



HAL
open science

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

Julio Cordero, Guruprasadh Swaminathan, Diana G Rogel-Ayala, Karla Rubio, Adel Elsherbiny, Stefan Günther, Thomas Braun, Gergana Dobрева, Guillermo Barreto

► **To cite this version:**

Julio Cordero, Guruprasadh Swaminathan, Diana G Rogel-Ayala, Karla Rubio, Adel Elsherbiny, et al.. 3D genome organization during TGFB-induced transcription requires nuclear microRNA and G-quadruplexes. 2024. hal-04798457

HAL Id: hal-04798457

<https://hal.science/hal-04798457v1>

Preprint submitted on 22 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

26 **ABSTRACT**

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

47

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

50

51 INTRODUCTION

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

74

75 Chromatin structure alone does not determine the functional status of a gene, but it effectively
76 enables RNA polymerase II (Poll II) recruitment to the promoters, as well as binding of
77 transcription factors, co-activators, co-repressors to DNA sequences that function as regulatory
78 elements controlling gene expression ¹¹. A promoter is a sequence of DNA to which proteins
79 bind to initiate transcription of RNA molecules that are usually complementary to the DNA
80 sequence that is located 3' of the promoter. On the other hand, enhancers are relatively short
81 (~100–1000 bp) DNA sequences that are bound by transcription factors and regulate gene
82 transcription independent of their distance, location, or orientation relative to their cognate
83 promoter ¹²⁻¹⁴. Super-enhancers (SE) have been proposed to be long genomic domains
84 consisting of clusters of transcriptional enhancers enriched with histone modification markers
85 (such as histone 3 mono-methylated at lysine 4 or acetylated at lysine 27, H3K4me1 and
86 H3K27ac respectively), cofactors (such as mediator of RNA polymerase II transcription
87 subunit 1, MED1, and components of the multimeric protein complex Cohesin), chromatin
88 modifying proteins (such as E1A Binding Protein P300, EP300) and cell-type-specific
89 transcription factors ¹⁵⁻¹⁷. One versatile feature of chromatin is its ability to form a loop,
90 mediating long-range interactions in which two distant sequences of DNA come to close
91 physical proximity. Chromatin looping has been broadly accepted as a means for enhancer-
92 promoter interactions ^{18, 19}.

93

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

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

106

107 The majority of the eukaryotic genome is transcribed into ncRNAs including microRNAs
108 (miRNAs, 21-25 nucleotides long) and long non-coding RNAs (lncRNAs, >200 nucleotides
109 long)³². LncRNAs are important regulators of different biological processes in the nucleus³³.
110 Together with other factors, lncRNAs provide a framework for the assembly of defined
111 chromatin structures at specific loci, thereby modulating gene expression, centromere function
112 and silencing of repetitive DNA elements³³⁻³⁶. Although miRNAs are assumed to act primarily
113 in the cytosol by inhibiting translation³⁷, mature miRNAs have been also reported in the nuclei
114 of different cells^{8, 38-41}. Interestingly, a hexanucleotide element has been reported to direct
115 miRNA nuclear import⁴², nevertheless, the function of miRNAs in the cell nucleus has been
116 sparsely studied. Here we report on microRNA-9 (*miR-9*), which even though its nucleotide
117 sequence is highly conserved across species, also shows high diversity in expression patterns
118 and biological functions depending on the cellular context^{43, 44}. For example, *miR-9* has been
119 reported to target the lncRNA *MALAT1* for degradation in the cell nucleus⁴¹. However, it has
120 not been linked to transcription regulation, chromatin structure nor 3D genome organization.
121 Here we propose a mechanism of transcriptional regulation of transforming growth factor beta
122 1 (TGFB1) responsive genes that requires nuclear *miR-9* and involves G4s and promoter-SE
123 looping.

124

125 RESULTS

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

127 A phylogenetic tree that was generated using the sequences of mature mouse miRNAs and a
128 heat map comparing their sequence similarity showed that *miR-9* clustered together with
129 miRNAs that have been functionally characterized in the cell nucleus, such as *miR-29b-3p*⁴²,
130 *miR-126-5p*⁴⁵ and *let-7d-5p*⁸ (Figure 1a, top). Sequence alignment between mouse *miR-126-*
131 *5p*, *miR-9* and *miR-29b-3p* showed that various nucleotides are conserved in a sequence stretch
132 that was reported as nuclear shuttling motif from *miR-126-5p*⁴⁵ (Figure 1a, bottom). We called
133 the partially conserved sequence 5'-AKYACCWUUUGRUWA-3' as expanded miRNA
134 nuclear shuttling motif. In addition, we found that the human orthologs of these mature
135 miRNAs, *hsa-miR-126-5P*, *hsa-miR-29B-3P* and *hsa-miR-9-5P*, also contain this expanded
136 miRNA nuclear shuttling motif (Supplementary Figure 1a), showing that it is conserved across
137 species. To confirm the nuclear localization of mature *miR-9*, we performed expression
138 analysis after cell fraction using TaqMan assays specific for mature *miR-9* and total RNA
139 isolated from the cytosolic and the nuclear fractions of different cells (Figure 1b and
140 Supplementary Figure 1b), including mouse lung fibroblasts (MLg and MFML4), mouse lung
141 epithelial cells (MLE-12), mouse mammary gland epithelial cells (NMuMG), and primary
142 human lung fibroblasts (hLF). We detected mature *miR-9* not only in the cytosolic fraction but
143 also in the nuclear fraction of all cells analyzed. Interestingly, the relative levels of nuclear
144 *miR-9* were higher in mouse fibroblasts when compared to epithelial cells. Further, the nuclear
145 localization of *miR-9* was confirmed by RNA fluorescence in situ hybridization (FISH) in MLg
146 cells (Figure 1c, top; Supplementary Figure 1c, left), which also showed that specific regions
147 of the nucleus were *miR-9* depleted. These results were further validated by loss-of-function
148 (LOF) experiments using unlabeled *miR-9*-specific antagomiR probes (Figure 1c, bottom;
149 Supplementary Figure 1c, right). All these results suggest that *miR-9* has a function in the cell

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

168

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

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

175 were transiently transfected with Ctrl or *miR9*-specific antagomiR to induce a *miR-9*-LOF.
176 Interestingly, *miR-9*-LOF significantly reduced the broad domains of H3K4me3 from 55% in
177 Ctrl transfected cells to 27% after *miR-9*-LOF ($P < 0.01$), whereas medium and narrow
178 H3K4me3 domains increased. The reduction of H3K4me3 broad domains after *miR-9*-LOF
179 was confirmed by confocal microscopy after H3K4me3-specific immunostaining in Ctrl- and
180 *miR-9*-antagomiR transfected MLg cells (Supplementary Figure 2b). Since broad domains of
181 H3K4me3 have been associated with increased transcription elongation⁴⁶, we analyzed the
182 transcriptome of MLg cells after *miR-9*-LOF by total RNA sequencing (RNA-seq, Figure 2b-
183 c, Supplementary Figure 2c). Remarkably, from the transcripts that were significantly affected
184 after *miR-9*-LOF ($n = 3,320$), only a minority ($n = 881$; 26.5%) showed increased expression
185 after *miR-9*-LOF, whereas 73.5% ($n = 2,439$) showed reduced expression with a median of
186 1.03 log₂ RPKM and an interquartile range (IQR) of 1.51 log₂ RPKM ($P = 5.34E-36$), when
187 compared to 1.50 log₂ RPKM (IQR = 1.81 log₂ RPKM) in Ctrl antagomiR transfected cells
188 (Figure 2c, top). The expression of genes without *miR-9*-enrichment (non-targets) as
189 determined by *miR-9* ChIRP-seq (Figure 1d-e and Supplementary Figure 1d) was non-
190 significantly affected after *miR-9*-LOF (Figure 2c, bottom) supporting the specificity of the
191 *miR9*-specific antagomiR. These results indicate that *miR-9* is required for high basal
192 transcriptional activity of its target genes. This interpretation was supported by ChIP-seq in
193 mouse embryonic fibroblasts (MEF) using antibodies specific for total RNA polymerase II (Pol
194 II) and serine 5 phosphorylated Pol II (Pol II S5p,⁹) showing transcription initiation. Heat maps
195 representing the results of the *miR-9* ChIRP-seq (Figure 2d), Pol II and Pol II S5p ChIP-seq
196 (Figure 2e) revealed that Pol II and Pol II S5p were enriched at the transcription start sites
197 (TSS) of the genes coding for the transcripts that were significantly affected after *miR-9*-LOF
198 (further referred to as *miR-9* target genes). Moreover, genome-wide precision nuclear run-on
199 assay (PRO-seq)⁴⁷ and global run-on sequencing (GRO-seq)⁴⁸, 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^{6, 20}. 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 promoters²⁷,
209²⁸, 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 specificity²⁴. 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 DNA²⁸. 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

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

234

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

236 Alluvial plot implementing the data from *miR-9* ChIRP-seq and G4P ChIP-seq²⁴ showed
237 enrichment of *miR-9* and G4 at loci that are also enriched with markers of active SE, such as
238 MED1 and H3K27Ac^{49, 50} (Figure 4a). Furthermore, a Venn diagram using the same data sets
239 together with the data from a GRO-seq experiment⁴⁸ showed 3,583 common loci (Figure 4b)
240 suggesting transcriptional activity from these loci. Remarkably, 95.5% ($n = 3,423$) of these
241 transcripts were found in the animal eRNA database⁵¹ as enhancer RNAs (eRNA) (Figure 4c).
242 Moreover, mapping the *miR-9* ChIRP-seq data to known loci of typical enhancer (TE) or SE
243 showed that *miR-9* is significantly enriched at SE (Figure 4d). These results were confirmed
244 by further analysis of the *miR-9* ChIRP-seq data together with publicly available ChIP-seq data
245 of proteins that have been related to SE (Figure 4e). Aggregate plots showed enrichment of
246 *miR-9* specifically at SE together with MED1, MYC, H3K27Ac, KLF4, MEIS1, EP300,
247 HDAC1, SMAD3, BRG1, HDAC2, EST1 and RAD21. In addition, we also observed
248 enrichment of CHD4 and SMARCA5 at the same SE as *miR-9*. Supporting these observations,
249 we found that *miR-9* pulled down endogenous CHD4 and SMARCA5 by miRNA pulldown

250 (miR-Pd) followed by Western Blot (Supplementary Figure 4a), whereas both proteins
251 precipitated endogenous *miR-9* by chromatin-RNA immunoprecipitation (Ch-RIP) followed
252 by *miR-9*-specific TaqMan assays (Supplementary Figure 4b). Our results suggest CHD4 and
253 SMARCA5 as components of SE. Remarkably, aggregate plots using G4-specific ChIP-seq
254 data²⁴ showed G4 enrichment at the same SE as *miR-9* in mouse lung fibroblasts (MLg cells)
255 and adenocarcinoma human alveolar basal epithelial cells (A549 cells) (Figure 4f).

256

257 Loci visualization of selected SE with *miR-9* enrichment using the IGV genome browser
258 (Figure 5a top and Supplementary Figure 5) confirmed the enrichment of G4 at the same loci,
259 as well as of MED1, KLF4 and H3K27Ac^{49, 50}, which are markers of active SE. Interestingly,
260 a zoom in to the loci revealed G-rich sequences that favor the formation of G4^{6, 20} (Figure 5a,
261 bottom, and Supplementary Figure 5). These results were confirmed by analysis of the loci of
262 these SE by ChIP followed by qPCR using H3K4me3- or G4-specific antibodies and chromatin
263 from MLg cells that were transiently transfected with Ctrl or *miR9*-specific antagomiR probes
264 (Figure 5b). We detected H3K4me3 and G4 enrichment at the loci of all analyzed SE in Ctrl
265 antagomiR-transfected cells, which was significantly reduced after *miR-9*-LOF. Our results
266 demonstrate that SE with *miR-9* enrichment are also enriched with H3K4me3 and G4 in a *miR*-
267 9-dependent manner.

