



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2023

Local compaction of the mir430 locus leads to merging of Nanog clusters and transcription activation

Chabot Noémie

Chabot Noémie , 2023, Local compaction of the mir430 locus leads to merging of Nanog clusters and transcription activation

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_FFEC1323FB512

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de la Section des sciences fondamentales
Centre intégratif génomique

**Local compaction of the *mir430* locus leads to merging of
Nanog clusters and transcription activation**

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Noémie CHABOT

Master de génétique de l'Université de Paris Diderot

Jury

Prof. Jan Roelof van der Meer, Président
Prof. Nadine Vastenhouw, Directrice de thèse
Prof. David Suter, Expert
Prof. Bart Deplancke, Expert

Lausanne
(2023)

Acknowledgements

I would like to first thank Nadine Vastenhouw for hosting me in her lab during my PhD. I have learnt a lot in her contact. More than science, she taught soft skills that will follow me my whole life: organization, communication, and project management.

A big up to the Vastenhouw lab, that everyone brought me something, present or former members.

-Edlyn, alias Yoda, my scientific pedestal since I am here in Lausanne. You are always here when I have a weird or useless question, and even when you don't understand, you always do your best to answer.

-Hann, my chaddi buddy, 'underwear friend', that had to stick with the past five years in the worst and best moments of my PhD. This person that translates your English to other people because he's the only one to understand you, this person that knows often what you have in mind before you could even say it. It is funny how destiny was always bringing us together on the most random thing the past five years! Let's create a saying: "If you don't like someone from the beginning, then you will become good friends!" :D.

Ksenia, alias "Beauty of the Lake" or "Sempai". My PhD model when I arrived in the lab as a scared -first year PhD student puppy.

-Gilles, the sassiest lab technician one could hire! But also, very observant, and empathic. Thank you to be there and listen to my problems, when you are already at work since 9:30 am and it's already 7 pm. Thank you for taking care of the fish for us for so many years, thank you for organizing the lab and being 'l'homme à tout faire' for so many years.

-Maciej, that decided one day that I should die every time I sneeze. Not only die, but also being called by very sweet names at the same time. Thank you to remind me that I exist as you seem to forget some often!

-Shivali, just for being yourself and make a relax atmosphere in the lab. And remind me the basic concepts of molecular biology: a STOP codon makes translation stop.

-Dora, for the walks that help us to survive over this last year. For feeding me with biscuits and protein bars.

-Jana, for your sweet smile and the good words that you always have ready in any situation.

-Ramya, for helping me with the project and giving me access to your science.

-Martino, for helping me with R and some mathematical advice.

-Arsenii, for the most comments about phase-separation I have got in my life.

-Alicia, for bearing me as a supervisor for five months and still be happy about it!

Thank you to the fourth floor at CIG and at MPI for creating a nice working environment. Especially thanks to Claire, Betül, Núria, Linh, Marina for CIG, and all the Norden lab in MPI.

Thanks to my TAC committee members in Dresden: Alf Honigmann, Jan Brugués and Simon Alberti for following my project for two years and giving me feedback. Thanks to my thesis committee here in Lausanne for taking time to read and judge my work: Jan Roelof van der Meer, Bart Deplancke and David Suter.

Thanks to Florian Jug, for giving his time and giving great comments on image analysis and more.

Thank you to my collaborators for making this project come true: Damian Dalle Nogare for the data analysis, Alessia Del Panta and Alma Dal Co for the transdisciplinary collaboration.

Words cannot expression my gratitude to Nathalie Clerc, secretary of the 4th floor and much more: shoemaker, saleswoman for cheese, coffee, oranges and honey, hug machine, cook, fish feeder, postwoman, grocer, seamstress, handywoman, policewoman, HR, event planners... You are always here to tell me that real hugs are done on the left side, that it's better to have a bigger butt than a bigger hurt and to remind me that you are going on vacation. For making me discovering ice hockey and how much we don't any team except Lausanne LHC!

Thank you to former and present members of the Roignant for the fun, especially Lina, Chiara, and Athena. Thanks to Edlyn, Maciej and Elisa for proof-reading my thesis.

Thank you to the Nanog fish for making my life so difficult since I am in Lausanne, but finally accepting to survive after gastrulation. Thank you for giving me so many eggs!

I am done.

Abstract

Transcription is an essential process implicated in the cell maintenance and homeostasis, but also specific cell processes like differentiation, stress adaptation and development. This process is regulated by many layers of control, from the transcriptional machinery, the chromatin landscape until the genome organization. However, the first events leading to transcription is the binding of transcription factors (TFs) on enhancers or at the distal promoter, with as a downstream consequence the recruitment of the RNA Polymerase II. In the recent years, TFs have been demonstrated to form cluster in vitro and in vivo at enhancers and target genes. Studies have suggested that clustering activators such as TFs could increase DNA binding by confinement, help to recruit other TFs or co-activators or bring enhancers and promoters together. Moreover, it has been suggested that the amount of TF inside these clusters is correlated with transcription initiation and bursting, but how this is regulated is still unclear. However, these studies are mostly done in cells, with artificially induced clustering or using artificial DNA array as a seed.

Here, I take the advantage of Nanog clusters colocalizing with the *mir430* locus transcription during zygotic genome activation in zebrafish embryo to study how the dynamics of Nanog clusters correlates with *mir430* transcription initiation. Using a live-imaging approach, I studied how the Nanog clusters positive for *mir430* transcription relates quantitatively and mechanistically to *mir430* transcription.

I have found that Nanog make multiple clusters in the nucleus, two of which only transcribed as such early stage and colocalize with *mir430* transcription. The Nanog clusters positive for *mir430* transcription contains more proteins than other Nanog clusters. Interestingly, the Nanog clusters regulating *mir430* transcription is formed by two mechanisms: growth and merging of smaller Nanog clusters. I observed that transcription at the *mir430* locus starts when the Nanog cluster is reaching a peak of Nanog amount, suggesting a threshold amount of Nanog for transcription initiation. Intriguingly, the merging of smaller Nanog clusters is important to increase Nanog amount locally and reach such a threshold. Using live DNA labelling, I found that the merging Nanog clusters are both seeded by the *mir430* locus and that merging of the Nanog was concomitant with *mir430* local compaction prior to transcription initiation. Notably, Nanog is not necessary for the *mir430* locus compaction but is important for keeping the DNA together once compaction occurred.

To conclude, these results suggests that the amount of Nanog above a critical threshold inside clusters is important for transcription initiation of the *mir430* locus. Local *mir430* compaction helps to reach such high levels of Nanog locally by merging of smaller clusters. These findings help to understand how TF clustering is influencing transcription initiation, by reaching a threshold amount of TFs.

Résumé

La transcription est un processus cellulaire de base qui implique la plupart des autres processus cellulaires. Chez les eucaryotes, la transcription se passe dans le noyau où l'ARN Polymérase II transcrit les gènes codant pour des protéines et des microARNs pour permettre l'expression du génome dans la cellule. Ce processus est régulé par de nombreux niveaux de régulation, de la machinerie de transcription, en passant par l'architecture de la chromatine jusqu'à l'organisation du génome. Cependant, la transcription commence toujours par la fixation de facteurs de transcription (FTs) sur des activateurs ou la partie distal du promoteur, ce qui a pour conséquence le recrutement de l'ARN Polymérase II. Récemment, il a été démontré que les FTs forment des accumulations *in vitro* et *in vivo*. Cependant, le rôle de ces accumulations dans la régulation de la transcription reste encore incertain.

Certaines récentes études ont montré que l'accumulation locale de FTs ou de co-activateurs augmentait la fixation sur l'ADN à cause du confinement, aidait à recruter d'autres FTs ou co-activateurs, ou permet de localiser les activateurs et les promoteurs localement. De plus, il a été suggéré que le niveau de FT dans ces accumulations est corrélée avec l'initiation de la transcription et son côté explosive. Cependant, ces études sont pour la plupart faites sur des cellules, avec des accumulations de FTs créées artificiellement ou en utilisant des bases ADN artificielles comme lieu de fixation. Dans cette thèse, on utilise à notre avantage l'activation du génome chez l'embryo du poisson zèbre pour étudier comment la dynamique des accumulations de Nanog impacte la transcription.

J'ai trouvé que Nanog fait de nombreuses accumulations dans le noyau, deux d'entre elles seulement colocalisant avec la transcription du groupe de gènes *mir430*. Les accumulations actives sont les plus volumineuses, brillantes et plus intenses comparées à toutes les autres dans le noyau. Les accumulations de Nanog régulant le locus *mir430* se forment de deux manières : soit par croissance, soit par croissance combinée avec la fusion de deux accumulations plus petites. J'ai observé que la transcription au locus *mir430* commence quand l'accumulation de Nanog atteint un pic de fluorescence. Fait intéressant, les accumulations de Nanog atteignent toujours ce pic si la transcription est inhibée, mais forment un plateau, suggérant que les accumulations ne peuvent contenir qu'une quantité maximale de Nanog. La fusion des plus petites accumulations de Nanog était nécessaire pour avoir assez de Nanog, suggérant la présence d'un niveau minimal à atteindre pour commencer à transcrire. En utilisant le marquage du locus *mir430* *in vivo*, j'ai trouvé que la fusion des accumulations de Nanog se forme toutes les deux sur le locus *mir430* et que la fusion de ces accumulations était concomitante avec la compaction locale de l'ADN du locus *mir430*, avant qu'il ne commence à transcrire. Surtout, Nanog n'est pas nécessaire pour la compaction du locus *mir430* mais est important pour garder l'ADN ensemble une fois compacté.

Pour conclure, ce projet de doctorat suggère que le niveau de Nanog au-dessus d'un niveau seuil à l'intérieur d'accumulations est important pour l'initiation de la

transcription du locus mir430. La compaction local de la chromatine au locus mir430 permet d'atteindre ces hauts niveaux de fluorescence par la fusion de plus petites accumulations. Ces données vont aider à préciser notre compréhension sur le rôle de la quantité de FTs en corrélation avec la transcription.

