NAR Genomics and Bioinformatics, 2024, **6**, lqae045
https://doi.org/10.1093/nargab/lqae045
Advance access publication date: 6 May 2024
Standard Article

Malat1 affects transcription and splicing through distinct pathways in mouse embryonic stem cells

Morteza Aslanzadeh[1], Laura Stanicek[1], Marcel Tarbier[2], Emilio Mármol-Sánchez[3], Inna Biryukova[1] and Marc R. Friedländer[1,*]

[1]Science for Life Laboratory, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, Sweden
[2]Science for Life Laboratory, Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Solna, Sweden
[3]Science for Life Laboratory and Center for Palaeogenetics. Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, Sweden

*To whom correspondence should be addressed. Tel: +46 737121558; Email: marc.friedlander@scilifelab.se

Abstract

Malat1 is a long-noncoding RNA with critical roles in gene regulation and cancer metastasis, however its functional role in stem cells is largely unexplored. We here perform a nuclear knockdown of Malat1 in mouse embryonic stem cells, causing the de-regulation of 320 genes and aberrant splicing of 90 transcripts, some of which potentially affecting the translated protein sequence. We find evidence that Malat1 directly interacts with gene bodies and aberrantly spliced transcripts, and that it locates upstream of down-regulated genes at their putative enhancer regions, in agreement with functional genomics data. Consistent with this, we find these genes affected at both exon and intron levels, suggesting that they are transcriptionally regulated by Malat1. Besides, the down-regulated genes are regulated by specific transcription factors and bear both activating and repressive chromatin marks, suggesting that some of them might be regulated by bivalent promoters. We propose a model in which Malat1 facilitates the transcription of genes involved in chromatid dynamics and mitosis in one pathway, and affects the splicing of transcripts that are themselves involved in RNA processing in a distinct pathway. Lastly, we compare our findings with Malat1 perturbation studies performed in other cell systems and *in vivo*.

Introduction

Complex gene regulation is critical for multicellular organisms, as the expression of thousands of genes need to be independently patterned across hundreds of distinct cell types. The combination of transcriptional and post-transcriptional regulation is important for exponentially increasing the regulatory complexity. Much regulation is conferred directly by proteins such as transcription factors and RNA-binding proteins, but regulatory RNAs also play key roles. With the completion of the ENCODE consortium effort, it has become clear that the human genome harbors tens of thousands of processed transcripts that have little or no protein coding capability (1). These long non-coding RNAs (lncRNAs) are defined as RNA molecules with a length of more than 200 nucleotides that do not encode proteins (2). Initially, it was believed that these long non-coding RNAs are transcriptional noise (3) but numerous studies have shown that many of these transcripts have complex expression patterns (4), play important roles in cellular processes (5), and their dysfunction can mediate different diseases (6). They have been demonstrated to play critical regulatory roles in gene expression, chromatin organization, as well as cellular processes in the cytoplasm (7). Many of the transcripts resemble mRNAs in that they are transcribed by RNA polymerase II and bear 5′ methyl-guanosine caps and 3′ poly(A) tails (7).

Metastasis Associated Lung Adenocarcinoma Transcript 1 (Malat1), also known as Nuclear Enriched Abundant Transcript 2 (NEAT2), is one of the most well-studied lncRNAs. Malat1 has become widely recognized in recent years as a result of its widespread expression, abundance, and apparent importance in a variety of molecular mechanisms and diseases. This lncRNA, which spans approximately 8 kb on human chromosome 11, is highly conserved in sequence across species, suggesting fundamental functions (8). Malat1 is mainly located in nuclear speckles and regulates different physiological and pathological processes by playing critical roles in diverse molecular processes, including alternative splicing, transcriptional regulation, and epigenetic control of gene expression (9). Alternative splicing is typically controlled by proteins known as serine/arginine-rich (SR) proteins in eukaryotic cells (10). Malat1 interacts with SR proteins to stabilize them and facilitate splicing (11), substantially affecting the precision and effectiveness of alternative splicing in eukaryotic cells. Furthermore, it has also been found that Malat1—like many other lncRNAs—is highly concentrated in the chromatin fraction, suggesting its involvement in the control of gene transcription (12,13). It has been shown that Malat1 can regulate gene expression by interacting with specific transcription factors and transcriptional repressors. For example, the interaction of Malat1 with SP1 and nuclear p65/p50 transcription factors affect the expression of LTBP3 and NF-κB regulated genes respectively (14,15). Finally, Malat1 is also capable of regulating the epigenetic landscape by interacting with different proteins such as EZH2, Suv39h1 and DBC1, affecting the expression of downstream genes (16–19).

Received: December 7, 2023. **Revised:** March 14, 2024. **Editorial Decision:** April 15, 2024. **Accepted:** April 30, 2024

Extensive research has been conducted on Malat1 to determine its role and implications in cancer progression and metastasis in patients, in mouse, and in cell migration studies in various cancer types (20,21). However, although Malat1 is highly abundant in mouse embryonic stem cells, no previous studies have investigated the impact of Malat1 perturbation in these cells. The advantage of using embryonic stem cells over other cell types is their ability to achieve relatively pure populations of undifferentiated cells after molecular and biochemical interventions due to their stable genome and high amenability to genome modifications (22). In addition, these cells can differentiate into a variety of other cell types and can undergo directed differentiation (22,23), making them an important model system for studying molecular mechanisms underlying embryogenesis. Examining the roles of essential factors such as Malat1 within this unique cellular context becomes especially relevant, given the importance of this lncRNA in gene expression regulation, contributing to a better understanding of Malat1's roles in processes related to early development and embryogenesis.

In this study, we apply LNA GapmeR antisense oligonucleotides to knock down Malat1 in the nucleus of mouse embryonic stem cells. We identify 122 transcripts that are up-regulated and 198 transcripts that down-regulated as a result of the perturbation as well as 90 transcripts that have altered splicing patterns. By overlaying these genes with available public multi-omics data from mouse embryonic stem cells, we stratify the transcripts into distinct pathways. Specifically, we find evidence that Malat1 interacts with enhancer elements of genes that are transcriptionally activated by Malat1, and that these genes have functions of mitosis, specifically in chromatid cohesion and microtubule organization. In addition, we find that Malat1 directly interacts with the differentially splicing transcripts at the RNA and DNA level, and that these targets themselves encode proteins that have functions in RNA processing. In summary, we apply gene perturbation, transcriptomics and multi-omics integration to propose a model where Malat1 affects splicing and transcriptional regulation through two distinct pathways in mouse embryonic stem cells.

Materials and methods

Cell culture and transfection

Drosha^{F} mouse embryonic stem cell line containing the tamoxifen-inducible LoxP—exon9—LoxP and a neomycin selection cassette, was provided by M Chong (60). Cells were cultured in serum-free Ndiff227 medium (Cellartis Takara Bio), supplemented with 1000 U/mL ESGRO mouse LIF (Millipore), differentiation inhibitors PD0325901 (1μM) and CHIR-99021 (3 μM, both from Selleckchem) and 1× penicillin–streptomycin (Gibco). For cell harvest, cells treated with Accutase (Sigma) at 37°C for 5 min followed by centrifugation. Mouse embryonic stem cells were transfected at 60–70% confluence with LNA GapmeRs using Lipofectamine RNAiMax (Thermo Scientific) according to the manufacturer's instructions. Pre-designed Negative control A (LG00000002-DDA) and mouse Malat1 (LG00000008-DDA) anti-sense LNA GapmeRs were purchased from Qiagen. Cells were grown onto feeder-free 0.1% EmbryoMax gelatine (Merck Millipore) coated flasks at 37 °C in a 5% CO_2 atmosphere and tested negative for mycoplasma.

RNA sequencing

Global gene expression was analyzed by RNA sequencing. Briefly, RNA was isolated from mouse embryonic stem cells 48h post-transfection with control or anti-MALAT1 GapmeR using the Quick-RNA Miniprep Kit (Zymo Research). RNA quality was assessed by Agilent RNA 6000 Pico Kit (Agilent Technologies). cDNA libraries were prepared using the TruSeq Stranded mRNA Library Prep kit (Illumina) and the average library size was assessed using the Agilent DNA 1000 Kit (Agilent Technologies). Libraries were normalized, pooled, denatured, and diluted for 80 bp single-end mRNA sequencing using Illumina NextSeq500 high-output sequencing. All kits were used according to the manufacturer's instructions.

Sequence quality control

The sequence quality of raw sequence data was checked using FastQC tool (version 0.11.9). The per-base sequence quality showed very good quality calls thus we included all reads in downstream analyses.

Read mapping and counting

The mouse genome assembly (GRCm38.p6) was indexed using the STAR alignment tool (version 2.7). STAR –runMode genomeGenerate –genomeFastaFiles GRCm38.p6.genome.fa –sjdbGTFfile gencode.vM25.annotation.gtf –sjdbOverhang 79. Reads were then mapped to the mouse genome using the same tool. STAR –genomeDir –readFilesIn raw_reads.fastq –outFilterMultimapNmax 1 –outSAMtype BAM SortedByCoordinate. The quantification of the RNA-seq data was performed using featureCounts (version 2.0). featureCounts -s 0 -t -g -a gencode.vM25.annotation.gtf -M -O –fraction –maxMOp 1000 -o output.txt input.bam. The -t option in featureCounts was specified based on the exon and intron level quantification. In order to quantify reads based on introns, a custom annotation file was generated from the original mouse annotation file using the makeEISAgtfs script from the EISAcompR package (61) available at https://github.com/emarmolsanchez/EISAcompR.

