Transcription bodies regulate gene expression by sequestering CDK9

Martino Ugolini[1,2], Ksenia Kuznetsova[2], Haruka Oda[3], Hiroshi Kimura[3], Nadine L. Vastenhouw[1,2*]

[1]University of Lausanne (UNIL), Center for Integrative Genomics (CIG); Lausanne, Switzerland. [2]Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG); Dresden, Germany. [3]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology; Yokohama 226-8503, Japan. [*]Corresponding author: nadine.vastenhouw@unil.ch

The localization of transcriptional activity in specialized transcription bodies is a hallmark of gene expression in eukaryotic cells. It remains unclear, however, if and how they affect gene expression. Here, we disrupted the formation of two prominent endogenous transcription bodies that mark the onset of zygotic transcription in zebrafish embryos and analysed the effect on gene expression using enriched SLAM-Seq and live-cell imaging. We find that the disruption of transcription bodies results in downregulation of hundreds of genes, providing experimental support for a model in which transcription bodies increase the efficiency of transcription. We also find that a significant number of genes are upregulated, counter to the suggested stimulatory effect of transcription bodies. These upregulated genes have accessible chromatin and are poised to be transcribed in the presence of the two transcription bodies, but they do not go into elongation. Live-cell imaging shows that the disruption of the two large transcription bodies enables these poised genes to be transcribed in ectopic transcription bodies, suggesting that the large transcription bodies sequester a pause release factor. Supporting this hypothesis, we find that CDK9, the kinase that releases paused

polymerase II, is highly enriched in the two large transcription bodies. Importantly, overexpression of CDK9 in wild type embryos results in the formation of ectopic transcription bodies and thus phenocopies the removal of the two large transcription bodies. Taken together, our results show that transcription bodies regulate transcription genome-wide: the accumulation of transcriptional machinery creates a favourable environment for transcription locally, while depriving genes elsewhere in the nucleus from the same machinery.

RNA polymerase II (RNAPII) and transcription factors are often concentrated in specialized transcription bodies [1–10], and it has been suggested that these can bring together multiple genes [11–14]. This has led to the hypothesis that transcription bodies increase the efficiency of transcription by promoting the biomolecular interactions underlying gene expression [15–17]. Analysis of gene expression in artificially induced condensates supports this model [18]. To understand how transcription bodies affect gene expression *in vivo*, however, it is necessary to analyse endogenous transcription bodies. This has been difficult because transcription bodies are often small, short-lived, and numerous. Here, we use the onset of transcription in zebrafish embryos to overcome this problem. In zebrafish embryos, development is initially driven by maternally loaded RNA and protein (Fig. 1a) [19]. Transcription gradually begins during a process called zygotic genome activation (ZGA). Before transcriptional activity can be seen throughout the nucleus, it is confined to two micron-sized transcription bodies (Fig. 1b) [9,20–23], nucleated by the *mir430* locus, which contains multiple copies of the *mir430* gene [21,24]. These bodies are enriched for RNAPII Serine 5 and Serine 2 phosphorylation (Ser5P, Ser2P), which mark the initiating and elongating form of RNAPII, respectively, as well as the transcription factors Nanog and Sox19b [9,20–22].

51 To investigate the role of transcription bodies in transcription, we generated a fish line in which
52 the *mir430* locus is deleted [9], resulting in the specific disruption of the two *mir430* transcription
53 bodies, as observed by the loss of Ser2P signal (Fig. 1c). *miR430* is required for downregulating
54 maternally loaded transcripts, and its absence results in severe developmental phenotypes [24,25].
55 Thus, effects on gene expression and development in the *mir430* mutant may be caused by the
56 loss of the *mir430* transcription bodies, or the loss of the mature microRNA MiR430. To
57 uncouple these two components and allow us to focus on the first, we supplemented mutant
58 embryos with the mature form of microRNA MiR430 (Fig. 1d) [24], which restored degradation
59 of mRNAs containing MiR430 binding sites (Fig. 1e and Extended Data Fig. 1). This rescued
60 the delay in epiboly (Fig. 1f-g) and the aberrant expression pattern of *goosecoid* (Extended
61 Data Fig. 2), which are characteristic for the loss of MiR430 [24,25]. Embryos supplemented with
62 mature microRNAs appeared normal, with the appearance of the Kupffer's vesicle (Fig. 1h),
63 and overall developmental progression (Fig. 1i and Extended Data Fig. 3) being
64 indistinguishable from wild type (WT) and heterozygous siblings, and embryos developed into
65 fertile fish. Thus, the disruption of two large transcription bodies does not result in obvious
66 developmental defects, and we can use it to analyse the effect of transcription bodies on gene
67 expression.
68
69 To assess the effect of the specific disruption of *mir430* transcription bodies on transcriptional
70 activity, we developed eSLAM-Seq, which uses a combination of protocols to label, enrich,
71 and detect nascent transcripts (Extended Data Fig. 4 and Methods), because the large amounts
72 of maternally loaded RNA in early embryos (Fig. 1a) masks nascent RNAs in total RNA
73 sequencing approaches. We used eSLAM-Seq to compare gene expression between WT and
74 *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage, when transcriptional activity is largely
75 restricted to the *mir430* transcription bodies in WT embryos (Fig. 1b). We observed a

76 significant number of genes that were downregulated in the absence of the *mir430* transcription
77 bodies (242) and a higher number of genes that were upregulated (716) (Fig. 2a-b and Extended
78 Data Table 1). Plotting the genomic distribution of genes that were up- and downregulated
79 showed that both classes of genes are distributed across the genome (Fig. 2c). We conclude
80 that the absence of two large transcription bodies has a widespread effect on transcription.
81
82 The identification of hundreds of up- and downregulated genes in *mir430*$^{-/-}$ + inj MiR430 *vs.*
83 WT embryos at the 256-cell stage contrasts with the observation that these embryos develop
84 normally (Fig. 1). To reconcile these two observations, we generated eSLAM-Seq data for WT
85 and *mir430*$^{-/-}$ + inj MiR430 embryos at the 1024-cell, Oblong, and Sphere stage and compared
86 it with the data obtained at the 256-cell stage. The expression levels of genes that are
87 misregulated in *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage (with a median of 3.5
88 $\log_2$ fold difference with WT for the upregulated genes, and 2.9 $\log_2$ fold for the downregulated
89 genes; Fig. 2d) gradually recover during consecutive developmental stages, reaching WT
90 expression levels at Sphere stage (with a median of 0.03 $\log_2$ fold difference with WT for the
91 upregulated genes, and 0.07 $\log_2$ fold for the downregulated genes; Fig. 2d). Looking at all
92 genes, we find that overall fewer genes are misregulated at later stages, from 716 up and 242
93 down at 256-cell to 51 up and 30 down at Sphere stage (Extended Data Fig. 5). We conclude
94 that in the absence of *mir430* transcription bodies, gene expression is misregulated, but that
95 this effect is short-lived as expression levels return to WT levels in a timeframe of a couple of
96 cell cycles. This likely explains the lack of obvious developmental defects in *mir430*$^{-/-}$ + inj
97 MiR430 embryos.
98
99 We then further analysed the genes that are misregulated at the 256-cell stage. The
100 downregulation of genes suggests that transcription bodies facilitate transcription of genes

101 other than *mir430* itself, consistent with work in which transcribed genes were shown to
102 colocalize [12–14,26], as well as the proposed role of transcription bodies in increasing
103 transcriptional efficiency [18,27,28]. We hypothesize that the genes that are downregulated in
104 *mir430*$^{-/-}$ + inj MiR430 embryos localize to the *mir430* transcription bodies in WT embryos
105 and are positively affected by this localization. The observation that downregulated genes
106 return to WT expression levels over time suggests that transcription bodies that form during
107 later stages in *mir430*$^{-/-}$ + inj MiR430 embryos (Extended Data Fig. 6) may compensate for the
108 loss of *mir430* transcription bodies. Here, we focus on the upregulated genes. Their high
109 number suggest that the *mir430* transcription bodies sequester components of the
110 transcriptional machinery, preventing genes elsewhere in the nucleus from being transcribed.
111
112 To investigate this sequestration model, we characterized the genes that are upregulated in
113 response to *mir430* transcription body removal. We first asked when these genes are induced
114 in WT embryos. We analysed their expression at the 1024-cell, Oblong, and Sphere stages, and
115 compared this to their expression at the 256-cell stage. We found that most of the genes (81%)
116 are activated during genome activation in WT embryos (Fig. 3a), indicating that loss of the
117 *mir430* transcription bodies specifically activates the expression of genes that are about to be
118 activated, with 62% (446/716) being advanced by 30 minutes (at the 1024-cell stage in WT,
119 Fig. 3a). When focusing on the genes that are advanced by 30 minutes, we observe a good
120 correlation between their degree of upregulation between *mir430*$^{-/-}$ + inj MiR430 and WT at
121 256-cell stage, and between 1024-cell and 256-cell stage in WT (Fig. 3b and Extended Data
122 Fig. 7). This further supports the notion that the effect on transcription activation is specific.
123
124 We then investigated the chromatin accessibility state of the genes that are upregulated. We
125 analysed an ATAC-Seq dataset we had previously generated in WT embryos [29] and found that

126 the transcription start sites of genes that are upregulated in the *mir430*$^{-/-}$ + inj MiR430 are
127 highly accessible in WT embryos (Fig. 3c), being as accessible as those of the most highly
128 expressed genes (Fig. 3c-d), even though they are hardly expressed (Fig. 3e). This suggests
129 that potentially sequestered factor(s) function downstream of transcription-factor binding and
130 chromatin opening. Consistent with this, motif analysis of the promoters of upregulated genes
131 did not identify transcription factor motifs that were both sufficiently enriched compared to
132 unexpressed genes and present in a sufficiently high fraction of the upregulated genes to
133 explain our observations (Extended Data Fig. 8). Even motifs of Nanog, Pou5f3 and Sox19b,
134 transcription factors that have been shown to be required for the activation of transcription
135 during ZGA in zebrafish [9,29–32] did not fulfil these two requirements. Taken together, our results
136 suggest that the presence of the *mir430* transcription bodies prevents a significant number of
137 genes that are poised for activation from going into elongation.
138
139 This reminded us of the observation that WT embryos from the 256-cell stage onward, in
140 addition to the *mir430* transcription bodies in which both transcription initiation and elongation
141 can be observed, display transcription bodies that are positive for transcription initiation, but
142 not elongation [9]. We were previously unable to explain this observation but now hypothesized
143 that the *mir430* transcription bodies might sequester a pause release factor, causing genes in
144 these transcription bodies to be arrested at the initiation phase. If this were true, these
145 transcription bodies would be predicted to acquire elongation signals in nuclei of *mir430*$^{-/-}$ +
146 inj MiR430 embryos. To investigate this, we injected 1-cell WT and *mir430*$^{-/-}$ + inj MiR430
147 embryos with fluorescently labelled antigen-binding fragments (Fabs) recognizing RNAPII-
148 Ser5P (transcription initiation) and RNAPII-Ser2P (transcription elongation), as described
149 before [9,22,23]. Consistent with our hypothesis, we saw more Ser2P-positive transcription bodies
150 in *mir430*$^{-/-}$ + inj MiR430 than in WT embryos during the 512- and 1024-cell stages (Fig. 3f,

151 and Extended Data Fig. 9-10). Quantification of initiation-only, initiation/elongation, and
152 elongation-only transcription bodies confirmed that more transcription bodies go into
153 productive elongation in *mir430*$^{-/-}$ + inj MiR430 than in WT embryos (with a maximum of 15%
154 elongation-positive transcription bodies in WT versus 33% in *mir430*$^{-/-}$ + inj MiR430 at the
155 256-cell stage, and a comparable difference at the 1024-cell stage; Fig. 3g). Thus, the removal
156 of the *mir430* transcription bodies causes transcription bodies that are arrested in the initiation
157 phase to transition to the transcription elongation phase.
158
159 Our data suggests that transcription bodies sequester a pause release factor. Pause release is
160 regulated by the pTEFb complex of which CDK9 is the catalytic subunit. We visualized CDK9
161 in WT embryos by immunofluorescence and found that it is strongly enriched in the two *mir430*
162 transcription bodies (Fig 4a). If the amount of CDK9 in the nucleus were limiting and its
163 sequestration in *mir430* transcription bodies would hamper transcription elongation elsewhere
164 in the nucleus, it would be predicted that overexpression of CDK9 in WT embryos would
165 induce the formation of ectopic transcription bodies. To directly test this, we compared the
166 occurrence of ectopic transcription bodies between WT embryos with and without injected
167 *cdk9* mRNA. We visualized RNAPII Ser2P to detect transcription bodies. To distinguish
168 transcription bodies that are nucleated by the *mir430* locus from ectopic transcription bodies,
169 we also visualized *miR430* transcripts using Morpholino Visualization of Expression (MOVIE
170 [21]). At the 128-cell stage, we detected two prominent *mir430* transcription bodies in nuclei of
171 WT embryos, as well as in WT embryos that were injected with *cdk9* mRNA (Fig. 4b and
172 Extended Data Fig. 11). In the latter, however, we detected additional transcription bodies.
173 Because these lack *miR430* MOVIE signal and we do not detect them in uninjected WT
174 embryos, we consider these to be ectopic transcription bodies. The ectopic bodies are small
175 and limited in number at the 128- and 256-cell stage, and more prominent at the 512- and 1024-

176 cell stage (Fig. 4b and Extended Data Fig. 11-14). The abundance of ectopic transcription
177 bodies in WT embryos in which CDK9 is overexpressed, phenocopies the abundance of ectopic
178 transcription bodies seen in *mir430*$^{-/-}$ + inj MiR430 embryos. The quantification of ectopic
179 transcription shows that at all four examined stages, ectopic transcription bodies are
180 significantly more abundant in nuclei in which CDK9 is overexpressed (WT + inj *cdk9*
181 mRNA), or in which the large *mir430* transcription bodies are disrupted (*mir430*$^{-/-}$ + inj
182 MiR430), than in WT embryos (Fig. 4c). This supports our hypothesis that CDK9 is
183 sequestered in *mir430* transcription bodies, limiting CDK9 availability elsewhere in the
184 nucleus.
185
186 Taken together, we show that the removal of the *mir430* transcription bodies results in a
187 genome-wide misregulation of gene expression. Many genes are downregulated, consistent
188 with a model where the bodies facilitate transcription by promoting the biomolecular
189 interactions underlying gene expression [3,16,18]. Unexpectedly, we also observed many genes
190 that are upregulated upon disruption of the *mir430* transcription bodies. We show that
191 transcription bodies sequester the pause release factor CDK9, preventing genes that are not
192 present in the transcription bodies from being transcribed (Extended Data Fig. 15). Factor
193 sequestration regulating biological processes has been described before [33,34] and a recent study
194 showed that the loss of the transcription factor Zelda in *Drosophila* embryos caused RNAPII
195 to relocate from Zelda-containing transcription bodies to the histone locus body [35]. More
196 generally, work in mouse embryos has shown that CDK9 abundance is an important factor in
197 the timing of zygotic genome activation[36]. Our work shows that endogenous transcription
198 bodies play an important role in controlling the availability of transcriptional machinery in the
199 nucleus.
200

201 **References (1-36 used in main text)**

202 1. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal
203 sites of transcription within human nuclei. *EMBO J.* **12**, 1059–1065 (1993).
204 2. Ghamari, A. *et al.* In vivo live imaging of RNA polymerase II transcription factories in
205 primary cells. *Genes Dev.* **27**, 767–777 (2013).
206 3. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity
207 enhancers. *Elife* **6**, 1–18 (2017).
208 4. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-
209 dependent condensates. *Science* **361**, 412–415 (2018).
210 5. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation
211 and gene control. *Science* **361**, eaar3958 (2018).
212 6. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions
213 that control gene transcription. *Science* **361**, eaar2555 (2018).
214 7. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation
215 Capacity of Their Activation Domains. *Cell* **175**, 1842–1855 (2018).
216 8. Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell
217 Identity Genes. *Mol. Cell* **76**, 753–766 (2019).
218 9. Kuznetsova, K. *et al.* Nanog organizes transcription bodies. *Curr. Biol.* **33**, 1–10
219 (2022).
220 10. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human
221 cells. *Science* **341**, 664–667 (2013).
222 11. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are
223 localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109**,
224 1427–1436 (1996).
225 12. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing

226 transcription. *Nat. Genet.* **36**, 1065–1071 (2004).

