (2013).
21. Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* **22**, 999–1007 (2015).
22. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chédin, F. GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**, 1590–1600 (2013).
23. Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
24. Richard, P. & Manley, J. L. R loops and links to human disease. *J. Mol. Biol.* **429**, 3168–3180 (2017).
25. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).
26. García-Muse, T. & Aguilera, A. R loops: from physiological to pathological roles. *Cell* **179**, 604–618 (2019).
27. Boque-Sastre, R. et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl Acad. Sci. USA* **112**, 5785–5790 (2015).
28. Tan-Wong, S. M., Dhir, S. & Proudfoot, N. J. R-loops promote antisense transcription across the mammalian genome. *Mol. Cell* **76**, 600–616.e6 (2019).
29. Beckedorff, F. C. et al. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *PLoS Genet.* **9**, e1003705 (2013).
30. Su, W. Y. et al. Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. *Cell Res.* **22**, 1374–1389 (2012).
31. Nojima, T. & Proudfoot, N. J. Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. *Nat. Rev. Mol. Cell Biol.* **23**, 389–406 (2022).
32. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2020).
33. Werner, A., Kanhere, A., Wahlestedt, C. & Mattick, J. S. Natural antisense transcripts as versatile regulators of gene expression. *Nat. Rev. Genet*. **25**, 730–744 (2024).
34. Herman, A. B., Tsitsipatis, D. & Gorospe, M. Integrated lncRNA function upon genomic and epigenomic regulation. *Mol. Cell* **82**, 2252–2266 (2022).
35. Hastings, M. L., Milcarek, C., Martincic, K., Peterson, M. L. & Munroe, S. H. Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. *Nucleic Acids Res.* **25**, 4296–4300 (1997).
36. Khochbin, S. & Lawrence, J. J. An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate. *EMBO J.* **8**, 4107–4114 (1989).
37. Werner, A. et al. Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. *BMC Genomics* **15**, 19 (2014).
38. Uchida, T. et al. Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J. Biol. Chem.* **279**, 14871–14878 (2004).
39. Faghihi, M. A. et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol*. **11**, R56 (2010).
40. Mahmoudi, S. et al. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol. Cell* **33**, 462–471 (2009).
41. Chen, C. H., Pan, C. Y. & Lin, W. chang. Overlapping protein-coding genes in human genome and their coincidental expression in tissues. *Sci. Rep*. **9**, 13377 (2019).
42. Rosikiewicz, W., Suzuki, Y. & Makałowska, I. OverGeneDB: a database of 5′ end protein coding overlapping genes in human and mouse genomes. *Nucleic Acids Res.* **46**, D186–D193 (2018).
43. Makhnovskii, P. A. et al. Alternative transcription start sites contribute to acute-stress-induced transcriptome response in human skeletal muscle. *Hum. Genomics* **16**, 24 (2022).
44. Tatip, S., Taggart, J., Wang, Y., MacDiarmid, C. W. & Eide, D. J. Changes in transcription start sites of Zap1-regulated genes during zinc deficiency: Implications for HNT1 gene regulation. *Mol. Microbiol.* **113**, 285–296 (2020).
45. Weber, R. et al. Monitoring the 5′UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene* **42**, 638–650 (2022).
46. Maruyama, R. et al. Altered antisense-to-sense transcript ratios in breast cancer. *Proc. Natl Acad. Sci. USA* **109**, 2820–2824 (2012).
47. Zhang, P. et al. Relatively frequent switching of transcription start sites during cerebellar development. *BMC Genomics* **18**, 1–14 (2017).
48. Ahmad, A., Strohbuecker, S., Tufarelli, C. & Sottile, V. Expression of a SOX1 overlapping transcript in neural differentiation and cancer models. *Cell. Mol. Life Sci.* **74**, 4245–4258 (2017).
49. Wei, W., Pelechano, V., Järvelin, A. I. & Steinmetz, L. M. Functional consequences of bidirectional promoters. *Trends Genet.* **27**, 267–276 (2011).
50. Bao, Y. & Shen, X. INO80 subfamily of chromatin remodeling complexes. *Mutat. Res. Mol. Mech. Mutagen.* **618**, 18–29 (2007).
51. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. **78**, 273–304 (2009).
52. Zhang, S. et al. INO80 is required for oncogenic transcription and tumor growth in non-small cell lung cancer. *Oncogene* **36**, 1430–1439 (2016).
53. Assrir, N., Filhol, O., Galisson, F. & Lipinski, M. HIRIP3 is a nuclear phosphoprotein interacting with and phosphorylated by the serine-threonine kinase CK2. *Biol. Chem.* **388**, 391–398 (2007).
54. Lorain, S. et al. Core histones and HIRIP3, a novel histone-binding protein, directly interact with WD repeat protein HIRA. **18**, 5546–5556 (2023).