Differential gene expression analysis

DESeq2 package (version 1.36) in R was used for performing differential gene expression analysis. Normalized read count filters were applied to only keep genes if at least half of the samples meet 50 and 25 counts at the exon and intron levels, respectively. The *P*-value cutoff of 0.01 was used to identify differentially expressed genes.

Alternative splicing analysis

We used rMATS tool (version 4.1) to detect differential alternative splicing events. Bam files of the control and knockdown samples were used as the main inputs for the tool to operate the analysis. *rmats.py –b1 b1_input –b2 b2_input –gtf gencode.vM25.annotation.gtf -t single –readLength 80 –variable-read-length*. We retrieved all significant (FDR < 0.05) retained intron events from the final output that considers both Junction and Exon counts (*RI.MATS.JCEC.txt*). No extra filters were applied.

Frameshift analysis

Retained intron lengths were extracted from the RI.MATS.JCEC.txt file (upstreamEE + 1 and downstreamES

as reported). Next, the farthest 5p and 3p coordinates were determined for each gene based on the information available in the annotation file, considering all available transcripts for each gene. Subsequently, the midpoint positions of the retained introns were calculated. For standardization, the 5p coordinates were adjusted based on the gene strand, setting them to 1. Moreover, the distances between the 3p location and the 5p location as well as the distance between the retained intron midpoint position and the 5p location were computed. Finally, the retained intron events were categorized as '5p' or '3p' based on their proximity to the respective transcripts ends. If the retained intron midpoint position was close to the 5p end (RI midpoint position < midpoint position of the gene), it was considered as '5p'. Conversely, if the retained intron midpoint position was close to the 3p end (RI midpoint position > midpoint position of the gene), it was labeled as '3p'.

Gene ontology analysis

Go analysis was performed using the R package Clusterprofiler (version 4.4). To select GO terms that pass a significant cutoff, *P*-values were adjusted using the Benjamini—Hochberg method. Only top 5 GO terms with a *P*-value and q-value less than 0.05 are shown as cnet plots. org.Mm.eg.db was used as the background for the enrichment analyses.

TF enrichment analysis

The public chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) data was used to identify transcription factor and cofactor targets. Transcription factor lists were obtained from the Cistrome database (44) using the species 'Mus musculus' and the biological source 'Embryonic Stem Cell'. Only sets that reached four out of the six quality scores listed were considered. The target rank was established using the assigned score and the top 300 targets of each transcription factor were considered in this analysis. After intersecting these targets with the deregulated and alternatively spliced genes, the number of remaining targets for each TF was calculated. The binomial test was used to calculate the probability of a set of targets being targeted by a specific TF. For the hypothesized probability of success in this test, we assumed that 300 genes are targeted by TFs out of 20000 genes in mice. To select the enriched TF candidates, the calculated *P*-values were adjusted using the Bonferroni method. The adjusted *P*-value cutoff of 0.05 was used to identify enriched TFs.

RNA half-life measurements

The half-life measurements used in our analysis were derived from a previously published study (42). We specifically used the half-life measurements of the MC1 cell line (129S6/SvEvTac) to maintain consistency with our cell line, which was derived from the 129Sv mouse strain. The plot was generated by intersecting our differentially expressed and RI genes with the half-life measurements of the MC1-LIF+ condition from the mentioned study. The background genes contained all genes reported in the MC1-LIF+ condition, excluding our differentially expressed and RI genes. *P*-values were calculated using the Wilcoxon rank sum test.

TSS analysis

Malat1-chromatin interaction data from the RADICL-seq study (39) were used to calculate the maximum distance from the TSS and TES of the group of genes of interest. Only interactions of the 1FA and 2FA datasets that were observed in four or more replicates in the RADICL-seq study were considered. The TSS and TES information of genes was retrieved using the same annotation file used in the RADICL-seq study (vM14). Distances were calculated considering if the Malat1 target gene is on the positive strand (*interaction position—TSS or TES*) or the negative strand (*TSS or TES—interaction position*). The final distribution plot was made considering only the distances that were 200 kb upstream or downstream of TSS and TES.

Epigenetic mark analysis

We retrieved the histone mark signals (GSM2417080, GSM2417100, GSM2417096, GSM2417108) for the mouse embryonic stem cells from the published study (47). These data were used to have a single value for each gene that represents the power of the epigenetic marks. For this purpose, wig files were converted to bed files using the bedops tool (version 2.4). *wig2bed –zero-indexed < input.wig > output.bed*. These bed files were intersected with the bed format of NCBI37/mm9 gtf (this genome version was used in the study mentioned above) to retrieve regions that cover our genes of interest using bedtools (version 2.30). *bedtools intersect -a gencode.vM1.annotation.bed -b input.bed -wo > output.csv*. The total sum of the peak values was calculated for each gene using the intersection output files. The background genes were set using our original count matrices at the exonic and intronic levels. To have the background genes, we applied the same filtering steps performed for the exonic and intronic read counts in the differential gene expression analyses and then we excluded the deregulated genes from this set. The Wilcoxon rank sum test was used to compare the distribution of the epigenetic marks between the deregulated groups and the background genes.

Results

Nuclear knockdown of Malat1 transcript with LNA GapmeRs

To gain insights into the functions of Malat1 in mouse embryonic stem cells, we performed a knock-down of this transcript (Figure 1). Importantly, since Malat1 is known to play important roles in the nucleus, we used anti-sense oligonucleotide LNA GapmeRs for the knockdown (Materials and Methods), since this RNaseH-mediated method efficiently targets nuclear transcripts (24). We performed the perturbation experiment using GapmeRs targeting Malat1 and a control sequence in three replicates. Each of the samples was then subjected to high-throughput RNA sequencing using a protocol that enriches for mRNA transcripts (Materials and Methods).

Malat1 affects abundances of hundreds of transcripts in mouse embryonic stem cells

Given the importance of lncRNAs in gene regulation at both the transcriptional and post-transcriptional level (9), we investigated the effect of Malat1 perturbation on gene expression in mouse embryonic stem cells as measured by sequenc-

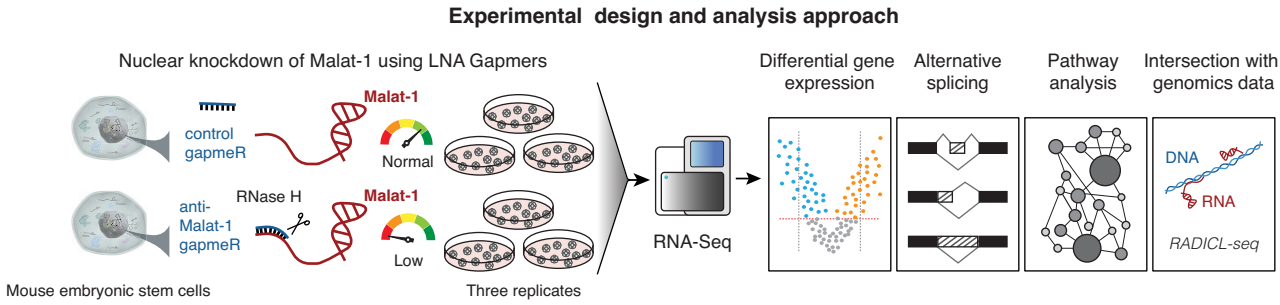

Figure 1. Overview of the experimental approach and computational analyses. Malat1 was knocked down using LNA-GapmeRs, followed by RNA-sequencing analysis with three replicates in the control condition and three replicates in the knockdown condition. Computational analyses were then performed on the RNA-seq data to detect differentially expressed genes, differentially spliced transcripts, functional pathway enrichment and to intersect the data with available complementary functional genomics resources.

ing (above). We performed the analysis at two levels: at the level of sequenced RNAs mapping to exons, to study the effect on mature transcripts; and at the level of sequences mapping to introns, as a proxy to estimate the effect on nascent transcription (i.e. the pre-mRNA level before splicing (25), Supplementary Tables S1 and S2). Differential gene expression analyses revealed significant deregulation in mRNA transcript abundance, with 81 genes being up-regulated and 172 genes being down-regulated when exonic sequence - corresponding to spliced transcripts - were considered (Figure 2A, Materials and Methods). Among the genes that were de-regulated were *Cdc7*, which plays a role in embryonic stem cell differentiation (26); *Gsc*, functioning as a transcription factor during early stages of development (27); *Dpp4* (also known as *CD26*), involved in cell surface receptor interactions relevant to immune responses and cell signaling (28); *Cep250*, an essential component of the centrosome; and *Smc1b*, a key constituent of the cohesin complex that is pivotal for chromosome segregation (29). As expected, we found the Malat1 transcript itself to be significantly and substantially down-regulated as a result of the applied knockdown (Figure 2A, left).