227 13. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a
228 transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).

229 14. Fanucchi, S., Shibayama, Y., Burd, S. & Weinberg, M. S. Chromosomal Contact
230 Permits Transcription between Coregulated Genes. *Cell* **155**, 606–620 (2013).

231 15. Sutherland, H. & Bickmore, W. A. Transcription factories: Gene expression in unions?
232 *Nat. Rev. Genet.* **10**, 457–466 (2009).

233 16. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase
234 Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).

235 17. Tsai, A., Galupa, R. & Crocker, J. Robust and efficient gene regulation through
236 localized nuclear microenvironments. *Development* **147**, dev161430 (2020).

237 18. Wei, M. T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat.*
238 *Cell Biol.* **22**, 1187–1196 (2020).

239 19. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition
240 revisited. *Dev.* **146**, dev161471 (2019).

241 20. Chan, S. H. *et al.* Brd4 and P300 Confer Transcriptional Competency during Zygotic
242 Genome Activation. *Dev. Cell* **49**, 867–881 (2019).

243 21. Hadzhiev, Y. *et al.* A cell cycle-coordinated Polymerase II transcription compartment
244 encompasses gene expression before global genome activation. *Nat. Commun.* **10**, 691
245 (2019).

246 22. Sato, Y. *et al.* Histone H3K27 acetylation precedes active transcription during
247 zebrafish zygotic genome activation as revealed by live-cell analysis. *Dev.* **146**,
248 dev179127 (2019).

249 23. Hilbert, L. *et al.* Transcription organizes euchromatin via microphase separation. *Nat.*
250 *Commun.* **12**, 1360 (2021).

251 24. Giraldez, A. J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science*
252 **308**, 833–838 (2005).
253 25. Liu, Y. *et al.* Genetic Deletion of miR-430 Disrupts Maternal-Zygotic Transition and
254 Embryonic Body Plan. *Front. Genet.* **11**, 853 (2020).
255 26. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans
256 interactions regulate olfactory receptor choice. *Nature* **565**, 448–453 (2019).
257 27. Ma, L. *et al.* Co-condensation between transcription factor and coactivator p300
258 modulates transcriptional bursting kinetics. *Mol. Cell* **81**, 1682–1697 (2021).
259 28. Sabari, B. R. Biomolecular Condensates and Gene Activation in Development and
260 Disease. *Dev. Cell* **55**, 84–96 (2020).
261 29. Pálfy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility
262 established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish
263 genome activation. *PLoS Genet.* **16**, e1008546 (2020).
264 30. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during
265 the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
266 31. Leichsenring, M., Maes, J., Mossner, R., Driever, W. & Onichtchouk, D. Pou5f1
267 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–
268 1009 (2013).
269 32. Miao, L. *et al.* The landscape of pioneer factor activity reveals the mechanisms of
270 chromatin reprogramming and genome activation. *Mol. Cell* **82**, 986–1002 (2022).
271 33. Klosin, A. *et al.* Phase separation provides a mechanism to reduce noise in cells.
272 *Science* **367**, 464–468 (2020).
273 34. Harigaya, Y. *et al.* Selective elimination of messenger RNA prevents an incidence of
274 untimely meiosis. *Nat. 2006 4427098* **442**, 45–50 (2006).
275 35. Huang, S. K., Whitney, P. H., Dutta, S., Shvartsman, S. Y. & Rushlow, C. A. Spatial

276 organization of transcribing loci during early genome activation in Drosophila. *Curr.*
277 *Biol.* **31**, 5102–5110 (2021).
278 36. Oqani, R. K., Kim, H. R., Diao, Y. F., Park, C. S. & Jin, D. I. The CDK9/Cyclin T1
279 subunits of P-TEFb in mouse oocytes and preimplantation embryos: A possible role in
280 embryonic genome activation. *BMC Dev. Biol.* **11**, 1–9 (2011).
281 37. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages
282 of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
283 38. Truett, G. E. *et al.* Preparation of PCR-quality mouse genomic DNA with hot sodium
284 hydroxide and tris (HotSHOT). *Biotechniques* **29**, 52–54 (2000).
285 39. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells
286 using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39**, 6475–
287 6488 (2011).
288 40. Stasevich, T. J. *et al.* Regulation of RNA polymerase II activation by histone
289 acetylation in single living cells. *Nature* **516**, 272–275 (2014).
290 41. Kimura, H. & Yamagata, K. Visualization of epigenetic modifications in
291 preimplantation embryos. *Methods Mol. Biol.* **1222**, 127–147 (2015).
292 42. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish
293 embryos. *Nat. Protoc.* **3**, 59–69 (2008).
294 43. Heyn, P. *et al.* The earliest transcribed zygotic genes are short, newly evolved, and
295 different across species. *Cell Rep.* **6**, 285–292 (2014).
296 44. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics.
297 *Nat. Methods* **14**, 1198–1204 (2017).
298 45. Holler, K. *et al.* Spatio-temporal mRNA tracking in the early zebrafish embryo. *Nat.*
299 *Commun.* **12**, 3358 (2021).
300 46. Bhat, P. *et al.* SLAMseq resolves the kinetics of maternal and zygotic gene expression

301 in early zebrafish embryogenesis. *bioRxiv* (2022) doi:10.1101/2022.06.01.494399.
302 47. Schofield, J. A., Duffy, E. E., Kiefer, L., Sullivan, M. C. & Simon, M. D. TimeLapse-
303 seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding.
304 *Nat. Methods* **15**, 221–225 (2018).
305 48. Reichholf, B. *et al.* Time-Resolved Small RNA Sequencing Unravels the Molecular
306 Principles of MicroRNA Homeostasis. *Mol. Cell* **75**, 756–768 (2019).
307 49. Duffy, E. E. *et al.* Tracking Distinct RNA Populations Using Efficient and Reversible
308 Covalent Chemistry. *Mol. Cell* **59**, 858–866 (2015).
309 50. Duffy, E. E., Canzio, D., Maniatis, T. & Simon, M. D. Solid phase chemistry to
310 covalently and reversibly capture thiolated RNA. *Nucleic Acids Res.* **46**, 6996–7005
311 (2018).
312 51. Sun, W. & Chen, W. Metabolic Labeling of Newly Synthesized RNA with 4sU to in
313 Parallel Assess RNA Transcription and Decay. in *Methods in molecular biology* vol.
314 1720 25–34 (Humana Press, New York, NY, 2018).
315 52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina
316 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
317 53. Zhang, Y., Park, C., Bennett, C., Thornton, M. & Kim, D. Rapid and accurate
318 alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.*
319 **31**, 1290–1295 (2021).
320 54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
321 2078–2079 (2009).
322 55. Broad Institute. Picard. http://broadinstitute.github.io/picard/ (2022).
323 56. Boileau, E., Altmüller, J., Naarmann-de Vries, I. S. & Dieterich, C. A comparison of
324 metabolic labeling and statistical methods to infer genome-wide dynamics of RNA
325 turnover. *Brief. Bioinform.* **22**, bbab219 (2021).

326 57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program
327 for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
328 58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
329 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
330 59. Bailey, T. L. & Grant, C. E. SEA: Simple Enrichment Analysis of motifs. *bioRxiv*
331 (2021) doi:10.1101/2021.08.23.457422.
332 60. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
333 analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
334 61. R Core Team. R: A Language and Environment for Statistical Computing. (2014).
335 62. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer International
336 Publishing, 2016). doi:10.1007/978-3-319-24277-4.
337 63. Yin, T., Cook, D. & Lawrence, M. ggbio: an R package for extending the grammar of
338 graphics for genomic data. *Genome Biol.* **13**, 1–14 (2012).
339 64. Herrgen, L., Schröter, C., Bajard, L. & Oates, A. C. Multiple embryo time-lapse
340 imaging of zebrafish development. *Methods Mol. Biol.* **546**, 243–254 (2009).
341 65. Boothe, T. *et al.* A tunable refractive index matching medium for live imaging cells,
342 tissues and model organisms. *Elife* **6**, e27240 (2017).
343 66. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat.*
344 *Methods* **9**, 676–682 (2012).
345 67. Katrukha, E. ComDet. https://github.com/ekatrukha/ComDet (2022).
346

347 **Acknowledgements**

348 We thank the members of the Vastenhouw lab for technical support, helpful feedback, and
349 stimulating discussions. We thank the Pauli and Ameres labs for technical advice on SLAM-
350 Seq. We thank Maciej Kerlin, Aleksandar Vještica, and Edlyn Wu for comments on the

351 manuscript, and the following facilities and services for their support: MPI-CBG - fish facility,
352 light microscopy facility, and scientific computing, UNIL - genomic technologies facility,
353 cellular imaging facility and the bioinformatics competence center. M.U. was supported by the
354 Boehringer Ingelheim Fonds. Research in N.L.V's laboratory was supported by the Max
355 Planck Society, the University of Lausanne, a European Research Council Consolidator Grant
356 (101003023), the Volkswagen Foundation (94773), and the German Research Foundation (VA
357 1209/2-1). Research in H.K.'s laboratory was supported by the Japan Society for the Promotion
358 of Science KAKENHI (JP18H05527, JP21H04764 and JP20K06484).
359

360 **Author contributions**

361 Conceptualization: M.U., N.L.V. Methodology: M.U., K.K., H.U. Investigation: M.U., K.K.
362 Data analysis: M.U. Writing – original draft: M.U. and N.L.V. Writing – reviewing & editing:
363 all authors. Funding acquisition: M.U., N.L.V., H.K. Supervision: N.L.V.
364

365 **Competing interests**

366 Authors declare that they have no competing interests.
367

368 **Data and materials availability**

369 Raw sequencing data will be uploaded on Gene Expression Omnibus (GEO) prior to
370 publication. Raw imaging data are available upon request. All other data is available in the
371 main text or the extended data.

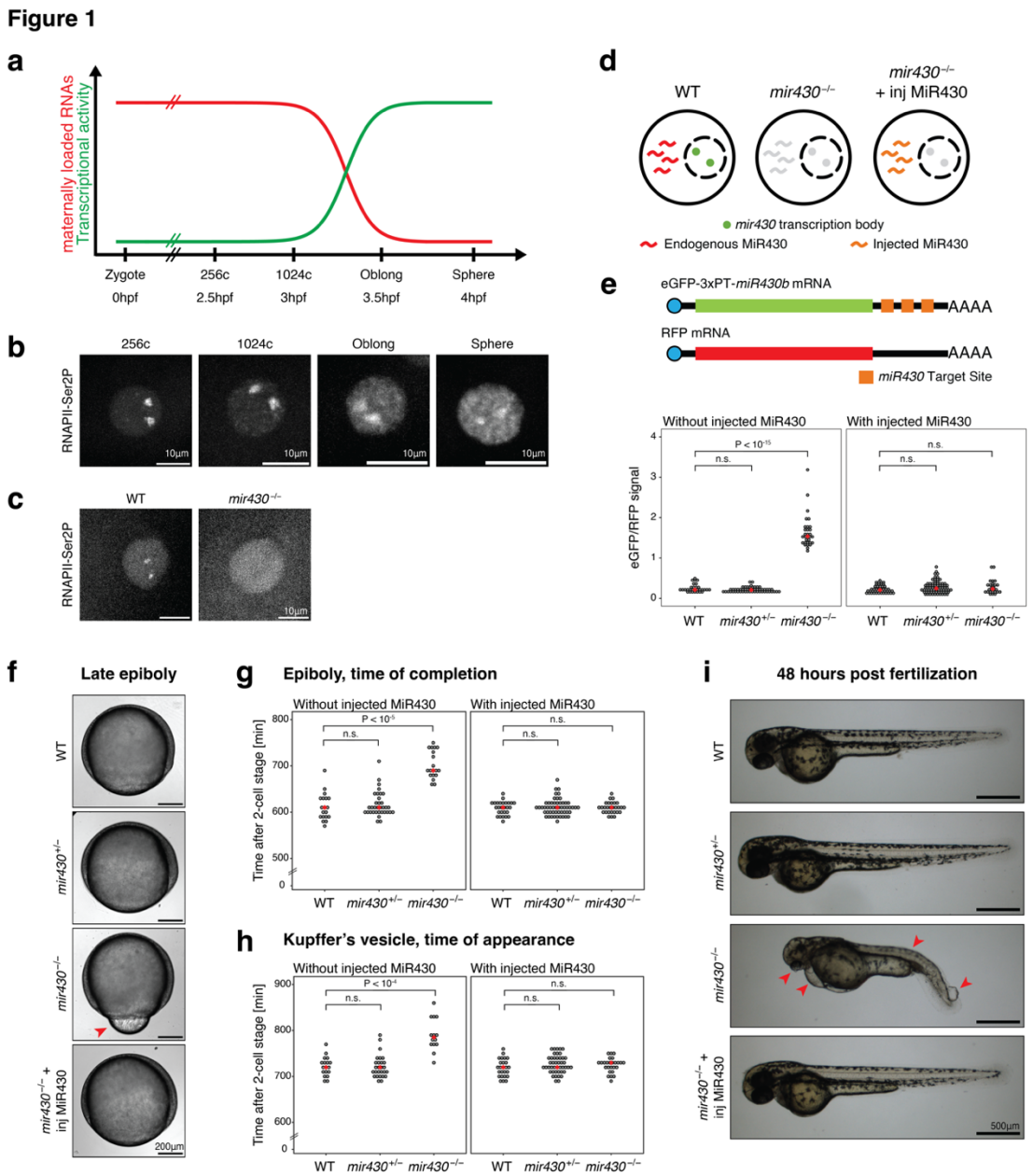
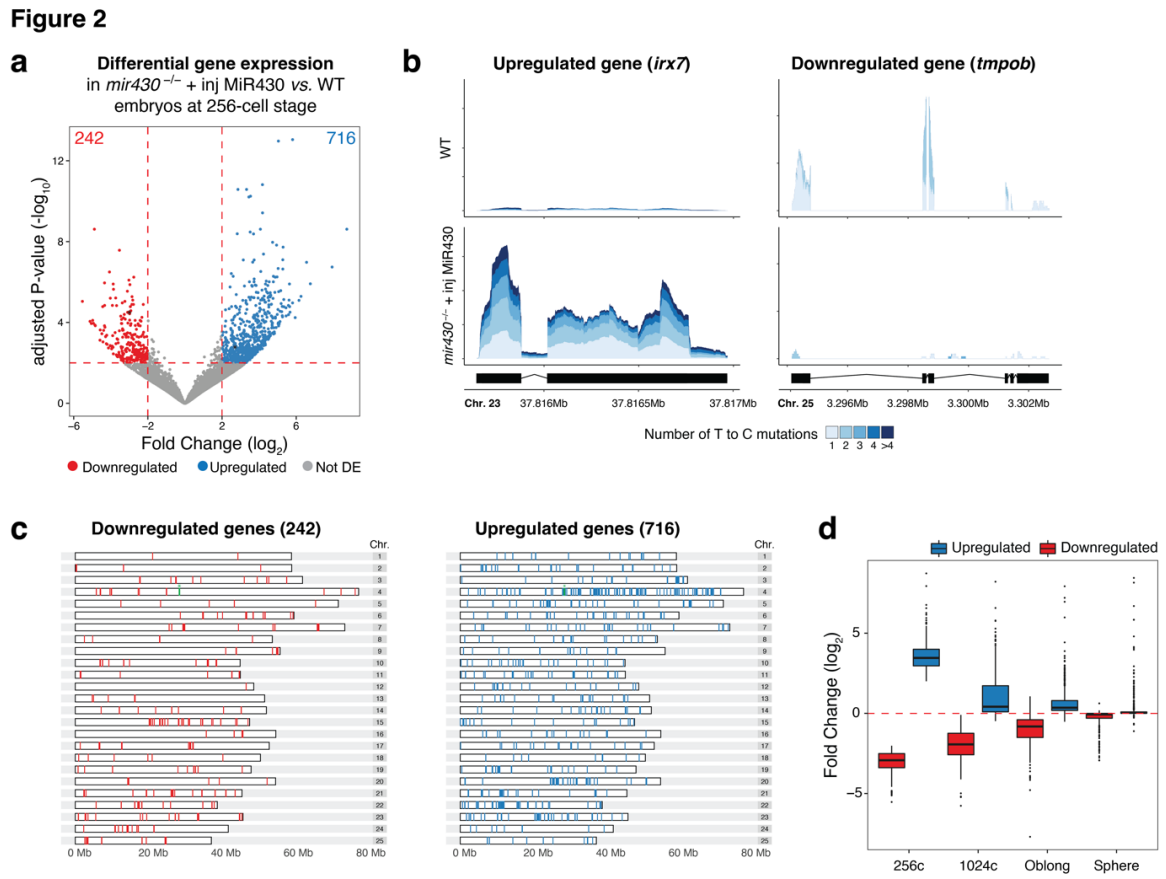