55. Siddaway, R. et al. The in vivo interaction landscape of histones H3.1 and H3.3. *Mol. Cell. Proteomics* **21**, 100411 (2022).
56. Martin, F. J. et al. Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
57. Sayers, E. W. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
58. Reese, F. et al. The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity. *bioRxiv* **18**, 23 (2023).
59. Lizio, M. et al. Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.* **47**, D752–D758 (2019).
60. Lizio, M. et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 1–14 (2015).
61. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
62. Szcześniak, M. W., Wanowska, E., Mukherjee, N., Ohler, U. & Makałowska, I. Towards a deeper annotation of human lncRNAs. *Biochim. Biophys. Acta Gene Regul. Mech* **1863**, 194385 (2020).
63. Lu, Z., Gong, J. & Zhang, Q. C. PARIS: psoralen analysis of RNA interactions and structures with high throughput and resolution. *Methods Mol. Biol.* **1649**, 59–84 (2018).
64. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global mapping of human RNA-RNA interactions. *Mol. Cell* **62**, 618–626 (2016).
65. Zhang, M. et al. Optimized photochemistry enables efficient analysis of dynamic RNA structuromes and interactomes in genetic and infectious diseases. *Nat. Commun.* **12**, 1–14 (2021).
66. Engreitz, J., Lander, E. S. & Guttman, M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol. Biol.* **1262**, 183–197 (2015).
67. Bakheet, T., Hitti, E., Al-Saif, M., Moghrabi, W. N. & Khabar, K. S. A. The AU-rich element landscape across human transcriptome reveals a large proportion in introns and regulation by ELAVL1/HuR. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1861**, 167–177 (2018).
68. Ma, W. J., Chung, S. & Furneaux, H. The Elav-like proteins bind to AU-rich elements and to the poly(A) tail of mRNA. *Nucleic Acids Res.* **25**, 3564 (1997).
69. Biswas, A. & Brown, C. M. Scan for motifs: a webserver for the analysis of post-transcriptional regulatory elements in the 3′ untranslated regions (3′ UTRs) of mRNAs. *BMC Bioinformatics* **15**, 1–6 (2014).
70. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–D308 (2011).
71. Lin, R. et al. R-loopBase: a knowledgebase for genome-wide R-loop formation and regulation. *Nucleic Acids Res.* **50**, D303–D315 (2022).
72. Boque-Sastre, R., Soler, M. & Guil, S. Detection and characterization of R loop structures. *Methods Mol. Biol.* **1543**, 231–242 (2017).
73. Rondón, A. G. & Aguilera, A. R-loops as promoters of antisense transcription. *Mol. Cell* **76**, 529–530 (2019).
74. Reske, J. J., Wilson, M. R. & Chandler, R. L. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin* **13**, 1–17 (2020).
75. Gontarz, P. et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. *Sci. Rep. 2020* **10**, 1–13 (2020). *101*.
76. Bendtsen, K. M. et al. Direct and indirect effects in the regulation of overlapping promoters. *Nucleic Acids Res.* **39**, 6879–6885 (2011).
77. Ning, Q. et al. The evolution and expression pattern of human overlapping lncRNA and protein-coding gene pairs. *Sci. Rep.***7**, 1–12 (2017).
78. Zhou, C. & Blumberg, B. Overlapping gene structure of human VLCAD and DLG4. *Gene* **305**, 161–166 (2003).
79. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl Acad. Sci. USA* **99**, 8796–8801 (2002).
80. Callen, B. P., Shearwin, K. E. & Egan, J. B. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* **14**, 647–656 (2004).
81. Johnsson, P. et al. Transcriptional kinetics and molecular functions of long noncoding RNAs. *Nat. Genet.* **54**, 306–317 (2022).
82. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
83. Chen, X. et al. 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat. Cell Biol.* **21**, 978–990 (2019).
84. Wang, I. X. et al. Human proteins that interact with RNA/DNA hybrids. *Genome Res.* **28**, 1405–1414 (2018).
85. Dou, P. et al. C1orf109L binding DHX9 promotes DNA damage depended on the R-loop accumulation and enhances camptothecin chemosensitivity. *Cell Prolif*. **53**, e12875 (2020).
86. Patel, P. S. et al. RNF168 regulates R-loop resolution and genomic stability in BRCA1/2-deficient tumors. *J. Clin. Invest*. **131**, e140105 (2021).
87. Pérez-Calero, C. et al. UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide. *Genes Dev.* **34**, 898–912 (2020).
88. Wolak, C. et al. Interaction with single-stranded DNA-binding protein localizes ribonuclease HI to DNA replication forks and facilitates R-loop removal. *Mol. Microbiol.* **114**, 495–509 (2020).
89. Wood, M. et al. TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. *J. Cell Sci*. **133**, jcs244129 (2020).
90. Hauksdottir, H., Farboud, B. & Privalsky, M. L. Retinoic acid receptors beta and gamma do not repress, but instead activate target gene transcription in both the absence and presence of hormone ligand. *Mol. Endocrinol.* **17**, 373–385 (2003).
91. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
92. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
93. Carey, M. The enhanceosome and transcriptional synergy. *Cell* **92**, 5–8 (1998).
94. Rieder, D., Trajanoski, Z. & McNally, J. G. Transcription factories. *Front. Genet.* **3**, 221 (2012).
95. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
96. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
97. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377 (2015).
98. Perez-Riverol, Y. et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553 (2025).
99. Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* **9**, 2045–2060 (2014).
100. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* **17**, 1518–1552 (2022).
101. Yoon, J. H. & Gorospe, M. Cross-Linking Immunoprecipitation and qPCR (CLIP-qPCR) analysis to map interactions between long noncoding RNAs and RNA-binding proteins. *Methods Mol. Biol.* **1402**, 11 (2016).
102. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*. 275–286 (2013).

