

# 3D genome organization during TGFB-induced transcription requires nuclear microRNA and G-quadruplexes

**Authors:** Julio Cordero<sup>1, 2, 3, \*</sup>, Guruprasadh Swaminathan<sup>4</sup>, Diana G Rogel-Ayala<sup>3, 4</sup>, Karla Rubio<sup>3, 4, 5, 6</sup>, Adel Elsherbiny<sup>1, 2</sup>, Stefan Günther<sup>7, 8</sup>, Thomas Braun<sup>8</sup>, Gergana Dobrev<sup>1, 2</sup>, and Guillermo Barreto<sup>3, 4, \*, †</sup>

## Affiliations:

<sup>1</sup>Anatomy and Developmental Biology, CBTM; 68167 Mannheim, Germany.

<sup>2</sup>European Center for Angioscience (ECAS), Medical Faculty Mannheim, Heidelberg University; 68167 Mannheim, Germany.

<sup>3</sup>Lung Cancer Epigenetics, Max-Planck-Institute for Heart and Lung Research; 61231 Bad Nauheim, Germany.

<sup>4</sup>Université de Lorraine, CNRS, Laboratoire IMoPA, UMR 7365; F-54000 Nancy, France.

<sup>5</sup>Massachusetts General Hospital and Harvard Medical School; Charlestown, MA, 02129, USA

<sup>6</sup>International Laboratory EPIGEN, Consejo de Ciencia y Tecnología del Estado de Puebla (CONCYTEP), Instituto de Ciencias, EcoCampus, Benemérita Universidad Autónoma de Puebla, 72570 Puebla, Mexico.

<sup>7</sup>ECCPS Bioinformatics and Deep Sequencing Platform, Max-Planck-Institute for Heart and Lung Research; 61231 Bad Nauheim, Germany.

<sup>8</sup>Department of Cardiac Development, Max-Planck-Institute for Heart and Lung Research; 61231 Bad Nauheim, Germany.

\* Correspondence to: [Guillermo.Barreto@univ-lorraine.fr](mailto:Guillermo.Barreto@univ-lorraine.fr) AND [Julio.Cordero@medma.uni-heidelberg.de](mailto:Julio.Cordero@medma.uni-heidelberg.de)

† Lead contact.

## ABSTRACT

Studying the dynamics of three-dimensional (3D) chromatin structure is essential to understand biological processes in the cell nucleus. Recent publications based on integrative analysis of multi-omics studies have provided comprehensive and multilevel insights into 3D genome organization emphasizing its role during transcriptional regulation. While enhancers are regulatory elements that play a central role in the spatiotemporal control of gene expression, chromatin looping has been broadly accepted as a means for enhancer-promoter interactions allowing them to establish cell-type-specific gene expression signatures. On the other hand, G-quadruplexes (G4s) are non-canonical DNA secondary structures that are both, enriched at promoters and related to increased gene expression. However, the role of G4s in promoter-distal regulatory elements, such as super-enhancers (SE), as well as in 3D genome organization and chromatin looping mediating long-range enhancer-promoter interactions has remained elusive. Here we show that mature microRNA 9 (*miR-9*) is enriched at promoters and SE of genes that are inducible by tissue growth factor beta 1 (TGFB1) signaling. Further, we found that nuclear *miR-9* is required for chromatin features related to increased transcriptional activity, such as broad domains of the euchromatin histone mark H3K4me3 (histone 3 trimethylated lysine 4) and G4s. Moreover, we show that nuclear *miR-9* is required for promoter-super-enhancer looping. Our study places a nuclear microRNA in the same structural and functional context with G4s and promoter-enhancer interactions during 3D genome organization and transcriptional activation induced by TGFB1 signaling, a critical regulator of proliferation programs in cancer and fibrosis.

**KEYWORDS:** nuclear miRNA, transcription, *miR-9*, G-quadruplex, promoter-enhancer looping, transforming growth factor, TGFB

## INTRODUCTION

The nuclear genome in eukaryotic cells consists of DNA molecules packaged into thread-like structures known as chromosomes, which are built of chromatin. Thus, chromatin is the physiological template for biological processes in the nucleus of eukaryotic cells. Studying how chromatin is folded inside the cell nucleus and its dynamic three-dimensional (3D) structure is essential to understand these biological processes that comprise transcription, RNA-splicing, -processing, -editing, DNA-replication, -recombination, and -repair. The chromatin is hierarchically organized at different levels including chromosomal territories, compartments, and self-interacting topologically associating domains (TADs) altogether giving rise to a highly dynamic 3D genome organization <sup>1,2</sup>. Remarkably, the structure of the genome is intrinsically associated with its function as shown by extensive correlations between chromatin condensation and related gene transcription. For example, chromatin shows condensed regions, referred to as heterochromatin (by convention, transcriptionally “inactive”), and less condensed regions, referred to as euchromatin (transcriptionally “active”). Transcriptional regulation directly corresponds to the mechanisms of how chromatin can be structurally arranged making it accessible for the transcription machinery <sup>3</sup>. These mechanisms regulating chromatin structure and transcription comprise histone modifications, histone deposition, nucleosome remodeling, DNA methylation, non-coding RNAs (ncRNA), and secondary structures of nucleic acids, among others <sup>4-9</sup>. In addition, an increasing number of recent publications based on integrative analysis of multi-omics studies implementing next-generation sequencing (NGS) technologies, chromosome conformation capture-based methods, and super-resolution microscopy have provided comprehensive and multilevel insights into 3D genome organization emphasizing its role during transcriptional regulation <sup>10</sup>.

Chromatin structure alone does not determine the functional status of a gene, but it effectively enables RNA polymerase II (Pol II) recruitment to the promoters, as well as binding of transcription factors, co-activators, co-repressors to DNA sequences that function as regulatory elements controlling gene expression<sup>11</sup>. A promoter is a sequence of DNA to which proteins bind to initiate transcription of RNA molecules that are usually complementary to the DNA sequence that is located 3' of the promoter. On the other hand, enhancers are relatively short (~100–1000 bp) DNA sequences that are bound by transcription factors and regulate gene transcription independent of their distance, location, or orientation relative to their cognate promoter<sup>12-14</sup>. Super-enhancers (SE) have been proposed to be long genomic domains consisting of clusters of transcriptional enhancers enriched with histone modification markers (such as histone 3 mono-methylated at lysine 4 or acetylated at lysine 27, H3K4me1 and H3K27ac respectively), cofactors (such as mediator of RNA polymerase II transcription subunit 1, MED1, and components of the multimeric protein complex Cohesin), chromatin modifying proteins (such as E1A Binding Protein P300, EP300) and cell-type-specific transcription factors<sup>15-17</sup>. One versatile feature of chromatin is its ability to form a loop, mediating long-range interactions in which two distant sequences of DNA come to close physical proximity. Chromatin looping has been broadly accepted as a means for enhancer-promoter interactions<sup>18, 19</sup>.

In addition to the predominant DNA double-helix structure, there are different non-canonical DNA secondary structures, including G-quadruplex (G4), R-loop, H-DNA, Z-DNA, etc.<sup>6</sup>. A G4 is a stable nucleic acid secondary structure formed by square planes, in which four guanines located in the same plane stabilized by a monovalent cation<sup>20, 21</sup>. While the early work about G4s mainly focused on their roles in telomeres<sup>22</sup>, recent studies demonstrated that G4s are both, enriched at promoters<sup>23, 24</sup> and related to increased gene expression<sup>25-29</sup>. On the other

hand, G4s were also located in gene bodies and related to reduced gene expression by inhibiting elongation of RNA polymerase<sup>30, 31</sup>. All these previous studies characterized the biological function of G4s in promoter or promoter-proximal regions enhancing or reducing gene expression depending on the relative position of G4s. Nevertheless, the role of G4s in promoter-distal regulatory elements, such as SE, as well as in chromatin looping mediating long-range enhancer-promoter interactions remains unclear.

The majority of the eukaryotic genome is transcribed into ncRNAs including microRNAs (miRNAs, 21-25 nucleotides long) and long non-coding RNAs (lncRNAs, >200 nucleotides long)<sup>32</sup>. lncRNAs are important regulators of different biological processes in the nucleus<sup>33</sup>. Together with other factors, lncRNAs provide a framework for the assembly of defined chromatin structures at specific loci, thereby modulating gene expression, centromere function and silencing of repetitive DNA elements<sup>33-36</sup>. Although miRNAs are assumed to act primarily in the cytosol by inhibiting translation<sup>37</sup>, mature miRNAs have been also reported in the nuclei of different cells<sup>8, 38-41</sup>. Interestingly, a hexanucleotide element has been reported to direct miRNA nuclear import<sup>42</sup>, nevertheless, the function of miRNAs in the cell nucleus has been sparsely studied. Here we report on microRNA-9 (*miR-9*), which even though its nucleotide sequence is highly conserved across species, also shows high diversity in expression patterns and biological functions depending on the cellular context<sup>43, 44</sup>. For example, *miR-9* has been reported to target the lncRNA *MALAT1* for degradation in the cell nucleus<sup>41</sup>. However, it has not been linked to transcription regulation, chromatin structure nor 3D genome organization. Here we propose a mechanism of transcriptional regulation of transforming growth factor beta 1 (TGFB1) responsive genes that requires nuclear *miR-9* and involves G4s and promoter-SE looping.

## RESULTS

### **Mature *miR-9* is detected in the cell nucleus enriched at promoters and introns.**

A phylogenetic tree that was generated using the sequences of mature mouse miRNAs and a heat map comparing their sequence similarity showed that *miR-9* clustered together with miRNAs that have been functionally characterized in the cell nucleus, such as *miR-29b-3p*<sup>42</sup>, *miR-126-5p*<sup>45</sup> and *let-7d-5p*<sup>8</sup> (Figure 1a, top). Sequence alignment between mouse *miR-126-5p*, *miR-9* and *miR-29b-3p* showed that various nucleotides are conserved in a sequence stretch that was reported as nuclear shuttling motif from *miR-126-5p*<sup>45</sup> (Figure 1a, bottom). We called the partially conserved sequence 5'-AKYACCWUUUGRUWA-3' as expanded miRNA nuclear shuttling motif. In addition, we found that the human orthologs of these mature miRNAs, *hsa-miR-126-5P*, *hsa-miR-29B-3P* and *hsa-miR-9-5P*, also contain this expanded miRNA nuclear shuttling motif (Supplementary Figure 1a), showing that it is conserved across species. To confirm the nuclear localization of mature *miR-9*, we performed expression analysis after cell fraction using TaqMan assays specific for mature *miR-9* and total RNA isolated from the cytosolic and the nuclear fractions of different cells (Figure 1b and Supplementary Figure 1b), including mouse lung fibroblasts (MLg and MFML4), mouse lung epithelial cells (MLE-12), mouse mammary gland epithelial cells (NMuMG), and primary human lung fibroblasts (hLF). We detected mature *miR-9* not only in the cytosolic fraction but also in the nuclear fraction of all cells analyzed. Interestingly, the relative levels of nuclear *miR-9* were higher in mouse fibroblasts when compared to epithelial cells. Further, the nuclear localization of *miR-9* was confirmed by RNA fluorescence in situ hybridization (FISH) in MLg cells (Figure 1c, top; Supplementary Figure 1c, left), which also showed that specific regions of the nucleus were *miR-9* depleted. These results were further validated by loss-of-function (LOF) experiments using unlabeled *miR-9*-specific antagomiR probes (Figure 1c, bottom; Supplementary Figure 1c, right). All these results suggest that *miR-9* has a function in the cell

nucleus. To investigate the role of *miR-9* in the cell nucleus we performed a sequencing experiment after chromatin isolation by miRNA purification (ChIRP-seq, Figure 1d and Supplementary Figure 1d) using chromatin from MLg cells and control (Ctrl) or *miR-9*-specific biotinylated antisense oligonucleotides for the precipitation of endogenous mature *miR-9* along with the chromatin bound to it. The genome-wide binding profile analysis of *miR-9* (Figure 1d) revealed over twofold increase in the number of *miR-9* peaks at promoters and intronic regions compared to the annotated promoters and intronic regions in the mouse genome. The loci of these putative *miR-9* targets were distributed on all chromosomes ( $n = 2,637$ ; Supplementary Figure 1e-f). From the *miR-9* ChIRP-seq results we selected putative *miR-9* target genes (*Zdhhc5*, *Ncl*, *Lzts2* and *Hdac7*) for further analysis. Visualization of the loci of these putative *miR-9* target genes using the integrative genomic viewer (IGV) (Figure 1e) showed *miR-9* enrichment at the promoters, whereas no *miR-9* enrichment was detected at the promoter of the negative control *Prap1*. These results were confirmed by quantitative PCR after ChIRP using promoter-specific primers, chromatin from MLg cells and Ctrl or *miR-9*-specific biotinylated antisense oligonucleotides for the precipitation of endogenous *miR-9*-chromatin complexes (Figure 1f). Our results together demonstrate that mature *miR-9* is in the cell nucleus and directly binds to promoters of putative *miR-9*-target genes suggesting a potential role in transcription regulation.