List of Abbreviations

RNA	Ribonucleic Acid
DNA	Deoxyribonucleic acid
TF	Transcription Factor
RNA Pol II	RNA Polymerase II
TBP	TATA box-binding protein
UTR	Untranslated region
ORF	Open Reading Frame
mRNA	Messenger RNA
TSS	Transcription Start Site
snRNA	Small Nuclear RNA
CTD	Carboxy-terminal domain
PIC	Pre-initiation complex
Cdk7	Cyclin-dependent kinase 7
Cdk9	Cyclin-dependent kinase 7
NELF	Negative Elongation Factor
ChIP-seq	Chromatin Immunoprecipitation
eRNAs	Enhancers RNAs
DBD	DNA-binding domain
ADs	Activator domain
IDRs	Intrinsically disordered regions
SMT	Single-molecule tracking
P300	Histone acetyltransferase p300
BRD4	Bromodomain-containing protein 4
FISH	Fluorescence in situ hybridization
mESC	Mouse embryonic stem cell
SEs	Super-enhancers
H3K27Ac	Histone 3 lysine 27 acetylation
ATAQ-seq	Assay for transposase-accessible chromatin using high-throughput sequencing
H3K27me3	Histone 3 lysine 27 methylation 3
H3K9me3	Histone 3 lysine 9 methylation 3
H3k36me2	Histone 3 lysine 36 methylation 2
H3k4me1	Histone 3 lysine 3 methylation 1
ATP	Adenosine triphosphate
CPB	CREB-binding protein
BET	Bromodomain and extra terminal domain
HMG	High mobility group
Hi-C	High-throughput chromosome conformation capture technique
PcG	Polycomb Group
TADs	Topologically associating domain
CTCF	CCCTC-binding factor
ESCs	Embryonic Stem Cell

LLPS	Liquid-liquid phase-separation
GR	Glucocorticoid Receptor
RBP2	Retinoblastoma-binding protein 2
S5P	Serine 5 Phosphorylation
S2P	Serine 2 Phosphorylation
HP1	Heterochromatin Protein 1
Ubx	Ubiquitin Regulatory X
YY1	Ying-Yang 1
TAF15	TATA binding associated factor 15
ZGA	Zygotic Genome Activation
MoVIE	Morpholino Visualization of Expression
Fab	Antibody fragment
XCI	X Chromosome Inactivation
MBT	Midblastula transition
AR	Aspect Ratio
NLS	Nuclear localization signal
NIS	Nuclear Export Signal
MCP	MS2 coat protein
dCas9	Catalytically dead Cas9
mNG	mNeonGreen
FRAP	Fluorescence Recovery After Photobleaching
TU	Tübingen
TL	Tupfel Long-Fin
CTD	C-terminal domain

List of Figures

FIGURE 1: TRANSCRIPTION REGULATION IN EUKARYOTES	19
FIGURE 2: REGULATION OF TRANSCRIPTION IN EUKARYOTES	21
FIGURE 3: THE DIFFERENT MOVEMENTS THAT TFS CAN HYPOTHETICALLY DO IN THE NUCLEUS	27
FIGURE 4: DNA LOOP BETWEEN AN ENHANCER AND A PROMOTER, MEDIATED BY THE MEDIATOR COMPLEX	29
FIGURE 5: STRUCTURE OF A NUCLEOSOME	31
FIGURE 6: POSSIBLE CHROMATIN MARKS AT HISTONE TAILS FOR ACETYLATION, METHYLATION AND PHOSPHORYLATION.	32
FIGURE 7: CHROMATIN WRITERS, CHROMATIN ERASERS AND CHROMATIN READERS FOR ACETYLATION AND METHYLATION AT HISTONE TAILS	34
FIGURE 8: GENOME ORGANIZATION AT MULTIPLE SCALES: FROM CHROMOSOMAL TERRITORIES TO DNA LOOPS	38
FIGURE 9: STRUCTURE AND COMPONENTS OF A CONVENTIONAL ENHANCER VERSUS A SUPER-ENHANCER	41
FIGURE 10: EXAMPLES OF STOICHIOMETRIC RECRUITMENT AND CO-CONDENSATION THROUGH MULTIVALENT INTERACTIONS IN THE CONTEXT OF TRANSCRIPTION	45
FIGURE 11: MECHANISMS BY WHICH TF CLUSTERING IS REGULATING TRANSCRIPTION	49
FIGURE 12: NANOG CLUSTERS ARE PRESENT AT DIFFERENT STAGES, DURING THE WHOLE CELL CYCLE, BUT ONLY TWO ARE POSITIVE FOR TRANSCRIPTION.	80
FIGURE 13: THE S_{5P}⁺ NANOG CLUSTERS SHOW DISTINCT FEATURES COMPARED TO THE S_{5P}⁻ NANOG CLUSTERS	84
FIGURE 14: THE MIR₄₃₀ S_{5P}⁺ NANOG CLUSTERS DISSOCIATE IN SMALLER CLUSTERS UPON MIR₄₃₀ TRANSCRIPTION BODY GROWTH ERROR! BOOKMARK NOT DEFINED.	
FIGURE 15: THE MIR₄₃₀ S_{5P}⁺ NANOG CLUSTERS CAN FORM BY GROWTH OR BY MERGING OF SMALLER CLUSTERS.	94
FIGURE 16: FREQUENCY OF THE DIFFERENT MIR₄₃₀ NANOG CLUSTER FORMATION MECHANISMS.	95
FIGURE 17: TRANSCRIPTION IS NOT RESPONSIBLE FOR THE MERGING OF NANOG CLUSTERS.	97
FIGURE 18: MIR₄₃₀ NANOG FLUORESCENCE LEVELS AND TIME LINE ARE NORMALIZED	99

FIGURE 19: MIR430 TRANSCRIPTION IS ACTIVATED WHEN MIR430 NANOG CLUSTERS REACH HIGH AMOUNT OF NANOG.	103
FIGURE 20: MIR430 NANOG CLUSTERS DO NOT GROW FURTHER IN TRANSCRIPTION-INHIBITED CONDITIONS.	105
FIGURE 21: THE MIR430 NANOG CLUSTERS DO NOT GROW DIFFERENTLY BASED ON THEIR MECHANISM OF FORMATION.	108
FIGURE 22: MIR430 NANOG CLUSTERS CAN START TRANSCRIPTION WITHOUT MERGING.	109
FIGURE 23: MERGING IS ESSENTIAL TO REACH HIGH LEVELS OF NANOG.	113
FIGURE 24: SCHEMATICS REPRESENTING DIFFERENT HYPOTHESES ABOUT MIR430 NANOG CLUSTER SEEDING REGIONS	114
FIGURE 25: NANOG MERGING CLUSTERS ARE SEEDED BY DNA SEQUENCES ON THE SAME CHROMOSOME.	116
FIGURE 26: BOTH NANOG MERGING CLUSTERS ARE POSITIVE FOR MIR430 TRANSCRIPTION.	117
FIGURE 27: BOTH NANOG MERGING CLUSTERS ARE SEEDED BY THE MIR430 LOCUS DNA.	121
FIGURE 28: NANOG AND MIR430 DNA SIGNAL ARE FADING OFF UPON DNA RELAXATION.	124
FIGURE 29: ANALYSIS OF THE MIR430 LOCUS COMPACTION BEFORE TRANSCRIPTION INITIATION.	131
FIGURE 30: THE AMOUNT OF MIR430 DNA CORRELATES WITH NANOG LEVELS.	135
FIGURE 31: IS NANOG IMPLICATED IN THE MIR430 LOCUS COMPACTION? HYPOTHESIS AND METHODS.	140
FIGURE 32: THE MIR430 LOCUS COMPACTS AS FAST IN PRESENCE AND ABSENCE OF NANOG	142
FIGURE 33: NANOG IS IMPORTANT TO KEEP THE MIR430 LOCUS TOGETHER.	144
FIGURE 34: NANOG PROTEINS IN CLUSTERS RECOVER FAST.	146
FIGURE 35: SOX19B PROTEINS IN CLUSTERS RECOVER FAST.	149
FIGURE 36: POU5F3 PROTEINS RECOVERS FAST IN CLUSTERS.	151
FIGURE 37: REDISTRIBUTION OF THE INJECTED COMPONENTS FROM THE YOLK TO THE CELL DURING THE FIRST CELL CYCLES	156
FIGURE 38: A HIGHER FLUORESCENCE NUCLEAR BACKGROUND IS CORRELATED WITH MORE NANOG CLUSTERS.	158
FIGURE 39: FEATURES OF NANOG CLUSTERS IN DIFFERENT EXPERIMENTAL CONDITIONS	160

FIGURE 40: NANOG, POU5F3 AND SOX19B BIND AT THE MIR430 LOCUS, AS REVEALED BY CHIP-SEQ EXPERIMENTS	163
FIGURE 41: VOLUME COMPARISON OF THE TWO MIR430 NANOG CLUSTERS INSIDE THE SAME NUCLEUS VERSUS OTHER NUCLEI	164
FIGURE 42: MIR430 LOCUS STRUCTURE ACROSS DIFFERENT GENOME ASSEMBLY	166
FIGURE 43: CROSSING OF DIFFERENT ZEBRAFISH STRAINS DOES NOT REVEAL ANY DIFFERENCES IN THE MIR430 NANOG ASSEMBLY.	173
FIGURE SUPPLEMENTAL 1: SEGMENTATION THRESHOLD AS A FUNCTION OF MEAN INTENSITY IN THE NUCLEUS	179
FIGURE SUPPLEMENTAL 2: FIRST SEGMENTATION THRESHOLD AS A FUNCTION OF NUCLEAR MEAN INTENSITY (NANOG MNEONGREEN)	179
FIGURE SUPPLEMENTAL 3: SECOND SEGMENTATION THRESHOLD AS A FUNCTION OF NUCLEAR MEAN INTENSITY (NANOG MNEONGREEN)	180
FIGURE SUPPLEMENTAL 4: SEGMENTATION THRESHOLDS FOR THE NUCLEUS AND NANOG CLUSTERS AS A FUNCTION OF NUCLEAR MEAN INTENSITY (NANOG-HALOTAG)	180
FIGURE SUPPLEMENTAL 5: SEGMENTATION THRESHOLD AS A FUNCTION OF NUCLEAR MEAN INTENSITY (RNA POL II S5P SIGNAL)	181
FIGURE SUPPLEMENTAL 6: SEGMENTATION THRESHOLD AS A FUNCTION OF NUCLEAR MEAN INTENSITY (MIR430 RNA)	181
FIGURE SUPPLEMENTAL 7: MIR430 DNA MASK ASPECT RATIO AS A FUNCTION OF TIME FOR ALL ANALYZED ALLELES	183
FIGURE SUPPLEMENTAL 8: STRATEGY FOR MIR430 DNA VISUALIZATION.	183
FIGURE SUPPLEMENTAL 9: THE MIR430 LOCUS EXPANDS UPON TRANSCRIPTION INITIATION.	184