Since Malat1 can regulate genes at the transcriptional level, we also investigated differentially expressed genes based on their intronic sequences. Recently transcribed pre-mRNAs still harbor the intronic segments before their removal during splicing (25), and thus reads mapping to introns could be used as a proxy to detect alterations in ongoing transcription. While most of our detected sequences were exonic, we did detect substantial amounts of intronic sequence in our measurements by RNA-seq (Supplementary Figures S1 and S2). In total, we found 47 genes to be up-regulated and 40 genes to be down-regulated when considering intronic sequences only (Figure 2A, right). Among the genes that were deregulated at the intronic levels were *Magi1*, implicated in cell adhesion and signaling (30); *Fgf17*, involved in developmental processes (31); *Ptprg*, known for its role in important process such as cell growth and differentiation (32); *Cxxc5*, associated with cell proliferation, differentiation, and apoptosis (33); and *Rbfox3*, involved in the regulation of alternative splicing of specific pre-mRNA molecules (34). Subsequently, we analyzed the functions of genes which were altered after knockdown of Malat1. Our results revealed that the deregulated genes included mostly protein-coding genes, as well as several non-coding RNAs, including lncRNAs such as *Eprn* which aids in the transition of embryonic stem cells from a naive pluripotent to a more differentiated state, and also *Xist* (35,36) (Figure 2B). Importantly, we found that genes that are down-regulated at the exonic level also tend to be down-regulated at the intronic level (Figure 2C, Supplementary Figure S3). The same holds for up-regulated genes. We, therefore, concluded that these genes are all transcriptionally regulated by Malat1 but are individually more or less robustly identified as being deregulated at either the exonic or intronic level. For this reason, we pooled the genes that were down-regulated at either the exonic or intronic level and considered them as one set of 198 non-redundant genes, and similarly for the 122 up-regulated genes (Figure 2C, Supplementary Table S3). In summary, we find evidence that Malat1 significantly and substantially alters the transcript levels of hundreds of protein-coding genes in mouse embryonic stem cells.

Malat1 perturbation is associated with altered intron retention

Given the central role of Malat1 in splicing, we evaluated the impact of reduced expression of Malat1 on splicing events. We applied the dedicated software MATS to identify introns that undergo differential splicing as a result of Malat1 perturbation (Methods). The software reported 231 unique genes with splicing instances, distributed among five different types of events (Skipped Exon; Alternative 5′ Splice Site; Alternative 3′ Splice Site; Mutually Exclusive Exons; and Retained Intron). However, when we performed computational control experiments by swapping control and knockdown samples, we found that only the intron retention events were enriched relative to permuted controls, suggesting that only these types of events are reliably reported (Figure 3A). In total, we found 96 instances of intron retention (in 90 unique genes), with 80 being retained in the control condition and 16 being retained in the knockdown (Figure 3B, Supplementary Table S3, Supplementary Add. Data S1). Since more introns are retained in the control than in the knockdown condition, this suggests that Malat1 predominantly facilitates intron retention - rather than intron removal - in mouse embryonic stem cells. We found that transcripts with retained introns (RI) did not significantly overlap with the deregulated transcripts (above) and also that they did not by themselves display any evidence of deregulation at the whole-transcript level (Supplementary Figure S4), suggesting that the differentially spliced and differentially expressed transcripts are affected by

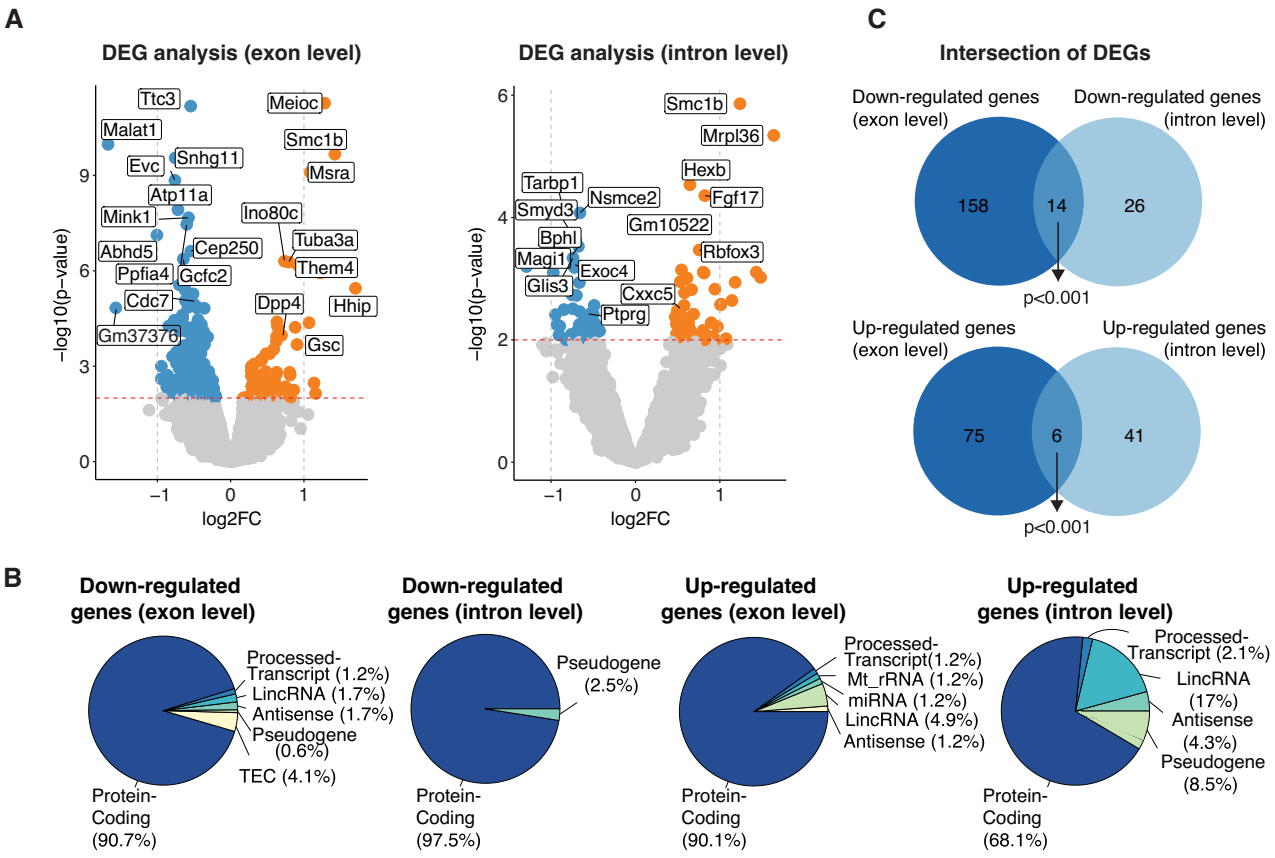

Figure 2. Differentially expressed genes upon Malat1 loss of function in mouse embryonic stem cells. (**A**) Volcano plots illustrating differentially expressed genes at the exon and intron levels. Genes selected for analysis are shown above the horizontal red dashed line, marked by a significance threshold of *P*-value = 0.01. (**B**) Pie charts depicting the proportions of each gene type among the up-regulated and down-regulated genes, separately for the perturbed genes at the exon and intron levels. (**C**) Venn diagram illustrating the overlapping genes among the up-regulated genes at the exon and intron levels, as well as among the down-regulated genes at the exon and intron levels. *P*-values were calculated using the hypergeometric distribution.

Malat1 through distinct pathways (Figure 3C). Among the transcripts that retain introns, we find transcripts involved in a variety of cellular functions and signaling pathways, including centrosome-related activities (for instance gene *Cep192*, Figure 3B), spindle assembly, acting as a molecular scaffold influencing ERK activity (for instance *Wdr83*), blastocyst development, protein localization and spliceosome assembly. Interestingly, most of the introns that are retained either in the control condition (54/80, Figure 3D) or in the knockdown condition (13/16, Supplementary Figure S5) would not preserve the open reading frame, suggesting that their inclusion could disrupt production of a functional protein. However, based on our findings, most of frameshifts (34/54) in the control group take place near the 3′ end of the transcripts. This observation suggests that these retained introns may not all affect the overall protein structure, impairing their functionality (Figure 3D). Most of the introns that would not cause a frameshift however introduce stop-codons towards the end of the open reading frame-frame, which would likely result in slightly truncated proteins (Supplementary Figure S6). We do not find any evidence that the amino acids translated from the introns would be biased towards hydrophobic residues, which could target them to proteasomal degradation (Supplementary Figure S7). In summary, we find evidence that Malat1 affects the splicing of close to one hundred transcripts in mouse embryonic stem cells, and that translation of these introns could disrupt or attenuate the production of functional proteins.