Figure 1. Disruption of *mir430* transcription bodies does not substantially impact development. a. Schematic representation of the maternal-to-zygotic transition in zebrafish embryos. Hpf, hours post fertilization. **b.** Visualization of elongating RNAPII (Ser2P) with antigen-binding fragments (Fabs) in WT embryos. Shown are representative micrographs of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. **c.** Visualization of elongating RNAPII (Ser2P) with antigen-binding fragments (Fabs) in WT and *mir430*$^{-/-}$ embryos at the 128-cell stage. Shown are representative micrographs of individual

380 nuclei, extracted from a spinning disk confocal microscopy time lapse. **d.** Schematic
381 representation of a nucleus in WT, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. **e.** Rescue of
382 *miR430* activity as assessed by comparing the expression of eGFP encoded by an mRNA with
383 three perfect target sites for *miR430,* and RFP encoded by an mRNA without such sites, in
384 different genotypes at 24hpf. Normalized eGFP signal in WT, *mir430*$^{+/-}$ and *mir430*$^{-/-}$ embryos
385 without (left) and with (right) injected MiR430. N=3, 27≤n≤74. See Extended Data Fig. 1 for
386 images. **f.** Rescue of *miR430* activity as assessed by epiboly progression. Shown are
387 representative micrographs of embryos at late epiboly stage in different genotypes. The
388 misregulation of yolk internalization in *mir430*$^{-/-}$ embryos is indicated. **g.** Time at which
389 epiboly is completed in different genotypes without (left), and with (right) injected MiR430
390 RNA. N=3, 19≤n≤53. **h.** Time at which Kupffer's vesicle appears in different genotypes
391 without (left), and with (right) injected MiR430. N=3, 14≤n≤44. **i.** Larvae at 48 hours post
392 fertilization for different genotypes. Representative micrographs are shown. The malformation
393 of trunk morphology and eye, the development of heart oedema, and the appearance of blisters
394 at the tail tip in *mir430*$^{-/-}$ embryos are indicated.
395



396

397 **Figure 2. Disruption of *mir430* transcription bodies causes widespread misregulation of**
398 **gene expression. a.** Volcano plot showing up- and downregulated genes in *mir430*$^{-/-}$ + inj
399 MiR430 *vs.* WT embryos at 256-cell stage. Genes whose coverage is shown in Fig. 2b are
400 shown in black. **b.** Coverage plot of *irx7* (upregulated) and *tmpob* (downregulated) in WT and
401 *mir430*$^{-/-}$ + inj MiR430. The single strata visualize the labelling degree of the reads. **c.**
402 Distribution of up- and downregulated genes across the genome. The *mir430* locus on
403 chromosome 4 is shown in green and labelled by an asterisk. **d.** Difference in average gene
404 expression between *mir430*$^{-/-}$ + inj MiR430 and WT embryos across stages for those genes that
405 were identified to be differentially expressed at 256-cell stage.

406

Figure 3

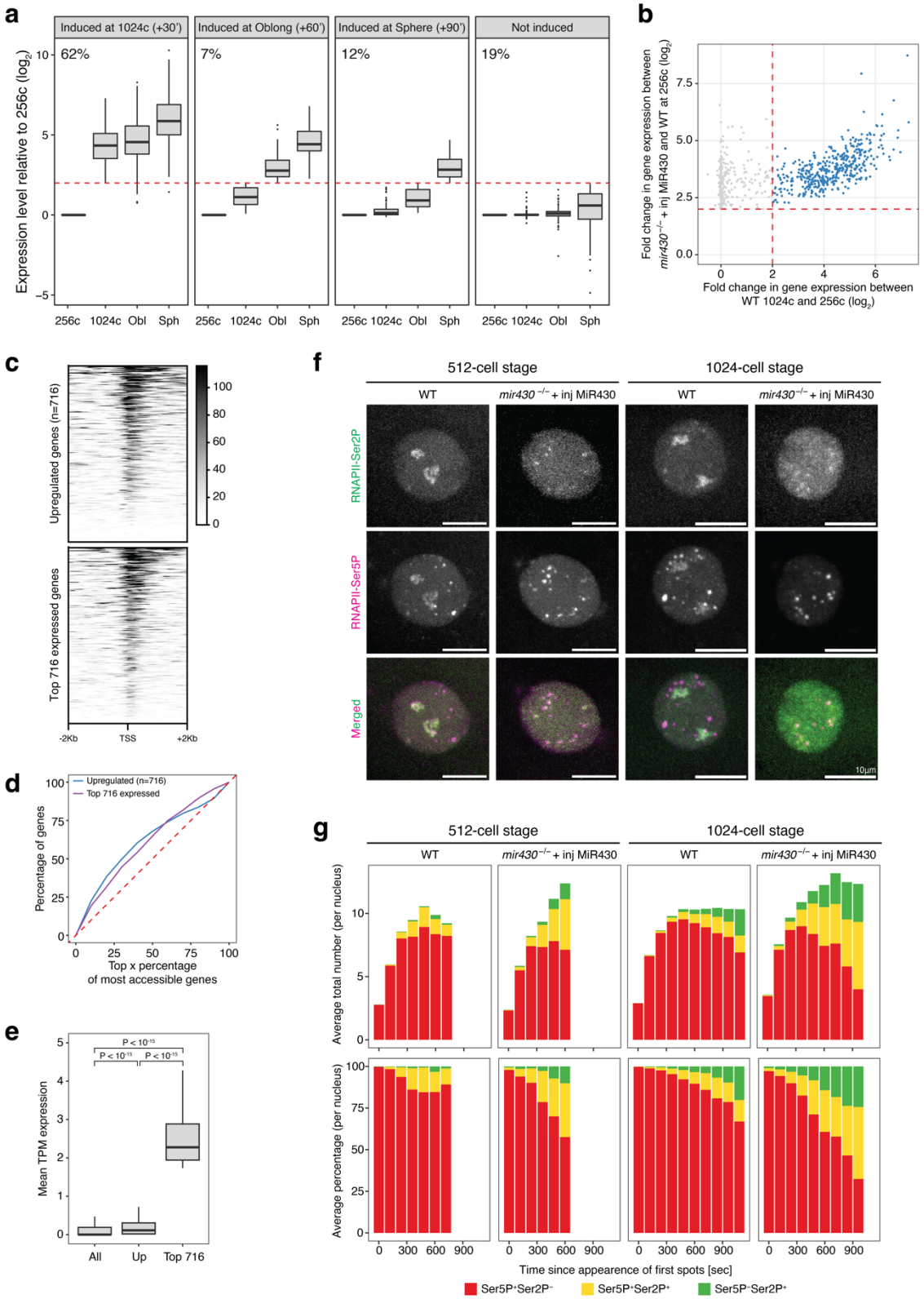
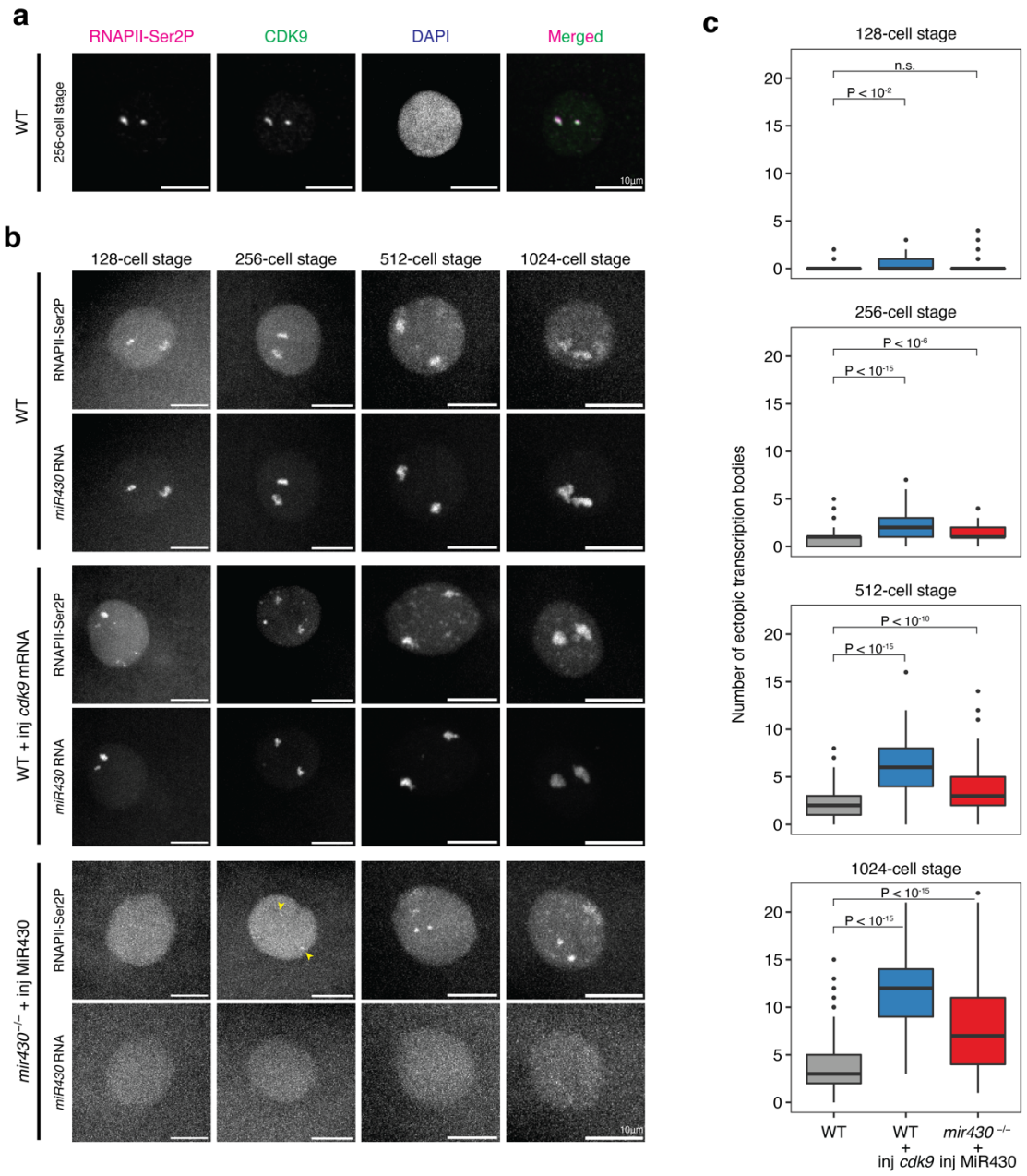


Figure 3. Loss of *mir430* transcription bodies causes premature gene activation. a. Time of induction in WT embryos for the 716 genes that are upregulated in *mir430*$^{-/-}$ + inj MiR430

410 *vs.* WT embryos (at 256-cell stage). Expression at 256-cell is used as a reference. The genes
411 are split in four groups based on when they are induced. The percentage of genes in each group
412 is indicated. **b.** Scatterplot representing the fold change in expression between *mir430*$^{-/-}$ + inj
413 MiR430 and WT embryos at 256-cell on the y-axis, and between 1024-cell and 256-cell in WT
414 embryos on the x-axis. All upregulated genes are shown in grey, genes induced at 1024-cell
415 (62%) are shown in blue. **c.** Heatmap of chromatin accessibility at 256-cell of the 716 genes
416 that are upregulated in *mir430*$^{-/-}$ + inj MiR430 embryos compared to WT embryos at 256-cell
417 stage, and – for comparison – the 716 most expressed genes in WT embryos at 256-cell stage.
418 Genes are ranked by accessibility. **d.** Graph showing the fraction of promoters of the 716
419 upregulated (in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT at 256-cell stage), and the 716 most expressed
420 genes (in WT at 256-cell) amongst the x% most accessible promoters at 256-cell in WT
421 embryos. **e.** Average expression level of the 716 upregulated genes (*mir430*$^{-/-}$ + inj MiR430 *vs.*
422 WT at 256-cell stage), the 716 most expressed genes, and all annotated genes, at the 256-cell
423 stage, in WT. Outliers are not shown. **f.** Visualization of RNAPII-Ser2P and RNAPII-Ser5P
424 with antigen-binding fragments (Fabs) in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 512- and
425 1024-cell stage. Shown are representative micrographs of individual nuclei. See Extended Data
426 Fig. 9-10 for complete cell cycles. **g.** Quantification of Ser5P-positive/Ser2P-negative (red),
427 Ser5P-positive/Ser2P-positive (yellow) and Ser5P-negative/Ser2P-positive (green)
428 transcription bodies during the cell cycle at 512-cell and 1024-cell, as shown in f. Number
429 (upper panel) and percentage (lower panel) are shown.
430