List of Tables

TABLEAU 1: INJECTION MIX AND IMAGING CONDITIONS FOR ALL IMAGING EXPERIMENTS IN THIS THESIS	68
TABLEAU 2: DISTANCE BETWEEN CLUSTERS AND SPEED OF COMPACTION (1K-CELL STAGE)	141
TABLEAU 3: STATISTICS ABOUT UNBOUND STATE (1K-CELL STAGE)	143
TABLEAU 4: THE DIFFERENT EXPERIMENTAL CONDITIONS TO RESCUE THE DEVELOPMENTAL PHENOTYPE OF NANOG -/- EMBRYOS IN THREE DIFFERENT STUDIES.	155
TABLEAU 5: NUCLEI, STAGE AND THRESHOLD VALUES FOR NUCLEUS AND NANOG-MNEONGREEN	175
TABLEAU 6: NUCLEI, STAGE AND THRESHOLD VALUES FOR NUCLEUS AND NANOG-HALOTAG (JFX650)	176
TABLEAU 7: NUCLEI, STAGE AND THRESHOLD VALUES FOR RNA POL II S₅P SIGNAL CALIBRATION CURVE	176
TABLEAU 8: NUCLEI, STAGE AND THRESHOLD VALUES FOR MIR430 RNA SIGNAL CALIBRATION CURVE	177
TABLEAU 9: DISTANCE BETWEEN CLUSTERS AND SPEED OF COMPACTION (ALL STAGES MIXED)	177
TABLEAU 10: STATISTICS ABOUT BOUND STATE (ALL STAGES MIXED)	178

Collaborations and work sharing

Parts of the work in this thesis are the result of collaborative work. Chapters 3.1, 3.2, 3.3 and 3.6 of this thesis are solely the work of the PhD candidate. Chapter 3.4 is in collaboration with post doc Ramya Purkanti, Vastenhouw lab, UNIL. Specifically, she developed the method to visualize the *mir430* locus. I produced all the data for this chapter and did all the analysis (image analysis, quantification, and calculations) except Figure 27C and Figure 27D. Chapter 3.5 is in collaboration with post doc Ramya Purkanti and Alessia Del Panta, PhD candidate, Dal Co lab, DBC, UNIL. Ramya Purkanti helped with the preparation of the embryos before imaging. I contributed to prepare embryos for imaging, performed the imaging experiments as well as most of the image analysis. Alessia Del Panta proposed the hypotheses as well as the method to answer the biological question and performed all the calculations, as well as the generation of all the plots.

Table of Contents

ACKNOWLEDGEMENTS	3
ABSTRACT.....	4
RÉSUMÉ	5
LIST OF ABBREVIATIONS.....	7
LIST OF FIGURES.....	9
LIST OF TABLES.....	12
COLLABORATIONS AND WORK SHARING	13
1. INTRODUCTION	19
1.1. TRANSCRIPTION REGULATION OF CODING-GENES AND MICRORNAS IN EUKARYOTES	19
1.1.1 – Basic unit of transcription : a promoter, a gene and a minimal set of transcriptional machinery.....	19
Prior to transcription initiation : the good elements at the right place	20
Initiation of transcription	21
The RNA Polymerase II:	21
The Pre-initiation complex (PIC):	22
Transcriptional pausing and elongation.....	23
Termination of transcription and poly-adenylation	23
RNA processing	24
1.1.2 – Gene regulation by more distant element : TFs, enhancers and the Mediator complex for DNA looping	24
Enhancers/Insulators	25
Transcription factors	25
Structure	25
On the way to work: Scanning DNA and binding to DNA motif	27
Functions: Recruitment of co-factors and PIC assembly by TFs.....	27
The Mediator complex:.....	28
Enhancer-promoter communication by DNA looping:.....	30
1.1.3. Another layer of complexity : Chromatin, chromatin marks and chromatin remodelers	31
Nucleosomes.....	31
Chromatin marks.....	33
DNA methylation.....	34
Chromatin remodelers and histone modifiers.....	34
Histones remodeling complexes.....	35
Histone modifiers	35
Non-histone architectural proteins.....	37
1.2 NUCLEAR ORGANIZATION: THE RUSSIAN DOLLS	37
1.2.1. Chromosome territories	38
1.2.2. Heterochromatin and euchromatin.....	38

1.2.3. A/B compartments	39
1.2.4. Topologically associated domains	40
1.2.5. Super-enhancers as a way to regulate master genes	41
1.2.6. Clustering of the transcriptional machinery in the nucleus	43
1.3. TRANSCRIPTIONAL MACHINERY CLUSTERING.....	44
1.3.1. FROM STOICHIOMETRIC INTERACTIONS TO LIQUID-LIQUID PHASE-SEPARATION	44
Classical view of the macromolecule complex formation: high-affinity interactions, specificity and stoichiometry	45
Low-affinity and non-stoichiometric interactions via low complexity domains	46
Liquid-liquid Phase-Separation	46
1.3.2. Mechanism and regulation of TF clustering.....	47
1.3.3 – Properties and role of TF clustering in transcription regulation	49
Bringing enhancers and genes together	50
Recruiting co-activators and the transcriptional machinery	51
Affecting binding properties: Decreasing search time, increasing binding rate or stabilize DNA binding ..	52
Transcription activation by high local amount of regulators: accelerates rate-limiting steps	54
Inhibition of transcription	55
1.4. THE ZYGOTIC GENOME ACTIVATION IN ZEBRAFISH AS A MODEL TO STUDY TRANSCRIPTION REGULATION.....	56
1.4.1. General principles underlying ZGA	56
The maternal-to-zygotic transition	56
Transcription landscape during early embryonic development.....	56
1.4.2. The timing of ZGA and proposed models	58
1.4.3. Chromatin landscape and genome architecture during ZGA in ZEBRAFISH	58
1.4.4. ZGA in Zebrafish as a model to study transcription regulation.....	59
1.5. AIM OF THE THESIS	60
2. MATERIALS AND METHODS	62
2.1 FISH WORK.....	62
2.1.1. Fish maintenance	62
2.1.2. Nanog fish line managing	62
2.2. PREPARATION OF EMBRYOS FOR IMAGING.....	62
2.2.1 – Preparation of the components for visualization	62
2.2.2 – Dechoronation, injection, health checking and soaking.....	66
2.2.3 – Mounting.....	66
2.2. IMAGING	67
2.2.1. Microscope.....	67
2.2.2. Imaging conditions :	68
2.3. IMAGE ANALYSIS.....	69
2.3.1. Pre-processing :.....	69
Time-lapse max projection and nucleus mapping.....	69
Cell-stage determination.....	69
Correction of channel registration	69
Isolation of nuclei	69
2.3.2. Segmentation of the signal in 3D :	70
General parameters for segmentation :.....	70
Specificities for each signal :	70
2.3.3. Tracking :.....	72
Manual tracking	72

Automatic tracking	72
2.3.4. Colocalization analysis in 3D :	72
2.3.5. Segmentation and image analysis in 2D :	73
2.4. Data processing and plotting	73
2.4.1. Association between Nanog clusters and transcription status :	73
2.4.1. Determine the category of Nanog tracks: growing or merging.....	73
2.4.2. Analysis of bound and unbound state for <i>mir430</i> DNA shape	75
Time-serie analysis of miR430 loci distances	75
Hypothesis testing.....	75
2.4.3. Data normalization before plotting.....	76
2.4.4. Plotting and graph construction.....	77
2.4.5. image analysis, normalization and fitting for frap data	77
3. RESULTS :	79
3.1. Nanog clusters positive for transcription initiation contain more Nanog than all other clusters	79
3.1.1 - Nanog makes multiple clusters in the nucleus, two of which colocalize with transcription initiation	80
3.1.2 – What makes the s5p+ Nanog clusters special compared to the s5p- Nanog clusters?	82
3.1.3 – S5P+ Nanog clusters dissociate in smaller clusters upon the growth of transcription body..	86
3. 2. The merging of smaller Nanog clusters increase TF amount locally and correlates with transcription initiation	91
3.2.1 – The <i>mir430</i> s5p+ Nanog cluster accumulate Nanog proteins by growth or by merging of smaller Nanog clusters	91
3.2.2 – Transcription is not responsible for the merging of Nanog clusters.....	97
3.2.3 – High levels of Nanog correlates with activation of transcription	99
3.2.4 – <i>mir430</i> Nanog clusters have a maximum TF loading capacity.....	104
3.2.5 – growing or merging: no difference in level of Nanog and time of activation	106
3.2.6 - Merging of smaller Nanog clusters is important to reach high enough levels of Nanog and start transcription.	109
3. 3. <i>mir430</i> Nanog subclusters are both seeded by the <i>mir430</i> locus	114
3.3.1 - Nanog merging clusters are seeded by DNA regions located on the same chromosome.....	115
3.3.2 - Both merging Nanog clusters colocalize with miR430 microRNA.....	117
3.3.3 - Both merging Nanog clusters are seeded by the <i>mir430</i> locus.....	118
3. 4. The <i>mir430</i> locus compacts as the <i>mir430</i> nanog clusters are merging	122
3.4.1. Nanog and <i>mir430</i> DNA signal disappear upon DNA relaxation during telophase.....	122
.....	123
3.4.2. The <i>mir430</i> locus compacts with a similar dynamics to merging Nanog clusters prior to transcription	126
3.4.3. The amount of <i>mir430</i> DNA correlates with Nanog levels.....	132
3.5. Nanog is not responsible for <i>mir430</i> compaction but help to keep the locus together	137
3.5.1 – Setting up the method to study the role of Nanog is <i>mir430</i> DNA compaction	137
.....	140
3.5.2 –Nanog itself is not responsible for <i>mir430</i> compaction.....	140
3.6.3 – Nanog acts as a glue to keep the <i>mir430</i> locus together.....	143
3.6. Dynamics of Nanog, Sox19b and Pou5f3 TFs in clusters during ZGA.....	145
Nanog.....	145
Sox19b.....	149