Transcripts perturbed by Malat1 knockdown have roles in mitosis and in mRNA processing

In the following, we will investigate three groups of genes that are affected by Malat1: (i) genes that are down-regulated; (ii) genes that are up-regulated and (iii) genes that retain or lose introns (henceforth referred to as the 'RI' genes) upon Malat1 perturbation. The putative functions of each of the three affected sets of genes were identified by Gene Ontology (GO) analysis (Materials and Methods). The results of the GO analysis revealed that the down-regulated genes were significantly enriched in several biological processes, such as the regulation of mRNA processing and involvement in mitotic processes, including sister chromatid cohesion, microtubule organization, and recombinational repair (Figure 4A). However, we did not observe any change in expression of cell-cycle stage marker genes upon the Malat1 knock-down, as might be expected if our experiment had caused substantial cell cycle perturbation (Supplementary Figure S8). The RI group was mainly enriched in RNA splicing, mRNA processing, cell growth, and non-membrane-bound organelle assembly (Figure 4B). On the contrary, the set of up-regulated genes did not appear to be enriched in any particular functions (not shown). In the next step, we aimed to investigate whether there were any potential networks of functional couplings or associations among the altered genes in each of the three gene sets. To this end, we performed network analyses using FunCoup (37). With this analysis, we again found coherent networks only among the genes of the down-regulated and RI groups, but not among the

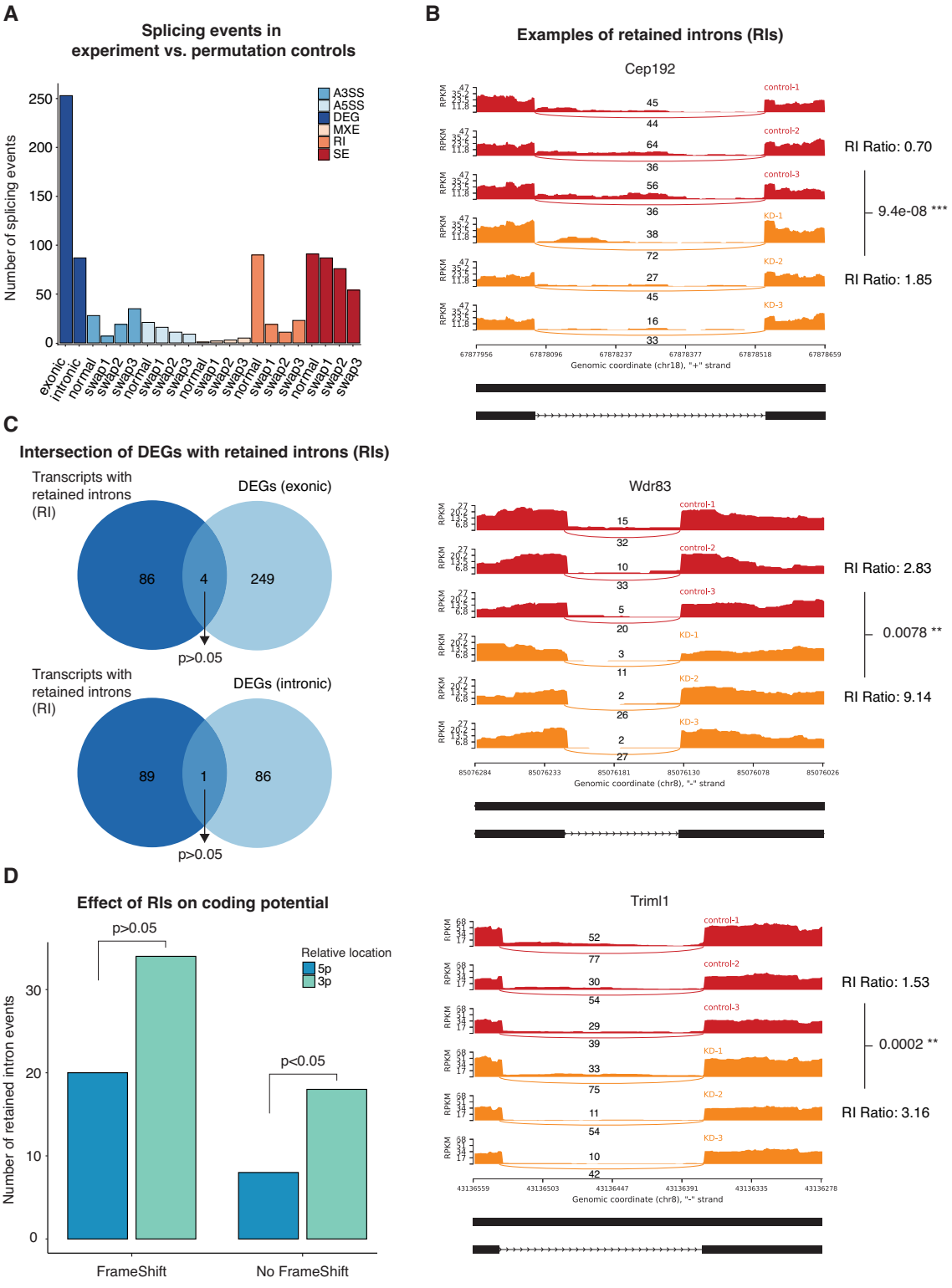

Figure 3. Transcripts with splice alterations upon Malat1 loss of function. (**A**) The bar plot shows the number of unique genes under different splicing conditions, including the normal comparison (3 control versus 3 knockdown samples) and the permutation controls, where samples of control or knockdown are randomly exchanged. The first two bars represent the number of differentially expressed genes at both the exon and intron levels, for reference. (**B**) Sashimi plots showing retained intron (RI) events, with numbers indicating the count of inclusion-supporting reads (IJC) located above the retained introns, and the count of splicing-supporting reads (SJC) positioned below the retained introns. The calculated RI ratios are derived by dividing the total sum of splicing-supporting reads by the total sum of inclusion-supporting reads. *P*-values were calculated using the Fischer's exact test. (**C**) Venn diagrams show the overlapping genes among the differentially expressed genes and the genes with retained introns. *P*-values were calculated using the hypergeometric distribution. (**D**) The bar plot shows the number of introns that are retained in the presence of Malat1, distinguishing between introns whose inclusion would cause a frameshift in the coding sequence and introns that would not cause a frameshift. The colors in the plot indicate the relative location of retained introns with respect to the 5-prime or 3-prime end of the transcript, with the 5-prime introns more likely to disrupt the translated protein sequence. Statistical significance was performed using the chi-square test.

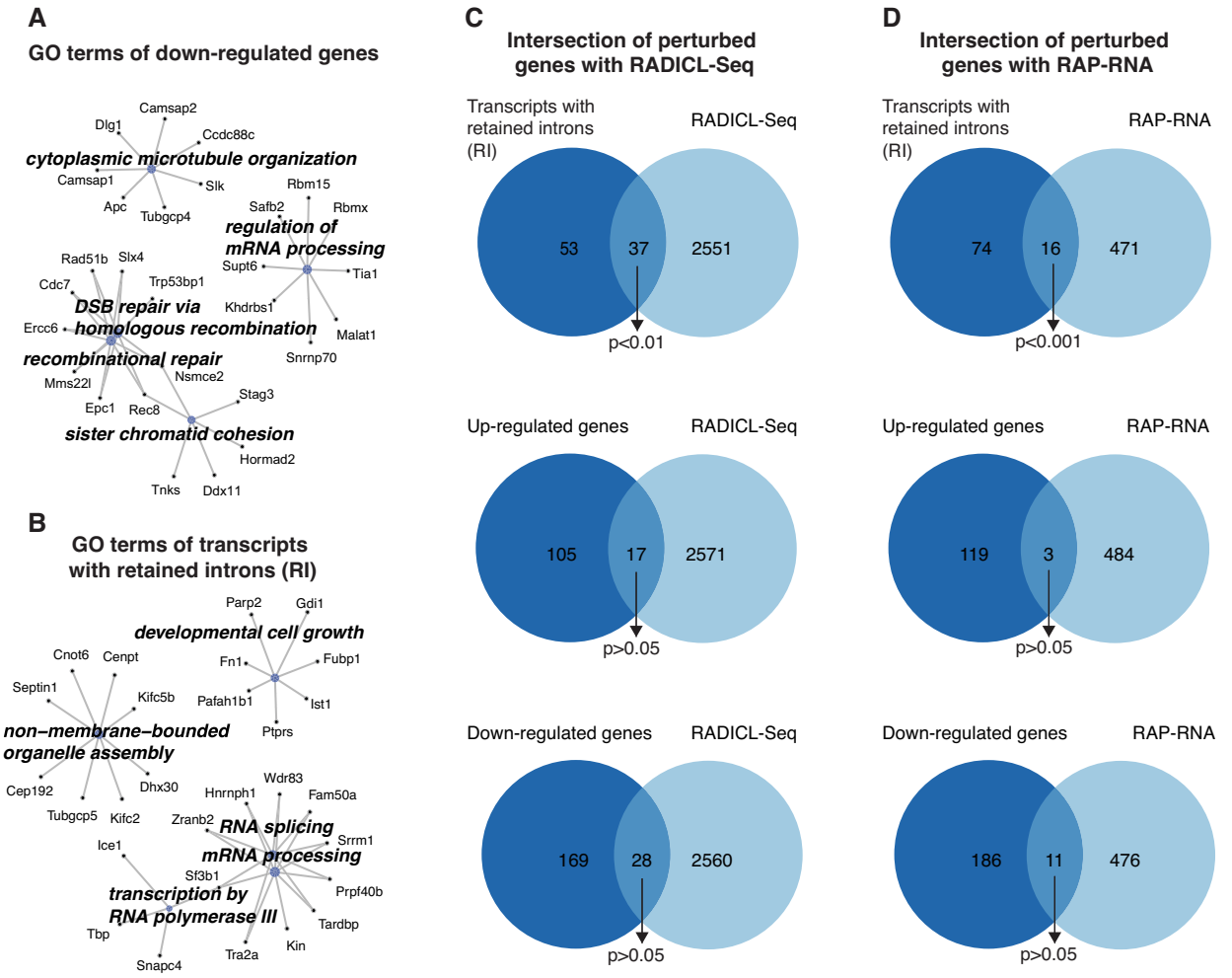

Figure 4. Functions of perturbed genes and direct interaction with Malat1. (**A, B**) Top 5 enriched biological processes based on gene ontology are shown for the down-regulated and retained intron groups. (**C, D**) Venn diagrams illustrating the overlapping genes among the altered genes identified with orthogonal datasets (RADICL-Seq and RAP–RNA). *P*-values were calculated using the hypergeometric distribution.