Figure 4



431

432 **Figure 4. CDK9 sequestration in *mir430* transcription bodies inhibits transcription**
433 **elongation elsewhere. a.** Visualization of CDK9 and RNAPII-Ser2P by immunofluorescence
434 in WT nuclei at 256c. Nuclei are labelled with DAPI. **b.** Visualization of RNAPII-Ser2P with
435 antigen-binding fragments (Fabs), and *miR430* RNA with MOVIE in WT, WT + inj *cdk9*
436 mRNA, and *mir430*$^{-/-}$ + inj MiR430 embryos across stages. Shown are representative
437 micrographs of individual nuclei, extracted from a spinning disk confocal microscopy time

438 lapse. See Extended Data Fig. 11-14 for complete cell cycles. **c.** Quantification of ectopic
439 transcription bodies as shown in b. N=3, 8≤n≤168. See Methods for a detailed description of
440 quantification.
441

442 **Methods**

443 Zebrafish and molecular biology approaches

444 *Zebrafish husbandry and manipulation*

445 Zebrafish were maintained and raised under standard conditions, and according to Swiss
446 regulations (canton Vaud, license number VD-H28). Wild type (TLAB) and mutant (*mir430*$^{-/-}$
447) [9] embryos were collected immediately upon fertilization and allowed to develop to the desired
448 stage at 28°C. *mir430*$^{-/-}$ embryos were rescued by injecting 2nL of rescue solution (10µM of
449 MiR430a, MiR430b and MiR430c duplex in 1x siRNA solution (300mM KCl, 30mM HEPES,
450 1mM $MgCl_2$, brought to pH 7.5 with KOH)), as described before [9,24]. If the chorion was
451 chemically removed, this was done by incubating embryos right after fertilization in 1.5 mg/mL
452 Pronase E (Sigma-Aldrich, #107433) for approximately five minutes. Developmental stage
453 was determined by morphology (according to [37]), the distance between nuclei, as well as cell
454 size, cell cycle duration, and synchronicity [9]. Fabs were injected into the cell at 0.44ng/embryo
455 (αRNAPII-Ser2P-Alexa488), or 0.66ng/embryo (αRNAPII-Ser5P-Cy5). Lissamine-labelled
456 anti-*miR430* morpholino (5'-TCTACCCCAACTTGATAGCACTTTC-3'-Lissamine, Gene
457 Tools) was injected into the cell at 14 fmole/embryo [21]. s^4UTP (TriLink, tebu-bio GmbH, #N-
458 1025-1, diluted to 50mM in 10mM Tris-HCl pH 7.5) was injected into the yolk at
459 50nM/embryo.

460

461 *Genotyping*

462 Embryos (live and fixed) were genotyped using the HotSHOT method [38].

463

464 *mRNA production and injection*

465 eGFP with three perfect *miR430* target sites in its 3'UTR (eGFP-3xPT-miR430b) was inserted
466 in the pCS2$^+$ plasmid (Addgene) as described before [24]. The coding sequence of *cdk9*

467 (ENSDART00000065859.8) was amplified from cDNA using SuperScript III (Invitrogen,
468 #18080085) and oligodT primers, and inserted in the pCS2^{+} plasmid (Addgene) using
469 restriction-enzyme cloning. mRNA was synthesized using the mMessage mMACHINE
470 Transcription Kit (ThermoFisher, #AM1340) and purified with the RNeasy MinElute Cleanup
471 Kit (QIAGEN, #74204). mRNA was injected into embryos at the 1-cell stage in the following
472 amounts: 50pg of eGFP mRNA, and 100pg of RFP or *cdk9* mRNA.
473
474 *Preparation of antigen-binding fragments (Fabs)*
475 Fluorescently labelled Fabs specific to RNAPII-Ser5P and RNAPII-Ser2P were prepared from
476 monoclonal antibodies specific to RNAPII-Ser5P and Ser2 phosphorylation [39–41]. Monoclonal
477 antibodies were digested with Ficin (ThermoFisher Scientific), and Fabs were purified through
478 protein A-Sepharose columns (Cytiva) to remove Fc and undigested IgG. After passing through
479 desalting columns (PD MiniTrap G25; Cytiva) to substitute the buffer with PBS, Fabs were
480 concentrated up to >1 mg/ml using 10 k cut-off filters (Amicon Ultra-0.5 10 k; Merck). Fabs
481 were conjugated with Alexa Fluor 488 (Sulfodichlorophenol Ester; ThermoFisher Scientific)
482 or Cy5 (N-hydroxysuccinimide ester monoreactive dye; Cytiv) to yield ~1:1 dye:protein ratio.
483 After the buffer substitution with PBS, the concentration was adjusted to ~1 mg/ml.
484
485 *Probe production*
486 A DIG-labeled antisense probe against *goosecoid* (*gsc*) was generated by *in vitro* transcription
487 from a linearized plasmid using 10x Transcription buffer (Roche, #11465384001), 10x DIG
488 RNA labelling mix (Roche, #11277073910), T7 RNA polymerase (Roche, #10881775001) and
489 RNase OUT (Invitrogen, #10777019). Probe was purified with the RNeasy MinElute Cleanup
490 Kit (QIAGEN, #74204), diluted in Hybridization buffer and stored at -20°C.
491

492 *Whole mount in situ hybridization*

493 Whole mount *in situ* hybridization (WM-ISH) was performed as described before [42], with some
494 differences. In brief, embryos were fixed at the desired stage in 4% (w/v) PFA in PBS pH 7.4
495 overnight at 4°C. Embryos were then manually dechorionated with sharp forceps (Dumont No.
496 5) in PBS with 0.1% Tween-20 (PBST) and dehydrated by rinsing them in 50% and then 100%
497 methanol in PBST on a nutator. At this point, embryos could be stored for several weeks at -
498 20°C. To continue, embryos were gradually rehydrated on a nutator, incubated in Hybridization
499 buffer (50% formamide, 1.3x SSC pH 5.0, 5mM EDTA, 50µg/mL torula yeast RNA, 0.2%
500 Tween-20, 0.1% SDS, 100µg/mL Heparin) for 1 hour at 70°C, and incubated with anti-
501 *goosecoid* DIG-labelled probe overnight at 70°C. The next day, the probe was washed off with
502 Hybridization buffer and embryos were incubated for 15 minutes in 50% (v/v) Hybridization
503 buffer in TBST (25mM Tris-HCl pH 7.8, 137mM NaCl, 2.7mM KCl, 0.1% Tween-20) at 70°C.
504 After rinsing with TBST, embryos were incubated for 1 hour in TBST and then in Blocking
505 Buffer (2% blocking reagent in maleic acid buffer pH 7.5 (Roche, #11096176001) and 10%
506 sheep serum (Sigma Aldrich, #S2263) in TBST) at room temperature on a nutator. anti-DIG
507 antibody (Roche, #11093274910) was added in a dilution of 1:2.000 and embryos incubated
508 overnight at 4°C on a nutator. The next day, embryos were washed with TBST and then NTMT
509 (100mM NaCl, 100mM Tris-HCl pH 9.5, 50mM MgCl2, 0.1% Tween-20), and stained in
510 NTMT-NBT-BCIP solution (4.5µL/mL of NBT (Roche, #11383213001) and 3.5µL/mL of
511 BCIP (Roche, #11383221001) in NTMT) in the dark. Once sufficiently stained, embryos were
512 washed with PBST, and then in 100%, 50% and 30% methanol in PBST at room temperature
513 on a nutator. Embryos were stored at 4°C in PBST for up to a few days before imaging.

514

515 *Immunofluorescence*

516 Dechorionated embryos were fixed at the desired stage in 4% (w/v) PFA in PBS pH 7.4
517 overnight at 4°C. Embryos were then dehydrated by incubating them in 30%, 50% and then
518 100% methanol in PBS and incubated overnight at -20°C. To continue, embryos were gradually
519 rehydrated. All following steps were performed while rocking. Embryos were washed multiple
520 times with PBST (0.8% Triton-X 100 in PBS) and incubated in blocking buffer (10% NGS
521 (Sigma-Aldrich, #G6767) and 4% BSA (Sigma Aldrich, #A8022) in PBST) for four hours at
522 room temperature. Primary antibodies were added in 1% NGS in PBST overnight at 4°C (anti-
523 CDK9, clone C12F7, Cell Signaling Technology, monoclonal IgG antibody generated in
524 rabbit, 1:200 dilution; anti-RNAPII-Ser2P, clone H5, ab24758, monoclonal IgM antibody
525 generated in mouse, Abcam, 1:1000 dilution). The next day, embryos were washed with PBST
526 and then incubated with secondary antibodies (anti-rabbit-IgG AlexaFluor488-conjugated,
527 Thermo Scientific, #A-21206, 1:500 dilution; anti-mouse-IgM AlexaFluor594-conjugated,
528 Thermo Scientific, #A-21044, 1:500 dilution) in 1% NGS in PBST overnight at 4°C. Embryos
529 were then washed with PBST, manually deyolked using sharp forceps (Dumont No. 5), and
530 mounted in Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories, VC-H-
531 1200-10) before imaging them on a Spinning Disk microscope.
532
533 Enriched SLAM-Sequencing (eSLAM-Seq)
534 *Rationale*
535 Early embryos contain large amounts of maternally loaded RNA [19] which mask the nascent
536 RNAs in total RNA sequencing approaches. We therefore developed a method to sensitively
537 detect nascent transcripts. Methods to enrich nascent transcripts have previously been
538 described. Typically, an uracil/UTP analog (EU/s^4UTP) is injected at the 1-cell stage and is
539 incorporated into newly made RNAs. Such labelled transcripts can then be biotinylated and
540 enriched by pull-down [20,43]. SLAM-Seq is another approach in which nascent transcripts are

541 labelled by s^4UTP incorporation, but here the label is used to convert 4-thio-uracils to
542 carboxyamido-methylated 4-thio-uracils, which will be misread as a cytosine during reverse
543 transcription, introducing specific point mutations that enable the identification of reads
544 derived from labelled transcripts by sequencing [44–46]. We combined these two approaches in
545 enriched SLAM-Seq (eSLAM-Seq), which is conceptually similar to TT-TimeLapse-Seq and
546 TT-SLAM-Seq [47,48]. Specifically, we used a combination of s^4UTP labelling [43] with improved
547 biotinylation and enrichment chemistry [49,50], followed by SLAM-Seq chemistry.
548
549 *Sample preparation*
550 Because of the light-sensitive nature of s^4UTP, all steps until the treatment of RNA with
551 iodoacetamide (IAA) were performed in the dark. When staging embryos at the
552 stereomicroscope, for example, we used the lowest light-setting that allowed us to still see the
553 embryos, and a UV filter was set between the injection plate and the light source to protect the
554 s^4UTP. Successive steps were performed protecting the samples from the light by using low
555 light settings in the laboratory room, by working inside a cardboard box, and by covering the
556 samples with aluminium foil.
557 Synchronized s^4UTP-injected embryos were collected and left to develop to the desired stage.
558 They were collected in as little 0.3x Danieau's Buffer as possible, homogenized in 500µL
559 Qiazol (QIAGEN, #79306) with a 5mm stainless steel bead (QIAGEN, #69989) using a
560 TissueLyser (30 seconds at 30Hz) and stored at -80°C. 100 embryos (256-cell stage) or 50
561 embryos (1024-cell, Oblong and Sphere stage) were used. Chloroform was added and RNA
562 was extracted by phase separation, using a MaXtract tube (QIAGEN, #129046), after which it
563 was precipitated with isopropanol supplemented with 0.2mM DTT and glycogen. The pellet
564 was washed with 80% ethanol supplemented with 1mM DTT, air-dried, and resuspended in
565 40µL of 1mM DTT. DNase treatment was performed in 100µL of 1x DNAse reaction buffer

566 with 4U of TURBO DNase (ThermoFisher, AM2238) for 30 minutes at 37°C. The volume was
567 brought to 250µL with nuclease-free water and RNA was extracted with 25:24:1 Phenol-
568 Chloroform-Isoamyl alcohol (Invitrogen, #15593-031) using a MaXtract tube, precipitated
569 with isopropanol supplemented with 0.2mM DTT, glycogen and sodium acetate. The pellet
570 was washed with 80% ethanol supplemented with 1mM DTT, air-dried, and resuspended in
571 40µL of nuclease-free water.
572
573 Labelled RNA was biotinylated in 100µL of 20mM HEPES pH 7.4, 1mM EDTA pH 8.0,
574 10ng/µL MTSEA-biotin-XX (Biotium, #90066) and 20% DMF) by incubation at room
575 temperature for 30 minutes while rotating. The volume was brought to 250µL with nuclease-
576 free water and RNA was extracted with 24:1 Chloroform-Isoamyl alcohol (Sigma-Aldrich,
577 #227056) using a MaXtract tube, and precipitated with isopropanol supplemented with
578 glycogen and NaCl. The pellet was washed with 80% ethanol, air-dried, and resuspended in
579 50µL of nuclease-free water.
580 To enrich for biotinylated RNA, Dynabeads MyOne Streptavidin C1 beads (Invitrogen,
581 #65001) were used. Per sample, 10µL of beads were washed twice in nuclease-free water and
582 twice in high-salt wash buffer (100mM Tris-HCl pH 7.4, 10mM EDTA pH 8.0, 1M NaCl,
583 0.05% Tween-20), incubated for at least 1 hour at room temperature in bead-blocking buffer
584 (40ng/µL glycogen in high-salt wash buffer) and washed twice with high-salt wash buffer.
585 RNA was denatured at 65°C for 10 minutes, incubated for 5 minutes on ice and then added to
586 the blocked beads. After incubation for 15 minutes while rotating at 30 rpm, the supernatant
587 was removed and beads were washed 3 times with high-salt wash buffer, twice with 5.9M
588 guanidinium chloride (with 5-minute incubation at room temperature between consecutive
589 washes), once with TE buffer and three times with TE buffer (with 5-minute incubation at 55°C
590 between consecutive washes). Bead-bound RNA was finally released by adding 25µL of