268

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

270 To further elucidate the biological relevance of our findings we performed gene set enrichment
271 analysis (GSEA)⁵² of the loci with *miR-9* enrichment and nascent RNA as determined by *miR*-
272 9 ChIRP-seq and GRO-seq, respectively (Figure 6a-b). We found significant enrichment of
273 genes related to the categories “TGFB cell response” ($P = 7.55E-09$), “TGFB” ($P = 2.61E-08$),

274 “Cell proliferation” ($P = 4.66E-08$) and “Fibroblasts proliferation” ($P = 1.61E-07$), suggesting
275 an involvement of the loci with *miR-9* enrichment and nascent RNA in these biological
276 processes. To link this observation with our results showing reduction of H3K4me3 after *miR-*
277 *9*-LOF at specific promoters (Figure 3b) and SE (Figure 5b) we did ChIP-seq using H3K4me3-
278 specific antibodies and chromatin from MLg cells that were transiently transfected with Ctrl or
279 *miR9*-specific antagomiR probes, and non-treated or treated with TGFB1 (Figure 6c and
280 Supplementary Figure 6a-b). Loci visualization of the selected *miR-9* target genes (*Zdhhc5*,
281 *Ncl*, *Lzts2*, *Hdac7*, *Gata6* and *Ep300*) using the IGV genome browser showed H3K4me3
282 enrichment at the promoters in non-treated, and Ctrl antagomiR transfected MLg cells that
283 increased after TGFB1 treatment. However, a combination of TGFB1 treatment and *miR9*-
284 specific antagomiR transfection showed that *miR-9*-LOF counteracted the effect caused by
285 TGFB1 demonstrating the requirement of *miR-9* for the chromatin changes induced by TGFB1
286 and suggesting its requirement for TGFB1-inducibility of the analyzed genes, as shown below.

287

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

299 the *miR-9* targets (*Zdhhc5*, *Ncl*, *Lzts2* and *Hdac7*) that were selected based on our *miR-9*
300 ChIRP-seq (Figure 1c-e). Further, we generated IGV genome browser snapshots to visualize
301 the enrichment of *miR-9*, G4, MED1, KLF4 and H3K27ac at the loci of promoters of *miR-9*
302 target genes and SE with *miR-9* enrichment (Figure 7a, top). In the same snapshots we
303 presented the results of the H3K4me3-specific HiChIP-seq (Figure 7a, bottom) showing
304 chromatin loops between the promoters of *miR-9* target genes and SE with *miR-9* enrichment
305 in Ctrl antagomiR-transfected cells. Strikingly, these promoter-SE-loops increased in TGFB1
306 treated cells in a *miR-9*-dependent manner, since *miR-9*-LOF counteracted the effect caused by
307 TGFB1. To correlate the results from H3K4me3-specific HiChIP-seq with changes in
308 chromatin, we analyzed the promoters and SE with *miR-9* enrichment by ChIP qPCR using
309 H3K4me3- and G4-specific antibodies and chromatin from MLg cells that were transfected
310 with Ctrl or *miR-9*-specific antagomiRs, and non-treated or treated with TGFB1 (Figure 7b).
311 TGFB1 treatment significantly increased H3K4me3 and G4 levels at the analyzed promoters
312 and SE. Further, *miR-9*-LOF significantly reduced the effect caused by TGFB1 treatment,
313 whereas *miR-9*-LOF alone significantly reduced H3K4me3 and G4 levels when compared to
314 Ctrl antagomiR transfected cells. To correlate these changes in chromatin structure with gene
315 expression, we analyzed the expression of *miR-9* target genes by qRT-PCR in MLg cells under
316 the same conditions as specified above (Figure 8a). The expression of all analyzed *miR-9* target
317 genes significantly increased after TGFB1 treatment in *miR-9*-dependent manner, correlating
318 with our chromatin structure analysis (Figures 6c and 7b). Further, *miR-9*-LOF alone
319 significantly reduced the basal transcription levels of the analyzed *miR-9* target genes when
320 compared to Ctrl antagomiR transfected cells confirming our RNA-seq results (Figure 2b-c
321 and Supplementary Figure 2c). Summarizing, our results support a model (Figure 8b), in which
322 G4s are formed in a *miR-9*-dependent manner at both, promoters of TGFB1-responsive genes,
323 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

328 **DISCUSSION**

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

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

354

355 Another interesting aspect of the model proposed here is the participation of G4s in 3D genome
356 organization during transcription activation. DNA G4s are stable four-stranded non-canonical
357 structures that are highly related to promoters and transcription activation²³⁻²⁶. Further, a recent
358 publication based on integrative analysis of multi-omics studies have provided comprehensive
359 mechanistic insights into the function of G4s as promoter elements that reduce nucleosome
360 density, increase the levels of active histone marks (H3K4me3 and H3K27ac), generate
361 nucleosome arrays by positioning nucleosomes at a periodic distance to each other and
362 facilitate pause release of Pol II into effective RNA production resulting in enhanced
363 transcriptional activity²⁹. Even though similar chromatin features were described to be induced
364 by G4s at promoter-distal regulatory elements, such as SE⁵⁸, no mechanistic insights were
365 provided prior to our study elucidating the role of G4s in chromatin looping mediating long-
366 range enhancer-promoter interactions. We found that G4s are formed in a *miR-9*-dependent
367 manner at both, promoters of TGFB1-responsive genes, as well as SE with which these
368 promoters form loops (Figure 7a-b). Moreover, our results showed promoter-SE looping
369 increase after TGFB1 treatment allowing these two regulatory elements to come to close
370 physical proximity and enhance transcription of the corresponding genes (Figure 8a). One can
371 hypothesize that G4 could stabilize the interaction between these two regulatory elements
372 allowing the increase of transcription. We will investigate this hypothesis in a future project
373 and whether the G4 are formed with DNA strands from the promoter and from the SE.

374

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

392

393 **MATERIALS AND METHODS**

394

395 **Cell culture**

396 Mouse lung fibroblast cells MLg (ATCC CCL-206) and MFML4⁶⁶ were cultured in complete
397 DMEM (4.5 g/L glucose, 10% FCS, 1% Penn-strep, 2 mM l-glutamine) at 37 °C in 5% CO₂.
398 Mouse lung epithelial cells MLE-12 (ATCC CRL-2110) were cultured in complete
399 DMEM/F12 (5% FCS, 1% Penn-strep) at 37°C in 5% CO₂. Mouse mammary gland epithelial
400 cells NMuMG (ATCC CRL-1636) were cultured in complete DMEM (4.5 g/L glucose,
401 10mcg/mL insulin (90%), 10% FCS, 1% Penn-strep) at 37 °C in 5% CO₂. Human primary lung
402 fibroblasts from control donors were cultured in complete MCDB131 medium (8% fetal calf
403 serum (FCS), 1% L-glutamine, penicillin 100 U/ml, streptomycin 0.1 mg/ml, epidermal growth
404 factor 0.5 ng/ml, basic fibroblast growth factor 2 ng/ml, and insulin 5 µg/ml)) at 37°C in 5%
405 CO₂. During subculturing, cells were 1x PBS washed, trypsinized with 0.25% (w/v) Trypsin
406 and split at the ratio of 1:5 to 1:10. The cell lines used in this paper were mycoplasma free.
407 They were regularly tested for mycoplasma contamination. In addition, they are not listed in
408 the database of commonly misidentified cell lines maintained by ICLAC.

409

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

411 Cells were transfected with antagomiR probes (Ambion) using Lipofectamine 2000
412 (Invitrogen) following the manufacturer's instructions, and harvested 24 h later for further
413 analysis. Anti-hsa-miR-*Mir9-5p* (Ambion, #17000), and Anti-miR negative control (Ambion,
414 #17010) were transfected at 60 nM final concentrations. Following 20 h after transfection,
415 TGFB1 signaling was induced with 5 ng/ml final concentration of human recombinant TGFB1
416 (Sigma-Aldrich) for 4 h.

417

418 **Bacterial culture and cloning**

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

422

423 **RNA isolation, reverse transcription, quantitative PCR and TaqMan assay**

424 Expression analysis by qRT-PCR were performed as previously described ⁶⁷. Briefly, total
425 RNA from cell lines was isolated using the RNeasy Mini kit (Qiagen) and quantified using a
426 Nanodrop Spectrophotometer (ThermoFisher Scientific). Synthesis of complementary DNA
427 was performed using 1–2 µg total RNA and the High Capacity cDNA Reverse Transcription
428 kit (Applied Biosystems). Quantitative real-time PCR reactions were performed using SYBR®
429 Green on the Step One plus Real-time PCR system (Applied Biosystems). Housekeeping gene
430 *Gapdh* was used to normalize gene expression. Primer pairs used for gene expression analysis
431 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
--	---------	---------------------

432

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

438

439 **MiRNA fluorescence in situ hybridization**

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

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

457

458 **Immunofluorescence and confocal microscopy**

459 Immunostaining was performed as previously described⁵³. Briefly, cells were grown on
460 coverslips, fixed with 4% PFA for 10min at RT and permeabilized with 0.4% Triton-
461 X100/1xPBS for 10min at RT. During immunostaining procedure, all incubations and washes
462 were performed with histobuffer containing 3% bovine serum albumin (BSA) and 0.2% Triton
463 X-100 in 1xPBS, pH7.4. Non-specific binding was blocked by incubating with BSA 5% serum
464 in 1x PBS, pH 7.4. Cells were then incubated with primary antibodies overnight at 4°C. After
465 3 washes with histobuffer (15 min each), secondary antibody was incubated at room
466 temperature for 1h followed by DAPI nuclear staining (Sigma, Germany). Immunostainings
467 were examined with an immunofluorescence microscope (Leica). Antibodies used were
468 specific anti-H3K4me3 (Abcam, ab8580). Alexa 488 or Alexa 594 tagged secondary antibodies
469 (Invitrogen, Germany, dilution 1:1000) were used. DAPI (Sigma, Germany) was used as
470 nuclear dye.

471

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

473 Chromatin isolation by miRNA purification (ChIRP) was performed as described⁶⁸, with slight
474 modifications. Briefly, MLE-12 cells were cross-linked by 1% formaldehyde for 10min, lysed,
475 and sonicated with a Diagenode Bioruptor to disrupt and fragment genomic DNA. After
476 centrifugation, chromatin was incubated with 150 pmol of biotin-labeled anti-sense LNA
477 probes (Exiqon) specific to mature Mirlet7d, or control, at 37 °C for 4h. Streptavidin-magnetic
478 C1 beads (Invitrogen) were blocked with 500ng/μL yeast total RNA and 1mg/mL BSA for 1h
479 at room temperature, and washed three times in nuclear lysis buffer (2mM Tris-HCl, pH 7.0,

480 250mM NaCl, 2mM EDTA, 2mM EGTA, 1% Triton X-100, 0.2mM DTT, 20mM NaF, 20mM
481 Na₃VO₄, 40µg/mL phenylmethylsulfonyl fluoride, protease inhibitor, and RNase inhibitor in
482 DEPC-treated water) and resuspended in its original volume. We added 100 µL of
483 washed/blocked C1 beads per 100 pmol of probes, and the whole reaction was mixed for
484 another 30min at 37 °C. Beads:biotinprobes:RNA:chromatin adducts were captured by
485 magnets (Invitrogen) and washed five times with 40×bead volume of wash buffer. DNA was
486 eluted with a cocktail of 100µg/mL RNase A (ThermoFisher Scientific). Chromatin was
487 reverse cross-linked at 65 °C overnight. Later, DNA was purified using the QIAquick PCR
488 Purification Kit (Qiagen) according to manufacturer’s instructions and used for single gene
489 promoter analysis by qPCR and for sequencing. Primer pairs used for qPCR analysis are
490 described in in the sub-section “Chromatin immunoprecipitation” below. ChIRP-seq was
491 performed by single-end sequencing on an Illumina HiSeq2500 machine at the Max Planck-
492 Genome-Centre Cologne. Raw reads were trimmed using Trimmomatic-0.36 with the
493 parameters (ILLUMINACLIP:\${ADAPTERS}:2:30:10 LEADING:3 TRAILING:3
494 SLIDINGWINDOW:4:15 MINLEN:20 CROP:70 HEADCROP:10
495 (<https://doi.org/10.1093/bioinformatics/btu170>). Trimmed reads were mapped to mouse
496 genome mm10 using Bowtie2 (default settings)⁶⁹. Next, PCR duplicates were removed from
497 the BAM files using the MarkDuplicates.jar tool from Picard (version 1.119). MCAS14⁷⁰ was
498 used (macs14 -t ChIRP -c Inp -f BAM -p 1e-3 -g 1.87e9 --nomodel --shiftsize 30 -n; and -p 1e-
499 3 -g 1.87e9 --nomodel --shiftsize 100) with the aim to capture genome areas with *miR-9*
500 enrichment. Both peak results were merged with bedtools merge (bedtools merge -d 100 -i
501 cat_PEAKS_s_2.bed) with different annotation with annotatePeaks.pl from homer⁷¹. Bam
502 files were converted to bigwig files by the help of bamCoverage from deptools (-bs 20 --
503 smoothLength 40 -p max --normalizeUsing RPKM -e 150)⁷². The cis-regulatory element
504 annotation system (CEAS)⁷³ was used to determine the distribution of the peaks from *miR-9*

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

507

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

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

516

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

518 To determine the potential super enhancer marked or not marked by *miR-9*, we crossed the
519 *miR-9* ChIRP-seq peaks list with the results from the peak list from H3K27ac peaks from
520 (GSM2417093) determined using the following parameters: findPeaks H3K27ac_maketaglib
521 -i Inp_makeTaglib -region -size 500 -minDist 20000 -fdr 0.05 -F 2 -style super -typical
522 output_enh -superSlope -1000 > OUPUT}_SE.txt.