### ***MiR-9* is required for H3K4me3 broad domains, high basal transcriptional activity, and G-quadruplex formation at promoters.**

To further investigate a potential role of nuclear *miR-9* in transcription regulation, we performed a sequencing experiment after chromatin immunoprecipitation assay (ChIP-seq, Figure 2a, Supplementary Figure 2a) using antibodies specific for the euchromatin histone mark, tri-methylated lysine 4 of histone 3 (H3K4me3), and chromatin from MLg cells that

were transiently transfected with Ctrl or *miR9*-specific antagomiR to induce a *miR-9*-LOF. Interestingly, *miR-9*-LOF significantly reduced the broad domains of H3K4me3 from 55% in Ctrl transfected cells to 27% after *miR-9*-LOF ( $P < 0.01$ ), whereas medium and narrow H3K4me3 domains increased. The reduction of H3K4me3 broad domains after *miR-9*-LOF was confirmed by confocal microscopy after H3K4me3-specific immunostaining in Ctrl- and *miR-9*-antagomiR transfected MLg cells (Supplementary Figure 2b). Since broad domains of H3K4me3 have been associated with increased transcription elongation<sup>46</sup>, we analyzed the transcriptome of MLg cells after *miR-9*-LOF by total RNA sequencing (RNA-seq, Figure 2b-c, Supplementary Figure 2c). Remarkably, from the transcripts that were significantly affected after *miR-9*-LOF ( $n = 3,320$ ), only a minority ( $n = 881$ ; 26.5%) showed increased expression after *miR-9*-LOF, whereas 73.5% ( $n = 2,439$ ) showed reduced expression with a median of 1.03 log<sub>2</sub> RPKM and an interquartile range (IQR) of 1.51 log<sub>2</sub> RPKM ( $P = 5.34E-36$ ), when compared to 1.50 log<sub>2</sub> RPKM (IQR = 1.81 log<sub>2</sub> RPKM) in Ctrl antagomiR transfected cells (Figure 2c, top). The expression of genes without *miR-9*-enrichment (non-targets) as determined by *miR-9* ChIRP-seq (Figure 1d-e and Supplementary Figure 1d) was non-significantly affected after *miR-9*-LOF (Figure 2c, bottom) supporting the specificity of the *miR9*-specific antagomiR. These results indicate that *miR-9* is required for high basal transcriptional activity of its target genes. This interpretation was supported by ChIP-seq in mouse embryonic fibroblasts (MEF) using antibodies specific for total RNA polymerase II (Pol II) and serine 5 phosphorylated Pol II (Pol II S5p,<sup>9</sup>) showing transcription initiation. Heat maps representing the results of the *miR-9* ChIRP-seq (Figure 2d), Pol II and Pol II S5p ChIP-seq (Figure 2e) revealed that Pol II and Pol II S5p were enriched at the transcription start sites (TSS) of the genes coding for the transcripts that were significantly affected after *miR-9*-LOF (further referred to as *miR-9* target genes). Moreover, genome-wide precision nuclear run-on assay (PRO-seq)<sup>47</sup> and global run-on sequencing (GRO-seq)<sup>48</sup>, both in MEF, showed nascent



200 RNAs at the TSS of the *miR-9* target genes demonstrating their basal transcriptional activity  
 201 (Figure 2f). Correlating with these results, we also observed at the TSS of the *miR-9* target  
 202 genes increased chromatin accessibility and increased H3K4me3 levels by assay for  
 203 transposase-accessible chromatin with sequencing (ATAC-seq) and ChIP-seq, respectively  
 204 (Figure 2g). To gain further insights into these results, we performed a motif analysis of the  
 205 *miR-9* target genes and identified significant enrichment of nucleotides motifs with high G  
 206 content (Figure 2h), as the ones that favor the formation of G4<sup>6, 20</sup>. Remarkably, we found  
 207 similar motifs significantly enriched in loci that form G4 as determined by G4 ChIP-seq.  
 208 Further, since G4 has been shown to cooperate with transcription factors at gene promoters<sup>27</sup>,  
 209<sup>28</sup>, we analyzed publicly available NGS data that were generated implementing different  
 210 methods for the assessment of G4 formation (Figure 2i). On one hand, we analyzed ChIP-seq  
 211 data that were generated using an artificial 6.7 kDa G4 probe (G4P) protein, which binds G4s  
 212 with high affinity and specificity<sup>24</sup>. On the other hand, we analyzed NGS data that were  
 213 generated by G4access, which is an antibody-independent method that relies on moderate  
 214 nuclease digestion of chromatinized DNA<sup>28</sup>. Remarkably, we found by both approaches  
 215 enrichment of G4 at the TSS of the *miR-9* target gene (Figure 2i). Our results demonstrate that  
 216 the TSS of *miR-9* target genes show (1) reduced nucleosome density, (2) increased levels of  
 217 the euchromatin histone mark H3K4me3, (3) enrichment of G4, *miR-9* and transcription  
 218 initiating Poll II S5p, and (4) nascent RNAs.

219  
 220 Loci visualization of selected *miR-9* target genes (*Zdhhc5*, *Ncl*, *Lzts2*, *Hdac7*, *Gata6* and  
 221 *Ep300*) using the IGV genome browser (Figure 3a top and Supplementary Figure 3) showed  
 222 enrichment of the euchromatin histone mark H3K4me3 at the promoters, which was reduced  
 223 after *miR-9*-LOF. We also observed at the same loci nascent RNA, supporting basal  
 224 transcriptional activity, and G4 enrichment. Interestingly, a zoom in to the loci revealed G-rich

sequences that favor the formation of G4<sup>6, 20</sup> (Figure 3a, bottom, and Supplementary Figure 3). These results were confirmed by promoter analysis of *Zdhhc5*, *Ncl* and *Lzts2* by ChIP using H3K4me3- or G4- specific antibodies and chromatin from MLg cells that were transiently transfected with Ctrl or *miR9*-specific antagomiR probes (Figure 3b). We detected H3K4me3 and G4 enrichment at the promoters of all analyzed *miR-9* target genes in Ctrl antagomiR-transfected cells, which was significantly reduced after *miR-9*-LOF. Our results demonstrate that the promoters of *miR-9* target genes are enriched with H3K4me3 and G4, correlating with the basal transcriptional activity detected by RNA-seq (Figure 2b-c), in a *miR-9*-dependent manner.

### **Nuclear *miR-9* is enriched at super-enhancers and is required for G-quadruplexes.**

Alluvial plot implementing the data from *miR-9* ChIRP-seq and G4P ChIP-seq<sup>24</sup> showed enrichment of *miR-9* and G4 at loci that are also enriched with markers of active SE, such as MED1 and H3K27Ac<sup>49, 50</sup> (Figure 4a). Furthermore, a Venn diagram using the same data sets together with the data from a GRO-seq experiment<sup>48</sup> showed 3,583 common loci (Figure 4b) suggesting transcriptional activity from these loci. Remarkably, 95.5% ( $n = 3,423$ ) of these transcripts were found in the animal eRNA database<sup>51</sup> as enhancer RNAs (eRNA) (Figure 4c). Moreover, mapping the *miR-9* ChIRP-seq data to known loci of typical enhancer (TE) or SE showed that *miR-9* is significantly enriched at SE (Figure 4d). These results were confirmed by further analysis of the *miR-9* ChIRP-seq data together with publicly available ChIP-seq data of proteins that have been related to SE (Figure 4e). Aggregate plots showed enrichment of *miR-9* specifically at SE together with MED1, MYC, H3K27Ac, KLF4, MEIS1, EP300, HDAC1, SMAD3, BRG1, HDAC2, EST1 and RAD21. In addition, we also observed enrichment of CHD4 and SMARCA5 at the same SE as *miR-9*. Supporting these observations, we found that *miR-9* pulled down endogenous CHD4 and SMARCA5 by miRNA pulldown

(miR-Pd) followed by Western Blot (Supplementary Figure 4a), whereas both proteins precipitated endogenous *miR-9* by chromatin-RNA immunoprecipitation (Ch-RIP) followed by *miR-9*-specific TaqMan assays (Supplementary Figure 4b). Our results suggest CHD4 and SMARCA5 as components of SE. Remarkably, aggregate plots using G4-specific ChIP-seq data<sup>24</sup> showed G4 enrichment at the same SE as *miR-9* in mouse lung fibroblasts (MLg cells) and adenocarcinoma human alveolar basal epithelial cells (A549 cells) (Figure 4f).

Loci visualization of selected SE with *miR-9* enrichment using the IGV genome browser (Figure 5a top and Supplementary Figure 5) confirmed the enrichment of G4 at the same loci, as well as of MED1, KLF4 and H3K27Ac<sup>49, 50</sup>, which are markers of active SE. Interestingly, a zoom in to the loci revealed G-rich sequences that favor the formation of G4<sup>6, 20</sup> (Figure 5a, bottom, and Supplementary Figure 5). These results were confirmed by analysis of the loci of these SE by ChIP followed by qPCR using H3K4me3- or G4-specific antibodies and chromatin from MLg cells that were transiently transfected with Ctrl or *miR9*-specific antagomiR probes (Figure 5b). We detected H3K4me3 and G4 enrichment at the loci of all analyzed SE in Ctrl antagomiR-transfected cells, which was significantly reduced after *miR-9*-LOF. Our results demonstrate that SE with *miR-9* enrichment are also enriched with H3K4me3 and G4 in a *miR-9*-dependent manner.

### **Promoter-super-enhancer looping of TGFB1-responsive genes requires *miR-9*.**

To further elucidate the biological relevance of our findings we performed gene set enrichment analysis (GSEA)<sup>52</sup> of the loci with *miR-9* enrichment and nascent RNA as determined by *miR-9* ChIRP-seq and GRO-seq, respectively (Figure 6a-b). We found significant enrichment of genes related to the categories “TGFB cell response” ( $P = 7.55E-09$ ), “TGFB” ( $P = 2.61E-08$ ),

“Cell proliferation” ( $P = 4.66\text{E-}08$ ) and “Fibroblasts proliferation” ( $P = 1.61\text{E-}07$ ), suggesting an involvement of the loci with *miR-9* enrichment and nascent RNA in these biological processes. To link this observation with our results showing reduction of H3K4me3 after *miR-9*-LOF at specific promoters (Figure 3b) and SE (Figure 5b) we did ChIP-seq using H3K4me3-specific antibodies and chromatin from MLg cells that were transiently transfected with Ctrl or *miR9*-specific antagomiR probes, and non-treated or treated with TGFB1 (Figure 6c and Supplementary Figure 6a-b). Loci visualization of the selected *miR-9* target genes (*Zdhhc5*, *Ncl*, *Lzts2*, *Hdac7*, *Gata6* and *Ep300*) using the IGV genome browser showed H3K4me3 enrichment at the promoters in non-treated, and Ctrl antagomiR transfected MLg cells that increased after TGFB1 treatment. However, a combination of TGFB1 treatment and *miR9*-specific antagomiR transfection showed that *miR-9*-LOF counteracted the effect caused by TGFB1 demonstrating the requirement of *miR-9* for the chromatin changes induced by TGFB1 and suggesting its requirement for TGFB1-inducibility of the analyzed genes, as shown below.

Since we observed enrichment of *miR-9*, H3K4me3 and G4 not only at promoters of *miR-9* target genes (Figures 1-3), but also in specific SE (Figures 4-5), we decided to analyze the genome-wide effect of TGFB1 on chromatin conformation by a technique that consists of an *in situ* Hi-C library preparation followed by a chromatin immunoprecipitation (HiChIP). For this HiChIP-seq we used H3K4me3-specific antibodies to precipitate active SE and chromatin from MLg cells that were transiently transfected with Ctrl or *miR9*-specific antagomiR probes, and non-treated or treated with TGFB1 (Figure 6d and Supplementary Figure 7). Analysis of the H3K4me3-specific HiChIP-seq data by k-means clustering revealed four clusters. We focused on clusters 1 and 3 since we observed in these two clusters an increase of chromatin interactions in response to TGFB1 treatment in a *miR-9*-dependent manner, since *miR-9*-LOF reduced the effect caused by TGFB1. Interestingly, in these two clusters we found the loci of

the *miR-9* targets (*Zdhhc5*, *Ncl*, *Lzts2* and *Hdac7*) that were selected based on our *miR-9* ChIRP-seq (Figure 1c-e). Further, we generated IGV genome browser snapshots to visualize the enrichment of *miR-9*, G4, MED1, KLF4 and H3K27ac at the loci of promoters of *miR-9* target genes and SE with *miR-9* enrichment (Figure 7a, top). In the same snapshots we presented the results of the H3K4me3-specific HiChIP-seq (Figure 7a, bottom) showing chromatin loops between the promoters of *miR-9* target genes and SE with *miR-9* enrichment in Ctrl antagomiR-transfected cells. Strikingly, these promoter-SE-loops increased in TGFB1 treated cells in a *miR-9*-dependent manner, since *miR-9*-LOF counteracted the effect caused by TGFB1. To correlate the results from H3K4me3-specific HiChIP-seq with changes in chromatin, we analyzed the promoters and SE with *miR-9* enrichment by ChIP qPCR using H3K4me3- and G4-specific antibodies and chromatin from MLg cells that were transfected with Ctrl or *miR-9*-specific antagomiRs, and non-treated or treated with TGFB1 (Figure 7b). TGFB1 treatment significantly increased H3K4me3 and G4 levels at the analyzed promoters and SE. Further, *miR-9*-LOF significantly reduced the effect caused by TGFB1 treatment, whereas *miR-9*-LOF alone significantly reduced H3K4me3 and G4 levels when compared to Ctrl antagomiR transfected cells. To correlate these changes in chromatin structure with gene expression, we analyzed the expression of *miR-9* target genes by qRT-PCR in MLg cells under the same conditions as specified above (Figure 8a). The expression of all analyzed *miR-9* target genes significantly increased after TGFB1 treatment in *miR-9*-dependent manner, correlating with our chromatin structure analysis (Figures 6c and 7b). Further, *miR-9*-LOF alone significantly reduced the basal transcription levels of the analyzed *miR-9* target genes when compared to Ctrl antagomiR transfected cells confirming our RNA-seq results (Figure 2b-c and Supplementary Figure 2c). Summarizing, our results support a model (Figure 8b), in which G4s are formed in a *miR-9*-dependent manner at both, promoters of TGFB1-responsive genes, as well as SE with which these promoters form loops (left). Further, H3K4me3, G4 and

324 promoter-SE looping increased after TGFB1 treatment allowing these two regulatory elements  
325 to come to close physical proximity and enhance transcription of the corresponding genes  
326 (middle) also in *miR-9*-dependent manner (right).

327

## DISCUSSION

Our study places a nuclear microRNA in the same structural and functional context with non-canonical DNA secondary structures and 3D genome organization during transcription activation. We uncovered a mechanism of transcriptional regulation of TFGFB1-responsive genes that requires nuclear *miR-9* and involves G4s and promoter-SE looping. Various aspects of the model proposed here are novel. For example, nuclear *miR-9* had not been related to transcription regulation nor chromatin structure prior to our study, even though *miR-9* participates in a wide spectrum of biological functions including AGO2-dependent degradation of the lncRNA *MALAT1* in the cell nucleus<sup>41, 43, 44</sup>. Previously, we have shown that other nuclear miRNA, lethal 7d (*Mirlet7d*, also known as *let-7d*), is part of the ncRNA-protein complex MiCEE that mediates epigenetic silencing of bidirectionally transcribed genes and nucleolar organization<sup>8, 53</sup>. In this context, nuclear *Mirlet7d* binds ncRNAs expressed from these genes, and mediates their degradation by the RNA exosome complex. It will be the scope of future work to determine whether nuclear *miR-9* also binds to ncRNA expressed from the *miR-9* target loci and mediates their degradation by a similar mechanism. Following this line of ideas, it has been reported that G4s are found in genomic regions containing R-loops<sup>54, 55</sup>, which are three-stranded nucleic acid structures consisting of a DNA-RNA hybrid and the associated non-template single-stranded DNA<sup>56</sup>. Supporting this line of ideas, we have previously reported that R-loops regulate transcription in response to TGFβ1 signaling<sup>9</sup>. Interestingly, when a G4 is formed opposite to the R-loop on the associated single-stranded DNA, a so called a G-loop structure is generated<sup>57</sup>. Sato and colleagues recently reported a mechanism involving G-loop structures, in which the transcripts stabilizing the R-loops are relevant for the controlled resolution of the G4s, thereby preventing mutagenic G4s and supporting genomic stability<sup>57</sup>. We will investigate in a future project whether nuclear *miR-9*

is involved in a similar mechanism targeting the ncRNAs in the R-loops and promoting controlled G4 resolution during transcription initiation.