591 freshly made elution buffer (100mM DTT, 20mM HEPES pH 7.4, 1mM EDTA pH 8.0,
592 100mM NaCl, 0.05% Tween-20) twice. Volume was brought to 100µL with nuclease-free
593 water and RNA was precipitated with isopropanol supplemented with 0.2mM DTT, glycogen
594 and NaCl. The pellet was washed with 80% ethanol supplemented with 1mM DTT, air-dried,
595 and resuspended in 20µL of 1mM DTT.
596 Iodoacetamide-mediated alkylation was performed in 10mM IAA (freshly dissolved in 100%
597 ethanol, BioUltra, #I1149), 50mM $NaPO_4$ pH 8.0 and 50% DMSO for 15 minutes at 50°C.
598 Samples were then transferred on ice and the reaction stopped by quickly adding 1.4µL of 1M
599 DTT. RNA was precipitated with 100% ethanol supplemented with glycogen and sodium
600 acetate overnight. The pellet was washed with 80% ethanol, air-dried, and resuspended in
601 8.5µL of nuclease-free water.
602 Libraries were prepared with the TruSeq stranded mRNA Kit (Illumina, #20020594) and
603 TruSeq RNA UD Indexes (Illumina, #20022371). The polyA enrichment step was not
604 performed. Fragmentation was performed for 1 minute at 94°C. Paired end 150bp sequencing
605 was performed on an Illumina NovaSeq 6000 System, loading each library on two lanes
606 (Illumina Reagent Kit v1.5). Sequencing data were demultiplexed with bcl2fastq 2.20.
607
608 *Quality controls*
609 To assess the integrity of total RNA (Extended Data Fig. 4c, left panel), 1µg/sample was diluted
610 in 2x RNA Gel Loading Dye (Thermo Scientific, #R0641), denatured for 90 seconds at 85°C,
611 chilled for 5 minutes on ice and loaded onto an agarose gel alongside RiboRuler High Range
612 RNA Ladder (ThermoFisher, #SM1823). The presence of two prominent bands (rRNA) are an
613 indicator of RNA integrity.
614 To test the specificity of s^4UTP biotinylation (Extended Data Fig. 4c, right panel), the RNA
615 was then transferred from the gel onto a Nylon Hybond-N+ membrane (Amersham,

616 #RPN203B) by capillary transfer using a Whatman Turboblotter in 20X SSC according to
617 manufacturer's instructions overnight. After drying, the membrane was crosslinked twice with
618 120mJ/cm^2 using a UV Stratalinker 1800 (Stratagene). The membrane was then incubated for
619 20min in Blocking Buffer (10% SDS and 1mM EDTA in 1xPBS pH 7.5), for 1h in 1:10.000
620 Streptavidin-HRP conjugate (Perkin, #NEL750) in Blocking buffer, 2x for 10min in 10% SDS
621 in 1xPBS pH 7.5, 2x for 10min in 1% SDS in 1xPBS pH 7.5, and 2x for 10min in 0.1% SDS
622 in 1xPBS pH 7.5, all while nutating [51]. Finally, the signal was visualized with ECL Western
623 Blotting Detection Reagents (Amersham, #RPN2106) according to manufacturer's
624 instructions, and documented with a Fusion FX (Vilber).
625
626 Bioinformatic analyses
627 *eSLAM-Seq*
628 To map the raw sequencing data we concatenated Fastq files that came from the same library
629 but that were run on different lanes with the 'cat' command (GNU core utilities, linux terminal),
630 trimmed them using Trimmomatic (Paired-end mode, LEADING:3 TRAILING:3
631 SLIDINGWINDOW:4:15 MINLEN:36) [52], and mapped them using HISAT-3N (--base-change
632 T,C --repeat --rna-strandness RF --fr --no-discordant) [53], which is a HISAT variant specifically
633 designed for methods like SLAM-Seq that introduce specific point-mutations. Uniquely
634 mapped fragments were filtered with Samtools [54] and mate information was fixed with Picard
635 FixMateInformation (--IGNORE_MISSING_MATES FALSE) [55].
636 To distinguish between naturally occurring SNPs and SLAM-Seq-specific mismatches, we first
637 identified naturally occurring SNPs by analysing four IAA-minus samples (one per stage) in
638 which the iodoacetamide treatment – and therefore the base conversion – was not performed.
639 The bam files were converted to a suitable format with the Samtools mpileup function and then
640 SNPs were identified by Varscan pileup2snp (--min-coverage 20 --min-reads2 5 --min-avg-

641 qual 15 --min-var-freq 0.25 --p-value 0.01). The union of the SNPs that were identified was
642 used as a consensus list. Then the splbam.py script from pulseR [56] was used to select fragments
643 with at least one SLAM-Seq-specific mismatch.
644 To assess gene expression levels of annotated genes, raw read counts were obtained using
645 FeatureCounts (-p --countReadPairs -B --minOverlap 10 -s 2 -t exon -g gene_id) [57].
646 Normalization, Principal Component Analysis (PCA) and call for differential expression (alpha
647 = 0.01 and $\log_2$(Fold Change) threshold = 2) was performed using DESeq2 [58]. When
648 performing PCA, only the top 500 genes (selected by highest variance in expression levels)
649 were considered. When counting the number of differentially expressed genes, mitochondrial
650 and *mir430*-coding genes were excluded.
651
652 *Motif analysis*
653 To assess the presence and enrichment of specific motifs in promoters of upregulated genes,
654 we started with the non-redundant JASPAR 2022 Vertebrate core database. From this, only
655 motifs whose cognate transcription factors are expressed during ZGA (Riboprofiling dataset,
656 >10rpkm, [30]) were kept. The Nanog motif [29] and Pou5f1 motif [31] were added, because it is
657 known that these factors are important during zebrafish ZGA. Next, we extracted promoter
658 sequences (4000bp centred around the TSS) for upregulated genes as well as non-expressed
659 genes (less than 10 raw reads in total across all samples, background group). Motif enrichment
660 was then assessed using SEA (MEME Suite, [59]). To estimate depletion of motifs, the
661 background and input sample were inverted. Only motifs with a P-value < 0.05 and an E-value
662 ≤ 10 were considered.
663
664 *Promoter accessibility analysis*

665 To analyse promoter accessibility of the upregulated genes in WT, we used an ATAC-Seq
666 dataset that we previously generated [29]. Replicates were merged and the signal in 4000bp
667 centred around the TSS was plotted using Deep Tools [60]. Promoter signal was used to rank all
668 promoters by accessibility and to assess their enrichment amongst the most accessible
669 promoters. The 716 most expressed (mean TPM) genes (excluding mitochondrial genes and
670 *mir430*-coding genes) were used as a comparison.
671
672 *Graph plotting for eSLAM-Seq*
673 Plots were generated using R version 4.0.5 [61] with the ggplot2 [62] and Ggbio package [63].
674
675 Microscopy
676 *Fluorescence microscopy*
677 To determine MiR430-activity using the GFP sensor assay, injected embryos were manually
678 dechorionated at 24hpf and kept in Petri dishes. Fluorescence signal was measured on an
679 OLYMPUS SZX16 stereo microscope with a MicroPublisher 5.0 RTV-R-CLR-10C colour
680 camera (5x magnification, 400ms exposure time).
681
682 *Bright field imaging*
683 For bright field time lapse imaging, a mold [64] was used to create agarose wells using as little
684 as possible 1% agarose in 0.3x Danieau's solution in an Ibidi µ-Dish 35mm, high (#81158).
685 The dish was mounted on a 28°C heated stage (by a Warner automatic temperature controller)
686 and dechorionated embryos were loaded to the single wells at the 2-cell stage. Movies with 10-
687 minute time intervals were taken on a Zeiss Axio Observer.Z1 microscope (Zeiss 5x/0.25 Fluar
688 air objective, Zeiss AxioCam MRm Monochrome CCD camera). A z-stack (400µm in 9 steps)
689 was taken for each embryo. A lid was used to reduce evaporation of buffer during imaging.

690 To assess developmental progression of live embryos past somitogenesis, snapshots of
691 manually dechorionated embryos were taken at 24hpf, 48hpf and 72hpf on a Zeiss Axio
692 Observer.Z1 microscope (Zeiss 2.5x/0.08 EC Plan-Neofluar air objective, Zeiss AxioCam ICc
693 1).
694 The gene expression pattern in fixed embryos after WM-ISH was visualized on a LEICA
695 M165C stereomicroscope with LEICA MC170 HD colour camera (4x magnification).
696 Embryos were oriented such that the dorsal region of the embryo was visible.
697
698 *Spinning Disk microscopy*
699 For imaging on a spinning disk microscope, dechorionated embryos were mounted at 32-cell
700 stage in small droplets of 1% low-melting agarose (Invitrogen, #16520-050) in 75% v/v 0.3x
701 Danieau's solution and 25% v/v Optiprep Density Gradient Medium (Sigma Aldrich, #D1556)
702 [65] on a Ibidi µ-Dish 35mm, high (#81158). Embryos were brought closer to the coverslip
703 surface by keeping the dish upside down until the agarose solidified (~20 minutes). More
704 agarose was then added to avoid drying out.
705 Transcription bodies in live embryos were imaged on a Nikon eclipse Ti2 with a Yokogava
706 CSU-W1 Spinning Disk unit at 28°C (Okolab temperature control and microscope enclosure).
707 Multi-position, z-stack (20µm in 41 steps, NIDAQ Piezo Z) time-lapse (2 minutes time
708 interval) imaging was performed with a Nikon 60x/1.2 Plan Apochromat VC water objective.
709 Images were captured in parallel on two Photometrics Prime 95B cameras with 100ms
710 exposure time at 10% laser intensity.
711 Immunofluorescence signal (CDK9, RNAPII-Ser2P and DAPI) was also imaged on a Nikon
712 eclipse Ti2 with a Yokogava CSU-W1 Spinning Disk unit. Images were taken with a Nikon
713 60x/1.2 Plan Apochromat VC water objective. Each channel was captured in series on a
714 Photometrics Prime 95B camera with 100ms exposure time at 10% laser intensity each.

715 Image processing and analysis

716 *Software used for image analysis*

717 Microscopy image handling was done using FIJI [66]. Further data processing was carried out
718 using R version 4.0.5 [61] and plots were generated with the ggplot2 package [62].

719

720 *Analysis of eGFP reporter and RFP control*

721 To assess *mir430* activity, mean fluorescence of eGFP (targeted by MiR430) and RFP (not
722 targeted, and used as a normalizer) were measured in a rectangle of equal size in the head
723 region and a background region outside of the embryo. The signal ratio was then calculated as
724 follows: $(eGFP_{head} - eGFP_{background}) / (RFP_{head} - RFP_{background})$.

725

726 *Detection and classification of mir430 and ectopic transcription bodies*

727 We identified transcription bodies by the presence of RNAPII-Ser2P signal. To classify them
728 into *mir430* and ectopic transcription bodies, we analysed *miR430* MOVIE signal. In case of
729 overlap between RNAPII-Ser2P and MOVIE signal, bodies were classified as *mir430*
730 transcription bodies. In case of no overlap between RNAPII-Ser2P and MOVIE signal, bodies
731 were classified as ectopic transcription bodies (Extended Data Fig. 16). We used FIJI plugin
732 ComDet [67], which detects spots using consistent intensity thresholding parameters that were
733 manually set and used for all experiments. We note that frames taken after the disassembly of
734 the nuclear envelope were not analysed because at that time transcription bodies also
735 disassembled.

736

737 *Detection and classification of transcription bodies by transcription state*

738 Nuclei were segmented using the RNAPII-Ser5P channel, and RNAPII-Ser5P and Ser2P
739 bodies detected using ComDet. This was used to classify transcription bodies into initiating,

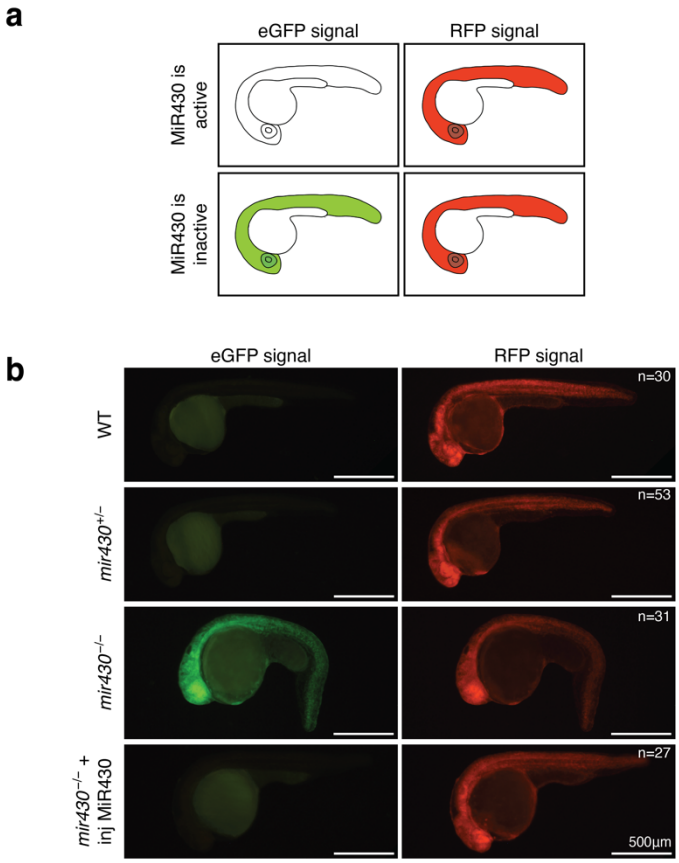
740 initiating-elongating and elongating bodies. We then identified and removed *mir430*
741 transcription bodies from the analysis based on their morphological features (early appearance
742 during cell cycle, larger size, growth behaviour) using the RNAPII-Ser2P channel. Here too,
743 frames taken after the disassembly of the nuclear envelope were not analysed. For each nucleus,
744 t=0 is set as the first frame at which spots appear, and the last time point is set as the last frame
745 before prophase. Nuclei in which spots were detected for only one frame were discarded.
746
747 Sample size and statistics
748 A minimum of 3 biological replicates (N) was acquired for each experimental condition.
749 Lower-case n (n) refers to the number of embryos or nuclei for each experimental condition.
750 Kruskal-Wallis tests were performed to determine if there is a statistically significant difference
751 between the three genotypes in the degree of MiR430 activity as determined by the eGFP-
752 Sensor assay (Fig. 1e), the time of completion of Epiboly (Fig. 1g) and the time of appearance
753 of Kupffer's vesicle (Fig. 1h). If this test was statistically significant (P-value < 0.05), pairwise
754 comparisons with Bonferroni correction were performed using a pairwise Wilcoxon rank-sum
755 test. A comparison was considered significant when adjusted P-value < 0.05. Dotplots were
756 used to display the data, and adjusted P-values were reported using WT as reference. Red
757 rhombuses represent the median.
758 A pairwise Wilcoxon rank-sum test with Bonferroni correction was performed to determine if
759 the upregulated genes and the top 716 most expressed genes are significantly higher expressed
760 compared to all genes (based on TPM expression values) (Fig. 3e). A comparison was
761 considered significant when adjusted P-value < 0.05.
762 A pairwise Wilcoxon rank-sum test with Bonferroni correction was performed at each stage to
763 determine if there is a statistically significant difference in the number of detected ectopic

764 transcription bodies between WT, *mir430*$^{-/-}$ + inj MiR430 and WT + inj *cdk9* mRNA (Fig. 4c).