523 Those significant SE with more than 100 normalized reads were further crossed with the *miR-*
524 *9* ChIRP-seq peak by the help of bedtools intersect (setings -wa). Those H3K27ac peaks not
525 overlapping the SE list and not overlapping the *miR-9* peaks were considered TE. All ChiP-seq
526 and *miR-9* ChIRP-seq were quantified from the center of both Peak list by the help of
527 annotatePeaks from homer with the settings: annotatePeaks.pl list.bed mm10 -size 4000 -norm
528 1000000 -hist 10 -d maketagLib_Samples > output_quanty.txt. The results of this command

529 was used as input in a custom R-script to produce the aggregate plots
530 (https://github.com/jcorder316/001Mir9_G4_3D_STRUC).

531

532 **Chromatin immunoprecipitation (ChIP)**

533 ChIP analysis was performed as described earlier^{9, 75} with minor adaptations. Briefly, cells
534 were cross-linked with 1% methanol-free formaldehyde (ThermoFisher Scientific) lysed, and
535 sonicated with Diagenode Bioruptor to an average DNA length of 300–600 bp. After
536 centrifugation, the soluble chromatin was immunoprecipitated with 3 µg of antibodies specific
537 for anti-H3K4me3 (Abcam, # ab8580), Anti-DNA G-quadruplex structures Antibody, clone
538 BG4 (Millipore, # MABE917), and anti-IgG (Santa Cruz, #sc-2027). Reverse crosslinked
539 immunoprecipitated chromatin was purified using the QIAquick PCR purification kit (Qiagen)
540 and subjected to ChIP-quantitative PCR. The primer pairs used for gene promoter and super-
541 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 miR-9 L	Forward	TTACAGCTTTCCGACCAGGC
	Reverse	CCTCTCAAGCCTCTGTCCAC

542

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

545 HiChIP experiments were performed as previously described ⁷⁶ using antibodies specific for
546 H3K4me3 (Abcam, # ab8580) with the following optimizations: 5-10 million cells were
547 crosslinked with 1% formaldehyde for 10 min at room temperature; prior to restriction
548 digestion, sodium dodecyl sulfate treatment at 62°C for 10 min; restriction digestion with *MboI*
549 (New England Biolabs France, R0147M) for 2 h at 37°C; prior to fill-in reaction, heat
550 inactivation of *MboI* at 62°C for 10 min followed by 2 washing steps of pelleted nuclei with
551 1x fill-in reaction buffer; after fill-in reaction, ligation at 4°C for 16 h.

552

553 **ChIP sequencing, HiChIP sequencing and data analysis**

554 HiChIP-seq paired-end reads were aligned to the mm10 genome, duplicate reads were
555 removed, reads were assigned to *MboI* restriction fragments, filtered into valid interactions,
556 and the interaction matrices were generated using the HiC-Pro pipeline default settings ⁷⁷. The
557 contig file of the HiC-Pro was set to allow validPairs at any distance from each other. HiC-Pro
558 valid interaction reads were then used to detect significant interactions using: (1) Maketag
559 libraries were made as followed (makeTagDirectory output_PCA_ucsc
560 smaple_R1_mm10.bwt2merged.bam,sample_R2_mm10.bwt2merged.bam -tbp 1 -genome
561 mm10 -checkGC -restrictionSite GATC. (2) We used (runHiCpca.pl sample_mer25_50
562 sample_PCA_ucsc -res 25000 -superRes 50000 -genome mm10 -cpu 16) to find the Principal
563 component analysis (PCA) of the data. (3) We used analyzeHiC as followed (analyzeHiC
564 sample_PCA_ucsc -res 1000000 -interactions sample_significantInteractions.txt -nomatrix).

565 Only Interaction looping between the H3K4me3 peaks and the TSS (+/-2kb) from *miR-9*
566 candidate genes were considered. The settings were (annotateInteractions.pl Sample_significan
567 mm10 Sample_output_filterk4m9 -filter H3k4me3_HICHIP-seq_peaks -filter2
568 TSS2kb_fMir9_genes -cpu 16 -washu -pvalue 0.01). (4) The mapped merge bam files ouput
569 from HiC-pro was processed as ChiP-seq to perform Peak calling, using (peak_call -i bam_file
570 -o ouput_peak -r MboI_mm10.txt -f 0.01 -a mm10_chr_size -w 8) from
571 (https://github.com/ChenfuShi/HiChIP_peaks).

572 To calculate the broadness of H3K4me3 peaks, we first performed summary statistics of the
573 peak size from the merged peak list. If the size of a peak was equal to or higher than the top 75
574 % of peaks quantile 3 (Q3) was consider wide (≥ 2.7 kb). If the peak was between Q3 and Q2
575 was consider medium size peak (≥ 2 kb and < 2.7 kb). The narrow Hek4me3 peaks were in the
576 bottom 25% peaks, Q1 or less (< 2 kb).

577

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

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

586 The enrichment of the factors and nascent RNAs (PRO-seq and GRO-seq) on the TSS of the
587 Down-regulated genes after *miR-9*-LOF were performed by the help of the library from R
588 profileplyr

589 (<https://www.bioconductor.org/packages/devel/bioc/vignettes/profileplyr/inst/doc/profileplyr>).

590 html) adapting to a custom R-script
591 (https://github.com/jcorder316/001Mir9_G4_3D_STRUC).

592

593 **MicroRNA pulldown**

594 For microRNA pulldown (miR-pd), MLE-12 cells were transfected with Biotinylated *mmu-*
595 *miR-9-5p*-RNA (Exiqon) or Biotinylated *Mirctrl* (Exiqon), at a final concentration of 20 nM.

596 After 48 hours, cells were fixed with 1% glutaraldehyde (Sigma-Aldrich) to preserve RNA-
597 Chromatin interactions. After washing three times with PBS, cell pellets were resuspended in

598 2 ml hypotonic cell lysis buffer (10 mM Tris-HCl (pH 7.4), 1.5 mM MgCl₂, 10 mM KCl, 1
599 mM DTT, 25 mM NaF, 0.5 mM Na₃VO₄, 40 µg/ml phenylmethylsulfonyl fluoride, protease

600 inhibitor (Calbiochem) and RNase inhibitor (Promega) in DEPC treated water) on ice for 10
601 min and then spun down at 700 g for 10 min at 4°C. Nuclear pellets were resuspended in 300

602 µL of nuclear lysis buffer (50 mM Tris-HCl (pH 7.4), 170 mM NaCl, 20% glycerol, 15 mM
603 EDTA, 0.1% (v/v) Triton X-100, 0.2 mM DTT, 20 mM NaF, 20 mM Na₃VO₄, 40 µg/ml

604 phenylmethylsulfonyl fluoride, protease inhibitor, RNase inhibitor in DEPC treated water).
605 Nuclear fraction was sonicated (Bioruptor NextGen, Diagenode) at high amplitude for 10

606 cycles of 30 sec (on/off) pulses. M280 streptavidin magnetic beads (ThermoFisher Scientific)
607 were blocked for 1 h at 4°C in blocking buffer (10 mM Tris-HCl pH 6.5, 1 mM EDTA, and 1

608 mg/ml BSA) and washed twice with 1 ml washing buffer (10 mM Tris-HCl (pH 7.0), 1 mM
609 EDTA, 0.5 M NaCl, 0.1% (v/v) Triton X-100, RNAase inhibitor and protease inhibitor). Beads

610 were resuspended in 0.5 ml washing buffer. The nuclear extract was then added to the beads
611 and incubated for 1 h at 4°C with slow rotation. The beads were then washed five times with 1

612 ml washing buffer. RNA bound to the beads (pulled down RNA) and from 10% of the extract
613 (input RNA) was isolated using Trizol reagent LS (Invitrogen) after DNase 1 (NEB) treatment.

614 For miRNA-protein interaction analysis, nuclear protein extract after DNase1 treatment
615 incubated with 500 pMol Biotinylated *miR-9* or Biotinylated *Mirctrl* for 1 h with shaking at
616 4°C. M280 streptavidin magnetic beads (ThermoFisher Scientific) were used for pulldown.
617 Protein bound to the beads (miR-Pd), was incubated with 30 µl 2x SDS samples loading buffer,
618 boiled at 95°C for 5min, spun down and loaded on SDS-PAGE for western blot analysis and/or
619 processed for proteomic analysis by mass spectrometry.

620

621 **Chromatin RNA immunoprecipitation**

622 Chromatin RNA immunoprecipitation (Ch-RIP) analysis was performed as described ⁸ with
623 minor adaptations. Briefly, cells were cross-linked by 1% formaldehyde for 10 min, lysed, and
624 sonicated with Diagenode Bioruptor to disrupt and fragment genomic DNA. After
625 centrifugation, the soluble chromatin was immunoprecipitated using antibodies. Precipitated
626 chromatin complexes were removed from the beads by incubating with 50 µl of 1% SDS with
627 0.1 M NaHCO₃ for 15 min, vortexing every 5 min and followed by treatment with DNase1.
628 The collected material was immunoprecipitated using IgG or an HA-specific antibody (Santa
629 Cruz, CatNo. 805). Reverse cross-linked immunoprecipitated materials were used for RNA
630 isolation by Trizol (Invitrogen). Isolated RNA was subjected to cDNA synthesis and further
631 for TaqMan assay specific for mature *miR-9* (Applied Biosystems, Cat. No. 4427975).

632

633 **Western blot**

634 Nuclear protein extracts from MLg cells were prepared by cell lysis and nuclei isolation.
635 Briefly, MLg cells were spooled down and washed with PBS. cell pellets were resuspended in
636 2 ml hypotonic cell lysis buffer (10 mM Tris-HCl (pH 7.4), 1.5 mM MgCl₂, 10 mM KCl, 1
637 mM DTT, 25 mM NaF, 0.5 mM Na₃VO₄, 40 µg/ml phenylmethylsulfonyl fluoride, protease
638 inhibitor (Calbiochem) and RNase inhibitor (Promega) in DEPC treated water) on ice for 10

639 min and then spun down at 700 g for 10 min at 4°C. Nuclear pellets were resuspended in 300
640 µL of nuclear lysis buffer (50 mM Tris-HCl (pH 7.4), 170 mM NaCl, 20% glycerol, 15 mM
641 EDTA, 0.1% (v/v) Triton X-100, 0.2 mM DTT, 20 mM NaF, 20 mM Na₃VO₄, 40 µg/ml
642 phenylmethylsulfonyl fluoride, protease inhibitor, RNase inhibitor in DEPC treated water).
643 Nuclear fraction was sonicated (Bioruptor NextGen, Diagenode) at high amplitude for 10
644 cycles of 30 sec (on/off) pulses. Detergent-insoluble material was precipitated by
645 centrifugation at 14,000 rpm for 30 min at 4 °C. The supernatant was transferred to a fresh tube
646 and stored at -20 °C. Protein concentration was estimated using Bradford assay, using serum
647 albumin as standard. 5 µl of serial dilutions of standard protein and samples were mixed with
648 250 µl of Bradford reagent (500-0205, BIO-RAD Quick Start™).