Another interesting aspect of the model proposed here is the participation of G4s in 3D genome organization during transcription activation. DNA G4s are stable four-stranded non-canonical structures that are highly related to promoters and transcription activation<sup>23-26</sup>. Further, a recent publication based on integrative analysis of multi-omics studies have provided comprehensive mechanistic insights into the function of G4s as promoter elements that reduce nucleosome density, increase the levels of active histone marks (H3K4me3 and H3K27ac), generate nucleosome arrays by positioning nucleosomes at a periodic distance to each other and facilitate pause release of Poll II into effective RNA production resulting in enhanced transcriptional activity<sup>29</sup>. Even though similar chromatin features were described to be induced by G4s at promoter-distal regulatory elements, such as SE<sup>58</sup>, no mechanistic insights were provided prior to our study elucidating the role of G4s in chromatin looping mediating long-range enhancer-promoter interactions. We found that G4s are formed in a *miR-9*-dependent manner at both, promoters of TGFB1-responsive genes, as well as SE with which these promoters form loops (Figure 7a-b). Moreover, our results showed promoter-SE looping increase after TGFB1 treatment allowing these two regulatory elements to come to close physical proximity and enhance transcription of the corresponding genes (Figure 8a). One can hypothesize that G4 could stabilize the interaction between these two regulatory elements allowing the increase of transcription. We will investigate this hypothesis in a future project and whether the G4 are formed with DNA strands from the promoter and from the SE.



Our *miR-9*-LOF experiments showed that *miR-9* is required for H3K4me3 broad domains (Figure 2a), basal transcriptional activity (Figure 2b-c) and TGFB1-inducibility (Figure 7c) of *miR-9* target genes. H3K4me3 is a well-characterized euchromatin histone mark related to genes with high transcriptional activity, probably contributing to release of Pol II pausing into elongation<sup>46, 59</sup>. Furthermore, it has been shown that genes with broad domains of H3K4me3 are transcriptionally more active than genes with narrow domains<sup>46, 60</sup>, which is consistent with our findings. Interestingly, H3K4me3 broad domains have been linked to genes that are critical to cellular identity and differentiation<sup>46, 60-62</sup>. Moreover, H3K4me3 broad domains are associated with increased transcription elongation and enhancer activity, which together lead to exceptionally high expression of tumor suppressor genes, including TP53 and PTEN<sup>46</sup>. On the other hand, TGFB signaling is one of the prominent pathways implicated in hyperproliferative disorders, including cancer<sup>36, 53, 63-65</sup>. In early stages of cancer, TGFB signaling exhibits tumor suppressive effects by inhibiting cell cycle progression and promoting apoptosis. However, in late stages, TGFB signaling exerts tumor promoting effects, increasing tumor invasiveness, and metastasis. It will be the scope of our future work to determine the clinical relevance of the model of transcription regulation proposed here within the context of hyperproliferative disorders with special attention on carcinogenesis.

## MATERIALS AND METHODS

### Cell culture

Mouse lung fibroblast cells MLg (ATCC CCL-206) and MFML4<sup>66</sup> were cultured in complete DMEM (4.5 g/L glucose, 10% FCS, 1% Penn-strep, 2 mM l-glutamine) at 37 °C in 5% CO<sub>2</sub>. Mouse lung epithelial cells MLE-12 (ATCC CRL-2110) were cultured in complete DMEM/F12 (5% FCS, 1% Penn-strep) at 37°C in 5% CO<sub>2</sub>. Mouse mammary gland epithelial cells NMuMG (ATCC CRL-1636) were cultured in complete DMEM (4.5 g/L glucose, 10mcg/mL insulin (90%), 10% FCS, 1% Penn-strep) at 37 °C in 5% CO<sub>2</sub>. Human primary lung fibroblasts from control donors were cultured in complete MCDB131 medium (8% fetal calf serum (FCS), 1% L-glutamine, penicillin 100 U/ml, streptomycin 0.1 mg/ml, epidermal growth factor 0.5 ng/ml, basic fibroblast growth factor 2 ng/ml, and insulin 5 µg/ml)) at 37°C in 5% CO<sub>2</sub>. During subculturing, cells were 1x PBS washed, trypsinized with 0.25% (w/v) Trypsin and split at the ratio of 1:5 to 1:10. The cell lines used in this paper were mycoplasma free. They were regularly tested for mycoplasma contamination. In addition, they are not listed in the database of commonly misidentified cell lines maintained by ICLAC.

### Cell transfection, treatment and antagomiR-mediated *miR-9* loss-function

Cells were transfected with antagomiR probes (Ambion) using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions, and harvested 24 h later for further analysis. Anti-hsa-miR-*Mir9-5p* (Ambion, #17000), and Anti-miR negative control (Ambion, #17010) were transfected at 60 nM final concentrations. Following 20 h after transfection, TGFB1 signaling was induced with 5 ng/ml final concentration of human recombinant TGFB1 (Sigma-Aldrich) for 4 h.

## Bacterial culture and cloning

For cloning experiments, chemically competent *E. coli* TOP10 (ThermoFisher Scientific) were used for plasmid transformation. TOP10 strains were grown in Luria broth (LB) at 37°C with shaking at 180rpm for 16h or on LB agar at 37°C overnight.

## RNA isolation, reverse transcription, quantitative PCR and TaqMan assay

Expression analysis by qRT-PCR were performed as previously described<sup>67</sup>. Briefly, total RNA from cell lines was isolated using the RNeasy Mini kit (Qiagen) and quantified using a Nanodrop Spectrophotometer (ThermoFisher Scientific). Synthesis of complementary DNA was performed using 1–2 µg total RNA and the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Quantitative real-time PCR reactions were performed using SYBR® Green on the Step One plus Real-time PCR system (Applied Biosystems). Housekeeping gene *Gapdh* was used to normalize gene expression. Primer pairs used for gene expression analysis are described in the following table:

Transcript	Primer sequence (5'→3')	
<i>Zdhhc5</i>	Forward	AGGTGCGAATGAAATGGTGT
	Reverse	TTTACCCAAGGGCAGTGATG
<i>Ncl</i>	Forward	TGTTGTGGATGTCAGAACTGGT
	Reverse	GCTCCAAGGCCTTTTCTAGG
<i>Lzts2</i>	Forward	GCTGTGCCACTTTTGAACG
	Reverse	TGAGATTCCTTCAGTTGCTGC
<i>Hdac7</i>	Forward	AGCTGGCTGAAGTGATCC
	Reverse	TCACCATCAGCCTCTGAG
<i>Prap1</i>	Forward	TGGGACCACTCAGCTACAAG

	Reverse	CCAGTTCTAAGACCTGCCC
--	---------	---------------------

For *miR-9* expression analysis, total RNA was isolated with Trizol (Invitrogen), quantified using a Nanodrop Spectrophotometer (ThermoFisher Scientific), 0.5–2 µg total RNA was used for reverse transcription (High-Capacity cDNA Reverse Transcription Kit, ThermoFisher Scientific) and subsequently *miR-9*-specific TaqMan assay (Applied Biosystems) in the Step One plus Real-time PCR system (Applied Biosystems).

### **MiRNA fluorescence in situ hybridization**

MiRNA Fluorescence in situ hybridization (MiRNA-FISH) was performed as described earlier<sup>8</sup> with minor adaptations. Briefly, cells were fixed with 4% PFA, dehydrated with 70% ethanol and incubated with pre-hybridization buffer (50% formamide, 5X SSC, 5X Denhardt's solution, 200 µg/ml yeast RNA, 500 µg/ml salmon sperm DNA and 2% Roche blocking reagent in DEPC treated water). Incubation with pre-hybridization buffer was carried out during 4h at room temperature. Pre-hybridization buffer was replaced with denaturing hybridization buffer (10% CHAPS, 20% Tween, 50% formamide, 5X SSC, 5X Denhardt's, 200 µg/ml yeast RNA, 500 µg/ml salmon sperm DNA and 2% Roche blocking reagent in DEPC treated water) containing biotin and/or digoxin labeled Locked Nucleic Acid (LNATM) probes (Exiqon) specific to mature *miR-9* and/or miR control respectively to a final concentration of 20 pM and incubated at 55°C overnight. Next day, cells were briefly washed with 5X SSC buffer pre-warmed to 60°C and then incubated with 0.2X SSC at 60°C for 1 h. Later, cells were incubated with B1 solution (0.1 M Tris pH 7.5, 0.15 M NaCl) at room temperature for 10 min. B1 solution was then replaced with blocking solution (10% FCS, 0.1 M Tris pH 7.5, 0.15 M NaCl) and incubated for 1 h at RT. Cells were then incubated with FITC labeled rabbit anti-Biotin

(Abcam) antibody overnight at 4°C. DAPI was used as nuclear dye. Cells were examined with a fluorescence microscope (Leica).

### **Immunofluorescence and confocal microscopy**

Immunostaining was performed as previously described<sup>53</sup>. Briefly, cells were grown on coverslips, fixed with 4% PFA for 10min at RT and permeabilized with 0.4% Triton-X100/1xPBS for 10min at RT. During immunostaining procedure, all incubations and washes were performed with histobuffer containing 3% bovine serum albumin (BSA) and 0.2% Triton X-100 in 1xPBS, pH7.4. Non-specific binding was blocked by incubating with BSA 5% serum in 1x PBS, pH 7.4. Cells were then incubated with primary antibodies overnight at 4°C. After 3 washes with histobuffer (15 min each), secondary antibody was incubated at room temperature for 1h followed by DAPI nuclear staining (Sigma, Germany). Immunostainings were examined with an immunofluorescence microscope (Leica). Antibodies used were specific anti-H3K4me3 (Abcam, ab8580). Alexa 488 or Alexa 594 tagged secondary antibodies (Invitrogen, Germany, dilution 1:1000) were used. DAPI (Sigma, Germany) was used as nuclear dye.

### **Chromatin isolation by miRNA purification and sequencing**

Chromatin isolation by miRNA purification (ChIRP) was performed as described<sup>68</sup>, with slight modifications. Briefly, MLE-12 cells were cross-linked by 1% formaldehyde for 10min, lysed, and sonicated with a Diagenode Bioruptor to disrupt and fragment genomic DNA. After centrifugation, chromatin was incubated with 150 pmol of biotin-labeled anti-sense LNA probes (Exiqon) specific to mature Mirlet7d, or control, at 37 °C for 4h. Streptavidin-magnetic C1 beads (Invitrogen) were blocked with 500ng/μL yeast total RNA and 1mg/mL BSA for 1h at room temperature, and washed three times in nuclear lysis buffer (2mM Tris-HCl, pH 7.0,

250mM NaCl, 2mM EDTA, 2mM EGTA, 1% Triton X-100, 0.2mM DTT, 20mM NaF, 20mM Na<sub>3</sub>VO<sub>4</sub>, 40µg/mL phenylmethylsulfonyl fluoride, protease inhibitor, and RNase inhibitor in DEPC-treated water) and resuspended in its original volume. We added 100 µL of washed/blocked C1 beads per 100 pmol of probes, and the whole reaction was mixed for another 30min at 37 °C. Beads:biotinprobes:RNA:chromatin adducts were captured by magnets (Invitrogen) and washed five times with 40×bead volume of wash buffer. DNA was eluted with a cocktail of 100µg/mL RNase A (ThermoFisher Scientific). Chromatin was reverse cross-linked at 65 °C overnight. Later, DNA was purified using the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions and used for single gene promoter analysis by qPCR and for sequencing. Primer pairs used for qPCR analysis are described in the sub-section "Chromatin immunoprecipitation" below. ChIRP-seq was performed by single-end sequencing on an Illumina HiSeq2500 machine at the Max Planck-Genome-Centre Cologne. Raw reads were trimmed using Trimmomatic-0.36 with the parameters (ILLUMINACLIP:\${ADAPTERS}:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:20 CROP:70 HEADCROP:10 (<https://doi.org/10.1093/bioinformatics/btu170>). Trimmed reads were mapped to mouse genome mm10 using Bowtie2 (default settings) <sup>69</sup>. Next, PCR duplicates were removed from the BAM files using the MarkDuplicates.jar tool from Picard (version 1.119). MCAS14 <sup>70</sup> was used (macs14 -t ChIRP -c Inp -f BAM -p 1e-3 -g 1.87e9 --nomodel --shiftsize 30 -n; and -p 1e-3 -g 1.87e9 --nomodel --shiftsize 100) with the aim to capture genome areas with *miR-9* enrichment. Both peak results were merged with bedtools merge (bedtools merge -d 100 -i cat\_PEAKS\_s\_2.bed) with different annotation with annotatePeaks.pl from homer <sup>71</sup>. Bam files were converted to bigwig files by the help of bamCoverage from deptools (-bs 20 --smoothLength 40 -p max --normalizeUsing RPKM -e 150) <sup>72</sup>. The cis-regulatory element annotation system (CEAS) <sup>73</sup> was used to determine the distribution of the peaks from *miR-9*

ChIRP-seq in different genomic areas (intron, exon, 5' UTR, 3' UTR, promoter, intergenic, and downstream).

### **Motif analysis of ChIRP and G4-ChIP-seq data**

MEME Suite (Motif-based sequence analysis tools)<sup>74</sup> was used for de-novo DNA motif search analysis. The *miR-9* ChIRP-seq data file containing peaks annotated near the promoter of the genes was used for the motif search. The G4 ChIP-seq<sup>24</sup> file containing peaks annotated near the promoter of the genes was used for the motif search was used for the motif search analysis. The settings in the de-novo DNA motifs search were: (1) a normal enrichment mode to search the motif, (2) length of the motifs allow between 10 and 25bp widths, (3) motif site distribution of any time of repetition on the fasta file and (4) three maximum motifs to report.

### **Enhancer areas enriched with *miR-9***

To determine the potential super enhancer marked or not marked by *miR-9*, we crossed the *miR-9* ChIRP-seq peaks list with the results from the peak list from H3K27ac peaks from (GSM2417093) determined using the following parameters: findPeaks H3K27ac\_maketaglib -i Inp\_makeTaglib -region -size 500 -minDist 20000 -fdr 0.05 -F 2 -style super -typical output\_enh -superSlope -1000 > OUPUT}\_SE.txt.