765 A comparison was considered significant when adjusted P-value < 0.05.

766

Extended Data Figure 1



767

768 **Extended Data Figure 1. MiR430 activity is restored by injecting MiR430 RNA in *mir430*-**

769 **/- embryos. a.** Schematic representation of expected eGFP and RFP expression in embryos

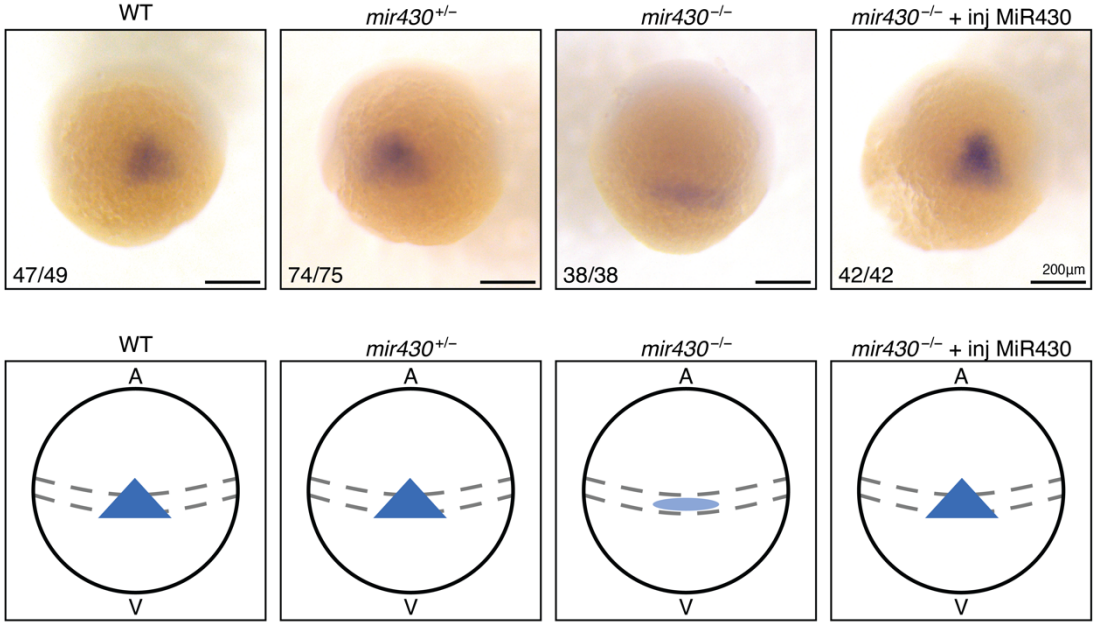
770 with active and inactive MiR430 microRNA activity. **b.** Representative micrographs showing

771 eGFP and RFP expression in WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos at

772 24hpf.

773

Extended Data Figure 2

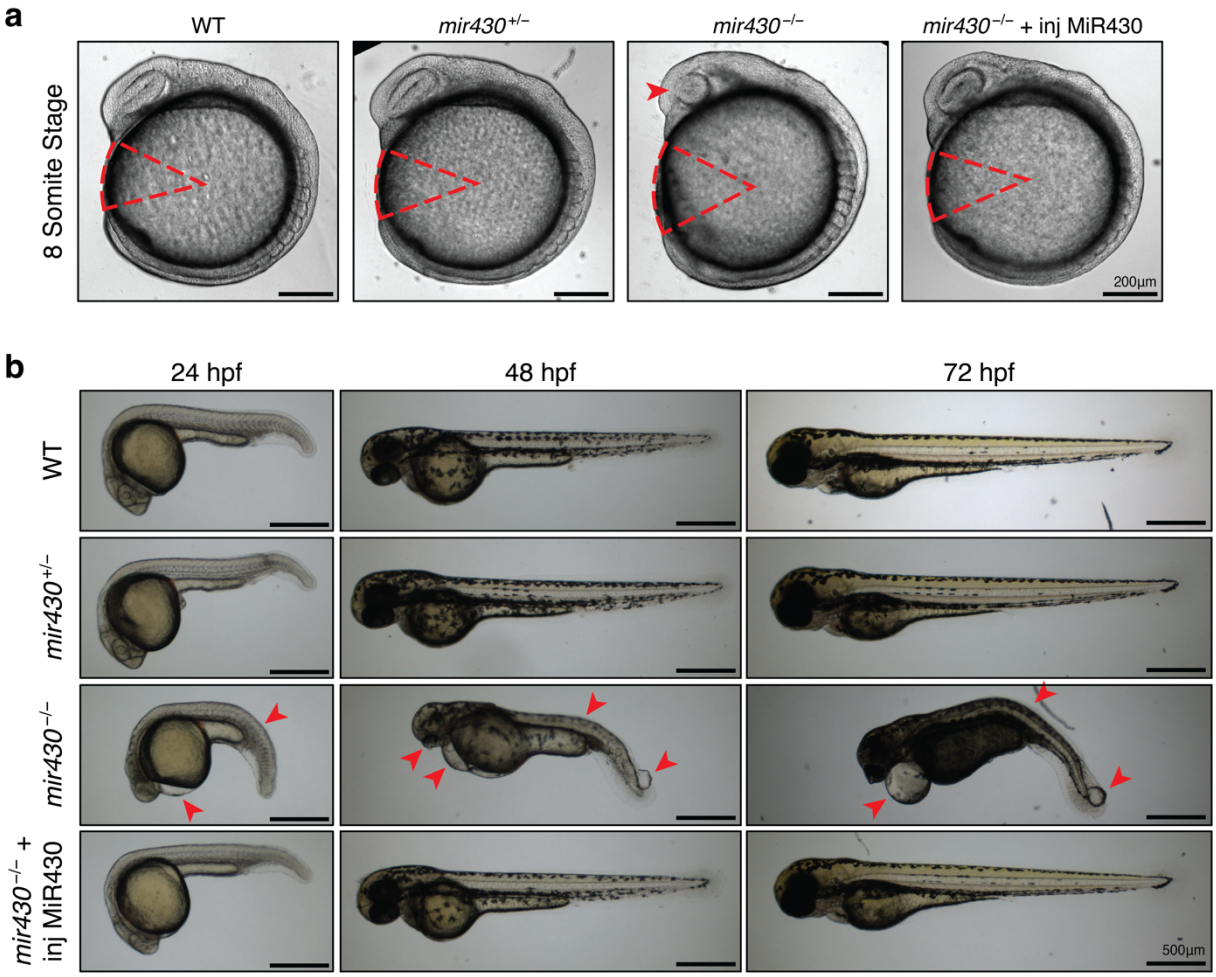


774

775 **Extended Data Figure 2. Phenotypic analysis of *mir430*$^{-/-}$ embryos with and without**
776 **injected MiR430 (*gsc* expression).** Whole mount *in situ* hybridization (WM-ISH) of
777 *Goosecoid (Gsc)* in WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos at Shield
778 stage. Representative micrographs show the embryos in dorsal view. Misexpression (reduced
779 expression levels and reduced anterior extension of expression pattern) of *Gsc* can be observed
780 in *mir430*$^{-/-}$ embryos. The fraction in the lower left corner represents the frequency with which
781 the shown phenotype was observed. A schematic representation of the phenotypes is shown
782 below.

783

Extended Data Figure 3

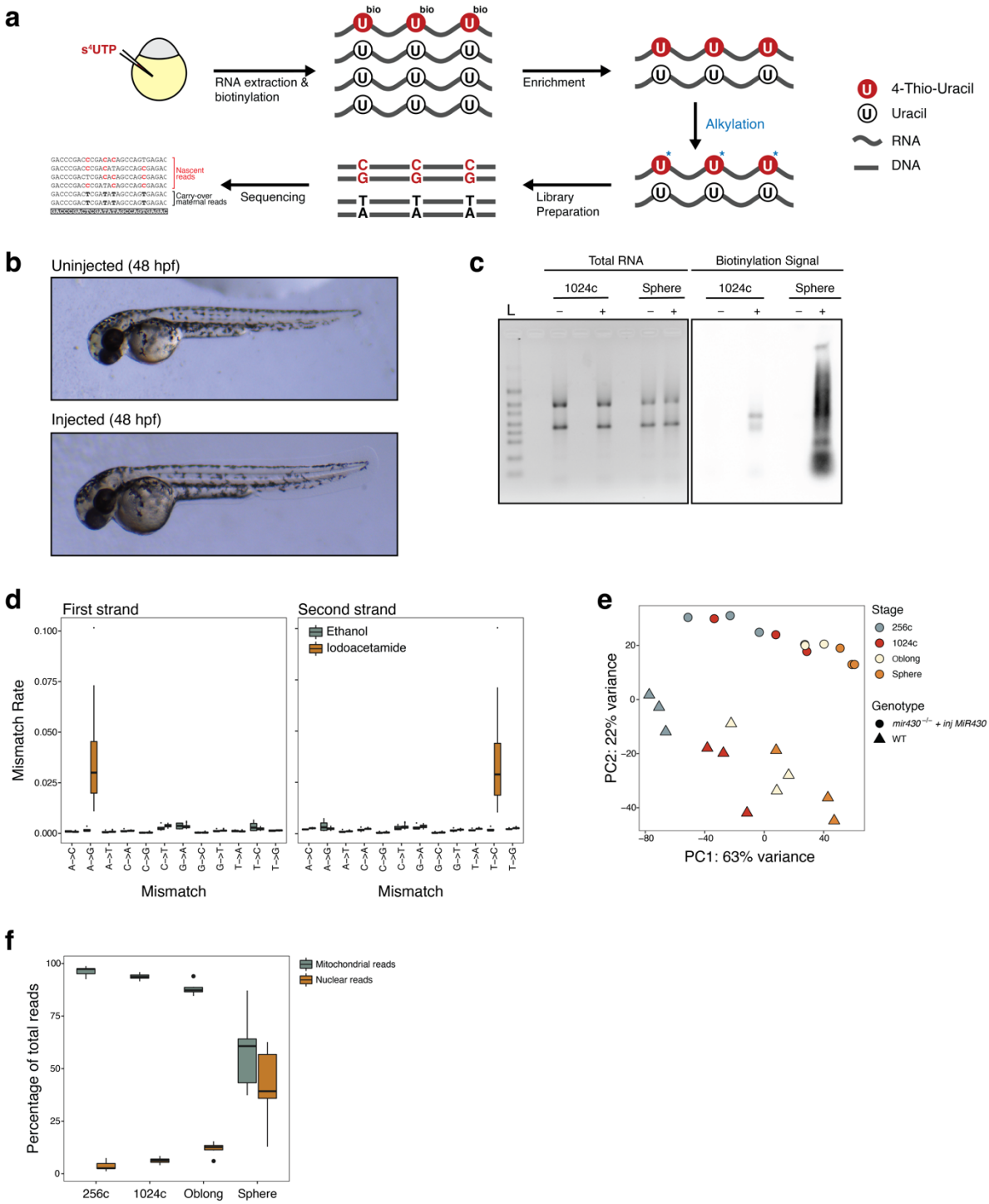


784

785 **Extended Data Figure 3. Phenotypic analysis of *mir430*$^{-/-}$ embryos with and without**
786 **injected MiR430 (developmental progress). a.** Embryos at 8-Somite stage for WT, *mir430*$^{+/-}$
787 , *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. The reduced body axis extension and the
788 malformation of the optic primordium in *mir430*$^{-/-}$ embryos are highlighted. Representative
789 micrographs are shown. **b.** Larvae at 24, 48 and 72 hours post fertilization (hpf) for WT,
790 *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. The malformation of trunk
791 morphology and of the eye, the development of heart oedema and the appearance of blisters at
792 the tail tip in *mir430*$^{-/-}$ embryos are highlighted. Representative micrographs are shown. The
793 48hpf stage is also shown in Fig. 1i.

794

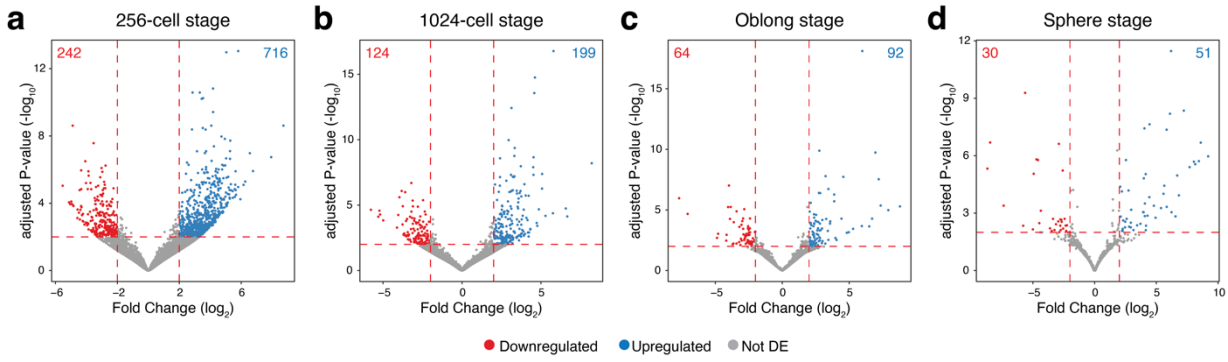
Extended Data Figure 4



Extended Data Figure 4. Establishment of enriched SLAM-Sequencing (eSLAM-Seq) in zebrafish embryos. a. Schematic representation of eSLAM-Seq protocol. See Methods for more detail. **b.** Representative micrographs of larvae that were/were not injected with 50nM/embryo of s^4UTP at the 1-cell stage. No developmental malformations were observed

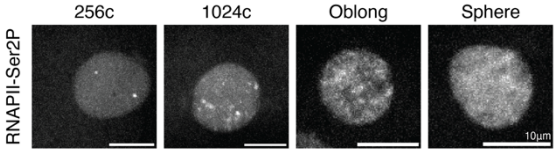
800 upon injection. **c.** Equal amounts of biotinylated and heat-denatured total RNA derived from
801 injected (+) and uninjected (-) embryos at 1024-cell and Sphere stage were run on an agarose
802 gel (left panel). The RNA was then transferred on a nylon membrane and biotinylation detected
803 by Streptavidin-HRP (right panel). Biotinylation was only observed in RNA derived from
804 injected embryos, showing that biotinylation is specific for s^4UTP-labeled RNA. L: RNA
805 Ladder. **d.** Box plot representation of mismatch rate in Iodoacetamide- (N=24) and Ethanol-
806 treated (N=4) samples. As expected, only Iodoacetamide-treated samples show a high level in
807 A>G (First strand) and T>C (Second strand) mismatch rate. **e.** Principal Component Analysis
808 (PCA) plot based on the top 500 genes (selected by highest variance in expression levels).
809 Samples separate according to their developmental stage along the first PC and according to
810 their experimental condition (WT and *mir430*$^{-/-}$ + inj MiR430) along the second PC. **f.** Box
811 plot representation of percentage of mitochondrial and nuclear reads, separated by
812 developmental stage. Most reads are of mitochondrial origin at early ZGA, as shown before [43],
813 with a gradual increase in the nuclear fraction during later stages.
814

Extended Data Figure 5



Extended Data Figure 5. The disruption of two large endogenous transcription bodies causes short-term misregulation of gene expression. Volcano plots showing up- and downregulated genes in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT embryos at 256-cell **(a)**, 1024-cell **(b)**, Oblong **(c)** and Sphere **(d)** stage. Panel A is also shown in Fig. 2a.