649 Western blotting was performed using standard methods and antibodies specific for
650 SMARCA5 (Invitrogen, MA5-35378), LMNB1 (Santa Cruz, sc-374015), CHD4 (Abcam,
651 70469), RAD21 (Abcam, ab992), GAPDH (Sigma, MFCD01322099). Immunoreactive
652 proteins were visualized with the corresponding HRP-conjugated secondary antibodies (Santa
653 Cruz) using the Super Signal West Femto detection solutions (ThermoFisher Scientific).
654 Signals were detected and analyzed with Luminescent Image Analyzer (Las 4000, Fujifilm).

655

656 **RNA sequencing and data analysis**

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

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

672

673 **Statistical Analysis**

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

684

685 **Data availability**

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

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

693

694 **Acknowledgments**

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

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

715

716 **Authors Contributions**

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

721

722 **Conflict of Interest**

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

725

726 **SUPPLEMENTARY INFORMATION**

727 Supplementary Figures 1 to 7

728 Supplementary Table 1

729 Source Data file 01

730

731 **REFERENCES**

- 732 1 Jerkovic I, Cavalli G. Understanding 3D genome organization by multidisciplinary methods.
733 *Nature reviews Molecular cell biology* 2021; **22**:511-528.
- 734 2 Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation.
735 *Nature reviews Molecular cell biology* 2019; **20**:535-550.
- 736 3 Misteli T. The Self-Organizing Genome: Principles of Genome Architecture and Function.
737 *Cell* 2020; **183**:28-45.
- 738 4 Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are
739 predictive for gene expression. *Proceedings of the National Academy of Sciences of the United*
740 *States of America* 2010; **107**:2926-2931.
- 741 5 Gohlke J, Scholz CJ, Kneitz S *et al.* DNA methylation mediated control of gene expression
742 is critical for development of crown gall tumors. *PLoS genetics* 2013; **9**:e1003267.
- 743 6 Georgakopoulos-Soares I, Morganella S, Jain N, Hemberg M, Nik-Zainal S. Noncanonical
744 secondary structures arising from non-B DNA motifs are determinants of mutagenesis.
745 *Genome research* 2018; **28**:1264-1271.
- 746 7 Singh I, Ozturk N, Cordero J *et al.* High mobility group protein-mediated transcription
747 requires DNA damage marker gamma-H2AX. *Cell research* 2015; **25**:837-850.
- 748 8 Singh I, Contreras A, Cordero J *et al.* MiCEE is a ncRNA-protein complex that mediates
749 epigenetic silencing and nucleolar organization. *Nature genetics* 2018; **50**:990-1001.
- 750 9 Dobersch S, Rubio K, Singh I *et al.* Positioning of nucleosomes containing gamma-H2AX
751 precedes active DNA demethylation and transcription initiation. *Nature communications* 2021;
752 **12**:1072.
- 753 10 Papantonis A, Cook PR. Genome architecture and the role of transcription. *Current opinion*
754 *in cell biology* 2010; **22**:271-276.
- 755 11 Jha RK, Levens D, Kouzine F. Mechanical determinants of chromatin topology and gene
756 expression. *Nucleus* 2022; **13**:94-115.
- 757 12 Buecker C, Wysocka J. Enhancers as information integration hubs in development: lessons
758 from genomics. *Trends Genet* 2012; **28**:276-284.
- 759 13 Schaffner W. Enhancers, enhancers - from their discovery to today's universe of
760 transcription enhancers. *Biol Chem* 2015; **396**:311-327.
- 761 14 Zaugg JB, Sahlen P, Andersson R *et al.* Current challenges in understanding the role of
762 enhancers in disease. *Nat Struct Mol Biol* 2022; **29**:1148-1158.
- 763 15 Whyte WA, Orlando DA, Hnisz D *et al.* Master transcription factors and mediator establish
764 super-enhancers at key cell identity genes. *Cell* 2013; **153**:307-319.
- 765 16 Hnisz D, Abraham BJ, Lee TI *et al.* Super-enhancers in the control of cell identity and
766 disease. *Cell* 2013; **155**:934-947.

- 767 17 Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome.
768 *Nucleic acids research* 2016; **44**:D164-171.
- 769 18 Hsieh TS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer-
770 promoter interactions and transcription are largely maintained upon acute loss of CTCF,
771 cohesin, WAPL or YY1. *Nature genetics* 2022; **54**:1919-1932.
- 772 19 Kyrchanova O, Georgiev P. Mechanisms of Enhancer-Promoter Interactions in Higher
773 Eukaryotes. *International journal of molecular sciences* 2021; **22**.
- 774 20 Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and
775 functions of DNA and RNA G-quadruplexes. *Nature reviews Molecular cell biology* 2020;
776 **21**:459-474.
- 777 21 Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic acids*
778 *research* 2015; **43**:8627-8637.
- 779 22 Henderson E, Hardin CC, Walk SK, Tinoco I, Jr., Blackburn EH. Telomeric DNA
780 oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs.
781 *Cell* 1987; **51**:899-908.
- 782 23 Hansel-Hertsch R, Beraldi D, Lensing SV *et al.* G-quadruplex structures mark human
783 regulatory chromatin. *Nature genetics* 2016; **48**:1267-1272.
- 784 24 Zheng KW, Zhang JY, He YD *et al.* Detection of genomic G-quadruplexes in living cells
785 using a small artificial protein. *Nucleic acids research* 2020; **48**:11706-11720.
- 786 25 Du Z, Zhao Y, Li N. Genome-wide colonization of gene regulatory elements by G4 DNA
787 motifs. *Nucleic acids research* 2009; **37**:6784-6798.
- 788 26 Renciuik D, Rynes J, Kejnovska I *et al.* G-quadruplex formation in the Oct4 promoter
789 positively regulates Oct4 expression. *Biochim Biophys Acta Gene Regul Mech* 2017; **1860**:175-
790 183.
- 791 27 Lago S, Nadai M, Cernilogar FM *et al.* Promoter G-quadruplexes and transcription factors
792 cooperate to shape the cell type-specific transcriptome. *Nature communications* 2021; **12**:3885.
- 793 28 Esnault C, Magat T, Zine El Aabidine A *et al.* G4access identifies G-quadruplexes and their
794 associations with open chromatin and imprinting control regions. *Nature genetics* 2023;
795 **55**:1359-1369.
- 796 29 Esnault C, Garcia-Oliver E, Aabidine AZE *et al.* G-quadruplexes are promoter elements
797 controlling nucleosome exclusion and RNA polymerase II pausing. *bioRxiv*
798 2023:2023.2002.2024.529838.
- 799 30 Agarwal T, Roy S, Kumar S, Chakraborty TK, Maiti S. In the sense of transcription
800 regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry*
801 2014; **53**:3711-3718.
- 802 31 Holder IT, Hartig JS. A matter of location: influence of G-quadruplexes on Escherichia coli
803 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.

- 841 47 Mahat DB, Salamanca HH, Duarte FM, Danko CG, Lis JT. Mammalian Heat Shock
842 Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation.
843 *Molecular cell* 2016; **62**:63-78.
- 844 48 Busslinger GA, Stocsits RR, van der Lelij P *et al.* Cohesin is positioned in mammalian
845 genomes by transcription, CTCF and Wapl. *Nature* 2017; **544**:503-507.
- 846 49 Buganim Y, Markoulaki S, van Wietmarschen N *et al.* The developmental potential of
847 iPSCs is greatly influenced by reprogramming factor selection. *Cell stem cell* 2014; **15**:295-
848 309.
- 849 50 Chronis C, Fiziev P, Papp B *et al.* Cooperative Binding of Transcription Factors
850 Orchestrates Reprogramming. *Cell* 2017; **168**:442-459 e420.
- 851 51 Jin W, Jiang G, Yang Y *et al.* Animal-eRNAdb: a comprehensive animal enhancer RNA
852 database. *Nucleic acids research* 2022; **50**:D46-D53.
- 853 52 Mootha VK, Lindgren CM, Eriksson KF *et al.* PGC-1alpha-responsive genes involved in
854 oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*
855 2003; **34**:267-273.
- 856 53 Rubio K, Singh I, Dobersch S *et al.* Inactivation of nuclear histone deacetylases by EP300
857 disrupts the MiCEE complex in idiopathic pulmonary fibrosis. *Nature communications* 2019;
858 **10**:2229.
- 859 54 De Magis A, Manzo SG, Russo M *et al.* DNA damage and genome instability by G-
860 quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National*
861 *Academy of Sciences of the United States of America* 2019; **116**:816-825.
- 862 55 Lyu J, Shao R, Kwong Yung PY, Elsasser SJ. Genome-wide mapping of G-quadruplex
863 structures with CUT&Tag. *Nucleic acids research* 2022; **50**:e13.
- 864 56 Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome
865 integrity and powerful regulators of gene expression. *Genes & development* 2014; **28**:1384-
866 1396.
- 867 57 Sato K, Hendrikx AGM, Knipscheer P. RNA transcripts suppress G-quadruplex structures
868 through G-loop formation. *bioRxiv* 2023:2023.2003.2009.531892.
- 869 58 Hou Y, Guo Y, Dong S, Yang T. Novel Roles of G-quadruplexes on Enhancers in human
870 chromatin. *bioRxiv* 2021:2021.2007.2012.451993.
- 871 59 Pokholok DK, Harbison CT, Levine S *et al.* Genome-wide map of nucleosome acetylation
872 and methylation in yeast. *Cell* 2005; **122**:517-527.
- 873 60 Li X, Wu F, Gunther S *et al.* Inhibition of fatty acid oxidation enables heart regeneration in
874 adult mice. *Nature* 2023; **622**:619-626.
- 875 61 Sze CC, Ozark PA, Cao K *et al.* Coordinated regulation of cellular identity-associated
876 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

917 **FIGURE LEGENDS**

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

919 (a) Phylogenetic tree (left) and a heat map (right) that were generated using the sequences of
920 the indicated mature mouse miRNAs. Numbers indicate distance score. Bottom, sequence
921 alignment of the indicated mature mouse miRNAs highlighting the published *miR-126* nuclear
922 shuttling motif (magenta) and the expanded miRNA nuclear shuttling motif (green square)
923 using the IUPAC nucleotide code. Pink letters are conserved among all sequences. Golden
924 letters are conserved in at least 2 sequences. Blue letters indicate conserved type of base (either
925 purine or pyrimidine). (b) Mature *miR-9*-specific TaqMan assay after cellular fractionation
926 (Supplementary Figure 1b) of indicated cell lines. Data are shown as means \pm s.e.m. ($n = 3$
927 independent experiments). (c) Fluorescence microscopy of MLg cells after RNA FISH
928 confirmed nuclear localization of endogenous *miR-9*. Cells were transiently transfected with
929 control (top) or *miR-9*-specific antagomiR probes (bottom) to induce a *miR-9* loss-of-function
930 (LOF). Representative images from three independent experiments. Squares are shown at
931 higher magnification. DAPI, nucleus. Scale bars, 10 μ m. (d) Genome-wide distribution of *miR-9*
932 peaks relative to gene structures by ChIRP-seq in MLg cells. Square on right shows *miR-9*
933 enrichment in different gene structure as Log₂ ratios. TTS, transcription termination sites;
934 3'UTR, 3'untranslated regions. (e) Visualization of selected *miR-9* target genes using IGV
935 genome browser showing *miR-9* enrichment in MLg cells. ChIRP-seq reads were normalized
936 using reads per kilobase per million (RPKM) measure and are represented as log₂ enrichment
937 over their corresponding inputs. Images show the indicated gene loci with their genomic
938 coordinates. Arrows, direction of the genes; blue boxes, exons; red line, regions selected for
939 single gene analysis in f. (f) Analysis of selected *miR-9* target genes by ChIRP using chromatin
940 from MLg cells and control (Ctrl) or *miR-9*-specific biotinylated antisense oligonucleotides for
941 the precipitation of endogenous mature *miR-9* along with the chromatin bound to it. Bar plots