Pou5f3.....	151
4. DISCUSSION.....	154
4.1 Are the Nanog clusters really present in a wild-type background?.....	154
4.1.1. An overexpression of Nanog could lead to artificially increase the number of Nanog cluster in the nucleus.	154
4.1.2. How to determine if the Nanog clusters are real?	158
4.1.3. The Nanog clusters at the mir430 locus form preferentially compared the other Nanog clusters.....	159
.....	159
4.1.4. Are Nanog clusters seeding by the mir430 locus real?	161
4.2 A threshold of Nanog proteins at the mir430 locus lead to transcription activation	163
4.3 Overlapping is not colocalization	167
4.4. Is the merging of Nanog clusters essential for mir430 activation ?.....	169
4.5. What is the mechanism for <i>mir430</i> locus compaction ?.....	170
4.5.1. Nanog is not important for mir430 locus compaction but help to maintain the locus compacted	170
4.5.2. DNA loop extrusion as a way to compact the mir430 locus	171
4.6. Why do we see sometimes one or two clusters Nanog clusters ?	172
4.7. Are Nanog clusters phase-separated?	174
5. SUPPLEMENTAL MATERIAL.....	175
5.1. Supplemental Tables.....	175
5.2. Supplemental Figures	179
.....	180
.....	181
6. BIBLIOGRAPHY.....	185
7. CURRICULUM VITAE	202

Transcription is a basic cell process that is involved in many important cellular functions like development, cell differentiation, responding to stress, and the development of cancer. Studying how transcription is controlled is essential because it is the first step that leads to making proteins, which are essential for all cell activities. This knowledge helps us understand and potentially influence these vital cellular processes.

1. Introduction

1.1. TRANSCRIPTION REGULATION OF CODING-GENES AND MICRORNAS IN EUKARYOTES

1.1.1 – BASIC UNIT OF TRANSCRIPTION : A PROMOTER, A GENE AND A MINIMAL SET OF TRANSCRIPTIONAL MACHINERY

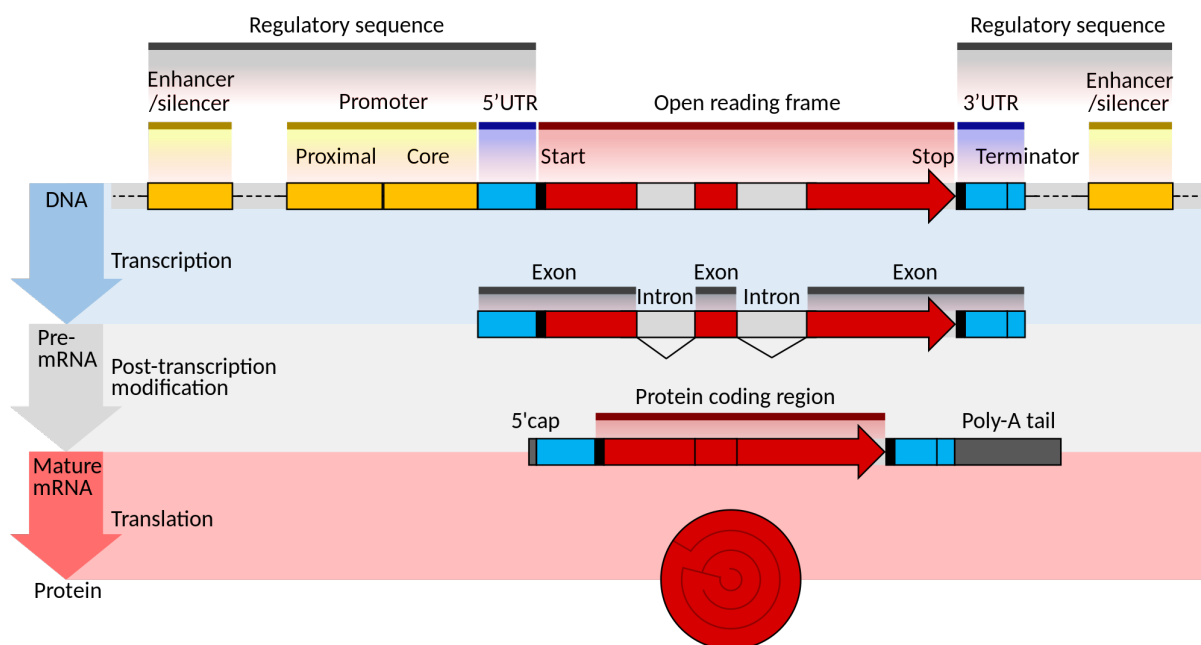


Figure 1: Transcription regulation in eukaryotes

From Eukaryotic and prokaryotic gene structure, by T. Shafee and R. Lowe, 2017, *SSRN Electronical Journal*, 4(1):2. Doi: 10.15347/wjm/2017.002.
Copyright C 2019 by the Authors. Creative Commons license CC BY.

Prior to transcription initiation : the good elements at the right place

Transcription is a fundamental genetic process, involving the conversion of specific nucleotide sequences within genes into RNA molecules, which can later be translated into proteins. This crucial process takes place within specific regions of the genome known as genes.

Genes are broadly categorized into two primary types: housekeeping genes, which encode proteins essential across various cell types, and cell-type-specific genes, responsible for producing proteins unique to specialized cell types or vital for specific cellular functions. While not all genes are transcribed in every cell, they share a similar structural framework.

A typical gene comprises two different functional parts: the transcribed part of the gene, that will be copied into RNA, and regulatory sequences, that control the start and level of transcription. In protein-coding genes, regulatory regions are predominantly situated at the 5' and 3' ends of the gene and comprised the promoter upstream of the transcribed part of the gene, and enhancers or insulators (**Figure 1**).

The promoter represents the initial section of a gene and is divided into two regions: the core or minimal promoter housing the transcription start site (TSS) and containing the minimal DNA necessary for initiating transcription, and the proximal promoter located upstream of the core promoter, harboring DNA elements that directly regulate gene expression, bound by specific transcription factors (TFs) (**Figure 1**).

Various types of promoters exist. For instance, housekeeping gene promoters often contain CpG islands susceptible to DNA methylation. Conversely, cell-type-specific genes often feature the TATA element upstream of their transcription start sites. However, many promoters lack specific DNA sequence elements for recognition as promoters and typically rely on an initiator element to recruit TATA-binding protein (TBP). Besides, promoters can be categorized as strong, resulting in high transcription rates, or weak, leading to lower transcription levels. Enhancer and silencer elements will discuss later in the next chapter.

On the other hand, in eukaryotes, the transcribed portion of a gene includes the transcription start site (TSS), 5' UTR (5' untranslated region), the open reading frame

(ORF), the stop codon sequence, the 3' UTR (3' untranslated region) and the transcriptional termination site. In particular, the ORF comprises exons, that will be translated into a protein, and introns, that will be removed from the pre-mRNA in a process called splicing.

Initiation of transcription

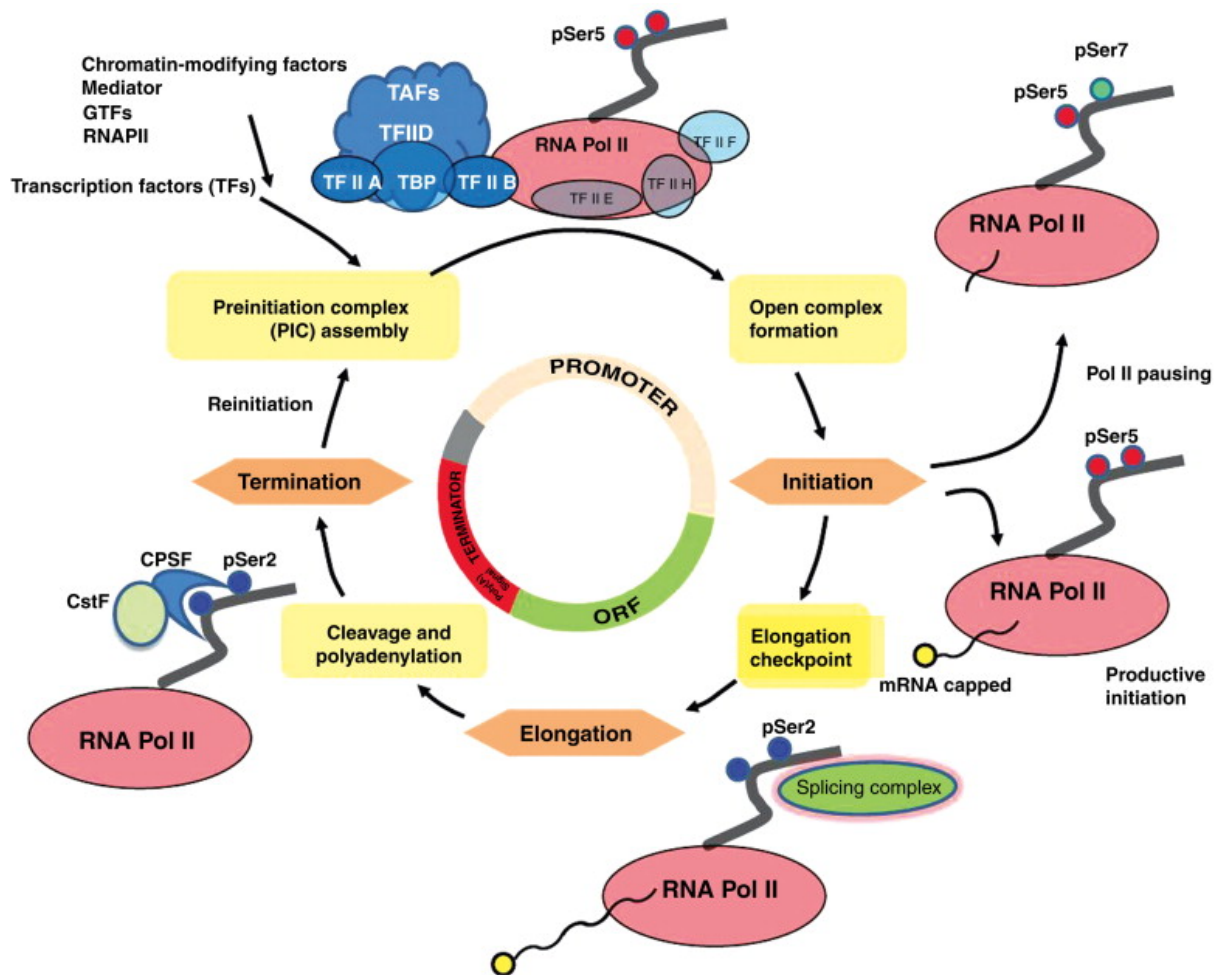


Figure 2: Regulation of transcription in eukaryotes

From Plant Transcription Factors, by Jong Chan Hong, 2016, *Academic Press*, Pages 35-36. Doi: <https://doi.org/10.1016/B978-0-12-800854-6.00003-8>. Copyright C 2016 Elsevier Inc. All rights reserved. Reused with permission.