Those significant SE with more than 100 normalized reads were further crossed with the *miR-9* ChIRP-seq peak by the help of bedtools intersect (setings -wa). Those H3K27ac peaks not overlapping the SE list and not overlapping the *miR-9* peaks were considered TE. All ChIP-seq and *miR-9* ChIRP-seq were quantified from the center of both Peak list by the help of annotatePeaks from homer with the settings: annotatePeaks.pl list.bed mm10 -size 4000 -norm 1000000 -hist 10 -d maketagLib\_Samples > output\_quanty.txt. The results of this command

was used as input in a custom R-script to produce the aggregate plots  
([https://github.com/jcorder316/001MIR9\\_G4\\_3D\\_STRUC](https://github.com/jcorder316/001MIR9_G4_3D_STRUC)).

### Chromatin immunoprecipitation (ChIP)

ChIP analysis was performed as described earlier<sup>9, 75</sup> with minor adaptations. Briefly, cells were cross-linked with 1% methanol-free formaldehyde (ThermoFisher Scientific) lysed, and sonicated with Diagenode Bioruptor to an average DNA length of 300–600 bp. After centrifugation, the soluble chromatin was immunoprecipitated with 3 µg of antibodies specific for anti-H3K4me3 (Abcam, # ab8580), Anti-DNA G-quadruplex structures Antibody, clone BG4 (Millipore, # MABE917), and anti-IgG (Santa Cruz, #sc-2027). Reverse crosslinked immunoprecipitated chromatin was purified using the QIAquick PCR purification kit (Qiagen) and subjected to ChIP-quantitative PCR. The primer pairs used for gene promoter and super-enhancer regions are described in the following table:

Gene	Primer sequence (5'→3')	
<i>Zdhhc5</i>	Forward	CCAATCACCGTTAAGCCTTT
	Reverse	TTTACCCCACTGACGCTTTC
<i>Ncl</i>	Forward	CACACCAGGAAGTCACCTCTC
	Reverse	TTCTTGGCTATGATGCGAGTC
<i>Lzts2</i>	Forward	CGAGCGCTACCGAGCACACAC
	Reverse	GCGCAAGTAACCCACGTCACGC
SE <i>miR-9 Z</i>	Forward	ACCAGGAAAGTTTGCGAGGG
	Reverse	CTCGAAGGAACAGACGGAGG
SE <i>miR-9 N</i>	Forward	AGTCCCAGTCGCCTTGTTTC
	Reverse	AATTCACCTGAGACTCCCGC



SE <i>miR-9 L</i>	Forward	TTACAGCTTTCCGACCAGGC
	Reverse	CCTCTCAAGCCTCTGTCCAC

## Chromatin conformation analysis by *in situ* Hi-C library preparation followed by chromatin immunoprecipitation (HiChIP)

HiChIP experiments were performed as previously described <sup>76</sup> using antibodies specific for H3K4me3 (Abcam, # ab8580) with the following optimizations: 5-10 million cells were crosslinked with 1% formaldehyde for 10 min at room temperature; prior to restriction digestion, sodium dodecyl sulfate treatment at 62°C for 10 min; restriction digestion with *MboI* (New England Biolabs France, R0147M) for 2 h at 37°C; prior to fill-in reaction, heat inactivation of *MboI* at 62°C for 10 min followed by 2 washing steps of pelleted nuclei with 1x fill-in reaction buffer; after fill-in reaction, ligation at 4°C for 16 h.

## ChIP sequencing, HiChIP sequencing and data analysis

HiChIP-seq paired-end reads were aligned to the mm10 genome, duplicate reads were removed, reads were assigned to *MboI* restriction fragments, filtered into valid interactions, and the interaction matrices were generated using the HiC-Pro pipeline default settings <sup>77</sup>. The contig file of the HiC-Pro was set to allow validPairs at any distance from each other. HiC-Pro valid interaction reads were then used to detect significant interactions using: (1) Maketag libraries were made as followed (makeTagDirectory output\_PCA\_ucsc smaple\_R1\_mm10.bwt2merged.bam,sample\_R2\_mm10.bwt2merged.bam -tbp 1 -genome mm10 -checkGC -restrictionSite GATC. (2) We used (runHiCpca.pl sample\_mer25\_50 sample\_PCA\_ucsc -res 25000 -superRes 50000 -genome mm10 -cpu 16) to find the Principal component analysis (PCA) of the data. (3) We used analyzeHiC as followed (analyzeHiC sample\_PCA\_ucsc -res 1000000 -interactions sample\_significantInteractions.txt -nomatrix ).

Only Interaction looping between the H3K4me3 peaks and the TSS (+/-2kb) from *miR-9* candidate genes were considered. The settings were (annotateInteractions.pl Sample\_significan mm10 Sample\_output\_filterk4m9 -filter H3k4me3\_HICHIP-seq\_peaks -filter2 TSS2kb\_fMir9\_genes -cpu 16 -washu -pvalue 0.01). (4) The mapped merge bam files ouput from HiC-pro was processed as ChIP-seq to perform Peak calling, using (peak\_call -i bam\_file -o ouput\_peak -r MboI\_mm10.txt -f 0.01 -a mm10\_chr\_size -w 8) from ([https://github.com/ChenfuShi/HiChIP\\_peaks](https://github.com/ChenfuShi/HiChIP_peaks)).

To calculate the broadness of H3K4me3 peaks, we first performed summary statistics of the peak size from the merged peak list. If the size of a peak was equal to or higher than the top 75 % of peaks quantile 3 (Q3) was consider wide ( $\geq 2.7$ kb). If the peak was between Q3 and Q2 was consider medium size peak ( $\geq 2$  kb and  $< 2.7$  kb). The narrow Hek4me3 peaks were in the bottom 25% peaks, Q1 or less ( $< 2$ kb).

### **Meta-analysis of ChIP-seq, PRO-seq and GRO-seq**

All published data from ChIP-seq, PRO-seq, GRO-seq and ATAC-seq used in this manuscript were listed in the Supplementary Table 1. All these NGS data were downloaded and processed according to the description in their respective publications. Briefly, PRO-seq, GRO-seq and ChIP-seq were processed according to (<https://doi.org/10.1038/s41588-018-0139-3>). ATAC-seq, G4 ChIP-seq and G4access were processed using Bowtie2 (default settings) and all mapped bam files were converted to bigwig files using the same parameters as for the ChIRP-seq mentioned above.

The enrichment of the factors and nascent RNAs (PRO-seq and GRO-seq) on the TSS of the Down-regulated genes after *miR-9*-LOF were performed by the help of the library from R profileplyr (<https://www.bioconductor.org/packages/devel/bioc/vignettes/profileplyr/inst/doc/profileplyr>).

html) adapting to a custom R-script  
([https://github.com/jcorder316/001MIR9\\_G4\\_3D\\_STRUC](https://github.com/jcorder316/001MIR9_G4_3D_STRUC)).

### MicroRNA pulldown

For microRNA pulldown (miR-pd), MLE-12 cells were transfected with Biotinylated *mmu-miR-9-5p*-RNA (Exiqon) or Biotinylated *Mirctrl* (Exiqon), at a final concentration of 20 nM. After 48 hours, cells were fixed with 1% glutaraldehyde (Sigma-Aldrich) to preserve RNA-Chromatin interactions. After washing three times with PBS, cell pellets were resuspended in 2 ml hypotonic cell lysis buffer (10 mM Tris-HCl (pH 7.4), 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 25 mM NaF, 0.5 mM Na<sub>3</sub>VO<sub>4</sub>, 40 µg/ml phenylmethylsulfonyl fluoride, protease inhibitor (Calbiochem) and RNase inhibitor (Promega) in DEPC treated water) on ice for 10 min and then spun down at 700 g for 10 min at 4°C. Nuclear pellets were resuspended in 300 µL of nuclear lysis buffer (50 mM Tris-HCl (pH 7.4), 170 mM NaCl, 20% glycerol, 15 mM EDTA, 0.1% (v/v) Triton X-100, 0.2 mM DTT, 20 mM NaF, 20 mM Na<sub>3</sub>VO<sub>4</sub>, 40 µg/ml phenylmethylsulfonyl fluoride, protease inhibitor, RNase inhibitor in DEPC treated water). Nuclear fraction was sonicated (Bioruptor NextGen, Diagenode) at high amplitude for 10 cycles of 30 sec (on/off) pulses. M280 streptavidin magnetic beads (ThermoFisher Scientific) were blocked for 1 h at 4°C in blocking buffer (10 mM Tris-HCl pH 6.5, 1 mM EDTA, and 1 mg/ml BSA) and washed twice with 1 ml washing buffer (10 mM Tris-HCl (pH 7.0), 1 mM EDTA, 0.5 M NaCl, 0.1% (v/v) Triton X-100, RNAase inhibitor and protease inhibitor). Beads were resuspended in 0.5 ml washing buffer. The nuclear extract was then added to the beads and incubated for 1 h at 4°C with slow rotation. The beads were then washed five times with 1 ml washing buffer. RNA bound to the beads (pulled down RNA) and from 10% of the extract (input RNA) was isolated using Trizol reagent LS (Invitrogen) after DNase 1 (NEB) treatment.

For miRNA-protein interaction analysis, nuclear protein extract after DNase1 treatment incubated with 500 pMol Biotinylated *miR-9* or Biotinylated *Mirctrl* for 1 h with shaking at 4°C. M280 streptavidin magnetic beads (ThermoFisher Scientific) were used for pulldown. Protein bound to the beads (miR-Pd), was incubated with 30 µl 2x SDS samples loading buffer, boiled at 95°C for 5min, spun down and loaded on SDS-PAGE for western blot analysis and/or processed for proteomic analysis by mass spectrometry.

### **Chromatin RNA immunoprecipitation**

Chromatin RNA immunoprecipitation (Ch-RIP) analysis was performed as described<sup>8</sup> with minor adaptations. Briefly, cells were cross-linked by 1% formaldehyde for 10 min, lysed, and sonicated with Diagenode Bioruptor to disrupt and fragment genomic DNA. After centrifugation, the soluble chromatin was immunoprecipitated using antibodies. Precipitated chromatin complexes were removed from the beads by incubating with 50 µl of 1% SDS with 0.1 M NaHCO<sub>3</sub> for 15 min, vortexing every 5 min and followed by treatment with DNase1. The collected material was immunoprecipitated using IgG or an HA-specific antibody (Santa Cruz, CatNo. 805). Reverse cross-linked immunoprecipitated materials were used for RNA isolation by Trizol (Invitrogen). Isolated RNA was subjected to cDNA synthesis and further for TaqMan assay specific for mature *miR-9* (Applied Biosystems, Cat. No. 4427975).

### **Western blot**

Nuclear protein extracts from MLg cells were prepared by cell lysis and nuclei isolation. Briefly, MLg cells were spooled down and washed with PBS. cell pellets were resuspended in 2 ml hypotonic cell lysis buffer (10 mM Tris-HCl (pH 7.4), 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 1 mM DTT, 25 mM NaF, 0.5 mM Na<sub>3</sub>VO<sub>4</sub>, 40 µg/ml phenylmethylsulfonyl fluoride, protease inhibitor (Calbiochem) and RNase inhibitor (Promega) in DEPC treated water) on ice for 10

min and then spun down at 700 g for 10 min at 4°C. Nuclear pellets were resuspended in 300 µL of nuclear lysis buffer (50 mM Tris-HCl (pH 7.4), 170 mM NaCl, 20% glycerol, 15 mM EDTA, 0.1% (v/v) Triton X-100, 0.2 mM DTT, 20 mM NaF, 20 mM Na<sub>3</sub>VO<sub>4</sub>, 40 µg/ml phenylmethylsulfonyl fluoride, protease inhibitor, RNase inhibitor in DEPC treated water). Nuclear fraction was sonicated (Bioruptor NextGen, Diagenode) at high amplitude for 10 cycles of 30 sec (on/off) pulses. Detergent-insoluble material was precipitated by centrifugation at 14,000 rpm for 30 min at 4 °C. The supernatant was transferred to a fresh tube and stored at -20 °C. Protein concentration was estimated using Bradford assay, using serum albumin as standard. 5 µl of serial dilutions of standard protein and samples were mixed with 250 µl of Bradford reagent (500-0205, BIO-RAD Quick Start™).

Western blotting was performed using standard methods and antibodies specific for SMARCA5 (Invitrogen, MA5-35378), LMNB1 (Santa Cruz, sc-374015), CHD4 (Abcam, 70469), RAD21 (Abcam, ab992), GAPDH (Sigma, MFCD01322099). Immunoreactive proteins were visualized with the corresponding HRP-conjugated secondary antibodies (Santa Cruz) using the Super Signal West Femto detection solutions (ThermoFisher Scientific). Signals were detected and analyzed with Luminescent Image Analyzer (Las 4000, Fujifilm).

## **RNA sequencing and data analysis**

RNA sequencing data for this paper were generated as previously described<sup>8,53</sup>. Briefly, total RNA from MLg cells that were transfected with control or *miR-9*-specific antagomiR was isolated using the Trizol method. RNA was treated with DNase (DNase-Free DNase Set, Qiagen) and repurified using the RNeasy micro plus Kit (Qiagen). Total RNA and library integrity were verified on LabChip Gx Touch 24 (Perkin Elmer). Sequencing was performed on the NextSeq500 instrument (Illumina) using v2 chemistry with 1x75bp single end setup. Raw reads were visualized by FastQC to determine the quality of the sequencing. Trimming

was performed using trimmomatic with the following parameters: LEADING:3 TRAILING:3  
SLIDINGWINDOW:4:15 MINLEN:15 CROP:60 HEADCROP:15. High quality reads were  
mapped to mm10 uscs genome using with bowtie2 , settings (-D 15 -R 3 -N 0 -L 20 -i S,1,0.56  
-k 1). Tag libraries were obtained with MakeTaglibrary from HOMER (default setting).  
Samples were quantified by using analyzeRepeats.pl with the parameters: mm10 –count exons  
–strand both –noad. Gene expression was quantified in reads per kilo base million (RPKM) by  
the help of rpkm.default from EdgeR. Down-regulated genes after *miR-9*-LOF were those  
genes with a Log2fc (LOF/Ctr) ≤ -0.58 and Up-regulated those genes with a log2FC ≥ 0.58.