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

945

946 **Figure 2: *MiR-9* is required for H3K4me3 broad domains, high basal transcriptional**
947 **activity, and G-quadruplex formation at promoters. (a)** Genome-wide distribution of
948 H3K4me3 peaks relative to broad (≥ 2.7 kb), medium (≥ 2 kb and < 2.7 kb) narrow (< 2 kb)
949 H3K4me3 domains in MLg cells that were transiently transfected with control (Ctrl) or *miR-*
950 *9*-specific antagomiR probes to induce *miR-9* loss-of-function (LOF). Square on right shows
951 H3K4me3 enrichment in different domains as Log2 ratios of MLg cells after *miR-9*-LOF versus
952 Ctrl transfected cells. Asterisks, P -values after two-tailed t-test, $**P \leq 0.01$. See also
953 Supplementary Figure 2. **(b)** RNA-seq using total RNA from MLg cells that were transfected
954 as in a. Pie chart shows the distribution of transcripts with significantly altered expression
955 levels ($n = 3,320$) between both conditions tested in decreased transcripts ($n = 2,439$) and
956 increased transcripts ($n = 881$) after *miR-9*-LOF. P -values after two-tailed t-test. **(c)** Box plots
957 of RNA-seq-based expression analysis of transcripts with significantly decreased levels after
958 *miR-9*-LOF (*miR-9* targets; $n = 2,439$) and with non-significantly (ns) changed levels (non-
959 targets; $n = 881$) after *miR-9*-LOF. Values were normalized using reads per kilobase per million
960 (RPKM) and represented as log2 RPKM. Box plots indicate median (middle line), 25th, 75th
961 percentile (box) and 5th and 95th percentile (whiskers); P -values after two-tailed t-test. Source
962 data are provided as Source Data file 01. **(d)** Heat map for *miR-9* enrichment at the TSS ± 2 kb
963 of *miR-9* target genes as determined by RNA-seq in b and c. **(e-g)** Heat maps for enrichment
964 of total Pol II and Pol II S5p (e), nascent RNA by precision nuclear run-on assay (PRO-seq)
965 and global run-on sequencing (GRO-seq) (f), chromatin accessibility by ATAC-seq and
966 H3K4me3 by CHIP-seq (g), at the TSS ± 5 kb of *miR-9* target genes as determined by RNA-

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

972

973 **Figure 3: H3K4me3, nascent RNA and G4 are enriched at promoters of selected *miR-9***
974 **target genes.** **(a)** Visualization of selected *miR-9* target genes using IGV genome browser
975 showing enrichment of *miR-9* by ChIRP-seq (green), H3K4me3 by ChIP-seq in Ctrl and *miR-*
976 *9*-specific antagomiR transfected MLg cells (black), nascent RNA by GRO-seq (brown) and G4
977 by G4P ChIP-seq (blue). Reads were normalized using reads per kilobase per million (RPKM)
978 after bamCoverage. Images show the indicated gene loci with their genomic coordinates.
979 Arrows, direction of the genes; blue boxes, exons; red lines, regions selected for single gene
980 analysis in b; green squares, regions with enrichment of *miR-9*, H3K4me3, nascent RNA and
981 G4; dotted lines, regions shown at the bottom with high G content. **(b)** Analysis of the promoter
982 of selected *miR-9* target genes by ChIP using chromatin from MLg cells transfected with
983 control (Ctrl) or *miR-9*-specific antagomiR to induce *miR-9* loss-of-function (LOF). Bar plots
984 presenting data as means; error bars, s.e.m ($n = 3$ biologically independent experiments);
985 asterisks, P -values after two-tailed t-test, $**P \leq 0.01$; $*P \leq 0.05$. See also Supplementary Figure
986 3. Source data are provided as a Source Data file 01.

987

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

991 *miR-9* enrichment. **(b)** Venn diagram showing the number of common loci ($n = 3,583$) with
992 enrichment of *miR-9* by ChIRP-seq (green), H3K27ac (purple) and MED1 (turquoise) by ChIP-
993 seq and nascent RNA (blue) by GRO-seq. **(c)** Venn diagram showing that 3,423 (95.5%) of the
994 transcripts related to the common loci in **b** were found in a database as enhancer RNAs. **(d)**
995 Genome-wide distribution of *miR-9* peaks by ChIRP-seq in MLg cells relative to super-
996 enhancers (SE) and typical enhancers (TE). **(e)** Aggregate plots showing at SE (green lines)
997 and TE (grey lines) ± 2 kb the enrichment of *miR-9* by ChIRP-seq and the indicated proteins
998 by ChIP-seq, which have been previously related to SE in various murine cells. Reads were
999 normalized using reads per kilobase per million (RPKM) measure. **(f)** Aggregate plots showing
1000 at SE (magenta and blue lines) and TE (grey lines) ± 2 kb the enrichment of G4 by ChIP-seq
1001 in mouse MLg and human A549 cells. Reads were normalized using RPKM. See also
1002 Supplementary Figure 4. Source data are provided as a Source Data file 01. Supplementary
1003 Table 1 contains all data sets used in this study.

1004

1005 **Figure 5: Nuclear *miR-9* is enriched at super-enhancers and is required for G-**
1006 **quadruplexes. (a)** Visualization of selected SE with *miR-9* enrichment using IGV genome
1007 browser showing enrichment *miR-9* by ChIRP-seq (green), G4 by G4P ChIP-seq (blue), MED1
1008 (turquoise), KLF4 (magenta) and H3K27Ac (purple) by ChIP-seq. Reads were normalized
1009 using reads per kilobase per million (RPKM). Images show the indicated gene loci with their
1010 genomic coordinates. Orange squares, regions with enrichment of *miR-9*, G4 and SE markers;
1011 red lines, regions selected for single gene analysis in **b**; dotted lines, regions shown at the
1012 bottom with high G content. **(b)** Analysis of selected SE with *miR-9* enrichment by ChIP using
1013 chromatin from MLg cells transfected with control (Ctrl) or *miR-9*-specific antagomiR to
1014 induce *miR-9* loss-of-function (LOF). Bar plots presenting data as means; error bars, s.e.m (n
1015 = 3 biologically independent experiments); asterisks, P -values after two-tailed t-test, $***P \leq$

1016 0.001; ** $P \leq 0.01$; * $P \leq 0.05$. See also Supplementary Figure 5. Source data are provided as a
1017 Source Data file 01.

1018

1019 **Figure 6: Nuclear *miR-9* is required for H3K4me3 enrichment and chromatin loops at**

1020 **promoters of TGF β 1-responsive genes. (a)** Venn diagram after cross analysis of *miR-9*

1021 ChIRP-seq and GRO-seq⁴⁸ showing loci with *miR-9* enrichment and nascent RNAs (n =

1022 4,189), in which the selected *miR-9* target genes are included (square). **(b)** Gene set enrichment

1023 analysis (GSEA) of the three loci groups identified in **a**. Resp, response; prol, proliferation;

1024 trans, transport. *P* values after two-tailed Fisher's exact test and represented as Log10. **(c)**

1025 Visualization of selected *miR-9* target genes using IGV genome browser showing enrichment

1026 H3K4me3 by ChIP-seq in MLg cells that were transfected with control (Ctrl) or *miR-9*-specific

1027 antagomiR to induce a loss-of-function (LOF), and non-treated or treated with TGF β 1, as

1028 indicated. Images show the indicated loci with their genomic coordinates. Arrows, transcription

1029 direction; green squares, promoter regions; dotted lines, regions selected for single gene

1030 analysis in Figure 3b. **(d)** Line charts showing the number of significant chromatin interactions

1031 at loci with *miR-9* and H3K4me3 enrichment in MLg cells treated as in **c**. Four clusters were

1032 generated using k-means algorithm. Data are represented as Log2 of the ratio relative to Ctrl-

1033 transfected, non-treated cells. See also Supplementary Figure 6 and 7. Source data are provided

1034 as a Source Data file 01.

1035

1036 **Figure 7: Promoter-super-enhancer looping of TGF β 1-responsive genes requires *miR-9*.**

1037 **(a)** Visualization of promoters of selected *miR-9* target genes (green squares) and SE with *miR-*

1038 *9* enrichment (orange squares) using IGV genome browser showing enrichment *miR-9* by

1039 ChIRP-seq (green), G4 by G4P ChIP-seq (blue), MED1 (turquoise), KLF4 (magenta) and

1040 H3K27Ac (purple) by ChIP-seq. Reads were normalized using reads per kilobase per million

1041 (RPKM) measure and are represented as log₂ enrichment over their corresponding inputs.
1042 Bottom, chromatin loops by HiChIP-seq in MLg cells that were transfected with control (Ctrl)
1043 or *miR-9*-specific antagomir (*miR-9*-LOF, loss-of-function), and non-treated or treated with
1044 TGFβ1, as indicated. Images show the indicated loci with their genomic coordinates. Arrows,
1045 transcription direction; red lines, regions selected for single gene analysis in b. **(b)** Analysis of
1046 the promoters and SE highlighted in a by ChIP using the indicated antibodies and chromatin of
1047 MLg cells treated as in a. All bar plots present data as means; error bars, s.e.m ($n = 3$
1048 biologically independent experiments); asterisks, P -values after two-tailed t-test, *** $P \leq 0.001$;
1049 ** $P \leq 0.01$; * $P \leq 0.05$; ns, non-significant. See also Supplementary Figure 7. Source data are
1050 provided as a Source Data file 01.

1051

1052 **Figure 8: Promoter-super-enhancer looping of TGFβ1-responsive genes requires *miR-9*.**

1053 **(a)** Expression analysis of the selected *miR-9* target genes by qRT-PCR in MLg cells that were
1054 transfected with control (Ctrl) or *miR-9*-specific antagomir (*miR-9*-LOF, loss-of-function), and
1055 non-treated or treated with TGFβ1, as indicated. All bar plots present data as means; error bars,
1056 s.e.m ($n = 3$ biologically independent experiments); asterisks, P -values after two-tailed t-test,
1057 *** $P \leq 0.001$; ** $P \leq 0.01$; * $P \leq 0.05$; ns, non-significant. See also Supplementary Figure 7.
1058 Source data are provided as a Source Data file 01. **(b)** Model summarizing the results presented
1059 in the manuscript. Left, G4 are formed in *miR-9* (red lines) -dependent manner at SE (orange
1060 box) and promoters (green box) of TGFβ1-responsive genes (blue box, coding region). Middle,
1061 TGFβ1 treatment increases euchromatin histone mark H3K4me₃, G4 and chromatin loops
1062 bringing SE and promoter to close physical proximity, thereby enhancing transcription (arrow)
1063 of the corresponding gene. Right, *miR-9*-LOF antagonizes the effects induced by TGFβ1.
1064 Created with BioRender.com