genes of the up-regulated group (Supplementary Figures S9-S11). Focusing on the top nodes of the networks in the down-regulated group (*Sdha, Dlg1, Apc, Ogt, Plcg1* and *Ptprg*), we concluded that the overall network could be related to stem cell vitality and control of cell division, which is consistent with our functional GO analyses for the same group of genes. In the case of the RI group, the top nodes (*Ipo7, Sf3b1, Hnrnph1* and *Gtpbp4*) were related to essential cellular processes that occur within the nucleus, such as RNA splicing and modification as also found in the GO analysis. Although there is evidence that Malat1 is involved in mammalian immunity (38), we did not observe any related GO terms (n.s.), however we did observe specific individual genes in the down-regulated group that have putative functions related to immunity (Supplementary Figure S12). In summary, we find that the down-regulated genes appear to be involved in chromatid dynamics and mitosis, while the genes whose transcripts are differentially spliced (RI genes) themselves have functions in mRNA processing.

Malat1 directly interacts with transcripts that are differentially spliced upon perturbation

Previous studies have applied genomics methods to show that the Malat1 transcript directly interacts with chromatin and various mRNAs in mouse embryonic stem cells (39,40). Using the RADICL-seq method, RNA and DNA molecules were ligated together and sequenced in an unbiased approach, yielding a transcriptome-wide and genome-wide map of RNA–DNA interactions (39). In a different study applying the RAP–RNA method, Malat1 was isolated using antisense purification and transcripts interacting with that lncRNA were identified using sequencing (40). We therefore re-analyzed these public genomics data to investigate if the genes we find to be affected by Malat1 knockdown directly interact with Malat1 at the DNA or RNA level. We first intersected each of our three sets of genes with the RADICL-seq data (Materials and Methods). We found that the RI genes are significantly enriched in interactions with Malat1, suggesting that this lncRNA is directly interacting with these genes at the chromatin level (Figure 4C). In contrast, there was no significant interaction between Malat1 and either the up- or down-regulated genes. We next re-analyzed the RAP–RNA data and found that the RI transcripts are also significantly enriched in interactions with Malat1, suggesting that the lncRNA binds to these transcripts (Figure 4D). Again, we found no significant interaction between the down- or up-regulated transcripts and Malat1. In summary, we find evidence that Malat1 directly interacts with transcripts whose splicing is altered upon Malat1 perturbation at the RNA level and right at the site of transcription.

Half-life profiles of transcripts affected by Malat1 perturbation

LncRNAs have been described to interact with mRNAs and influence their half-lives. One way in which lncRNAs can do this is by promoting the stabilization or decay of mRNAs by specific base-pairing patterns (41). In light of this, we investigated the steady-state half-lives of the transcripts that were affected by the Malat1 perturbation in mouse embryonic stem cells using available data (42) (Materials and Methods). We observed that the transcripts of up-regulated genes show longer half-lives relative to background transcripts (Supplementary Figure S13), even when normalizing for their expression level (Supplementary Figures S14 and S15). In contrast, the down-regulated genes and the genes with retained introns did not exhibit any significant differences. Based on these observations, we cannot infer that Malat1 directly influences the stability of the transcripts, since it is possible that different sets of transcripts have distinct half-lives as a result of their functions—for instance the specific transcripts that have functions in mitosis might be relatively short-lived. However, the influence of Malat1 on the stability of these transcripts could be the focus of future studies.

Evidence that specific perturbed genes are transcriptionally regulated by Malat1

Given that Malat1 has been shown to regulate gene expression both transcriptionally and post-transcriptionally, we sought to determine which level of transcriptional regulation is affected in the three groups of genes when Malat1 expression is knocked down. Thus, we examined the gene expression fold change measures of the down-regulated, up-regulated, and RI groups at the exon and intron levels (Figure 5A). Interestingly, we found that the down-regulated genes decrease in expression at both the exon and intron level to comparable degrees. Similarly, the up-regulated genes increase in expression at both the exon and intron levels. This suggests that the perturbation happens at the transcriptional level, since perturbation that affects only spliced transcripts would only be observable at the exonic but not intronic level. As expected, the RI transcripts did not substantially increase or decrease overall at either the exonic or intronic level, consistent with only specific exonic structures being affected (Figure 5A). Our findings confirm that Malat1 can influence specific genes at the transcriptional level in mouse embryonic stem cells.

Malat1 localizes upstream of genes that are down-regulated upon Malat1 perturbation

Given that we find that Malat1 regulates specific genes at the transcriptional level, we looked to available functional genomics data to understand the mode of regulation. Studies have indicated that Malat1 interacts with active chromatin sites, primarily in promoter regions and over gene bodies, suggesting its direct involvement in transcriptional regulation (12). Therefore, we reanalyzed public RNA–DNA (RADICL-seq) interaction data to identify where Malat1 is binding in the gene bodies of our three groups of genes (Materials and Methods). We found that Malat1 distributed relatively evenly over the bodies of the up-regulated and RI genes (Figure 5B). For the RI genes, this would be expected if Malat1 is affecting co-transcriptional splicing of introns while the nascent transcript is still located at the gene bodies (43). In contrast, Malat1 tended to bind upstream of the gene bodies of the down-regulated genes (Figure 5B), with a mean distance of ∼27 000 nucleotides upstream of the transcription start site. This could be consistent with Malat1 binding upstream regulatory regions such as enhancer sites. In summary, we find that Malat1 binds the RI group genes in the gene bodies, while it tends to bind down-regulated genes upstream, possibly at enhancer sites.

Perturbed genes are enriched for binding sites for specific transcription factors and repressors

As previously mentioned, Malat1 has been described to further regulate transcription through interaction with transcriptional activators and repressors. Therefore, to investigate if each of the three groups of genes were enriched in binding sites for specific transcription factors, we studied 172 distinct transcription factors for which mouse embryonic stem cells ChIP-seq data were available in the Cistrome database (44) (Materials and Methods). Based on the top 300 ranked target genes of each of these transcription factors, our analysis showed the enrichment of distinct transcription factors for each set of our genes (Table 1). For the down-regulated genes, the CBX8 transcriptional repressor was enriched. This transcription factor is part of the PRC1-like complex that is required to maintain the transcriptionally repressive state of many genes. In the case of the up-regulated genes we observed SALL1 as an enriched protein factor that is involved in transcriptional repression. There were several transcription factors that target the genes showing RI. For this group, we found six enriched transcription factors (EZH2, KAT8, LEO1, TBP, RARA and USP7) that are involved in both transcriptional repression and activation (Table 1). Similarly, it has been proposed that Malat1 can interact with the RNA-binding protein Nucleolin in nuclear speckles (45), and consistent with this we find a significant enrichment of Nucleolin binding motifs in down-regulated transcripts and in transcripts with retained introns (Supplementary Figure S16). In summary, we find several transcription factors and repressors that are enriched in binding sites in the three groups of genes.

Down-regulated genes are enriched for both activating and repressive chromatin marks

Our findings indicate that Malat1 plays a role in both upregulating and downregulating the expression of specific transcripts, with distinctive differences in its binding positions in these genes (see above). Previous studies have reported that Malat1 can interact with the Polycomb repressive complex 2 (PRC2) and promote the repressive H3K27me3 mark (46). Therefore, to investigate if there are differences in chromatin marks between our groups of genes, we analyzed the steady-state profiles of four histone marks for these genes. ChIP-Seq data for the active (H3K27ac, H3K36me3, H3K4me3) and repressive (H3K27me3) chromatin marks in mouse embryonic stem cells were retrieved from a previously published dataset (47). Our analysis revealed that the down-regulated group showed enrichment in all chromatin marks compared to the background and up-regulated genes (Figure 5C), suggesting that some of the genes may have bivalent chromatin marks, which are commonly found in embryonic stem cells and help to maintain their pluripotency (48). Although it is well-established that Malat1 can interact with the PRC2 complex to promote the repressive H3K27me3 mark on gene bodies, we did not observe an enrichment of the up-regulated