The RNA Polymerase II:

The RNA Polymerase II (RNA Pol II) is an essential enzyme responsible for catalyzing the synthesis of RNA using a DNA template. In eukaryotes exist three distinct types of RNA Polymerases, each dedicated to transcribing different sets of

genes. The RNA Polymerase II is primarily responsible for transcribing messenger RNA (mRNA), a substantial portion of small nuclear RNA (snRNA), and microRNAs¹.

This complex enzyme is composed of 12 subunits in human². The largest subunit, RPB1, possesses a unique disordered carboxyl-terminal domain (CTD). This CTD consists of repeats of a heptapeptide sequence with the consensus Y¹S²P³T⁴S⁵P⁶S⁷. While the CTD sequence is highly conserved among eukaryotes, the actual number of repeats can vary between different species^{3,4}. Functionally, the CTD plays a crucial role in co-transcriptional processing due to its ability to undergo phosphorylation at various serine residues, depending on which step of transcription the RNA Pol II is engaged ^{3,4} (**Figure 2**).

The Pre-initiation complex (PIC):

It was established long ago that RNA Polymerase II (RNA Pol II) cannot transcribe DNA templates *in vitro* without the addition of cellular extracts. This observation emphasizing the significance of other protein factors in the process of RNA synthesis ⁵. Indeed, a fixed number of initiation factors assemble with the RNA Polymerase II to start transcription. The assembly of RNA Pol II and these initiation factors is known as the pre-initiation complex (PIC). In the case of genes transcribed by the RNA Pol II, the minimal PIC consists of six class II general initiation factors, including TFIID and TFIIH ^{1,6-8} (**Figure 2**).

The order of events for initiation of transcription relies on consecutive binding and interactions between the different initiation factors. Indeed, the RNA Pol II itself cannot independently recognize the promoter DNA. It relies first on the specific initiation factors, such as TFIID, which binds to the TATA element at the promoter ⁹. Subsequently, TFIID assembles with TFIIB, which in turn binds to RNA Pol II and recruits the Pol II-TFIIF complex. Finally, TFIIE and H are brought into the complex, a process facilitated by the DNA translocase XPB, a subunit of the general initiation factor TFIIH. The TFIIH initiation factor include an ATP-dependent helicases that unwind the DNA at the transcription start site, causing the initiation of transcription¹. The final step prior to transcription initiation is the phosphorylation of the RNA Pol II CTD is phosphorylated on 5th serine by the CDK7 subunit of the TFIIH factors¹⁰. Then the RNA Pol II can start to produce a small pre-mRNA transcript of typically 2-15 nucleotides via abortive initiation (**Figure 2**).

Transcriptional pausing and elongation

The RNA Pol II is however not directly transcribing the whole gene after transcription initiation, but it is pausing after a short number of transcribed bases. Indeed, pausing of the RNA Pol II can lead to arrest and termination of the transcription¹¹. It is thus an important step of transcription regulation.

The RNA Pol II pauses generally 50 bp downstream of the transcriptional start^{12–14}. At this moment of transcription pausing, the complex DRB-sensitivity inducing factor (DSIF), composed of the subunits HSPT4 and HSPT5, and NELF (Negative Elongation Factor) are associated with the RNA Pol II. NELF maintains the RNA Pol II in the pausing state, locking the RNA Pol II-TFIIF complex to the DNA to avoid dissociation of the elongation complex¹⁵. As indicated by its name, NELF is negatively regulating the transcription elongation. To proceed into transcription, the RNA Pol II needs to release this factor. The positive transcription elongation factor b (P-TEFb) contains a kinase subunit, CDK9, that phosphorylates the DSIF elongation factor, to avoid the binding of the Negative transcription regulator NELF during elongation (Negative Elongation Factor)¹⁶. CDK9 phosphorylates the RNA Pol II on its CTD serine 2 (S2P) as well to release the paused Pol II^{16,17}.

Termination of transcription and poly-adenylation

Once the RNA Pol II arrives at the end of the genes, several processes are occurring to stop transcription and prepare the pre-mRNA for translation (**Figure 2** **Figure 1**). As the RNA Pol II arrives at the end of the genes, the factor 7SK snRNP is recruited to inhibit the P-TEFb factor, leading to dephosphorylation of the DSIF. The NELF factor then binds on DSIF. During the process, the S5P-Phosphatase dephosphorylates the RNA Pol II S5P CTDs. Once the RNA Pol II goes through the 5'-AAUAAA-3', the transcription termination signal, the Cleavage and Polyadenylation Specificity Factor (CPSF) binds the pre-mRNA on the polyadenylation signal. The CPSF contains subunits that can binds on the polyadenylation signal and cleaves the pre-mRNA downstream of the signal, releasing the pre-mRNA from the processing RNA Pol II. To stop the RNA Pol II to synthesize more RNA, the enzyme XRN2 in human degrades the remaining uncapped mRNA, causing the release of the RNA Pol II from the pre-mRNA.

RNA processing

To protect the mRNA and prepare it for translation, the pre-mRNA goes through different steps during and after transcription generally called RNA processing. It includes three important steps: capping, splicing and tailing^{18,19}. Capping and tailing help in the transportation of the mRNA to ribosome and protects it from degradation²⁰, while slicing will remove unnecessary part of the pre-mRNA for translation.

The pre-mRNA capping is the addition of a 7-methylguanosine nucleotide to the 5' end of the pre-mRNA. The first phosphorylation event, mediated by CDK7, enables the recruitment of the capping enzyme¹⁸.

The pre-mRNA Tailing is the attachment of a poly-A tail at the 3' end the mRNA. Once the pre-mRNA is released from the processing RNA Pol II, The CstF complex binds on U or G/U rich regions downstream of the AAUAAA polyadenylation signal on the pre-mRNA, cleaves of the pre-mRNA between those two signals, bound by the two protein complexes. Once the pre-mRNA is cleaved, the enzyme Poly-A-Polymerase adds a stretch of A ribonucleotides to the 3' tail of the pre-mRNA, adding until 250 adenosines²¹. Longer poly-A tails are associated with more translation, showing how RNA processing can also influence gene expression post-transcriptionally.

The pre-mRNA splicing consists of the removal of the introns of the pre-mRNA and joining together the exons. This process is performed by the spliceosome, a big complex of proteins and small nuclear RNAs.

1.1.2 – GENE REGULATION BY MORE DISTANT ELEMENT : TFs, ENHANCERS AND THE MEDIATOR COMPLEX FOR DNA LOOPING

The simplest transcription event relies on a minimal core complex comprising just six essential proteins and the RNA Pol II in vitro^{1,6-8}. However, transcription undergoes more complex regulation through remote DNA regulatory elements in vivo, known as enhancers or silencers. These elements are bound by specific transcription factors, which in turn recruit both the Mediator complex and RNA Pol II to target

genes²². This additional layer of regulation results in the formation of DNA loops, facilitating enhancer-promoter interactions.

Enhancers/Insulators

The first basic element necessary for DNA looping are enhancers or insulators. On one side, enhancers are short DNA elements, containing binding motifs for TFs, with permissive chromatin. Unlike promoters, enhancers are not restricted to proximity to the genes, but they can be situated at huge distances, sometimes megabases away, upstream, or downstream of the gene they regulate^{23,24}. However, enhancers need to get in close proximity to their target genes to activate transcription by DNA looping²⁵. Surprisingly, enhancers can be transcribed into eRNAs before transcription of their target genes, in both directions, and eRNAs can be polyadenylated^{26,27}. This provides a good approach to detect which genomic regions serve as enhancers^{28,29}. Functionally, enhancers are the place of cooperation between TFs and co-activators through the binding of multiple different factors³⁰. On the other side, insulators are like enhancers, short regulatory regions bound by TFs, but block the interactions between enhancers and promoters³¹.

Transcription factors

Structure

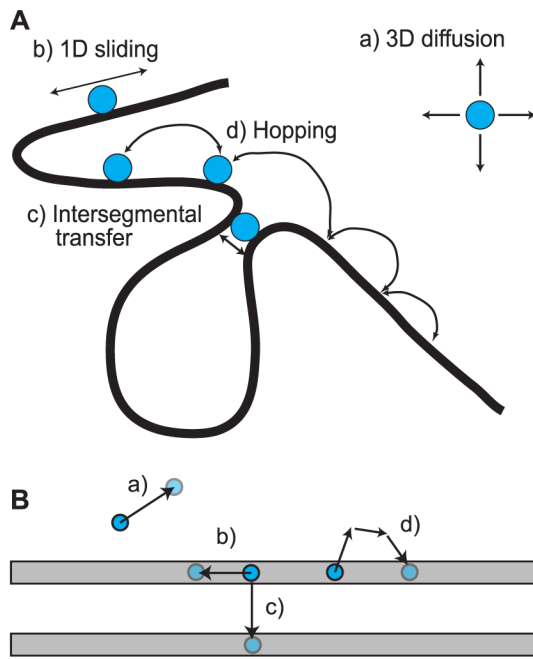
Enhancers regions cannot act on their own to regulate transcription and needs the binding of transcription factors (TFs) to get in closer contact with the promoter³². TFs are proteins that can bind DNA in a sequence-specific fashion and can regulate transcription³³. TFs can be ubiquitous, such as the general transcription factors, like TBP and the general transcription factors of class II^{1,6-8} or gene-specific, such as the pluripotency factors, Nanog, Sox2 and Oct4³⁴.