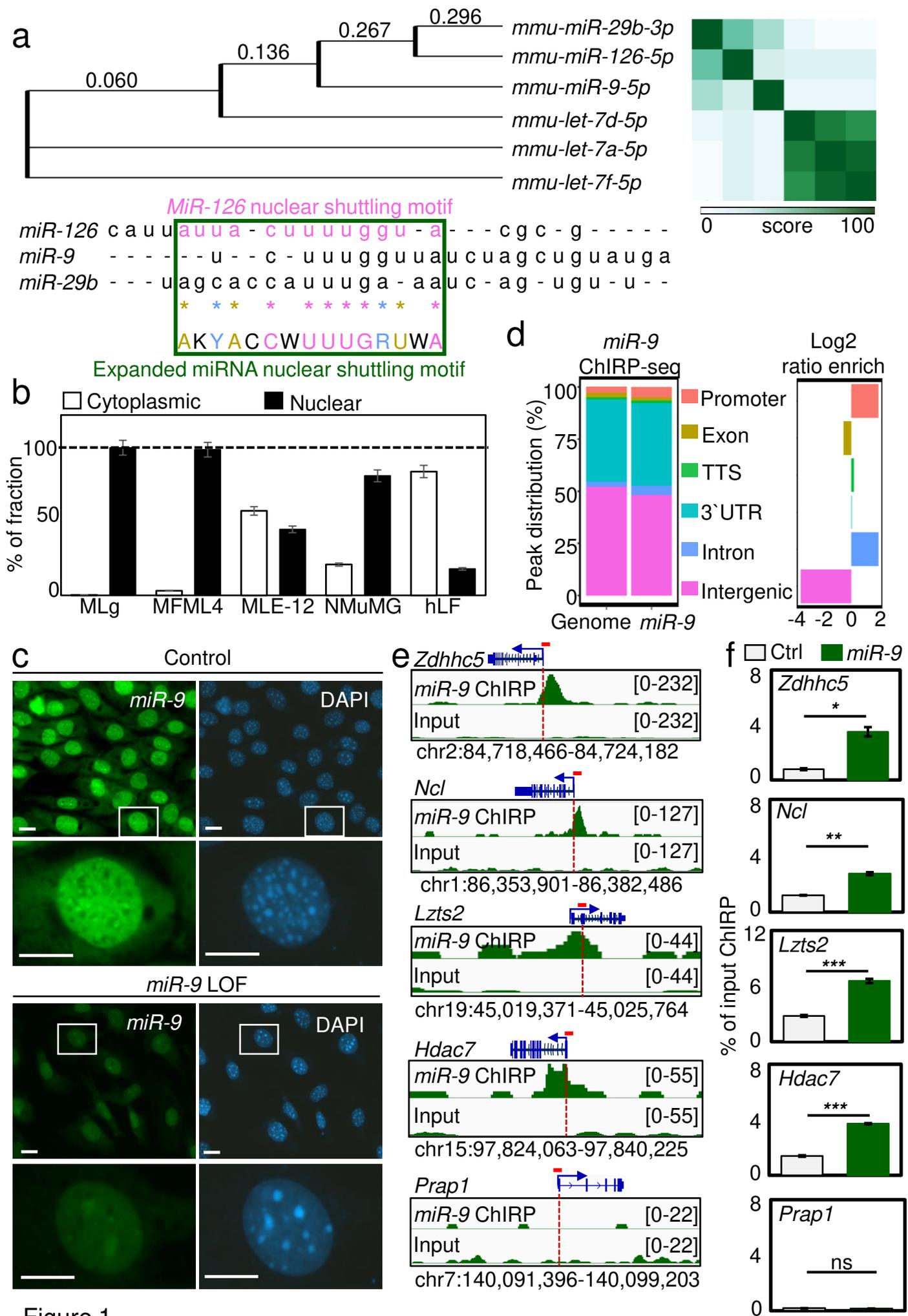


Figure 1

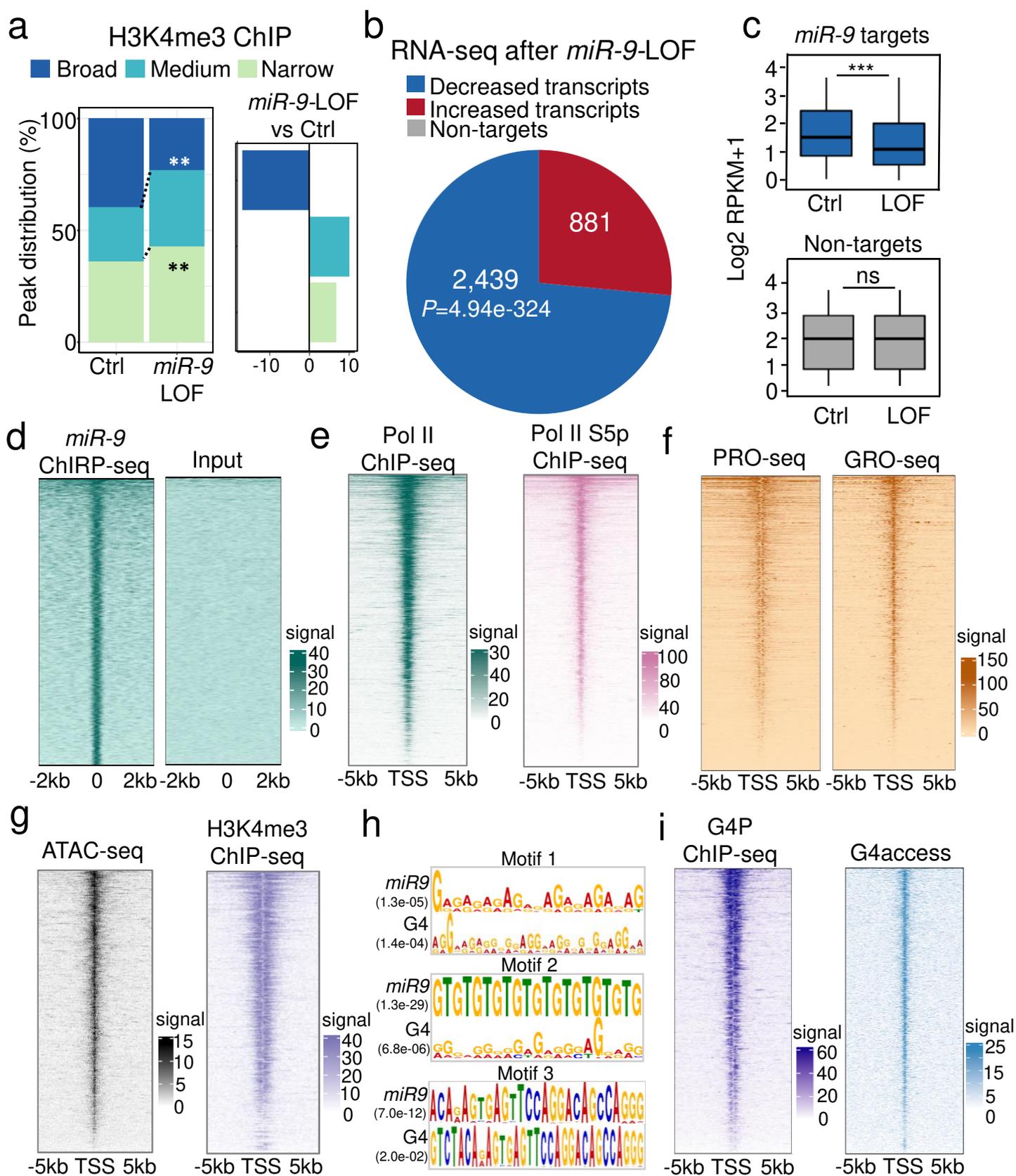


Figure 2

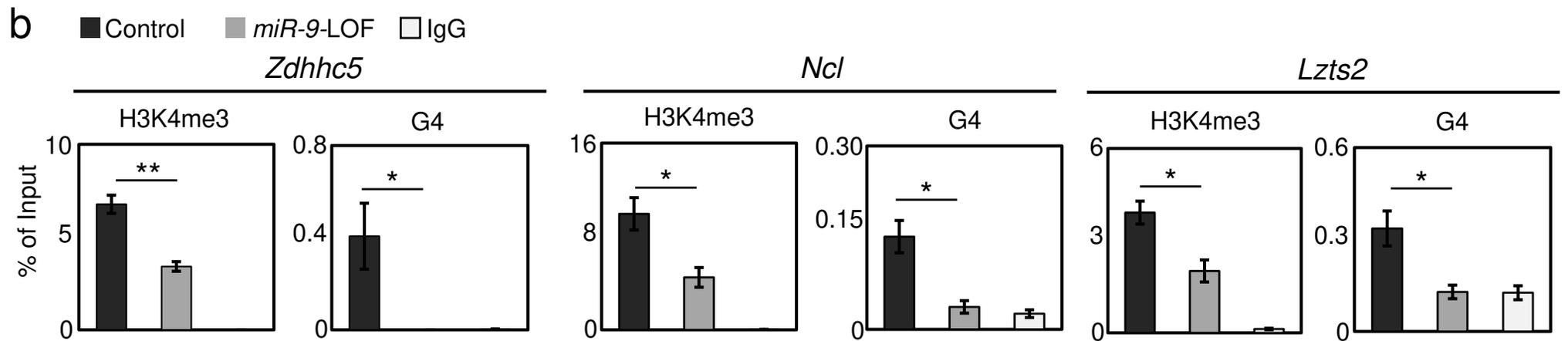
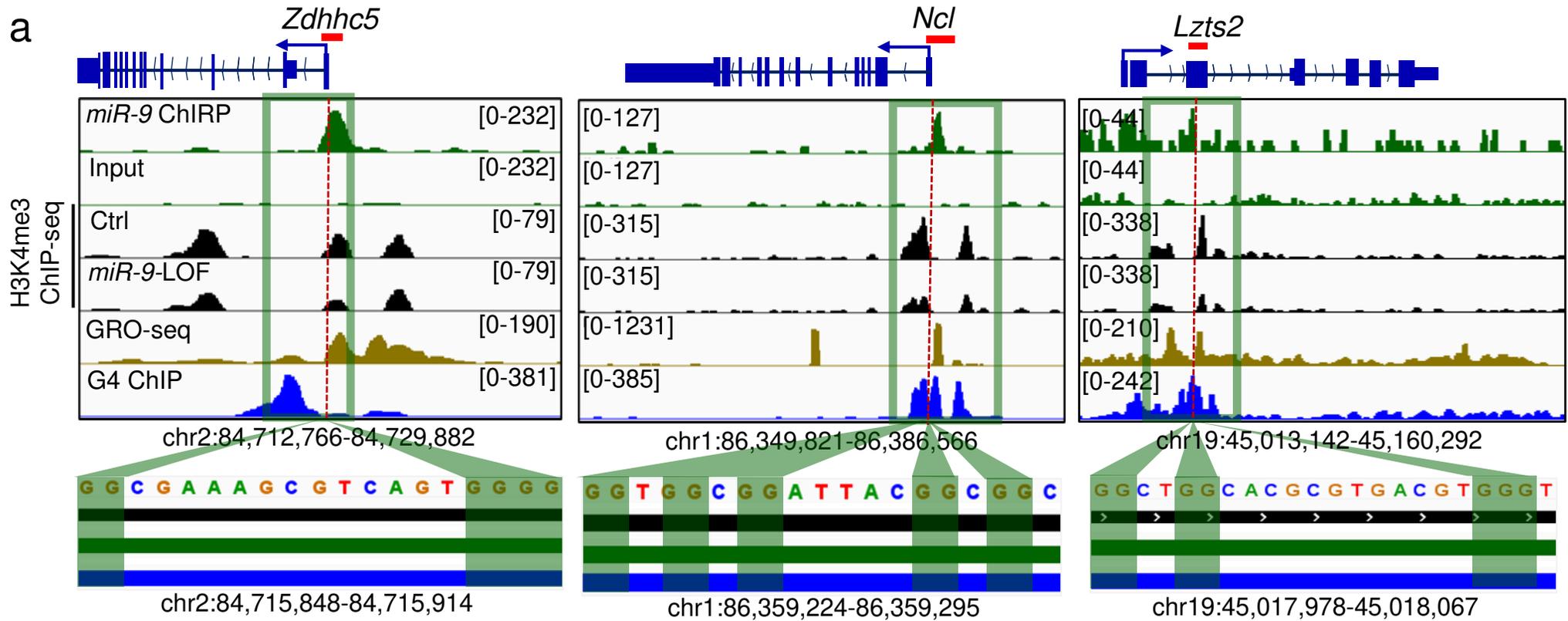