### Statistical Analysis

Depending on the data, different tests were performed to determine the statistical significance  
of the results. The values of the statistical tests used in the different experiments can be found  
in Source Data file 01. Further details of statistical analysis in different experiments are  
included in the figures and figure legends. Briefly, one set of ChIRP, RNA, ChIP and HiChIP  
samples were analyzed by deep sequencing. For the rest of the experiments presented here,  
samples were analyzed at least in triplicates and experiments were performed three times.  
Statistical analysis was performed using Excel Solver and Prism9. Data in bar plots are  
represented as mean ± standard error (mean ± s.e.m.). Two-tailed t-tests were used to determine  
the levels of difference between the groups and *P*-values for significance. *P*-values after two-  
tailed t-test, \**P* ≤ 0.05; \*\**P* < 0.01, and \*\*\**P* < 0.001.

### Data availability

The data that support this study are available from the corresponding author upon reasonable  
request. The source data are provided with this paper as a Source Data file 01. In addition,  
sequencing data of ChIRP, HiChIP and RNA have been deposited in NCBI's Gene Expression

Omnibus<sup>78</sup> and is accessible through SRA Sequence Read Archives NCBI with accession number GSE244952. In addition, we retrieved and used publicly available datasets to aid analysis of our data. Supplementary Table 1 contains all data sets used in this study. The model in Figure 8b was Created with BioRender.com.

## Acknowledgments

We thank Roswitha Bender for technical support; Kerstin Richter and Alessandro Ianni for antibodies; Dominique Dumas for support with confocal microscopy; Amine Armich, Sylvie Fournel-Gigleux, Mohamed Ouzzine, Jean-Baptiste Vincourt, Sandrine Gulberti, Catherine Bui, Lydia Barré and Nick Ramalanjaona for comments.

**Funding:** Guillermo Barreto was funded by the “Centre National de la Recherche Scientifique” (CNRS, France), “Délégation Centre-Est” (CNRS-DR6) and the “Lorraine Université” (LU, France) through the initiative “Lorraine Université d’Excellence” (LUE) and the dispositive “Future Leader”, the Max-Planck-Society (MPG, Munich, Germany) and the “Deutsche Forschungsgemeinschaft” (DFG, Bonn, Germany) (BA 4036/4-1). Gergana Dobрева and Julio Cordero are supported by the CRC 1366 (Projects A03, A06), the CRC 873 (Project A16), the CRC1550 (Project A03) funded by the DFG, the DZHK (81Z0500202), funded by BMBF and the Baden-Württemberg foundation special program “Angioformatics single cell platform”. Guruprasadh Swaminathan receive a doctoral fellowship through the initiative “Lorraine Université d’Excellence” (LUE). Diana G. Rogel-Ayala receive a doctoral fellowship from the DAAD (57552340). Karla Rubio was funded by the “Consejo de Ciencia y Tecnología del Estado de Puebla” (CONCYTEP, Puebla, Mexico) through the initiative International Laboratory EPIGEN. Work in the lab of Thomas Braun is supported by the Deutsche Forschungsgemeinschaft, Excellence Cluster Cardio-Pulmonary Institute (CPI), Transregional Collaborative Research Center TRR81, TP A02, SFB1213 TP B02, TRR 267 TP A05 and the German Center for Cardiovascular Research.

## Authors Contributions



JC, GS, DGRA, KR, AE, SG and GB designed and performed the experiments; TB and GD were involved in study design; GB and JC designed the study; JC, GB, GS, DGRA, KR and SG analyzed the data; GB, JC, GS and DGRA wrote the manuscript. All authors discussed the results and commented on the manuscript.

## **Conflict of Interest**

All authors declare that they have no competing financial interests in relation to the work described in the manuscript.

## **SUPPLEMENTARY INFORMATION**

Supplementary Figures 1 to 7

Supplementary Table 1

Source Data file 01

## REFERENCES

- 1 Jerkovic I, Cavalli G. Understanding 3D genome organization by multidisciplinary methods. *Nature reviews Molecular cell biology* 2021; **22**:511-528.
- 2 Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. *Nature reviews Molecular cell biology* 2019; **20**:535-550.
- 3 Misteli T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* 2020; **183**:28-45.
- 4 Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**:2926-2931.
- 5 Gohlke J, Scholz CJ, Kneitz S *et al.* DNA methylation mediated control of gene expression is critical for development of crown gall tumors. *PLoS genetics* 2013; **9**:e1003267.
- 6 Georgakopoulos-Soares I, Morganella S, Jain N, Hemberg M, Nik-Zainal S. Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome research* 2018; **28**:1264-1271.
- 7 Singh I, Ozturk N, Cordero J *et al.* High mobility group protein-mediated transcription requires DNA damage marker gamma-H2AX. *Cell research* 2015; **25**:837-850.
- 8 Singh I, Contreras A, Cordero J *et al.* MiCEE is a ncRNA-protein complex that mediates epigenetic silencing and nucleolar organization. *Nature genetics* 2018; **50**:990-1001.
- 9 Dobersch S, Rubio K, Singh I *et al.* Positioning of nucleosomes containing gamma-H2AX precedes active DNA demethylation and transcription initiation. *Nature communications* 2021; **12**:1072.
- 10 Papantonis A, Cook PR. Genome architecture and the role of transcription. *Current opinion in cell biology* 2010; **22**:271-276.
- 11 Jha RK, Levens D, Kouzine F. Mechanical determinants of chromatin topology and gene expression. *Nucleus* 2022; **13**:94-115.
- 12 Buecker C, Wysocka J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet* 2012; **28**:276-284.
- 13 Schaffner W. Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. *Biol Chem* 2015; **396**:311-327.
- 14 Zaugg JB, Sahlen P, Andersson R *et al.* Current challenges in understanding the role of enhancers in disease. *Nat Struct Mol Biol* 2022; **29**:1148-1158.
- 15 Whyte WA, Orlando DA, Hnisz D *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013; **153**:307-319.
- 16 Hnisz D, Abraham BJ, Lee TI *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 2013; **155**:934-947.

- 17 Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic acids research* 2016; **44**:D164-171.
- 18 Hsieh TS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nature genetics* 2022; **54**:1919-1932.
- 19 Kyrchanova O, Georgiev P. Mechanisms of Enhancer-Promoter Interactions in Higher Eukaryotes. *International journal of molecular sciences* 2021; **22**.
- 20 Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. *Nature reviews Molecular cell biology* 2020; **21**:459-474.
- 21 Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic acids research* 2015; **43**:8627-8637.
- 22 Henderson E, Hardin CC, Walk SK, Tinoco I, Jr., Blackburn EH. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* 1987; **51**:899-908.
- 23 Hansel-Hertsch R, Beraldi D, Lensing SV *et al.* G-quadruplex structures mark human regulatory chromatin. *Nature genetics* 2016; **48**:1267-1272.
- 24 Zheng KW, Zhang JY, He YD *et al.* Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic acids research* 2020; **48**:11706-11720.
- 25 Du Z, Zhao Y, Li N. Genome-wide colonization of gene regulatory elements by G4 DNA motifs. *Nucleic acids research* 2009; **37**:6784-6798.
- 26 Renciuik D, Rynes J, Kejnovska I *et al.* G-quadruplex formation in the Oct4 promoter positively regulates Oct4 expression. *Biochim Biophys Acta Gene Regul Mech* 2017; **1860**:175-183.
- 27 Lago S, Nadai M, Cernilogar FM *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nature communications* 2021; **12**:3885.
- 28 Esnault C, Magat T, Zine El Aabidine A *et al.* G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nature genetics* 2023; **55**:1359-1369.
- 29 Esnault C, Garcia-Oliver E, Aabidine AZE *et al.* G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. *bioRxiv* 2023:2023.2002.2024.529838.
- 30 Agarwal T, Roy S, Kumar S, Chakraborty TK, Maiti S. In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry* 2014; **53**:3711-3718.
- 31 Holder IT, Hartig JS. A matter of location: influence of G-quadruplexes on Escherichia coli gene expression. *Chemistry & biology* 2014; **21**:1511-1521.

804 32 Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*  
805 2012; **489**:57-74.

806 33 Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function.  
807 *Nat Rev Genet* 2015; **17**:47-62.

808 34 Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L *et al.* LncRNA Khps1 Regulates  
809 Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin  
810 Structure. *Molecular cell* 2015; **60**:626-636.

811 35 Rubio K, Dobersch S, Barreto G. Functional interactions between scaffold proteins,  
812 noncoding RNAs, and genome loci induce liquid-liquid phase separation as organizing  
813 principle for 3-dimensional nuclear architecture: implications in cancer. *FASEB journal :*  
814 *official publication of the Federation of American Societies for Experimental Biology*  
815 2019:fj201802715R.

816 36 Rubio K, Castillo-Negrete R, Barreto G. Non-coding RNAs and nuclear architecture during  
817 epithelial-mesenchymal transition in lung cancer and idiopathic pulmonary fibrosis. *Cell*  
818 *Signal* 2020; **70**:109593.

819 37 Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional  
820 regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 2008; **9**:102-114.

821 38 Park CW, Zeng Y, Zhang X, Subramanian S, Steer CJ. Mature microRNAs identified in  
822 highly purified nuclei from HCT116 colon cancer cells. *RNA Biol* 2010; **7**:606-614.

823 39 Rasko JE, Wong JJ. Nuclear microRNAs in normal hemopoiesis and cancer. *J Hematol*  
824 *Oncol* 2017; **10**:8.

825 40 Huang V. Endogenous miRNAa: miRNA-Mediated Gene Upregulation. *Adv Exp Med Biol*  
826 2017; **983**:65-79.

827 41 Leucci E, Patella F, Waage J *et al.* microRNA-9 targets the long non-coding RNA MALAT1  
828 for degradation in the nucleus. *Scientific reports* 2013; **3**:2535.

829 42 Hwang HW, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear  
830 import. *Science* 2007; **315**:97-100.

831 43 Yuva-Aydemir Y, Simkin A, Gascon E, Gao FB. MicroRNA-9: functional evolution of a  
832 conserved small regulatory RNA. *RNA Biol* 2011; **8**:557-564.

833 44 Coolen M, Katz S, Bally-Cuif L. miR-9: a versatile regulator of neurogenesis. *Front Cell*  
834 *Neurosci* 2013; **7**:220.

835 45 Santovito D, Egea V, Bidzhekov K *et al.* Noncanonical inhibition of caspase-3 by a nuclear  
836 microRNA confers endothelial protection by autophagy in atherosclerosis. *Science*  
837 *translational medicine* 2020; **12**.

838 46 Chen K, Chen Z, Wu D *et al.* Broad H3K4me3 is associated with increased transcription  
839 elongation and enhancer activity at tumor-suppressor genes. *Nature genetics* 2015; **47**:1149-  
840 1157.

- 47 Mahat DB, Salamanca HH, Duarte FM, Danko CG, Lis JT. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular cell* 2016; **62**:63-78.
- 48 Busslinger GA, Stocsits RR, van der Lelij P *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* 2017; **544**:503-507.
- 49 Buganim Y, Markoulaki S, van Wietmarschen N *et al.* The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell stem cell* 2014; **15**:295-309.
- 50 Chronis C, Fiziev P, Papp B *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* 2017; **168**:442-459 e420.
- 51 Jin W, Jiang G, Yang Y *et al.* Animal-eRNAdb: a comprehensive animal enhancer RNA database. *Nucleic acids research* 2022; **50**:D46-D53.
- 52 Mootha VK, Lindgren CM, Eriksson KF *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* 2003; **34**:267-273.
- 53 Rubio K, Singh I, Dobersch S *et al.* Inactivation of nuclear histone deacetylases by EP300 disrupts the MiCEE complex in idiopathic pulmonary fibrosis. *Nature communications* 2019; **10**:2229.
- 54 De Magis A, Manzo SG, Russo M *et al.* DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 2019; **116**:816-825.
- 55 Lyu J, Shao R, Kwong Yung PY, Elsasser SJ. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic acids research* 2022; **50**:e13.
- 56 Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & development* 2014; **28**:1384-1396.
- 57 Sato K, Hendrikx AGM, Knipscheer P. RNA transcripts suppress G-quadruplex structures through G-loop formation. *bioRxiv* 2023:2023.2003.2009.531892.
- 58 Hou Y, Guo Y, Dong S, Yang T. Novel Roles of G-quadruplexes on Enhancers in human chromatin. *bioRxiv* 2021:2021.2007.2012.451993.
- 59 Pokholok DK, Harbison CT, Levine S *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 2005; **122**:517-527.
- 60 Li X, Wu F, Gunther S *et al.* Inhibition of fatty acid oxidation enables heart regeneration in adult mice. *Nature* 2023; **622**:619-626.
- 61 Sze CC, Ozark PA, Cao K *et al.* Coordinated regulation of cellular identity-associated H3K4me3 breadth by the COMPASS family. *Sci Adv* 2020; **6**:eaaz4764.

877 62 Benayoun BA, Pollina EA, Ucar D *et al.* H3K4me3 breadth is linked to cell identity and  
878 transcriptional consistency. *Cell* 2014; **158**:673-688.

879 63 Syed V. TGF-beta Signaling in Cancer. *J Cell Biochem* 2016; **117**:1279-1287.

880 64 Dobersch S, Rubio K, Barreto G. Pioneer Factors and Architectural Proteins Mediating  
881 Embryonic Expression Signatures in Cancer. *Trends Mol Med* 2019.

882 65 Rubio K, Romero-Olmedo AJ, Sarvari P *et al.* Non-canonical integrin signaling activates  
883 EGFR and RAS-MAPK-ERK signaling in small cell lung cancer. *Theranostics* 2023; **13**:2384-  
884 2407.

885 66 Akeson AL, Wetzel B, Thompson FY *et al.* Embryonic vasculogenesis by endothelial  
886 precursor cells derived from lung mesenchyme. *Developmental dynamics : an official*  
887 *publication of the American Association of Anatomists* 2000; **217**:11-23.

888 67 Mehta A, Dobersch S, Dammann RH *et al.* Validation of Tuba1a as Appropriate Internal  
889 Control for Normalization of Gene Expression Analysis during Mouse Lung Development.  
890 *International journal of molecular sciences* 2015; **16**:4492-4511.

891 68 Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA  
892 occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* 2011; **44**:667-678.

893 69 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*  
894 2012; **9**:357-359.

895 70 Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome*  
896 *biology* 2008; **9**:R137.

897 71 Heinz S, Benner C, Spann N *et al.* Simple combinations of lineage-determining transcription  
898 factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular*  
899 *cell* 2010; **38**:576-589.

900 72 Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for  
901 exploring deep-sequencing data. *Nucleic acids research* 2014; **42**:W187-191.

902 73 Ji X, Li W, Song J, Wei L, Liu XS. CEAS: cis-regulatory element annotation system.  
903 *Nucleic acids research* 2006; **34**:W551-554.

904 74 Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets.  
905 *Bioinformatics* 2011; **27**:1696-1697.