103. Ahlmann-Eltze, C. & Patil, I. ggsignif: R package for displaying significance brackets for 'ggplot2'. Preprint at PsyArxiv https://doi.org/10.31234/OSF.IO/7AWM6 (2021).
104. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
105. Brand, T. van den. ggh4x: Hacks for 'ggplot2'. R package version 0.2.3. https://cran.r-project.org/package=ggh4x (2022).
106. Seung, H. M. smin95/smplot: smplot2 - a package for statistical data visualization version 0.1.0. https://rdrr.io/github/smin95/smplot/ (2023).
107. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
108. Kassambara, A. 'ggplot2' based publication ready plots [R package ggpubr version 0.6.0]. (2023).

Acknowledgements

We thank Prof. Krzysztof Sobczak for his very valuable remarks and comments regarding the manuscript. This research was supported by the National Science Center Poland (grant number: 2017/25/B/NZ2/01519).

Author contributions

I.M. and N.R. conceptualized and conceived the study. N.R., A.Ł., E.W., and J.K.-M. performed the experiments. I.M., N.R., and E.W. analyzed and interpreted the data. I.M. and N.R. supervised the experiments. N.R. and I.M. wrote the paper. I.M. obtained funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07815-4.

Correspondence and requests for materials should be addressed to Izabela Makałowska.

Peer review information *Communications Biology* thanks Elizabeth Tran and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Rosie Bunton-Stasyshyn & Ophelia Bu. A peer review file is available.