Figure 3

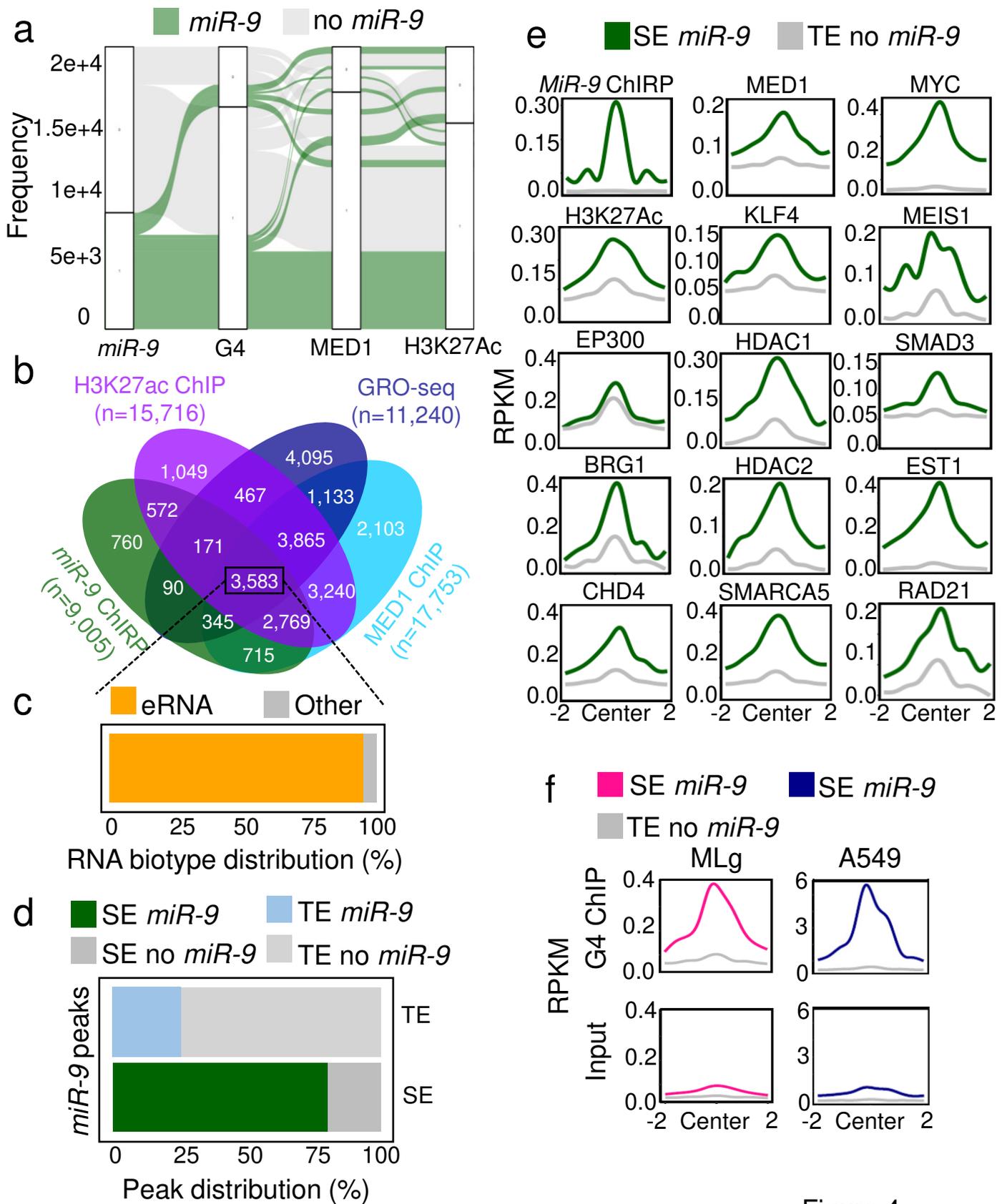


Figure 4

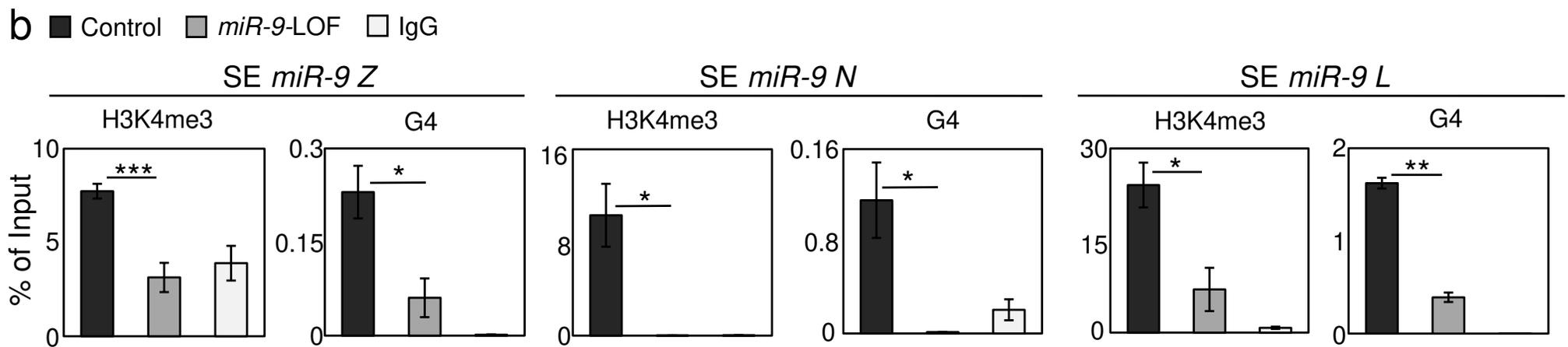
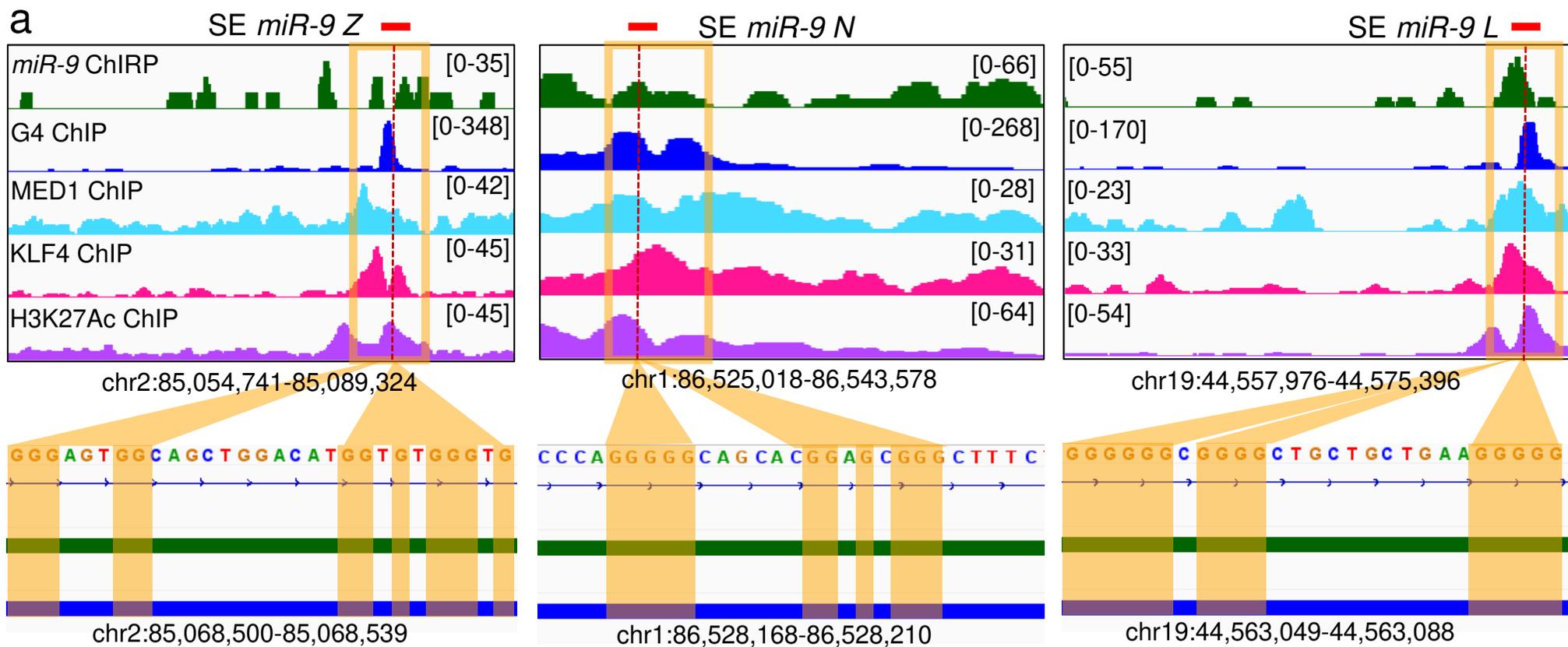


Figure 5

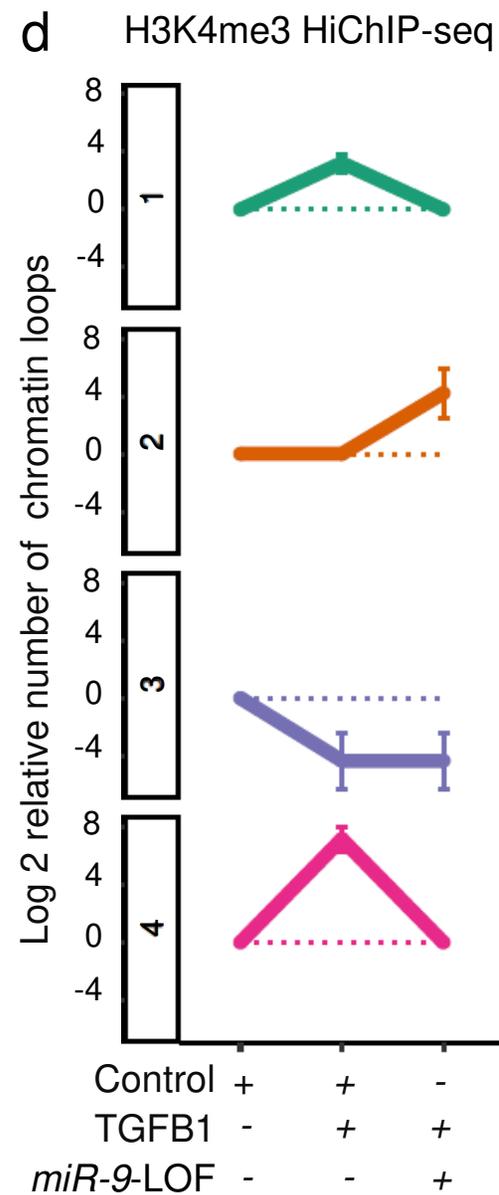
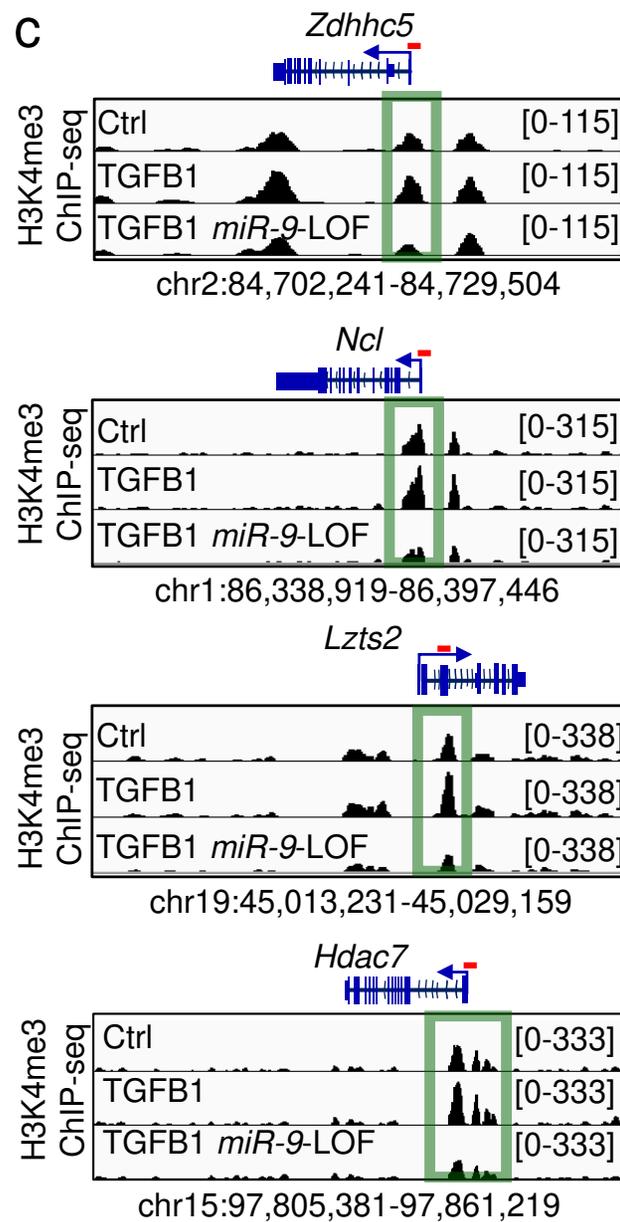
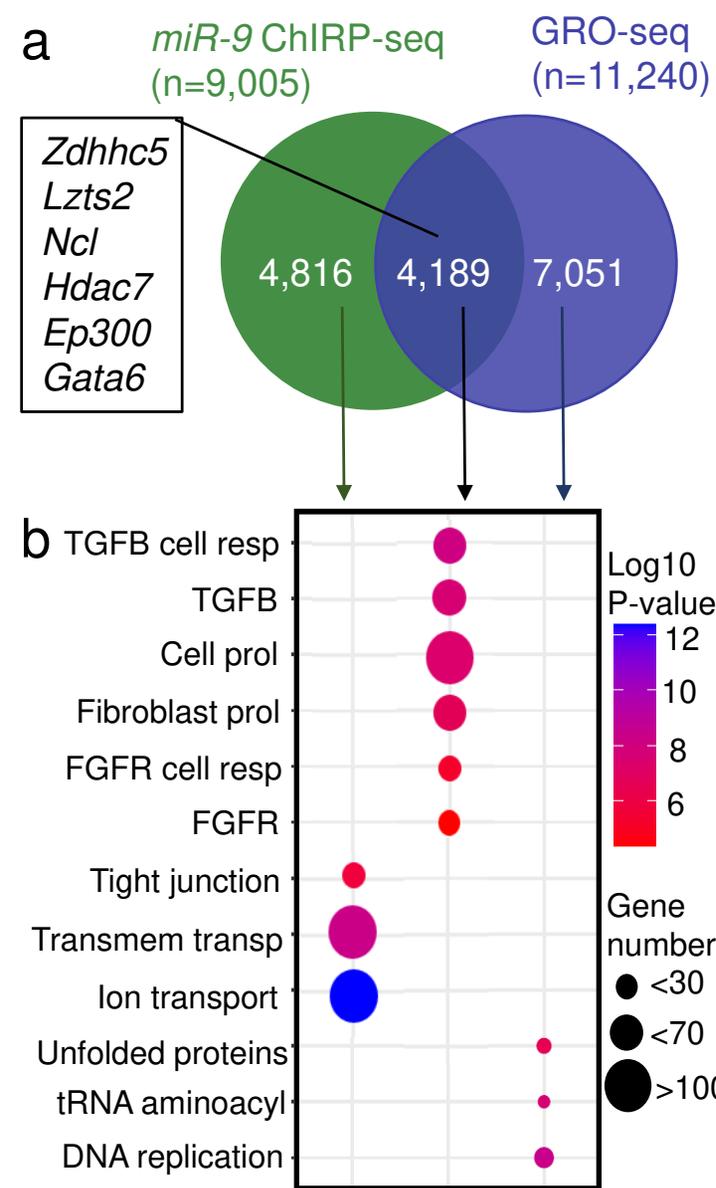