The structure of TFs is quite simple. In eukaryotes, TFs contain at least one DNA-binding domain (DBD) which binds specifically to short specific DNA motifs, found either in promoter or regulatory regions³⁵. Based on the structure of their DBDs, TFs are grouped into different families^{36 37,38}.

The binding of TFs on DNA is not straightforward to understand, especially in eukaryote genomes. While we could already determine TF motifs and binding positions in the genome early on, molecular biology techniques like ChIP-seq

(Chromatin ImmunoPrecipitation sequencing) have led to great improvements in the field by allowing identification of all binding sites for a specific factor in the genome^{33,39}. However, determine the optimal motif of TF is not enough to know if the TF will bind there, as TF can bind to suboptimal site in vivo⁴⁰. Indeed, TFs can recognize many motifs that are different combinations close to the optimal motif⁴¹. Moreover, some TFs can bind only partially these motifs when they are cooperative binding with another TFs, such as Sox2 and Oct4^{40,42}, or in the case of pioneering factors, that can recognize partial motifs on DNA-bound nucleosome⁴³. Finally, only a small fraction of the motifs recognizable by a specific transcription factor are indeed bound by this specific TF⁴⁴.

TFs have generally a DNA binding consensus sequence, but they do not all bind on it. This raised the question of how a specific TF knows where to bind. Recent advances in the field have shown that another part of the TFs could participate in the target specificity⁴⁵. Indeed, in addition to DBDs, TFs also contains one or more transactivation domain (ADs) that can interacts with co-activators and chromatin remodelers to initiate transcription ⁴⁶⁻⁴⁹. While the structure of DBDs in many transcription factors has been well characterized and classified, ADs are less studied due the lack of defined structures ^{50,51}. Indeed, they are mostly composed of intrinsically disordered regions (IDRs), that takes most part of TF amino-acid sequence⁵². However, even if DBDs are thought to be the element recognizing primarily the DNA binding motif, recent studies suggest that IDRs conduct the location of TFs to their gene target ⁵³⁻⁵⁵. The role of IDRs will be discussed further in a specific subchapter about TF clustering.



On the way to work: Scanning DNA and binding to DNA motif

It is possible to look at the function of TF and their search of a DNA target by a biophysical point of view. In general, transcription factor binding is very dynamics and only interact transiently with DNA ^{56,57}. Different types of mobility for TFs exist in the nucleus including nucleoplasm diffusion, sliding and hops along DNA and longer jump ^{58–60} (**Figure 3**). More specifically, to find their target, TFs can slide on DNA laterally or jump from DNA piece to another locally or further away. ^{57,61–63}.

Figure 3: The different movements that TFs can hypothetically do in the nucleus

Adapted from Schmidt HG, Sewitz S, Andrews SS, Lipkow K (2014) An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding. *PLoS ONE* 9(10): e108575. <https://doi.org/10.1371/journal.pone.0108575> Copyright C 2014 by the Authors. Published by PLoS ONE. Creative Commons license CC BY 4.0.

The same TF can also have different binding patterns temporally ^{61,64}. Indeed, using single-molecule tracking (SMT), studies have quantified two different populations for the same TF: short-term binding events, displayed as non-specific binding and TF “exploration of the genome” and long-term binding events, displayed as specific binding to their motif ^{61,65,66}. However, more recent studies have suggested more complex interactions than

specific and non-specific binding to DNA ^{67,68}. For example, self-interactions of TFs can cause “trapping” of TFs in clusters, slowing down diffusion in the cluster ⁶⁹. Moreover, multiple TFs usually bind in a cooperative manner to enhancers, changing their dynamics⁷⁰.

Functions: Recruitment of co-factors and PIC assembly by TFs

Adding their roles in the general transcription machinery for ubiquitous TFs, gene-specific TFs are the main players regulating cell differentiation⁷¹, reprogramming (in the case of induced pluripotent stem cells) ³⁴, developmental patterning (Hox genes

⁷², stripe-like pattern in fly embryos ⁷³) or specific cellular pathways like the immune response ⁷⁴ or cellular response to specific ligand, like the hormone receptors ⁷⁵.

TFs can work on their own or recruit other factors to help them in their function. To act on such general aspects of the biology of the cell, TFs can work on their own to regulate transcription. For example, some TFs can directly recruit the RNA Polymerase (like TBP) or block directly other proteins from binding ⁷⁶. However, most of TFs recruit other TFs, co-factors, and chromatin remodelers to form multimeric complexes at enhancers and promoters ³⁰. Through these diverse co-regulators, they are involved in chromatin binding, nucleosome remodeling, covalent modifications of histones or other TFs ⁷⁷. A most common recruited complex is the Mediator complex or the histone acetyltransferase p300. Moreover, TFs can also act as several players: they can help each other to bind target sequences. This is called as cooperative binding ⁷⁸.

On the other hand, TFs can have a positive or an inhibitory role. Indeed, TFs can stabilize (activating) or destabilize (inhibiting) RNA Pol II binding directly or indirectly through other factors. However, the notion of positive or inhibitory role of one specific TF is decadent: the same TF can recruit cofactors that either activating or repressing ⁷⁹.

The Mediator complex:

One of the most general co-activators to be recruited by TFs is the Mediator complex. The Mediator complex is a complex of proteins acting as a transcriptional co-activator, discovered in 1990 in the lab of Roger D. Kornberg ⁸⁰. They isolated this complex, as necessary to lead to RNA Pol II transcription in an *in vitro* assay from *S. cerevisiae* compounds ⁸¹. The complex is composed of 21 subunits in yeast and up to 26 subunits in mammals. The Mediator complex is a very conserved complex and is further distinguished in three different functional modules: “head”, “middle” and “tail”. While the tail modules bind activating TFs, the head module is in contact with the PIC and the RNA Pol II ⁴⁸. Interestingly, different TFs bind to different subunit of the Mediator complex ^{82–84}.

The Mediator complex has multiple roles in transcription regulation: PIC assembly, transcription initiation, pausing and elongation, re-initiation, DNA looping and heterochromatin regulation ⁸⁵ (**Figure 4**). However, one of the main roles of the

Mediator complex is to be a bridge between enhancers, bound by TFs and promoters, by direct interactions between the AD domains of TFs and specific subunits of the Mediator complex⁸⁵ (**Figure 4**). A such bridge helps to directly regulate initiation of transcription⁸⁶ through direct interactions with the initiation factors TFIIB and TFIID^{87–89}. Indeed, The Mediator complex assembles extensively with the RNA Pol II^{90–92} around which other factors of the PIC assembles and is required for Pol II transcription^{93,94}. More specifically, the Mediator complex helps in the RNA Pol II recruitment via interactions between the CTD of the Pol II^{81,95–97}.

Although the Mediator complex is a very large complex, changes in Mediator function correlate with structural changes, like the binding with a TF to a specific subunit⁹⁵ and correlates with RNA Pol II transcription⁹⁸. The IDRs of the Mediator complex is an important part of its function, as they are very conserved⁹⁹ and could be part of the reason why the Mediator complex has a such flexibility within its interactions and conformations⁸⁵.

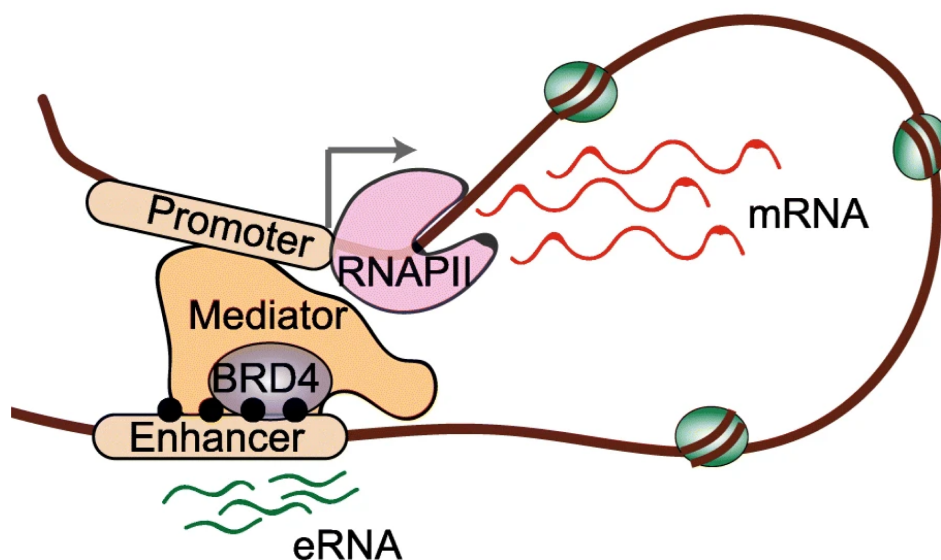


Figure 4: DNA loop between an enhancer and a promoter, mediated by the Mediator complex

Adapted from Circadian rhythms in the three-dimensional genome: implications of chromatin interactions for cyclic transcription, I. Pacheco-Bernal, F. Becerril-Pérez and L. Aguilar-Arnal, 2019, *Clinical Epigenetics*, 11, Article number: 79.

<https://doi.org/10.1186/s13148-019-0677-2>. Copyright C 2019 by the Authors. Creative Commons license CC BY.

Enhancer-promoter communication by DNA looping:

Enhancers can be megabases away from their target genes ¹⁰⁰. DNA looping has been proposed long back as a way of explaining how distant enhancers can influence transcription of genes ¹⁰¹. Experiments have later showed how bound TFs on enhancers could contact the RNA Pol II distantly through DNA looping by electron microscopy in bacteria in vitro ¹⁰². DNA loops became later evident with the emergence of molecular biology methods based on chromosome conformation technologies (the first being 3C and develop in 2002) ¹⁰³ and all following derivatives and improvements, as well as improvements in imaging and especially targeting of specific DNA sequences in FISH ¹⁰⁴.