Extended Data Figure 6



822

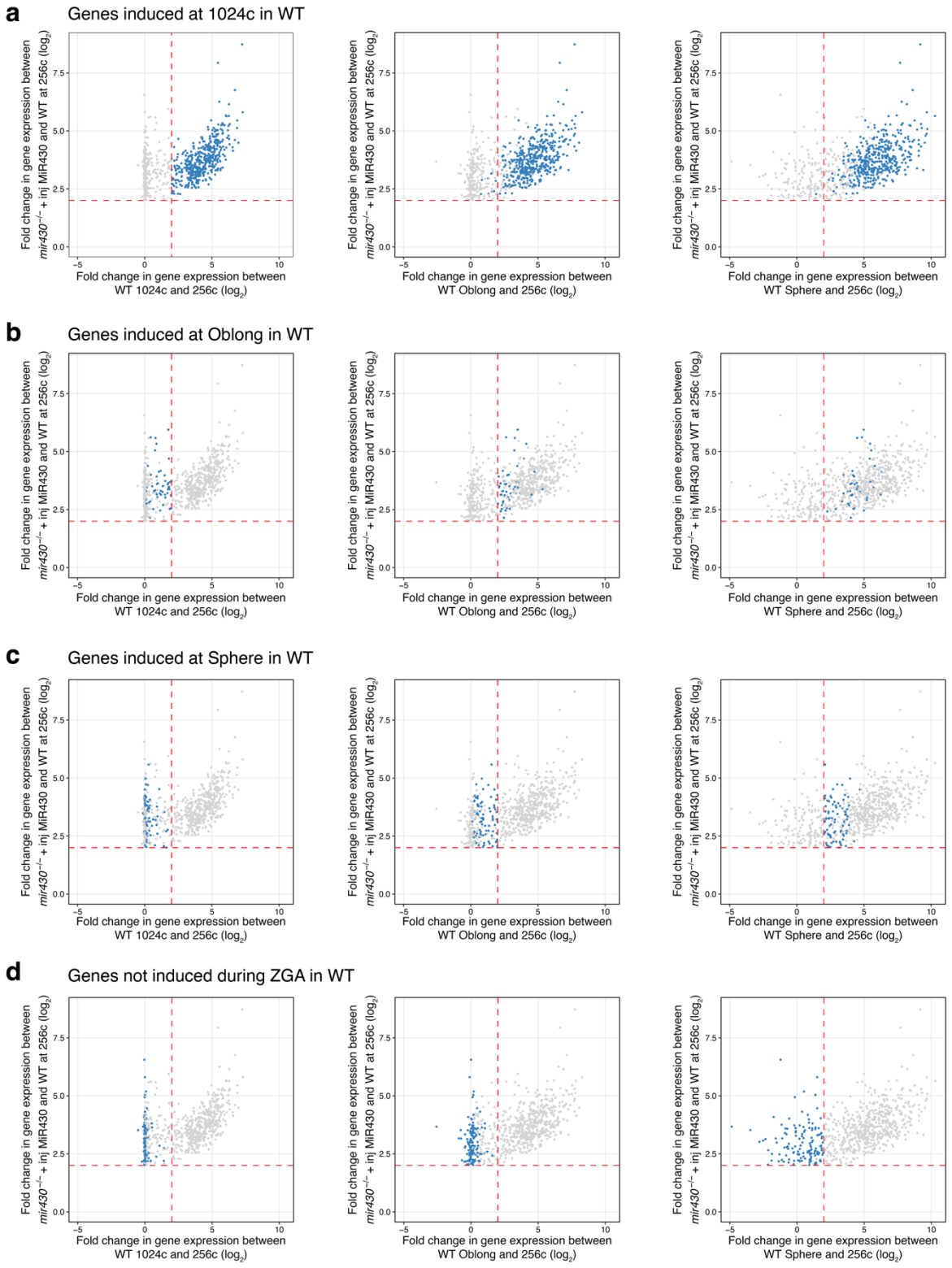
823 **Extended Data Figure 6. RNAPII-Ser2P signal at 256-cell, 1024-cell, Oblong and Sphere**

824 **stage in *mir430*$^{-/-}$ + inj MiR430 embryos.** Visualization of RNAPII-Ser2P signal with Fabs.

825 Representative micrographs are shown.

826

Extended Data Figure 7



827

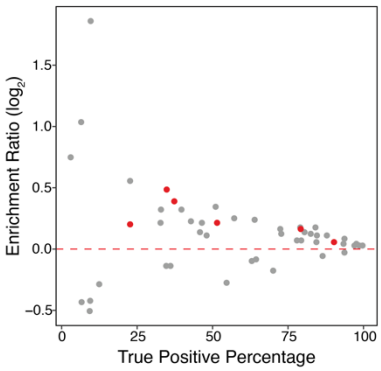
828 **Extended Data Figure 7. Correlation between expression level in *mir430*$^{-/-}$ + inj MiR430**

829 ***vs.* WT and 1024-cell, Oblong and Sphere *vs.* 256-cell in WT embryos. a.** Scatterplots

830 representing the fold change in gene expression ($\log_2$ scale) between *mir430*$^{-/-}$ + inj MiR430

831 and WT at 256-cell (y-axis), and the fold change in gene expression (log_2 scale) between 1024-
832 cell and 256-cell (left), Oblong and 256-cell (central) and Sphere and 256-cell (right) in WT
833 embryos (x-axis). All upregulated genes are shown in grey, and genes induced at 1024-cell are
834 highlighted in blue. The left graph is also shown in Fig. 3b. **b-d.** Same as in (A) but genes that
835 are highlighted represent genes that are induced at Oblong **(b)**, Sphere **(c)**, or not induced
836 during ZGA **(d)**.
837

Extended Data Figure 8



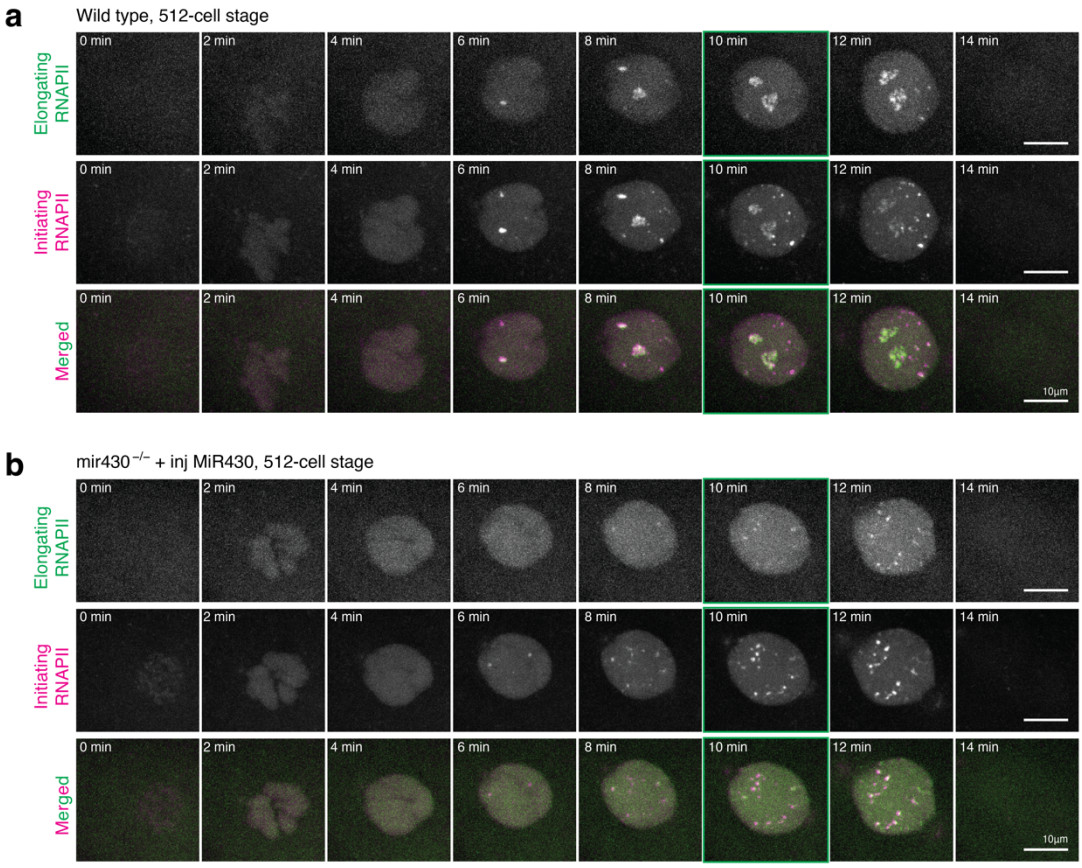
838

839 **Extended Data Figure 8. Transcription factor motif analysis of upregulated genes**. Scatter
840 plot representing the Enrichment Ratio (log_2) of transcription factor motifs in the promoters
841 (TSS ± 2Kb) of upregulated genes compared to the promoters of not expressed genes (y-axis),
842 and the percentage of promoters of the upregulated genes that have the motif (x-axis). Only
843 motifs whose corresponding protein is translated during early development [30] were considered,
844 and only motifs with a p-value < 0.05 and E-value ≤ 10 are shown. Motifs corresponding to
845 the three pluripotency factors Nanog, Sox19b and Pou5f3 (POU5F1, Pou5f1, Pou5f1::Sox2,
846 POU5F1B, POU2F1, POU2F1::SOX2) are labelled in red.

847

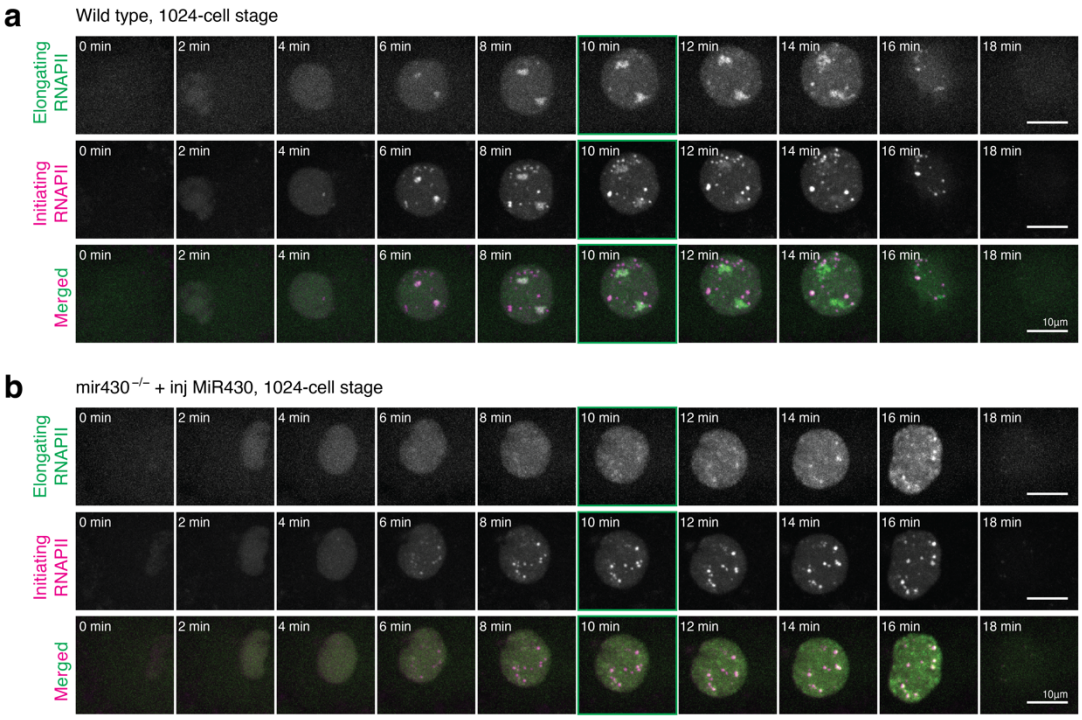
848

Extended Data Figure 9



Extended Data Figure 9. Stalled transcription bodies go into productive transcription in *mir430*$^{-/-}$ + inj MiR430 embryos (512-cell stage). Visualization of initiating (Ser5P) and elongating (Ser2P) RNAPII with Fabs in WT **(a)** and *mir430*$^{-/-}$ + inj MiR430 **(b)** embryos at 512-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Extended Data Figure 10



858

859 **Extended Data Figure 10. Stalled transcription bodies go into productive transcription in**

860 ***mir430*$^{-/-}$ + inj MiR430 embryos (1024-cell stage).** Visualization of initiating (Ser5P) and

861 elongating (Ser2P) RNAPII with Fabs in WT **(a)** and *mir430*$^{-/-}$ + inj MiR430 **(b)** embryos at

862 1024-cell stage. Shown are representative images of individual nuclei, extracted from a

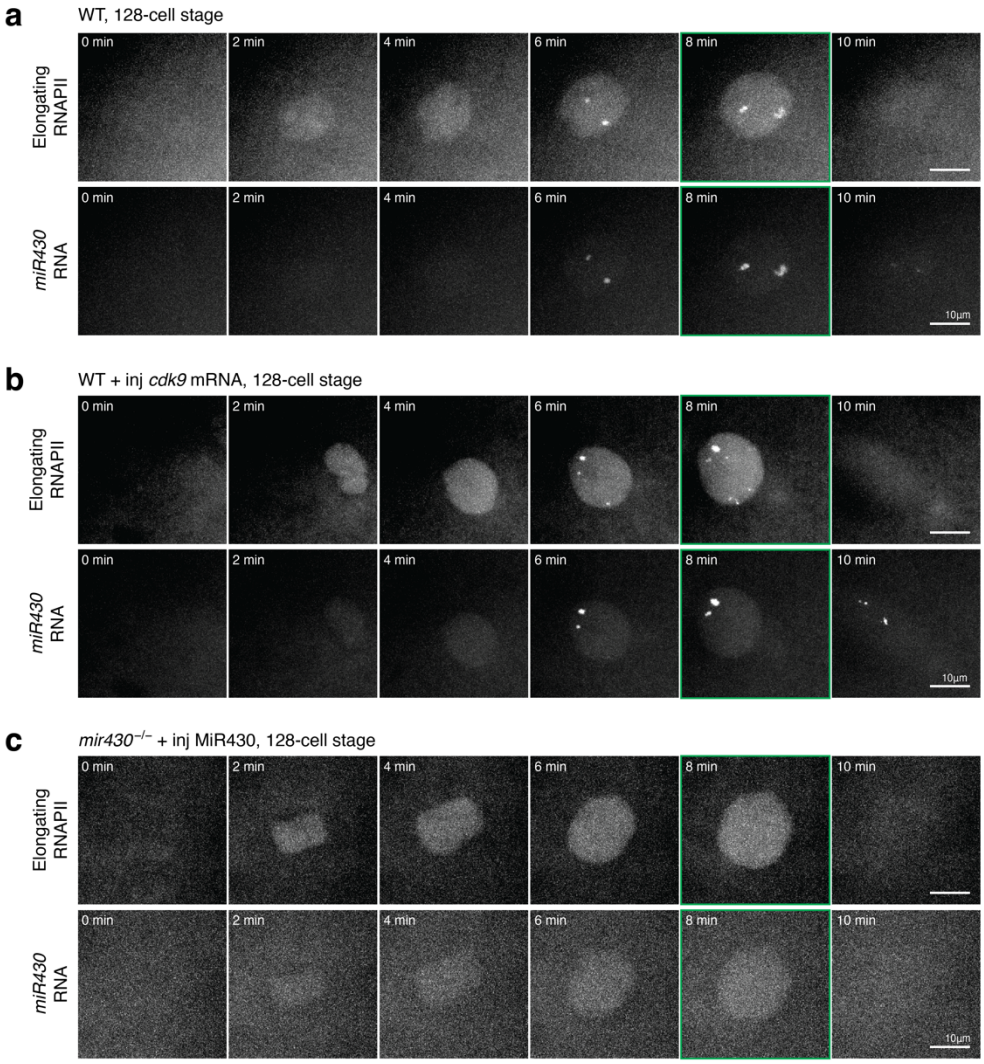
863 spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are

864 highlighted in green.