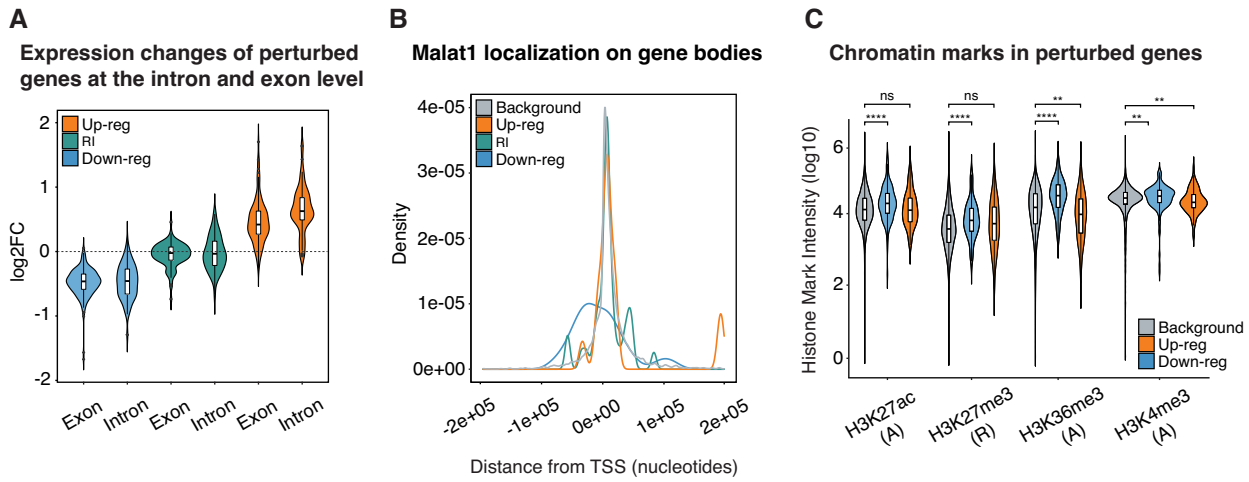

Figure 5. Evidence of transcriptional regulation by Malat1. (**A**) Violin plot showing the gene expression change levels of the altered genes at the exon and intron levels. Each group is represented by a distinct color. (**B**) Density plot presenting the distribution of the distance between Malat1-chromatin interaction positions and transcriptional start sites (TSS; Methods). (**C**) Violin plot representing the distribution of the four different chromatin marks for the background, up-regulated, and down-regulated genes (A = activating mark; R = repressive mark).

Table 1. Transcription factors enriched in binding to promoter regions of perturbed genes

TF	Function	Targeting group
CBX8	A key component of a Polycomb group (PcG) complex, essential for maintaining the transcriptionally repressive state of numerous genes	Down-regulated
EZH2	A Polycomb group (PcG) protein, it catalyzes the methylation of 'Lys-9' (H3K9me) and 'Lys-27' (H3K27me) on histone H3, leading to transcriptional repression of the target gene	RI
KAT8	A histone acetyltransferase that may be involved in transcriptional activation	RI
LEO1	A PAF1 complex component, plays multiple roles in development, maintenance of embryonic stem cell pluripotency, and transcriptional elongation	RI
TBP	Serves a central role in the DNA-binding multiprotein factor TFIID	RI
RARA	A nuclear retinoic acid receptor, the encoded protein, retinoic acid receptor alpha, regulates transcription in a ligand-dependent manner	RI
USP7	This protein deubiquitinates different target proteins and plays important roles in DNA damage, cell growth, and apoptosis	RI
SALL1	A transcription factor involved in embryonic development	Up-regulated

genes for this mark. However, the up-regulated genes were depleted in the active H3K36me3 and H3K4me3 marks relative to background genes (Figure 5C). In summary, we find that the down-regulated genes are enriched for both activating and repressive chromatin marks, while the up-regulated genes are depleted for some activating marks.

Comparison to previous Malat1 perturbation studies

Given that we find evidence that Malat1 has functions related to mitosis (Figure 4A) and mRNA processing (Figure 4B), we set out to compare our results with previous perturbation studies. Upon Malat1 constitutive genetic deletion in mice (49), the authors found that the development of mice was not visibly disrupted and the only clear transcriptomic effect in liver and brain was that relatively few transcripts changed in steady-state levels in the absence of Malat1. The transcripts that changed in expression in the mouse brain were associated with extracellular structure organization, which may be related to the well-described role of Malat1 in cell migration and cancer (20). The lack of clear developmental phenotypes and strong transcriptomic effect in this study may be due to redundant regulatory systems during development (see Discussion below). Interestingly, transcripts related to extracellular structure organization have also been found to show intron retention or to decrease in expression in cell lines upon Malat1 repression, specifically in HeLa cells (from our functional term analyses of these published data, (50)) and in human lung adenocarcinoma cells (51). In other cell lines, transcripts related to chromosome segregation and mitosis were found to be down-regulated in response to Malat1 knockdowns, specifically in mouse cultured hippocampal neurons (52) and in human lung fibroblasts (53). This is similar to what we observe in mouse embryonic stem cells (Figure 4A). In summary, we find that the functions of Malat1 in chromosome dynamics have also been described in previous loss-of-function experiments. Malat1 roles in cell migration has also been described in several previous studies, although we do not observe this in mouse embryonic stem cells. It appears that Malat1 may have both cell-specific and more ubiquitous functions.

Discussion

In this study, we investigated the impact of the Malat1 lncRNA on gene expression and splicing in mouse embryonic stem cells using a nuclear knock-down system. Our results revealed that 198 genes—including Malat1—were down-

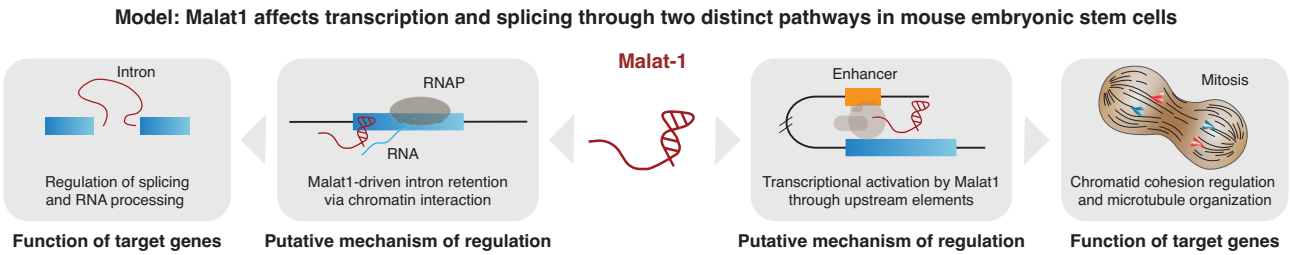

Figure 6. Model of Malat1 functions in mouse embryonic stem cells through two distinct pathways. In the first pathway, Malat1 interacts with putative enhancer regions upstream of the target genes, facilitating transcription of genes involved in mitosis – specifically in chromatid dynamics, microtubule organization and recombinational repair. In the second pathway, Malat1 interacts directly with gene bodies and nascent transcripts to inhibit splicing of specific introns. These target transcripts themselves have functions in regulation of splicing and RNA processing, suggesting a possible feedback mechanism.

regulated and 122 genes were up-regulated, while 80 introns were spliced out and 16 introns were aberrantly retained upon Malat1 loss of function. While there seemed to be an avoidance of splicing alterations that would change coding sequence, there was a subset of retained introns that would affect the translated protein sequence. We found that the down-regulated genes are enriched in functions related to chromatid cohesion, microtubule organization, and recombinational repair, while genes with retained introns are enriched in functions related to RNA splicing, mRNA processing, cell growth, and non-membrane-bound organelle assembly. We find evidence that Malat1 directly interacts with gene bodies and transcripts of the targets that retain introns, and that Malat1 binds upstream of genes that are down-regulated, at putative enhancer sites. Consistent with binding at enhancer regions, we find that genes that are down-regulated are affected at both exon and intron level, suggesting that they are regulated by Malat1 transcriptionally. The groups of down-regulated genes and genes with retained introns are both affected by specific groups of transcription factors, and the down-regulated genes are enriched for both activating and repressive marks, suggesting that some may be bivalent 'poised' genes. Lastly, we have made a systematic comparison of de-regulated genes and their functional terms with previous Malat1 perturbation studies, finding both similarities and differences.

Based on our findings, we propose a model in which Malat1 affects transcription and splicing through two distinct pathways (Figure 6). In the first pathway (Figure 6, right), Malat1 interacts with enhancer regions and putative protein factors to facilitate transcription of genes involved in mitosis—specifically in chromatid dynamics, microtubule organization and recombinational repair. In the second pathway (Figure 6, left), Malat1 interacts directly with gene bodies and nascent transcripts to inhibit splicing of specific introns. The targeted transcripts interestingly themselves have functions in regulation of splicing and RNA processing, suggesting a possible feedback mechanism: when Malat1 is abundant, it will inhibit splicing of transcripts that are themselves translated to splicing factors. In contrast, when Malat1 is lowly expressed, these transcripts will undergo splicing, thus increasing production of splicing factors and facilitating global RNA processing. While this model is consistent with our observations, more studies will be needed to test its validity and scope.

An important part of our study here is the integration of public RADICL-seq and RAP-RNA data, which allows us to detect direct interactions between Malat1 and DNA and RNA targets, giving both information about enrichment of such interactions and also the positions of the interactions relative to gene bodies. While the RAP-RNA data analyzed only yields information about Malat1 – since that particular transcript was isolated using antisense purification—the RADICL-seq data yields unbiased information about transcriptome-wide and genome-wide RNA-DNA interactions. Thus, these specific datasets could be used to profile interactions also of other lncRNAs and regulatory RNAs. While it is not certain that the interactions detected are direct physical interactions—they could for instance also be mediated by proteins—they do provide evidence that the detected molecules have been in relatively close proximity in the nucleus.