906 75 Hansel-Hertsch R, Spiegel J, Marsico G, Tannahill D, Balasubramanian S. Genome-wide  
907 mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and  
908 high-throughput sequencing. *Nature protocols* 2018; **13**:551-564.

909 76 Mumbach MR, Satpathy AT, Boyle EA *et al.* Enhancer connectome in primary human cells  
910 identifies target genes of disease-associated DNA elements. *Nature genetics* 2017; **49**:1602-  
911 1612.

912 77 Servant N, Varoquaux N, Lajoie BR *et al.* HiC-Pro: an optimized and flexible pipeline for  
913 Hi-C data processing. *Genome biology* 2015; **16**:259.

914 78 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and  
915 hybridization array data repository. *Nucleic acids research* 2002; **30**:207-210.  
916



## FIGURE LEGENDS

### **Figure 1: Mature *miR-9* is detected in the cell nucleus enriched at promoters and introns.**

(a) Phylogenetic tree (left) and a heat map (right) that were generated using the sequences of the indicated mature mouse miRNAs. Numbers indicate distance score. Bottom, sequence alignment of the indicated mature mouse miRNAs highlighting the published *miR-126* nuclear shuttling motif (magenta) and the expanded miRNA nuclear shuttling motif (green square) using the IUPAC nucleotide code. Pink letters are conserved among all sequences. Golden letters are conserved in at least 2 sequences. Blue letters indicate conserved type of base (either purine or pyrimidine). (b) Mature *miR-9*-specific TaqMan assay after cellular fractionation (Supplementary Figure 1b) of indicated cell lines. Data are shown as means  $\pm$  s.e.m. ( $n = 3$  independent experiments). (c) Fluorescence microscopy of MLg cells after RNA FISH confirmed nuclear localization of endogenous *miR-9*. Cells were transiently transfected with control (top) or *miR-9*-specific antagomiR probes (bottom) to induce a *miR-9* loss-of-function (LOF). Representative images from three independent experiments. Squares are shown at higher magnification. DAPI, nucleus. Scale bars, 10  $\mu$ m. (d) Genome-wide distribution of *miR-9* peaks relative to gene structures by ChIRP-seq in MLg cells. Square on right shows *miR-9* enrichment in different gene structure as Log2 ratios. TTS, transcription termination sites; 3'UTR, 3'untranslated regions. (e) Visualization of selected *miR-9* target genes using IGV genome browser showing *miR-9* enrichment in MLg cells. ChIRP-seq reads were normalized using reads per kilobase per million (RPKM) measure and are represented as log2 enrichment over their corresponding inputs. Images show the indicated gene loci with their genomic coordinates. Arrows, direction of the genes; blue boxes, exons; red line, regions selected for single gene analysis in f. (f) Analysis of selected *miR-9* target genes by ChIRP using chromatin from MLg cells and control (Ctrl) or *miR-9*-specific biotinylated antisense oligonucleotides for the precipitation of endogenous mature *miR-9* along with the chromatin bound to it. Bar plots



presenting data as means; error bars, s.e.m ( $n = 3$  biologically independent experiments); asterisks,  $P$ -values after two-tailed t-test,  $***P \leq 0.001$ ;  $**P \leq 0.01$ ;  $*P \leq 0.05$ . See also Supplementary Figure 1. Source data are provided as a Source Data file 01.

**Figure 2: *MiR-9* is required for H3K4me3 broad domains, high basal transcriptional activity, and G-quadruplex formation at promoters.** (a) Genome-wide distribution of H3K4me3 peaks relative to broad ( $\geq 2.7$  kb), medium ( $\geq 2$  kb and  $< 2.7$  kb) narrow ( $< 2$  kb) H3K4me3 domains in MLg cells that were transiently transfected with control (Ctrl) or *miR-9*-specific antagomiR probes to induce *miR-9* loss-of-function (LOF). Square on right shows H3K4me3 enrichment in different domains as Log2 ratios of MLg cells after *miR-9*-LOF versus Ctrl transfected cells. Asterisks,  $P$ -values after two-tailed t-test,  $**P \leq 0.01$ . See also Supplementary Figure 2. (b) RNA-seq using total RNA from MLg cells that were transfected as in a. Pie chart shows the distribution of transcripts with significantly altered expression levels ( $n = 3,320$ ) between both conditions tested in decreased transcripts ( $n = 2,439$ ) and increased transcripts ( $n = 881$ ) after *miR-9*-LOF.  $P$ -values after two-tailed t-test. (c) Box plots of RNA-seq-based expression analysis of transcripts with significantly decreased levels after *miR-9*-LOF (*miR-9* targets;  $n = 2,439$ ) and with non-significantly (ns) changed levels (non-targets;  $n = 881$ ) after *miR-9*-LOF. Values were normalized using reads per kilobase per million (RPKM) and represented as log2 RPKM. Box plots indicate median (middle line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers);  $P$ -values after two-tailed t-test. Source data are provided as Source Data file 01. (d) Heat map for *miR-9* enrichment at the TSS  $\pm 2$  kb of *miR-9* target genes as determined by RNA-seq in b and c. (e-g) Heat maps for enrichment of total Pol II and Pol II S5p (e), nascent RNA by precision nuclear run-on assay (PRO-seq) and global run-on sequencing (GRO-seq) (f), chromatin accessibility by ATAC-seq and H3K4me3 by ChIP-seq (g), at the TSS  $\pm 5$  kb of *miR-9* target genes as determined by RNA-

seq in b and c. **(h)** Motif analysis of the *miR-9* target genes showed significant enrichment of nucleotide motifs that are similar to the motifs found in loci that form G4 as determined by G4P ChIP-seq. **(i)** Heat maps for G4 enrichment at the TSS  $\pm$  5 kb of the *miR-9* target genes by G4P ChIP-seq (left) or G4access (right). See also Supplementary Figure 2. Source data are provided as a Source Data file 01.

**Figure 3: H3K4me3, nascent RNA and G4 are enriched at promoters of selected *miR-9* target genes.** **(a)** Visualization of selected *miR-9* target genes using IGV genome browser showing enrichment of *miR-9* by ChIRP-seq (green), H3K4me3 by ChIP-seq in Ctrl and *miR-9*-specific antagomiR transfected MLg cells (black), nascent RNA by GRO-seq (brown) and G4 by G4P ChIP-seq (blue). Reads were normalized using reads per kilobase per million (RPKM) after bamCoverage. Images show the indicated gene loci with their genomic coordinates. Arrows, direction of the genes; blue boxes, exons; red lines, regions selected for single gene analysis in b; green squares, regions with enrichment of *miR-9*, H3K4me3, nascent RNA and G4; dotted lines, regions shown at the bottom with high G content. **(b)** Analysis of the promoter of selected *miR-9* target genes by ChIP using chromatin from MLg cells transfected with control (Ctrl) or *miR-9*-specific antagomiR to induce *miR-9* loss-of-function (LOF). Bar plots presenting data as means; error bars, s.e.m ( $n = 3$  biologically independent experiments); asterisks,  $P$ -values after two-tailed t-test,  $**P \leq 0.01$ ;  $*P \leq 0.05$ . See also Supplementary Figure 3. Source data are provided as a Source Data file 01.

**Figure 4: Nuclear *miR-9* is enriched at super-enhancers.** **(a)** Alluvial plot showing the proportion of loci with *miR-9* enrichment by ChIRP-seq (green) also with enrichment of G4 by G4P ChIP-seq and the SE markers MED1 and H3K27Ac by ChIP-seq. No-*miR-9*, loci without

*miR-9* enrichment. **(b)** Venn diagram showing the number of common loci ( $n = 3,583$ ) with enrichment of *miR-9* by ChIRP-seq (green), H3K27ac (purple) and MED1 (turquoise) by ChIP-seq and nascent RNA (blue) by GRO-seq. **(c)** Venn diagram showing that 3,423 (95.5%) of the transcripts related to the common loci in **b** were found in a database as enhancer RNAs. **(d)** Genome-wide distribution of *miR-9* peaks by ChIRP-seq in MLg cells relative to super-enhancers (SE) and typical enhancers (TE). **(e)** Aggregate plots showing at SE (green lines) and TE (grey lines)  $\pm 2$  kb the enrichment of *miR-9* by ChIRP-seq and the indicated proteins by ChIP-seq, which have been previously related to SE in various murine cells. Reads were normalized using reads per kilobase per million (RPKM) measure. **(f)** Aggregate plots showing at SE (magenta and blue lines) and TE (grey lines)  $\pm 2$  kb the enrichment of G4 by ChIP-seq in mouse MLg and human A549 cells. Reads were normalized using RPKM. See also Supplementary Figure 4. Source data are provided as a Source Data file 01. Supplementary Table 1 contains all data sets used in this study.

**Figure 5: Nuclear *miR-9* is enriched at super-enhancers and is required for G-quadruplexes.** **(a)** Visualization of selected SE with *miR-9* enrichment using IGV genome browser showing enrichment *miR-9* by ChIRP-seq (green), G4 by G4P ChIP-seq (blue), MED1 (turquoise), KLF4 (magenta) and H3K27Ac (purple) by ChIP-seq. Reads were normalized using reads per kilobase per million (RPKM). Images show the indicated gene loci with their genomic coordinates. Orange squares, regions with enrichment of *miR-9*, G4 and SE markers; red lines, regions selected for single gene analysis in **b**; dotted lines, regions shown at the bottom with high G content. **(b)** Analysis of selected SE with *miR-9* enrichment by ChIP using chromatin from MLg cells transfected with control (Ctrl) or *miR-9*-specific antagomiR to induce *miR-9* loss-of-function (LOF). Bar plots presenting data as means; error bars, s.e.m ( $n = 3$  biologically independent experiments); asterisks,  $P$ -values after two-tailed t-test,  $***P \leq$

0.001;  $**P \leq 0.01$ ;  $*P \leq 0.05$ . See also Supplementary Figure 5. Source data are provided as a Source Data file 01.

**Figure 6: Nuclear *miR-9* is required for H3K4me3 enrichment and chromatin loops at promoters of TGF $\beta$ 1-responsive genes.** (a) Venn diagram after cross analysis of *miR-9* ChIRP-seq and GRO-seq<sup>48</sup> showing loci with *miR-9* enrichment and nascent RNAs (n = 4,189), in which the selected *miR-9* target genes are included (square). (b) Gene set enrichment analysis (GSEA) of the three loci groups identified in a. Resp, response; prol, proliferation; trans, transport. *P* values after two-tailed Fisher's exact test and represented as Log10. (c) Visualization of selected *miR-9* target genes using IGV genome browser showing enrichment H3K4me3 by ChIP-seq in MLg cells that were transfected with control (Ctrl) or *miR-9*-specific antagomiR to induce a loss-of-function (LOF), and non-treated or treated with TGF $\beta$ 1, as indicated. Images show the indicated loci with their genomic coordinates. Arrows, transcription direction; green squares, promoter regions; dotted lines, regions selected for single gene analysis in Figure 3b. (d) Line charts showing the number of significant chromatin interactions at loci with *miR-9* and H3K4me3 enrichment in MLg cells treated as in c. Four clusters were generated using k-means algorithm. Data are represented as Log2 of the ratio relative to Ctrl-transfected, non-treated cells. See also Supplementary Figure 6 and 7. Source data are provided as a Source Data file 01.

**Figure 7: Promoter-super-enhancer looping of TGF $\beta$ 1-responsive genes requires *miR-9*.** (a) Visualization of promoters of selected *miR-9* target genes (green squares) and SE with *miR-9* enrichment (orange squares) using IGV genome browser showing enrichment *miR-9* by ChIRP-seq (green), G4 by G4P ChIP-seq (blue), MED1 (turquoise), KLF4 (magenta) and H3K27Ac (purple) by ChIP-seq. Reads were normalized using reads per kilobase per million

(RPKM) measure and are represented as log2 enrichment over their corresponding inputs. Bottom, chromatin loops by HiChIP-seq in MLg cells that were transfected with control (Ctrl) or *miR-9*-specific antagomir (*miR-9*-LOF, loss-of-function), and non-treated or treated with TGFB1, as indicated. Images show the indicated loci with their genomic coordinates. Arrows, transcription direction; red lines, regions selected for single gene analysis in b. **(b)** Analysis of the promoters and SE highlighted in a by ChIP using the indicated antibodies and chromatin of MLg cells treated as in a. All bar plots present data as means; error bars, s.e.m ( $n = 3$  biologically independent experiments); asterisks,  $P$ -values after two-tailed t-test,  $***P \leq 0.001$ ;  $**P \leq 0.01$ ;  $*P \leq 0.05$ ; ns, non-significant. See also Supplementary Figure 7. Source data are provided as a Source Data file 01.