865

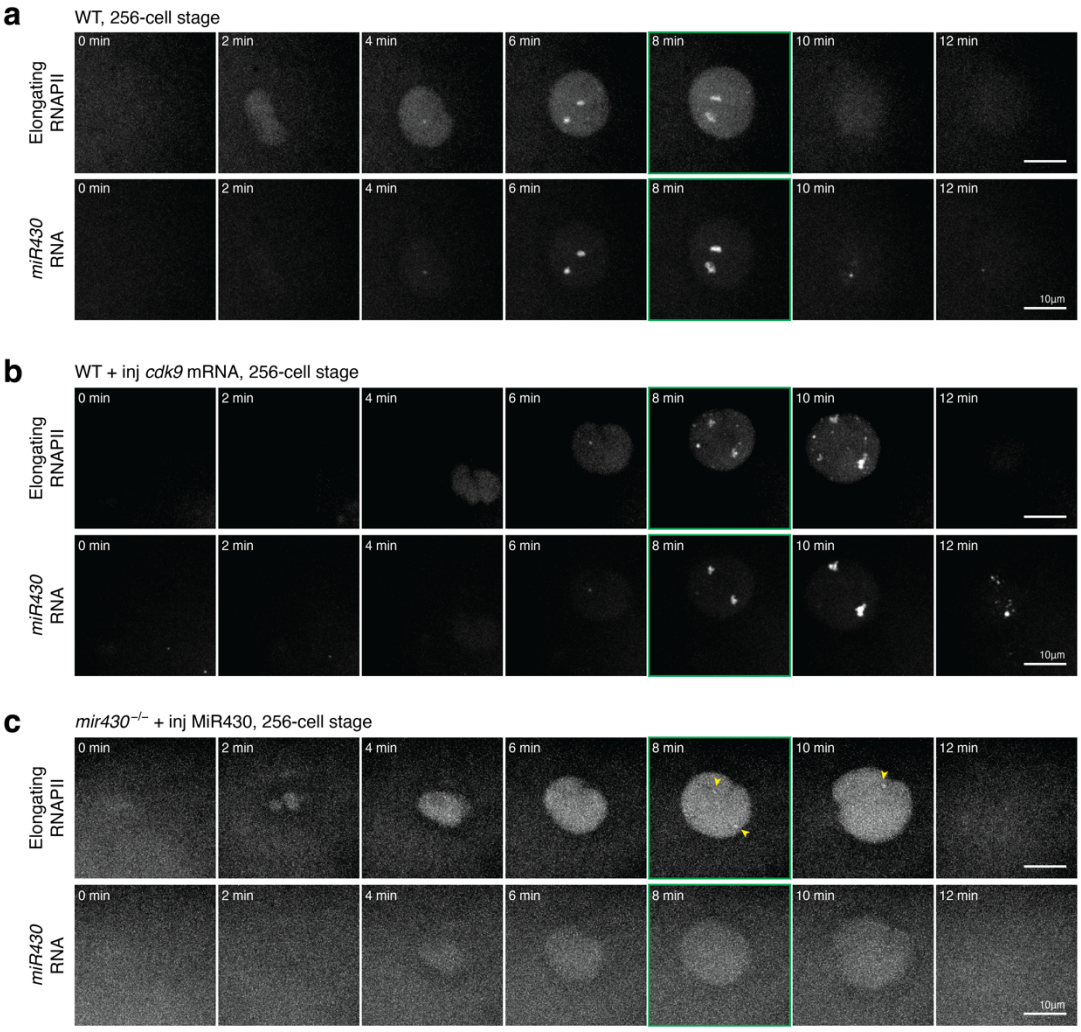
866

Extended Data Figure 11



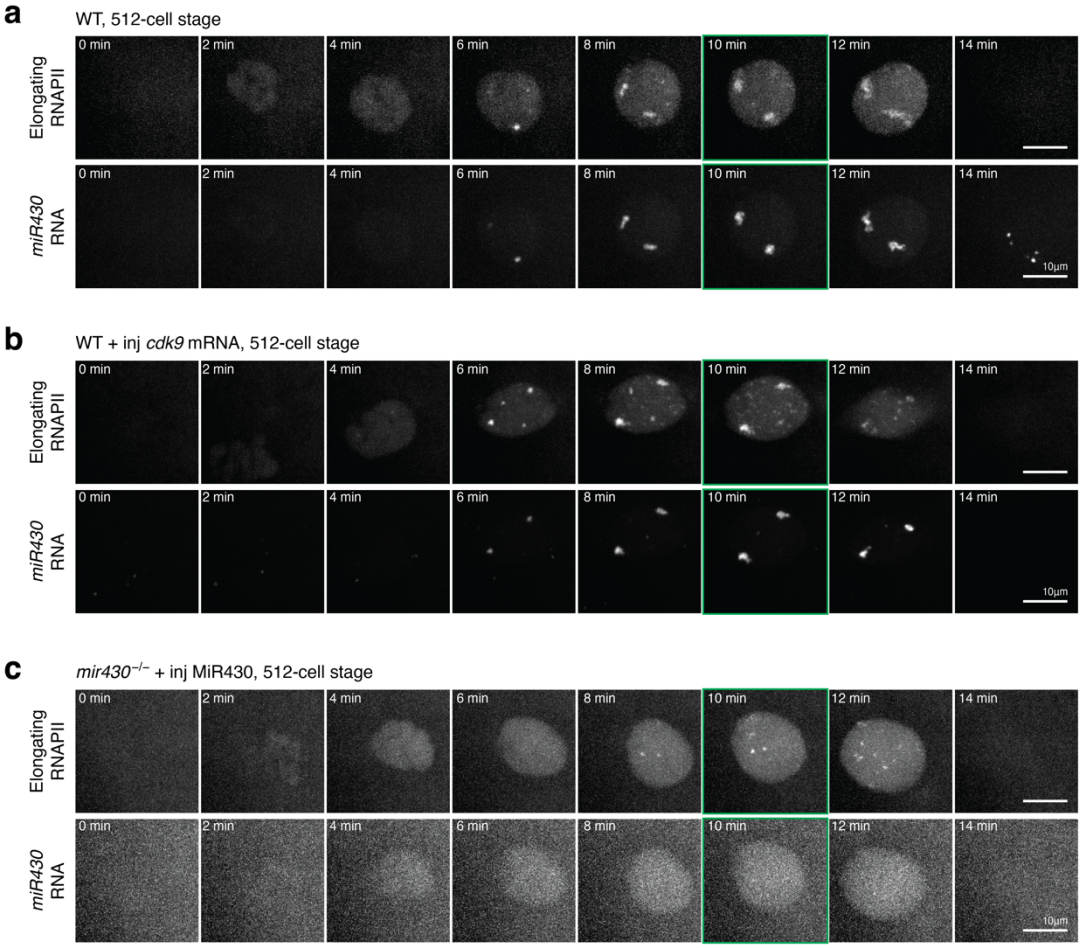
Extended Data Figure 11. RNAPII-Ser2P signal in WT, WT + inj *cdk9* mRNA and *mir430*$^{-/-}$ + inj MiR430 embryos at 128-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(a)**, WT + inj *cdk9* mRNA **(b)**, and *mir430*$^{-/-}$ + inj MiR430 **(c)** embryos at 128-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Extended Data Figure 12



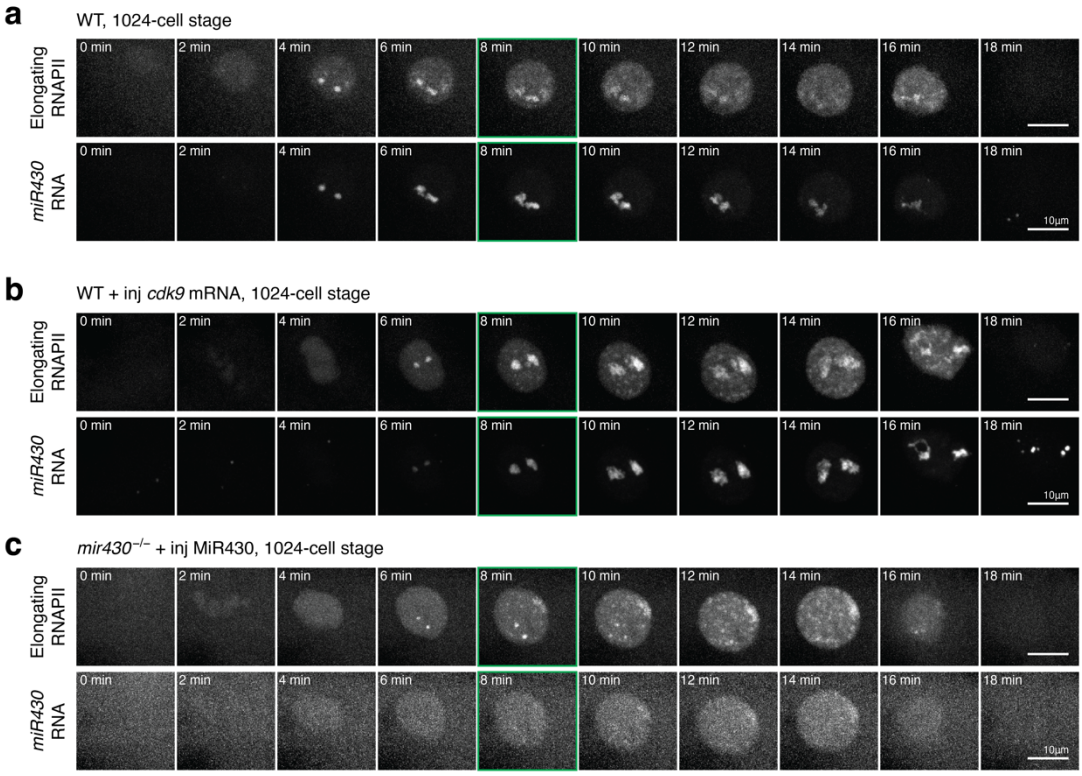
Extended Data Figure 12. RNAPII-Ser2P signal in WT, WT + inj *cdk9* mRNA and *mir430*$^{-/-}$ + inj MiR430 embryos at 256-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(a)**, WT + inj *cdk9* mRNA **(b)**, and *mir430*$^{-/-}$ + inj MiR430 **(c)** embryos at 256-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Extended Data Figure 13



Extended Data Figure 13. RNAPII-Ser2P signal in WT, WT + inj *cdk9* mRNA and *mir430*$^{-/-}$ + inj MiR430 embryos at 512-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(a)**, WT + inj *cdk9* mRNA **(b)**, and *mir430*$^{-/-}$ + inj MiR430 **(c)** embryos at 512-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Extended Data Figure 14



894

895 **Extended Data Figure 14. RNAPII-Ser2P signal in WT, WT + inj *cdk9* mRNA and**

896 ***mir430*$^{-/-}$ + inj MiR430 embryos at 1024-cell stage.** Visualization of RNAPII Ser2P (with

897 Fabs) and *miR430* RNA (with MOVIE) in WT **(a)**, WT + inj *cdk9* mRNA **(b)**, and *mir430*$^{-/-}$ +

898 inj MiR430 **(c)** embryos at 1024-cell stage. Shown are representative images of individual

899 nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown

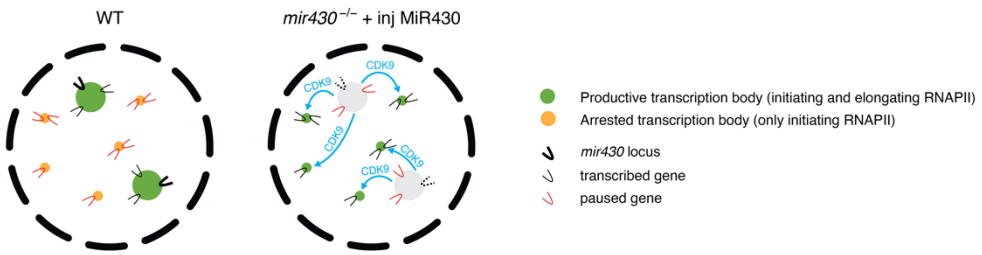
900 in the main figure are highlighted in green.

901

902

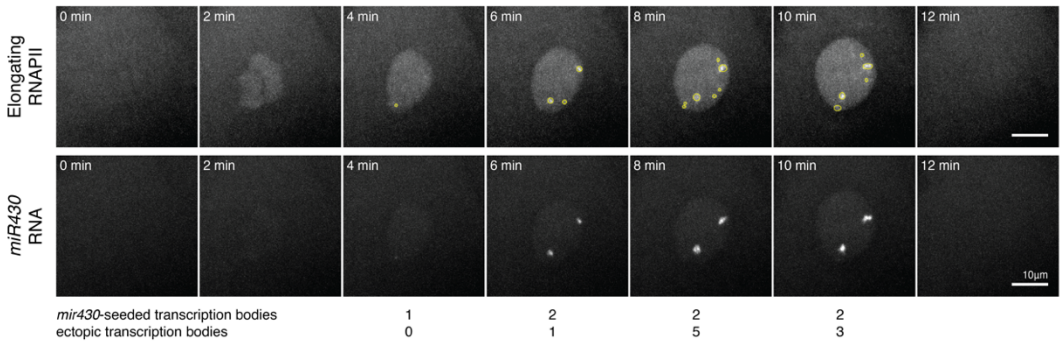
903

Extended Data Figure 15



Extended Data Figure 15. Model for the role of transcription bodies in transcription regulation. In WT nuclei, two large transcription bodies are nucleated by the *mir430* locus. These facilitate the expression of genes that localize there. They also sequester CDK9 (and potentially other factors) which is required for pause release, thereby stalling transcription elsewhere in the nucleus in the initiation state. The specific disruption of the two *mir430* transcription bodies leads to a redistribution of the transcriptional machinery (among which CDK9) which results in the downregulation of genes that localize to the *mir430* transcription bodies and the upregulation of genes elsewhere in the nucleus.

Extended Data Figure 16



914

915 **Extended Data Figure 16. Classification of *mir430* and ectopic transcription bodies.**
916 Transcription bodies are identified by detecting elongating RNAPII signal. If this overlaps with
917 MOVIE signal (detecting *miR430* RNA) they are classified as *mir430*, and if they do not
918 overlap with MOVIE signal, they are classified as ectopic transcription bodies. See Methods
919 for more detail.

920

921

922

923 **Extended Data Table 1. List of differentially expressed genes in mir430$^{-/-}$ + inj MiR430**
924 ***vs.* WT at all stages.**