While we in our model above suggest how Malat1 regulates the genes that were down-regulated or had altered splicing patterns in our perturbation experiment, we do not include the up-regulated genes in our model. For these genes, we did not find any enrichment in functional terms, nor did we find any evidence for direct interaction with Malat1 at either the DNA or RNA level. It seems plausible that most of these genes are up-regulated as a result of secondary effects—that Malat1 knockdown perturbs the expression of one or more other regulators, that might in turn cause the upregulation of these genes in a cascading manner. Considering the ability of Malat1 to regulate the epigenetic landscape (16–19,54), it is conceivable that the up-regulated genes could be a consequence of the loss of repressive marks or even the gain of activating marks, given that these genes are in a steady-state condition depleted in activating marks (Figure 5C).

Connected with this, we measured RNA changes 48 hours after the knockdown by LNA GapmeRs. Shorter or longer waiting times would likely have resulted in somewhat different lists of perturbed genes, as some changes might not yet be observable at the level of steady-state RNAs, or some changes might have returned to an equilibrium state. Still, we did observe two groups of perturbed genes that each have distinct functions and where their interaction with Malat1 is supported by omics data. Another limitation of our study is that we used steady-state data for transcript half-lives and chromatin marks rather than measure changes before and after the Malat1 perturbation. This could be the focus of future studies. Although we have integrated omics data from several sources, it is important to highlight that these data were all obtained from mouse embryonic stem cells, which facilitates the direct comparison across data modalities.

Based on multiple lines of evidence (12,14,50,52) Malat1 has a function in regulating gene expression and alternative splicing, however other studies conducted with Malat1 knock-

out animal models (49,55) show that Malat1 is not essential for normal physiology and development. Consistent with studies conducted in other cell lines, we find that Malat1 has roles in transcriptional regulation and intron retention in mouse embryonic stem cells. This raises the question: why does Malat1 clearly have regulatory functions in mouse embryonic stem cells, yet lacks critical importance for overall mouse development and normal physiological processes? One explanation could be that there is functional redundancy for Malat1 RNA (12,49) where molecular function of a gene is substituted by a second gene as it is for many other RNAs (56). Another explanation could be that there exist compensatory mechanisms for specific important transcriptional and post-transcriptional processes (57) for Malat1 in knockout mice as it has been described for other genes (57–59). These apparently contradictory findings confirm that functional redundancy or compensatory mechanisms of Malat1 are underexplored and the significance of mouse knockout models cannot be underestimated in addressing these intricate questions.

Taken together, while previous studies have indicated that Malat1 is not critical for mouse development, our study explores the role of Malat1 in gene expression and splicing within mouse embryonic stem cells, revealing its specific regulatory functions in this cellular context. This investigation holds significant importance as it allows us to identify altered genes and biological properties resulting from Malat1 knockdown in these cells, providing valuable insights into its significance in stem cell biology. Considering the pivotal role of embryonic stem cells in developmental biology and disease research, our study offers essential and relevant insights into the relevance of Malat1 in these critical fields and our work significantly advances our understanding of Malat1-mediated gene regulation in stem cells, thereby opening new doors for further research and potential therapeutic applications.

Data availability

Sequencing data have been deposited at SRA under this accession number: PRJNA1101455.

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

We acknowledge the following funding sources: ERC Starting Grant 758397, 'miRCell'; Swedish Research Council (VR) Grant 2019-05320, 'MioPec'; VR Consolidator Grant 2022-03953 'InSync'; and funding from the Strategic Research Area (SFO) program of the Swedish Research Council through Stockholm University to the Friedländer lab. The computation was enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at Uppmax partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973. We thank members of the Friedländer, Kutter and Pelechano labs for insightful suggestions, and Dr Rory Johnson and Prof. Erik Sonnhammer for comments and helpful suggestions on our study.

Author contributions: M.R.F, M.A. and L.S. conceived the study and designed the experiments. M.A. performed all computational analyses under supervision of M.R.F. L.S. performed all lab experiments. E.M-S. co-supervised intron annotation and splicing analyses. M.T. co-supervised half-life and TF ChIP-seq enrichment analyses. I.B. advised on lab experiments. M.A. and M.R.F. wrote the manuscript with input from all authors.

Funding

H2020 European Research Council [ERC Starting Grant 758397]; Vetenskapsrådet [Consolidator Grant 2022-03953, Project Grant 2019-05320].

Conflict of interest statement

None declared.

References

1. Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F., *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
2. Mattick,J.S., Amaral,P.P., Carninci,P., Carpenter,S., Chang,H.Y., Chen,L.-L., Chen,R., Dean,C., Dinger,M.E., Fitzgerald,K.A., *et al.* (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.*, **24**, 430–447.
3. Gibb,E.A., Brown,C.J. and Lam,W.L. (2011) The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer*, **10**, 38.
4. Uszczynska-Ratajczak,B., Lagarde,J., Frankish,A., Guigó,R. and Johnson,R. (2018) Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.*, **19**, 535–548.
5. Ulitsky,I. and Bartel,D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell*, **154**, 26–46.
6. Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
7. Statello,L., Guo,C.-J., Chen,L.-L. and Huarte,M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, **22**, 96–118.
8. Arun,G., Aggarwal,D. and Spector,D.L. (2020) MALAT1 Long non-coding RNA: functional implications. *Non-Coding RNA*, **6**, 22.
9. Zhang,X., Hamblin,M.H. and Yin,K.-J. (2017) The long noncoding RNA Malat1: its physiological and pathophysiological functions. *RNA Biol*, **14**, 1705–1714.
10. Long,J.C. and Caceres,J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
11. Miao,H., Wu,F., Li,Y., Qin,C., Zhao,Y., Xie,M., Dai,H., Yao,H., Cai,H., Wang,Q., *et al.* (2022) MALAT1 modulates alternative splicing by cooperating with the splicing factors PTBP1 and PSF. *Sci. Adv.*, **8**, eabq7289.
12. West,J.A., Davis,C.P., Sunwoo,H., Simon,M.D., Sadreyev,R.I., Wang,P.I., Tolstorukov,M.Y. and Kingston,R.E. (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell*, **55**, 791–802.
13. Chu,C., Qu,K., Zhong,F.L., Artandi,S.E. and Chang,H.Y. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell*, **44**, 667–678.
14. Li,B., Chen,P., Qu,J., Shi,L., Zhuang,W., Fu,J., Li,J., Zhang,X., Sun,Y. and Zhuang,W. (2014) Activation of LTBP3 gene by a long noncoding RNA (lncRNA) MALAT1 transcript in mesenchymal stem cells from multiple myeloma. *J. Biol. Chem.*, **289**, 29365–29375.
15. Zhao,G., Su,Z., Song,D., Mao,Y. and Mao,X. (2016) The long noncoding RNA MALAT1 regulates the