More specifically, to get those genomic regions close, the linear DNA is thought to make loops in the 3D nuclear space (**Figure 4**). Although the Mediator complex is not required for the formation of DNA loops, it is very important for their formation and stabilization ^{105,106}. Other than the Mediator complex, numerous other factors are thought to mediate close contacts between enhancers and promoters ¹⁰⁷. The beta-globin locus is a typical example of DNA looping between enhancers and promoters and these loops are promoted by the factors EKLF and GATA-1, through the recruitment of BRG1 ¹⁰⁸.

Other than protein factors, gene looping may also be correlated with a specific type of long intergenic ncRNA called activator RNA (aRNA) ¹⁰⁹. Moreover, transcription of enhancer sequences into eRNAs correlates with enhancer-promoter looping as well (**Figure 4**) ²⁷. In both cases, Mediator would be necessary for enhancer-promoter looping dependent on RNA. ¹¹⁰

One of the main ideas about enhancer-promoter communication using DNA looping is that both needs to be in close contact. However, it is unclear if the distance between promoters and enhancers is correlated with transcription. ³² Among all the studies, conclusions are divergent for different genes, or even for the same gene in different developmental context. ³²

The vision in which promoters and enhancers communicate has evolve quite a lot in the last two decades. Not only one enhancer controls on gene: multiples enhancers can regulate one single gene and one enhancer can regulates multiple genes. ^{111,112} Moreover, more complex types of interactions have been unraveled, in structures called super-enhancers (SEs), that will discuss in a specific chapter ¹¹².

1.1.3. ANOTHER LAYER OF COMPLEXITY : CHROMATIN, CHROMATIN MARKS AND CHROMATIN REMODELERS

In eukaryotes, the DNA is not naked but wrapped around histones proteins. It is a natural way of inhibited transcription as the PIC cannot form on nucleosomal DNA, in the contrary of naked DNA ¹¹³. Indeed, promoter access by RNA Polymerases and the pre-initiation complex is blocked by histone complexes.

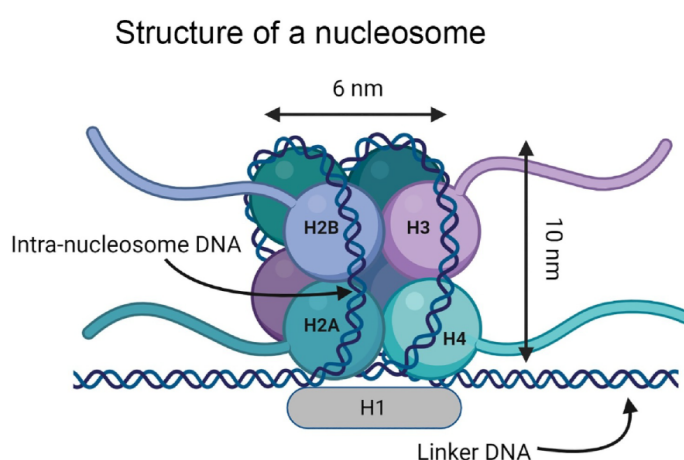


Figure 5: Structure of a nucleosome

Adapted from Histone post-translational modifications as potential therapeutic targets for pain management, J. V. Torres-Perez, 2021, Trends in Pharmacological Sciences, Volume 42, Issue 11, P897-911. <https://doi.org/10.1016/j.tips.2021.08.002>. Copyright C 2021 by the Authors. Published by Elsevier Ltd. Creative Commons license CC BY 4.0.

Nucleosomes

In eukaryotes, DNA is wrapped up around histone complexes, association of both elements being called nucleosomes and consist of the basic repeating subunit of the chromatin (**Figure 5**). On one side, histones are composed of 4 core units, two H2A-H2B dimers and a H3-H4 tetramers. Histones have tails consisting of amino acids that are unstructured and exposed, allowing interactions with regulatory proteins. On the other side, DNA is tightly wrapped around each nucleosome by 1.7 super helical turns (around 147 pb of DNA) ¹¹⁴. Some naked DNA between each nucleosomes complexes is called the linker DNA, which length vary. Nucleosomal arrays or « beads-on-a-

string » are the first level of chromatin structure, referred as the 11 nm filament (**Figure 5**).¹¹⁴. The histone H1 is outside of the core histones, bound to the linker DNA and organized another level of compaction, the 30 nm diameter chromatin structure, resulting from the compaction of several nucleosomes together¹¹⁵. Other chromatin fiber organization were proposed but their universality and arrangements are still debated¹¹⁶. In addition to the consensus core histone H2A, H2B, H3 and H4, some other types of histones can be switched and are called histone variants. Histone variants can bring another layer of regulation and can be specifically associated to specific functions. For example, the histone variant H2A.Z promotes the formation of higher-order chromatin structure and are associated with constitutive heterochromatin domains¹¹⁷. Another example is the histone variant H3.3, that is generally correlated with gene activation¹¹⁸.

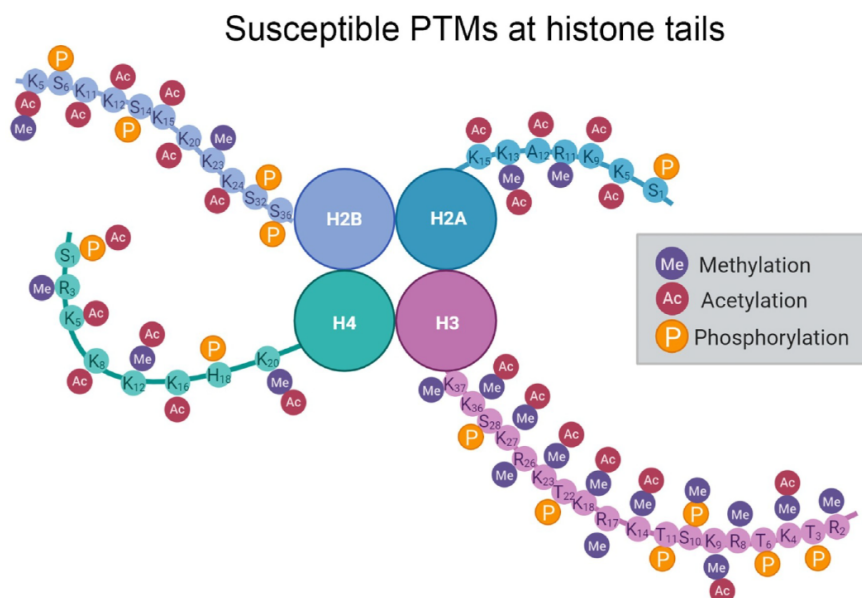


Figure 6: Possible chromatin marks at histone tails for acetylation, methylation and phosphorylation.

Adapted from Histone post-translational modifications as potential therapeutic targets for pain management, J. V. Torres-Perez, 2021, Trends in Pharmacological Sciences, Volume 42, Issue 11, P897-911. <https://doi.org/10.1016/j.tips.2021.08.002>. Copyright C 2021 by the Authors. Published by Elsevier Ltd. Creative Commons license CC BY 4.0.

Chromatin marks

Histones can be modified on their free terminal amino-acid regions (histone tail), their surface and their lateral side by a large variety of post-translational modifications, among the most common being acetylation, methylation, phosphorylation, and ubiquitination ¹¹⁹ (**Figure 6**). New categories of histone modifications are constantly discovered. Adding the number of possible modifications and the number of possible modifiable amino acid residues, more 500 possible modifications have been described ^{119,120}. As one histone can be modified on several amino acids residues and nucleosome counts 8 histones, the number of possible combinations is almost unlimited (**Figure 6**). With such possible modulations, it is thought that the chromatin marks bring another layer of complexity and increase greatly the coding potential of DNA code only. This additional information is wrapped under the hypothesis of the « histone code » ^{119,121}.

Each modification can change differently the chemical nature and the conformation of histones, as well as the level of DNA wrapping around the histones. Moreover, it can also serve as a binding platform to numerous regulatory proteins ¹¹⁹. Modifying the histone tails can lead to new interactions with other proteins, that can directly regulate gene expression and change the chemical properties of the chromatin, for example by changing the level of compactness.

Chromatin marks can change the chemical properties of the chromatin and ARE tightly correlated to transcription regulation. For example, histone acetylation loosens the chromatin and increase DNA accessibility. It is thus a chromatin mark related to transcriptionally active regions. Enhancers are wrapped around nucleosomes that are highly acetylated, especially on the Lysine 27 of Histone 3 (H3K27Ac). Due to this loose chromatin, enhancers are more accessible as proved by their sensitivity to Dnase I treatment (Dnase-I hypersensitive site sequencing) ¹²² and transposase-accessible chromatin using sequencing (ATAC-seq) ¹²³.

Repressed chromatin states are created by methylation. The genome is organized in two distinctive phases: euchromatin and heterochromatin. The heterochromatin can be facultative or constitutive. Each of those two subsets of heterochromatin is associated with a specific histone modification: facultative: H3K27me3 and constitutive: H3K9me3. The presence of these marks on the chromatin leads to the recruitment of specific factors, leading to chromatin

compaction. By contrast, euchromatin intergenic regions are enriched for H3K36me2 and genes bodies for H3K36me3 ¹²⁴.

Certain chromatin marks are specifically associated with specific functional DNA regions. Enhancers are enriched for H3K4me1, H3K4me3 and H3K27Ac ¹²⁵. On the other hand, H3 lysine 4 di and tri-methylation (H3K4me2/3) are strongly correlated with TSS of active genes. Finally, H3K79me2/3 nucleosomes are associated with active chromatin domains. Certain marks of histones have a direct role in chromatin compaction. For example, acetylation relaxes the nucleosomes and avoid the folding into the 30-nm fiber type. Acetylation of the histone H4 on lysine 16 (H4K16ac) can inhibit the formation of higher-order chromatin structures in specific conditions ¹²⁶.

Chromatin marks can be passed on during cell division. During cell division, new histones are being added to the newly synthesized DNA. Chromatin marks are stably inherited after DNA replication to reconstitute heterochromatin domains ¹²⁷.

DNA methylation

Although not a chromatin marks, DNA methylation is an essential epigenetic modification. Cytosine on CpG islands can be methylated. It was shown to cause closed and rigid nucleosome structures and to influence dynamics of mononucleosomes ¹²⁸.