**Figure 8: Promoter-super-enhancer looping of TGFB1-responsive genes requires *miR-9*.**

**(a)** Expression analysis of the selected *miR-9* target genes by qRT-PCR in MLg cells that were transfected with control (Ctrl) or *miR-9*-specific antagomir (*miR-9*-LOF, loss-of-function), and non-treated or treated with TGFB1, as indicated. All bar plots present data as means; error bars, s.e.m ( $n = 3$  biologically independent experiments); asterisks,  $P$ -values after two-tailed t-test,  $***P \leq 0.001$ ;  $**P \leq 0.01$ ;  $*P \leq 0.05$ ; ns, non-significant. See also Supplementary Figure 7. Source data are provided as a Source Data file 01. **(b)** Model summarizing the results presented in the manuscript. Left, G4 are formed in *miR-9* (red lines) -dependent manner at SE (orange box) and promoters (green box) of TGFB1-responsive genes (blue box, coding region). Middle, TGFB1 treatment increases euchromatin histone mark H3K4me3, G4 and chromatin loops bringing SE and promoter to close physical proximity, thereby enhancing transcription (arrow) of the corresponding gene. Right, *miR-9*-LOF antagonizes the effects induced by TGFB1. Created with BioRender.com





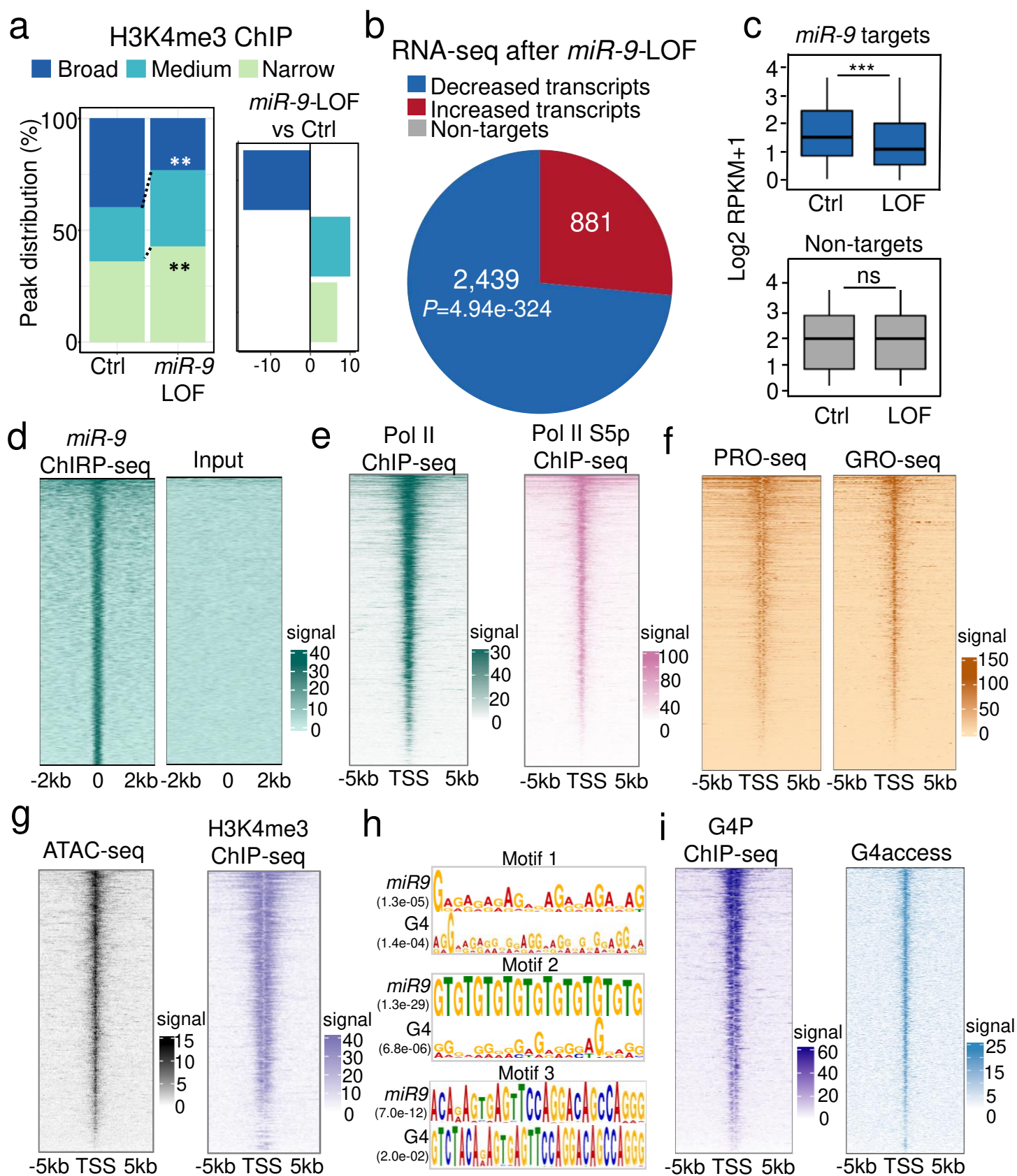


Figure 2

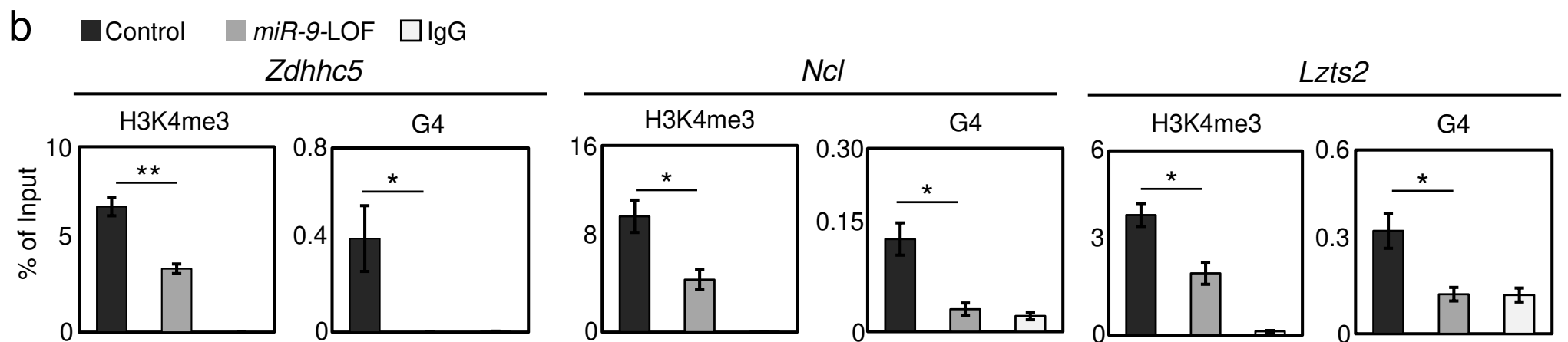
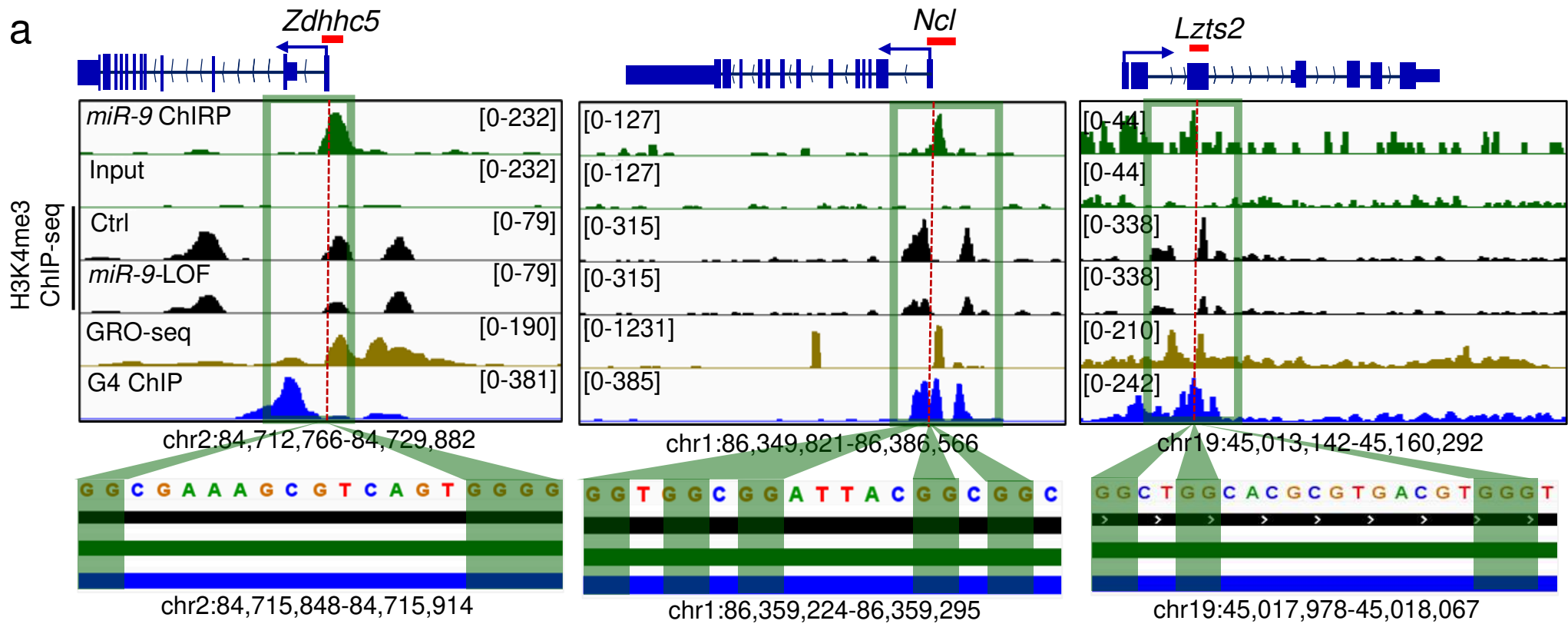


Figure 3



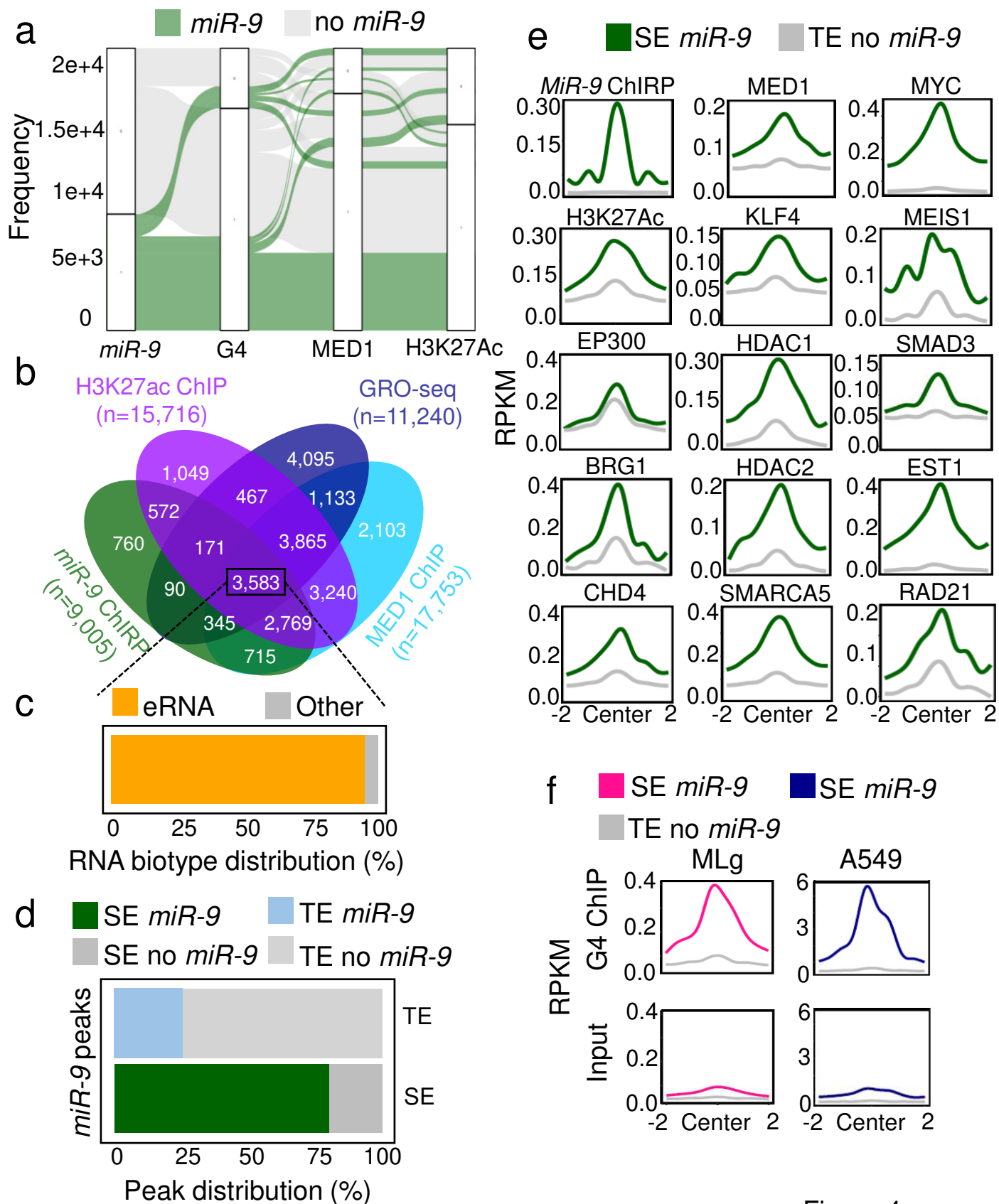


Figure 4

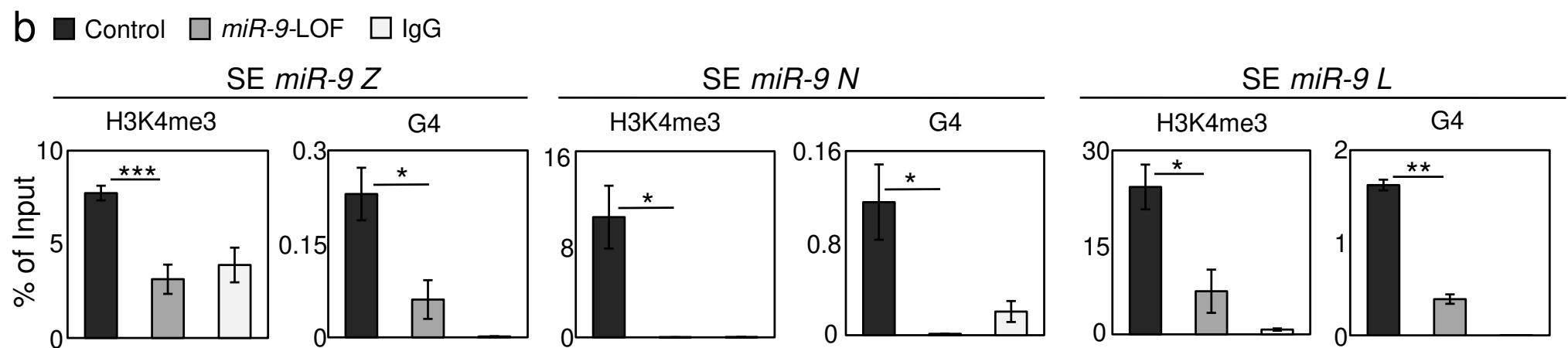
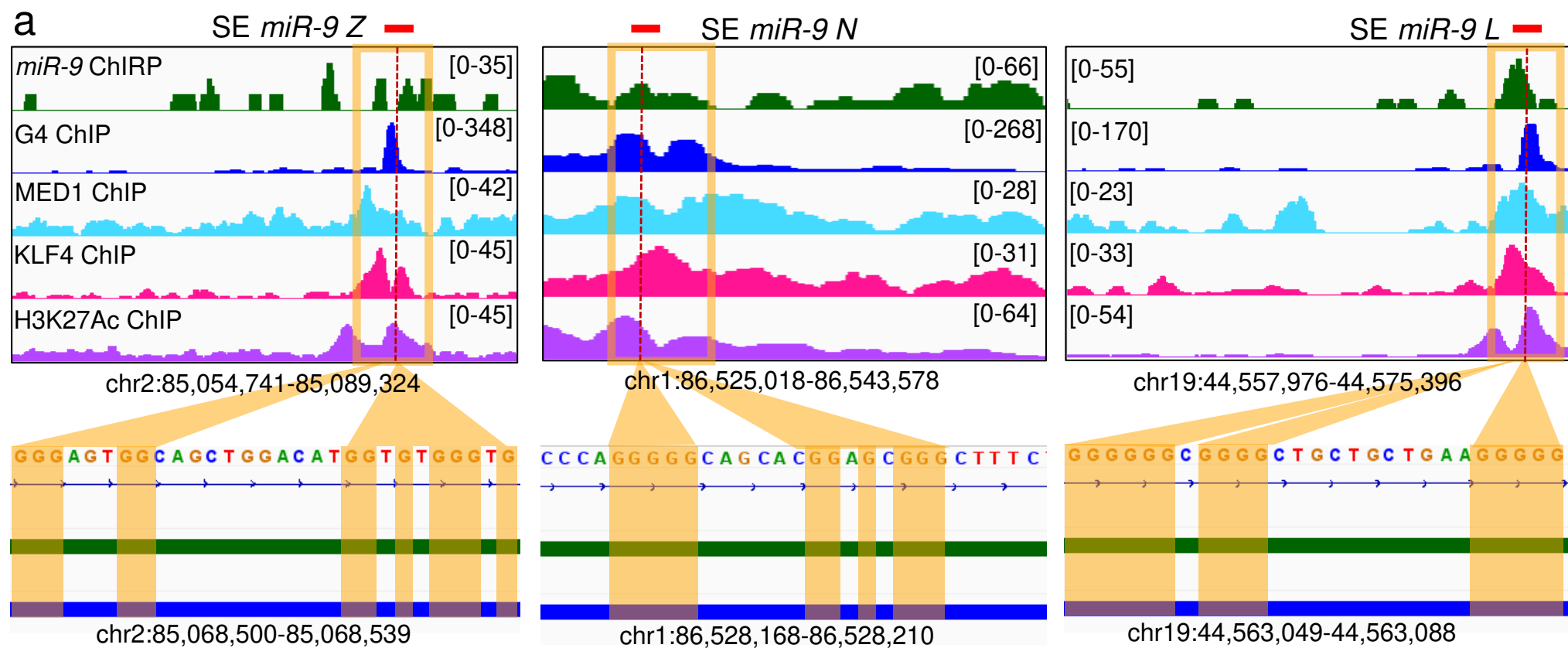