lipopolysaccharide-induced inflammatory response through its interaction with NF-κb. *FEBS Lett.*, **590**, 2884–2895.
16. Hirata,H., Hinoda,Y., Shahryari,V., Deng,G., Nakajima,K., Tabatabai,Z.L., Ishii,N. and Dahiya,R. (2015) Long noncoding RNA MALAT1 promotes aggressive renal cell carcinoma through Ezh2 and interacts with miR-205. *Cancer Res.*, **75**, 1322–1331.
17. Wang,D., Ding,L., Wang,L., Zhao,Y., Sun,Z., Karnes,R.J., Zhang,J. and Huang,H. (2015) LncRNA MALAT1 enhances oncogenic activities of EZH2 in castration-resistant prostate cancer. *Oncotarget*, **6**, 41045–41055.
18. Chen,X., He,L., Zhao,Y., Li,Y., Zhang,S., Sun,K., So,K., Chen,F., Zhou,L., Lu,L., *et al.* (2017) Malat1 regulates myogenic differentiation and muscle regeneration through modulating MyoD transcriptional activity. *Cell Discov.*, **3**, 17002.
19. Chen,R., Liu,Y., Zhuang,H., Yang,B., Hei,K., Xiao,M., Hou,C., Gao,H., Zhang,X., Jia,C., *et al.* (2017) Quantitative proteomics reveals that long non-coding RNA MALAT1 interacts with DBC1 to regulate p53 acetylation. *Nucleic Acids Res.*, **45**, 9947–9959.
20. Schmidt,L.H., Spieker,T., Koschmieder,S., Schäffers,S., Humberg,J., Jungen,D., Bulk,E., Hascher,A., Wittmer,D., Marra,A., *et al.* (2011) The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J. Thorac. Oncol.*, **6**, 1984–1992.
21. Lai,M., Yang,Z., Zhou,L., Zhu,Q., Xie,H., Zhang,F., Wu,L., Chen,L. and Zheng,S. (2012) Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med. Oncol.*, **29**, 1810–1816.
22. Smith,A.G. (2001) Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.*, **17**, 435–462.
23. Fairchild,P.J., Brook,F.A., Gardner,R.L., Graça,L., Strong,V., Tone,Y., Tone,M., Nolan,K.F. and Waldmann,H. (2000) Directed differentiation of dendritic cells from mouse embryonic stem cells. *Curr. Biol. CB*, **10**, 1515–1518.
24. Dhuri,K., Bechtold,C., Quijano,E., Pham,H., Gupta,A., Vikram,A. and Bahal,R. (2020) Antisense oligonucleotides: an emerging area in drug discovery and development. *J. Clin. Med.*, **9**, 2004.
25. Gaidatzis,D., Burger,L., Florescu,M. and Stadler,M.B. (2015) Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.*, **33**, 722–729.
26. Kim,J.M., Yamada,M. and Masai,H. (2003) Functions of mammalian Cdc7 kinase in initiation/monitoring of DNA replication and development. *Mutat. Res.*, **532**, 29–40.
27. Blum,M., De Robertis,E.M., Kojis,T., Heinzmann,C., Klisak,I., Geissert,D. and Sparkes,R.S. (1994) Molecular cloning of the human homeobox gene goosecoid (GSC) and mapping of the gene to human chromosome 14q32.1. *Genomics*, **21**, 388–393.
28. Klemann,C., Wagner,L., Stephan,M. and von Hörsten,S. (2016) Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin. Exp. Immunol.*, **185**, 1–21.
29. Revenkova,E., Eijpe,M., Heyting,C., Hodges,C.A., Hunt,P.A., Liebe,B., Scherthan,H. and Jessberger,R. (2004) Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nat. Cell Biol.*, **6**, 555–562.
30. Wörthmüller,J. and Rüegg,C. (2021) MAGI1, a scaffold protein with tumor suppressive and vascular functions. *Cells*, **10**, 1494.
31. Xu,J., Lawshe,A., MacArthur,C.A. and Ornitz,D.M. (1999) Genomic structure, mapping, activity and expression of fibroblast growth factor 17. *Mech. Dev.*, **83**, 165–178.
32. Boni,C. and Sorio,C. (2021) The role of the tumor suppressor gene protein tyrosine phosphatase gamma in cancer. *Front. Cell Dev. Biol.*, **9**, 768969.
33. Xiong,X., Tu,S., Wang,J., Luo,S. and Yan,X. (2019) CXXC5: a novel regulator and coordinator of TGF-β, BMP and wnt signaling. *J. Cell. Mol. Med.*, **23**, 740–749.
34. Dredge,B.K. and Jensen,K.B. (2011) NeuN/Rbfox3 nuclear and cytoplasmic isoforms differentially regulate alternative splicing and nonsense-mediated decay of Rbfox2. *PLoS One*, **6**, e21585.
35. Li,M.A., Amaral,P.P., Cheung,P., Bergmann,J.H., Kinoshita,M., Kalkan,T., Ralser,M., Robson,S., von Meyenn,F., Paramor,M., *et al.* (2017) A lncRNA fine tunes the dynamics of a cell state transition involving Lin28, let-7 and de novo DNA methylation. *eLife*, **6**, e23468.
36. Wang,W., Min,L., Qiu,X., Wu,X., Liu,C., Ma,J., Zhang,D. and Zhu,L. (2021) Biological function of long non-coding RNA (LncRNA) xist. *Front. Cell Dev. Biol.*, **9**, 645647.
37. Persson,E., Castresana-Aguirre,M., Buzzao,D., Guala,D. and Sonnhammer,E.L.L. (2021) FunCoup 5: functional association networks in all domains of life, supporting directed links and tissue-specificity. *J. Mol. Biol.*, **433**, 166835.
38. Hewitson,J.P., West,K.A., James,K.R., Rani,G.F., Dey,N., Romano,A., Brown,N., Teichmann,S.A., Kaye,P.M. and Lagos,D. (2020) Malat1 Suppresses immunity to infection through promoting expression of Maf and IL-10 in Th cells. *J. Immunol.*, **204**, 2949–2960.
39. Bonetti,A., Agostini,F., Suzuki,A.M., Hashimoto,K., Pascarella,G., Gimenez,J., Roos,L., Nash,A.J., Ghilotti,M., Cameron,C.J.F., *et al.* (2020) RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.*, **11**, 1018.
40. Engreitz,J.M., Sirokman,K., McDonel,P., Shishkin,A.A., Surka,C., Russell,P., Grossman,S.R., Chow,A.Y., Guttman,M. and Lander,E.S. (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell*, **159**, 188–199.
41. Yoon,J.-H., Abdelmohsen,K. and Gorospe,M. (2013) Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.*, **425**, 3723–3730.
42. Sharova,L.V., Sharov,A.A., Nedorezov,T., Piao,Y., Shaik,N. and Ko,M.S.H. (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes*, **16**, 45–58.
43. Tilgner,H., Knowles,D.G., Johnson,R., Davis,C.A., Chakrabortty,S., Djebali,S., Curado,J., Snyder,M., Gingeras,T.R. and Guigó,R. (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*, **22**, 1616–1625.
44. Mei,S., Qin,Q., Wu,Q., Sun,H., Zheng,R., Zang,C., Zhu,M., Wu,J., Shi,X., Taing,L., *et al.* (2017) Cistrome Data Browser: a data portal for ChIP-seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.*, **45**, D658–D662.
45. Ghosh,A., Pandey,S.P., Joshi,D.C., Rana,P., Ansari,A.H., Sundar,J.S., Singh,P., Khan,Y., Ekka,M.K., Chakraborty,D., *et al.* (2023) Identification of G-quadruplex structures in MALAT1 lncRNA that interact with nucleolin and nucleophosmin. *Nucleic Acids Res.*, **51**, 9415–9431.
46. Amodio,N., Raimondi,L., Juli,G., Stamato,M.A., Caracciolo,D., Tagliaferri,P. and Tassone,P. (2018) MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J. Hematol. Oncol. J Hematol Oncol*, **11**, 63.
47. Chronis,C., Fiziev,P., Papp,B., Butz,S., Bonora,G., Sabri,S., Ernst,J. and Plath,K. (2017) Cooperative binding of transcription factors orchestrates reprogramming. *Cell*, **168**, 442–459.
48. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K., *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
49. Zhang,B., Arun,G., Mao,Y.S., Lazar,Z., Hung,G., Bhattacharjee,G., Xiao,X., Booth,C.J., Wu,J., Zhang,C., *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.*, **2**, 111–123.

50. Tripathi,V., Ellis,J.D., Shen,Z., Song,D.Y., Pan,Q., Watt,A.T., Freier,S.M., Bennett,C.F., Sharma,A., Bubulya,P.A., *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, **39**, 925–938.
51. Gutschner,T., Hämmerle,M., Eissmann,M., Hsu,J., Kim,Y., Hung,G., Revenko,A., Arun,G., Stentrup,M., Gross,M., *et al.* (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.*, **73**, 1180–1189.
52. Bernard,D., Prasanth,K.V., Tripathi,V., Colasse,S., Nakamura,T., Xuan,Z., Zhang,M.Q., Sedel,F., Jourdren,L., Coulpier,F., *et al.* (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.*, **29**, 3082–3093.
53. Tripathi,V., Shen,Z., Chakraborty,A., Giri,S., Freier,S.M., Wu,X., Zhang,Y., Gorospe,M., Prasanth,S.G., Lal,A., *et al.* (2013) Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.*, **9**, e1003368.
54. Li,X., Zhou,B., Chen,L., Gou,L.-T., Li,H. and Fu,X.-D. (2017) GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.*, **35**, 940–950.
55. Eißmann,M., Gutschner,T., Hämmerle,M., Günther,S., Caudron-Herger,M., Groß,M., Schirmacher,P., Rippe,K., Braun,T., Diederichs,S., *et al.* (2012) Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol*, **9**, 1076–1087.
56. Barbaric,I., Miller,G. and Dear,T.N. (2007) Appearances can be deceiving: phenotypes of knockout mice. *Brief. Funct. Genomic. Proteomic.*, **6**, 91–103.
57. El-Brolosy,M.A. and Stainier,D.Y.R. (2017) Genetic compensation: a phenomenon in search of mechanisms. *PLoS Genet.*, **13**, e1006780.
58. Dawlaty,M.M., Ganz,K., Powell,B.E., Hu,Y.-C., Markoulaki,S., Cheng,A.W., Gao,Q., Kim,J., Choi,S.-W., Page,D.C., *et al.* (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, **9**, 166–175.
59. Daude,N., Wohlgemuth,S., Brown,R., Pitstick,R., Gapeshina,H., Yang,J., Carlson,G.A. and Westaway,D. (2012) Knockout of the prion protein (PrP)-like Sprn gene does not produce embryonic lethality in combination with PrP(C)-deficiency. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9035–9040.
60. Chong,M.M.W., Rasmussen,J.P., Rudensky,A.Y. and Littman,D.R. (2008) The RNAseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. *J. Exp. Med.*, **205**, 2005–2017.
61. Mármol-Sánchez,E., Cirera,S., Zingaretti,L.M., Jacobsen,M.J., Ramayo-Caldas,Y., Jørgensen,C.B., Fredholm,M., Cardoso,T.F., Quintanilla,R. and Amills,M. (2022) Modeling microRNA-driven post-transcriptional regulation using exon-intron split analysis in pigs. *Anim. Genet.*, **53**, 613–626.

Received: December 7, 2023. **Revised:** March 14, 2024. **Editorial Decision:** April 15, 2024. **Accepted:** April 30, 2024