Figure 6

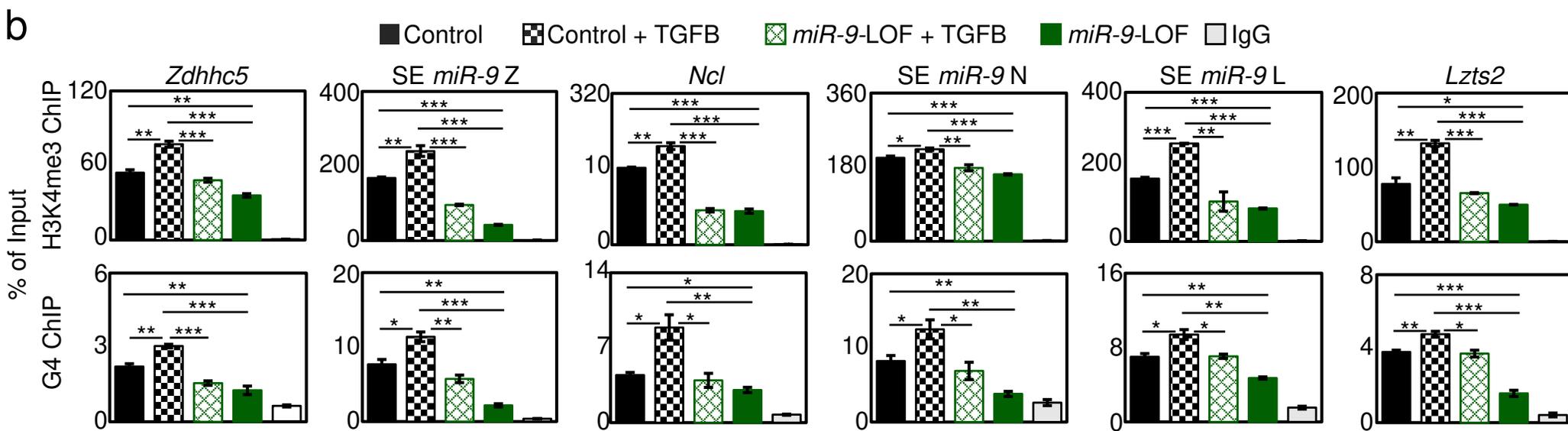
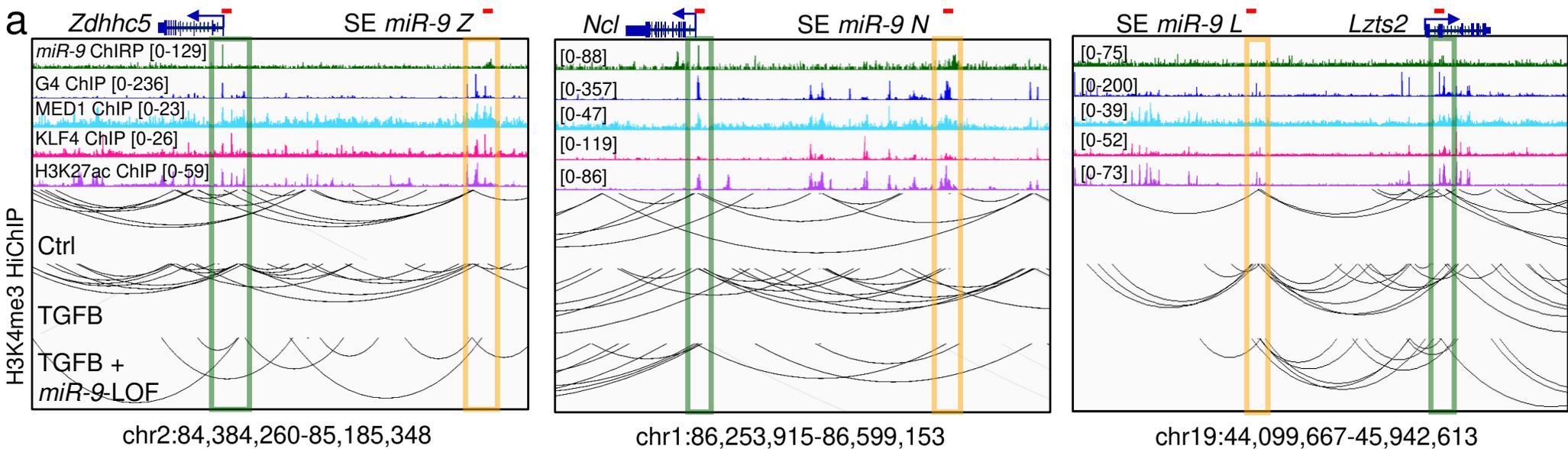


Figure 7

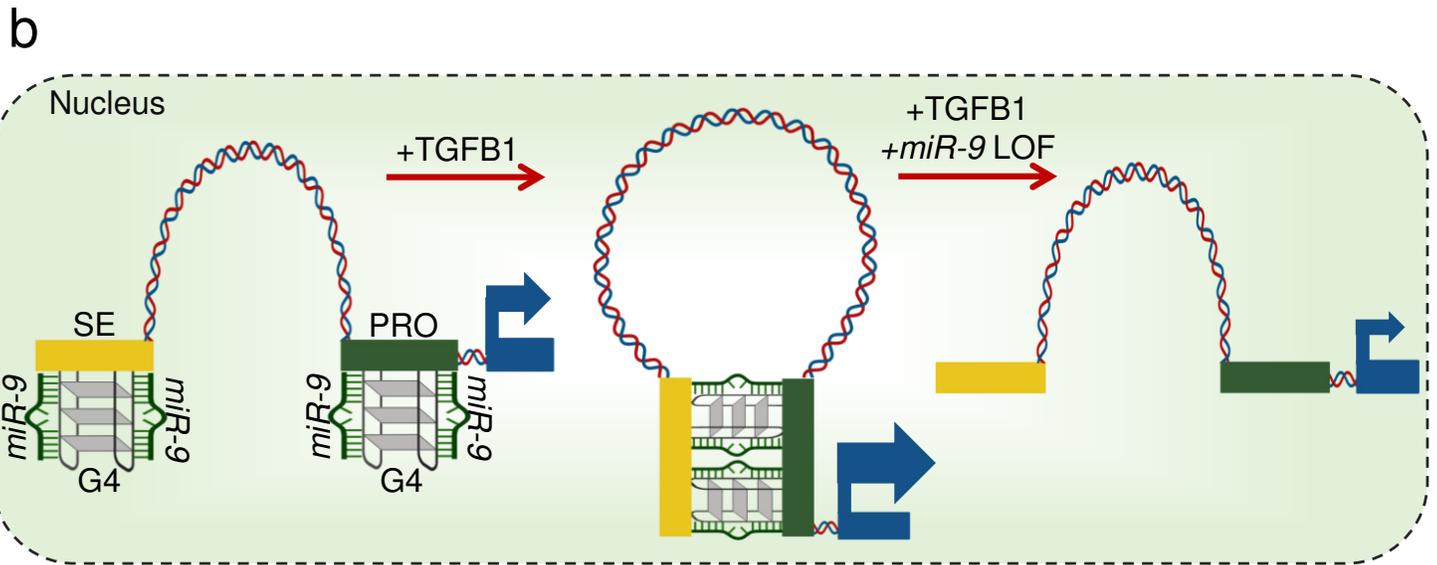
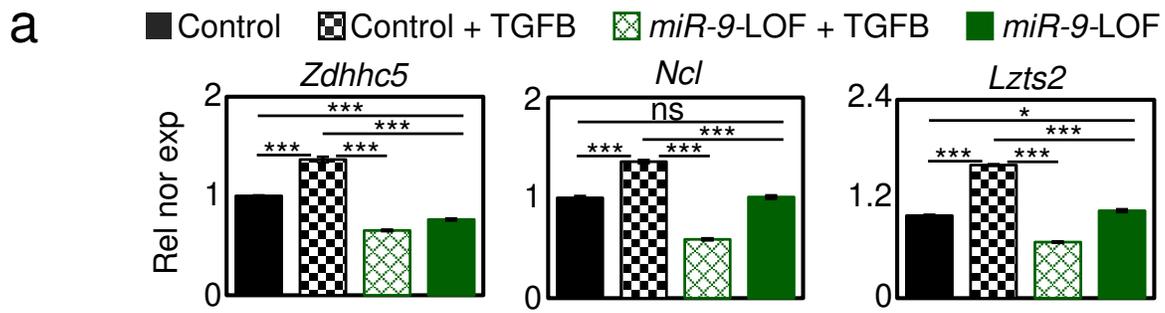


Figure 8