Figure 5

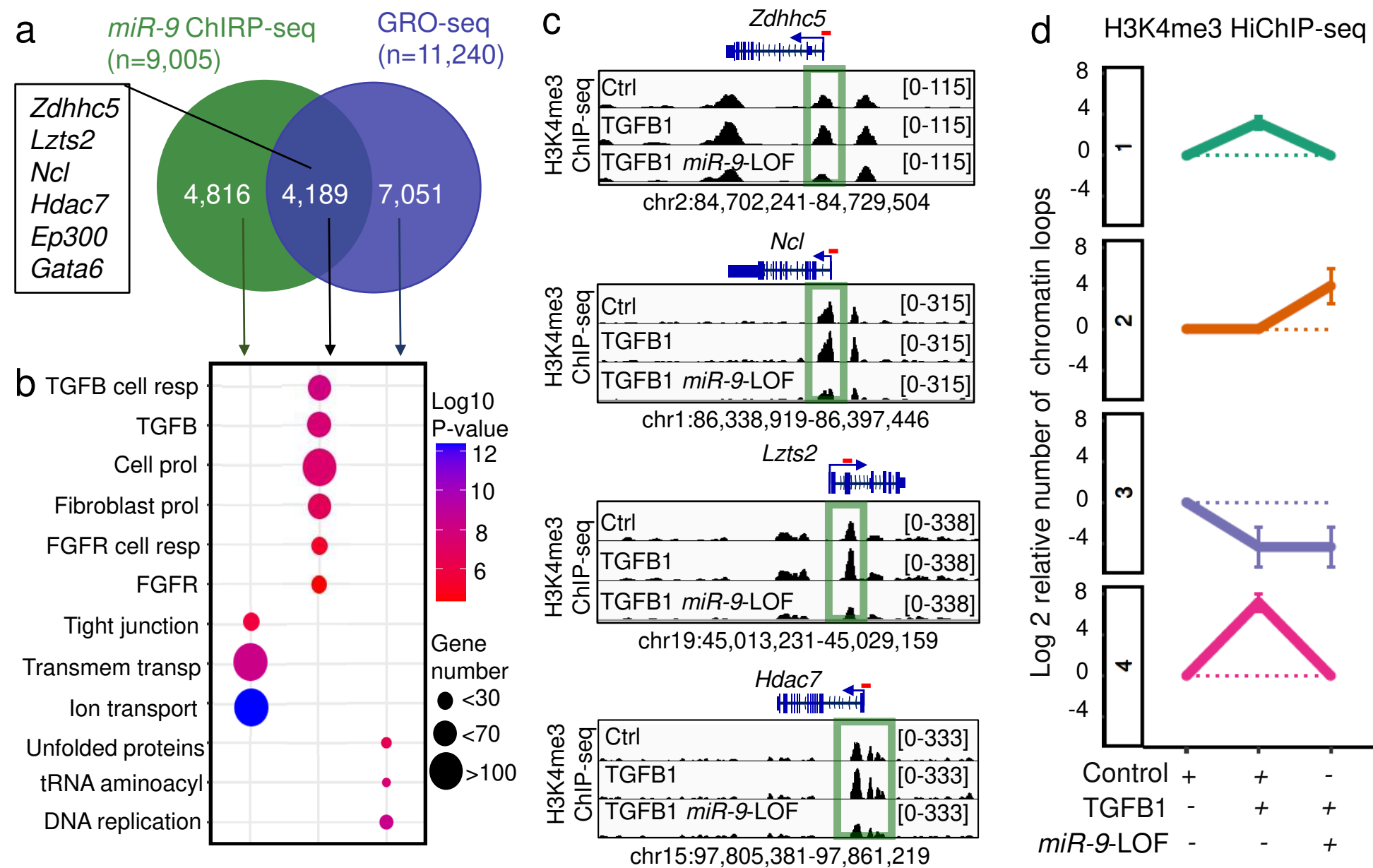


Figure 6

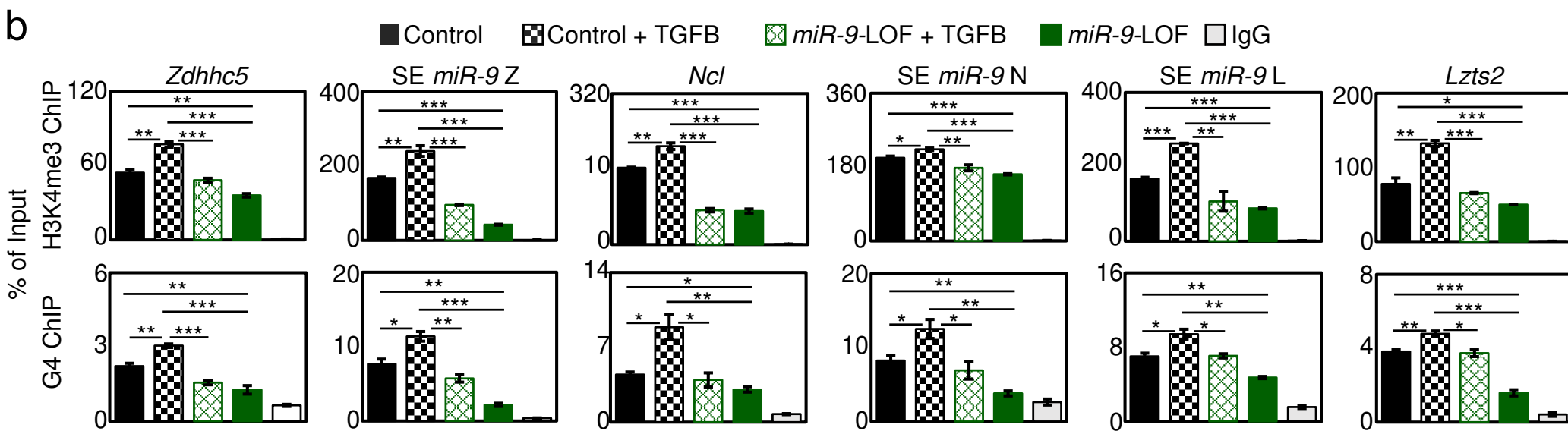
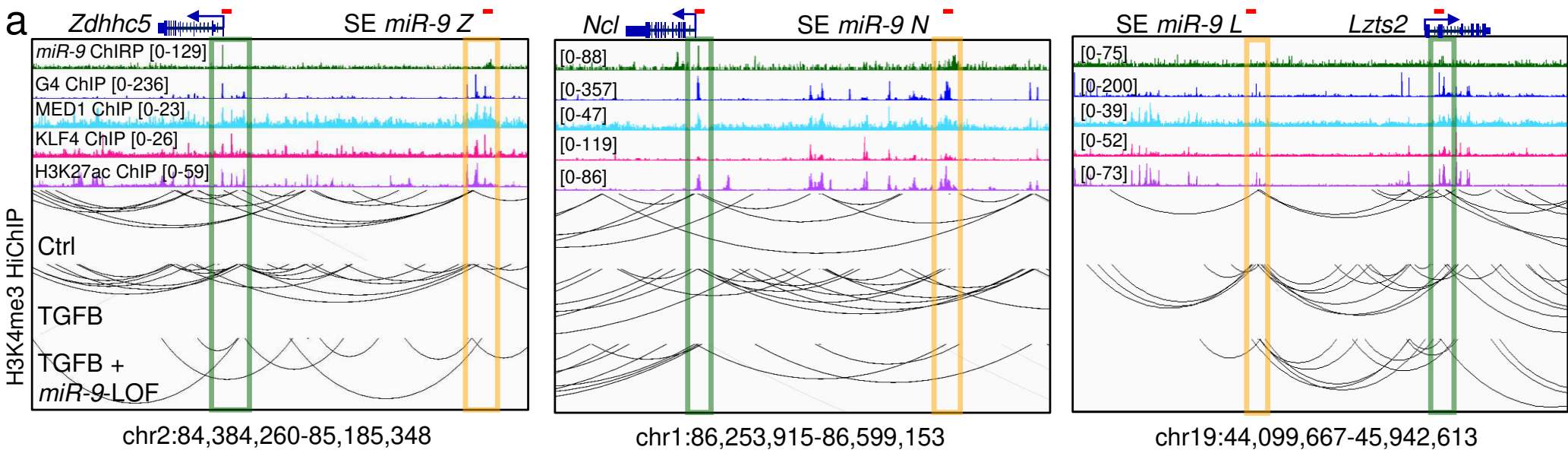


Figure 7

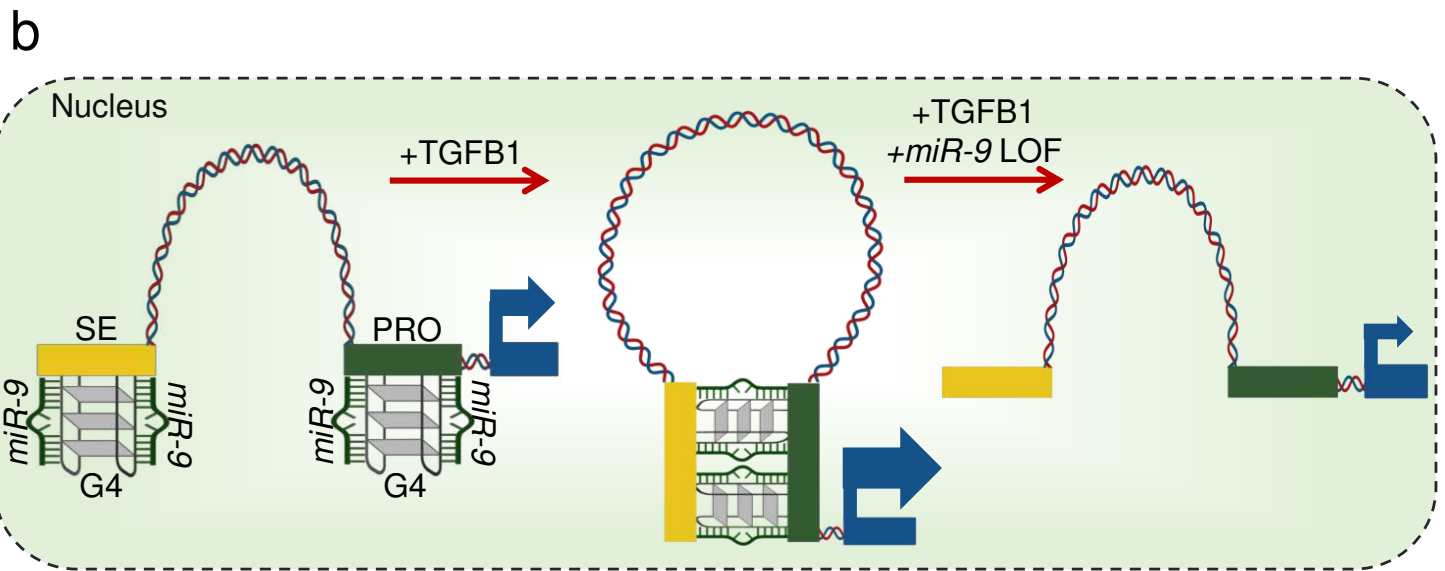
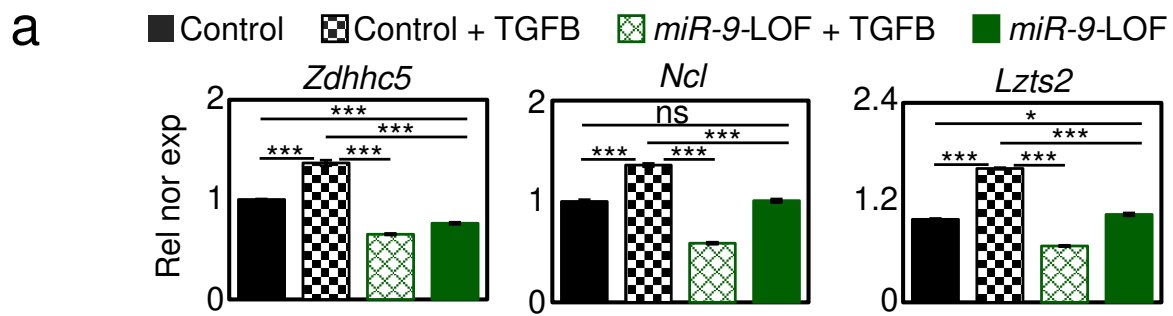


Figure 8