Chromatin remodelers and histone modifiers

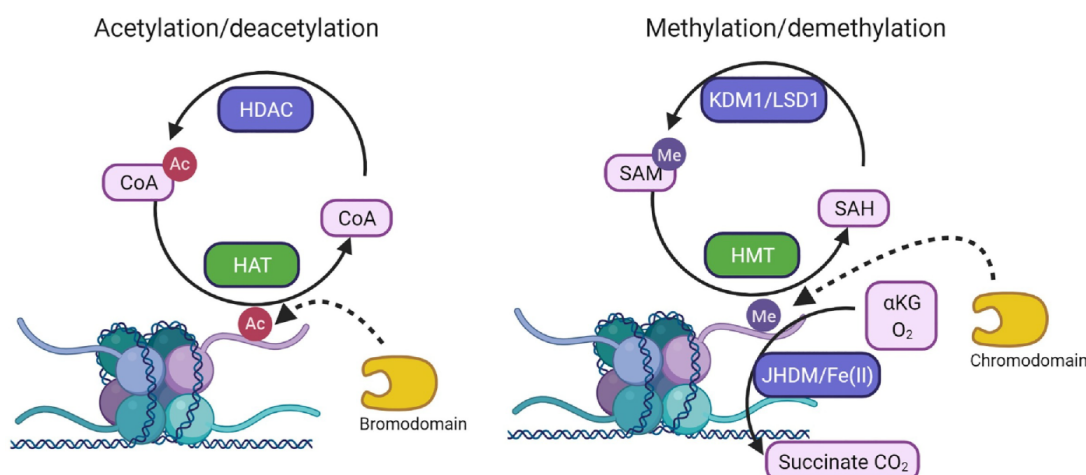


Figure 7: Chromatin writers, chromatin erasers and chromatin readers for acetylation and methylation at histone tails

Adapted from Histone post-translational modifications as potential therapeutic targets for pain management, J. V. Torres-Perez, 2021, Trends in Pharmacological Sciences, Volume 42, Issue 11, P897-911. <https://doi.org/10.1016/j.tips.2021.08.002>. Copyright C 2021 by the Authors. Published by Elsevier Ltd. Creative Commons license CC BY 4.0.

Several kinds of enzymes can act on histones at three different levels: 1) change histone composition, for example exchanging canonical histone for histone variants, 2) change the chemical composition of histones, such as adding or removing histones marks, and 3) change their position, such as sliding, add or evict nucleosomes¹²⁹. Such modifications are carried out by two functional types of enzymes: the histones modifiers and the ATP-dependent chromatin remodeling complexes. Enzyme complexes that can change the histone composition or to change their position are called chromatin remodeling complexes¹²⁹. On the other hand, enzyme complexes that are changing the chemical nature of the histones, by adding or removing chromatin marks are called chromatin writers, chromatin erasers and chromatin readers¹²⁹.

Histones remodeling complexes

DNA being wrapped around histones, it is necessary to have protein complexes that can act directly on histones to enable diverse nuclear function such as transcription, DNA repair or DNA duplication. Such protein complexes are called chromatin remodeling complexes, and all contain a subunit with an ATPase domain¹³⁰. They are responsible for the removal, slicing or remodeling of nucleosomes and allow creation of nucleosomes-free regions, that can be later bound by TF and RNA Pol II and thus transcribed. They are addressed to specific part of the chromatin by recognizing specific signals, like histone modifications, specific DNA sequence, non-coding RNAs, histone variants or other DNA bound proteins like TFs¹³¹. In eukaryotes, there are four families that can be defined by the specific domains they are composed: SWI/SNF (switching defective/sucrose non-fermenting), CHD, ISWI and INO80 families. Specifically, the SWI/SNF families contain respectively a bromodomain and chromodomain, that recognize post-translational modifications on the histones¹³¹.
(Figure 7).

Histone modifiers

As previously described, the chromatin can be in different state of compaction. Histone modifications are one of the regulation mechanisms by which chromatin is more or less compacted. More precisely, three kind of enzyme complexes play a role in histone modification regulation: chromatin writers, chromatin erasers, chromatin readers that can respectively add, remove, or recognizes histone post-translational modifications (**Figure 7**).

Chromatin writers are classified based on the specific histone substrates they add on histones, such as histone acetyltransferases (HATs) for acetylation or histones methyltransferases for histone methylation ¹³². HATs include for example the TFIID subunit TAF_{II}250 ¹³³ or p300/ CREB-binding protein (CBP) ¹³⁴. Other than adding histone marks, histone writers can also be very important regulation of transcription. For example, the co-activator p300/CBP can regulate genes by helping the assembly of the PIC and by acetylating histones to open chromatin. First, p300/CBP acts like the Mediator complex as a bridge to connect TFs and other co-activators or the PIC. This leads to a stabilization of the PIC and promotes transcription. Second, p300 is a HAT and can recognize the H3K27Ac mark. It is interesting to notice that p300/CBP is a highly disordered protein ¹³⁵ . We will discuss later the role of protein disorder in transcription regulation.

Chromatin erasers are, like chromatin writers, classified upon the histone substrates they remove from histones, such as histone deacetylases for acetylation or histone demethylases for methylation (**Figure 7**).

Finally, chromatin readers mediate the action of chromatin writers and erasers by recognizing the histone modifications and activate the following consequences. Chromatin readers have in finally many diverse functions like chromatin remodeling or direct transcription regulation. For example, the chromatin remodeler complex SWI/SNF contains bromodomain, that are important for proper binding on the chromatin and subsequent remodeling activity ¹³⁶. In term of transcription regulation, the BET (bromodomain and extra terminal domain) family, including its member BRD4, can target histone modification, especially H3K27Ac, and directly regulation transcription by interacting with the transcription machinery ⁴⁹.

Non-histone architectural proteins

Other protein complexes, while directly being involved in histone modifications, are important for changing the structure of the chromatin and have a direct role on many nuclear processes. The HMG (high mobility group) superfamily is composed of three families: HMGA, HMGB and HMGN¹³⁷. HMG proteins do not have a specific binding site. They all have a unique functional protein motif, leading to specific nuclear functions. Such examples of nuclear function are altering the chromatin structure to facilitate binding of other factors¹³⁷ through DNA bending activity, formation of enhanceosomes¹³⁸ or compete with other proteins for chromatin binding, for example with linker histones¹³⁹. They can also decrease the compactness of the chromatin fiber, therefore increasing the access to transcription factors¹⁴⁰.

1.2 NUCLEAR ORGANIZATION: THE RUSSIAN DOLLS

Some researchers have asked how transcription managed to stay organized when we observe the high number of factors required for its process to happen¹⁴¹. Inside this closed compartment of bilayer membrane, it is organized even more in an infinite number of smaller compartments without physical membranes.

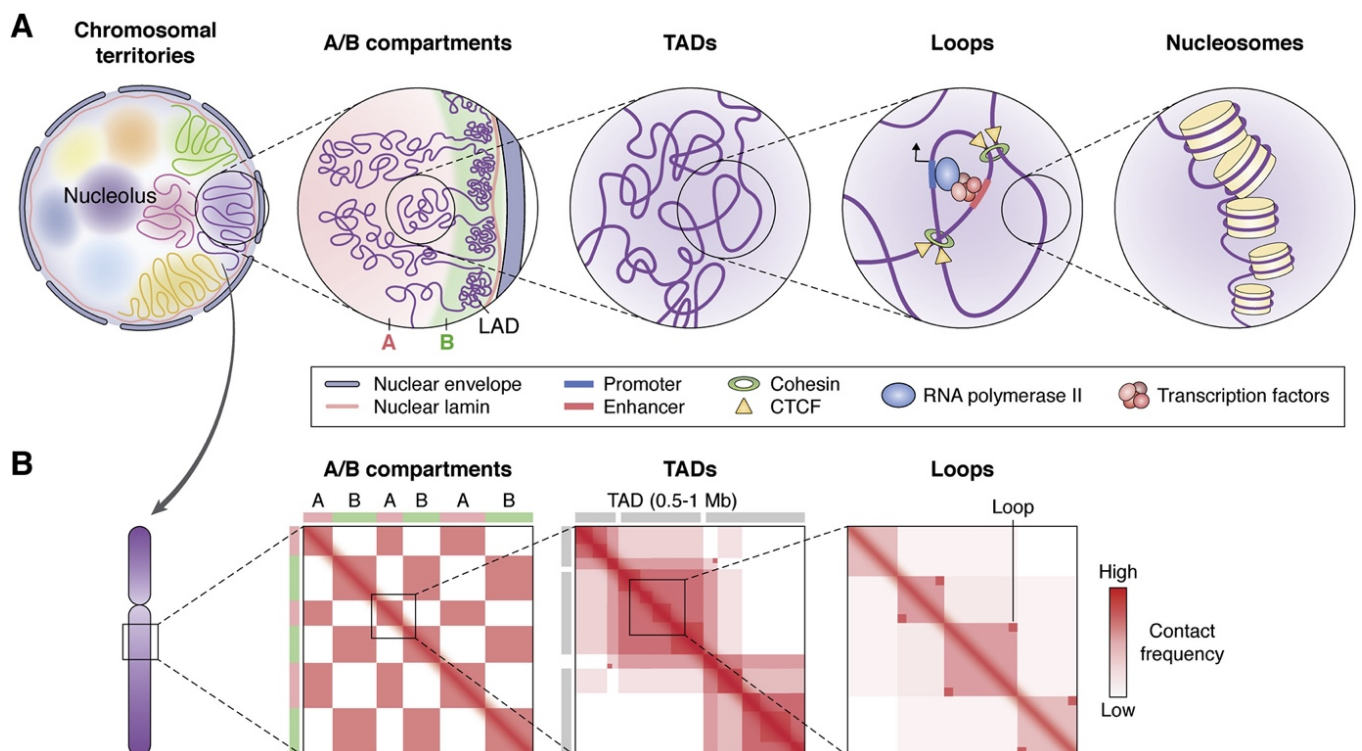


Figure 8: Genome organization at multiple scales: from chromosomal territories to DNA loops

From Emerging roles of epigenetic regulation in obesity and metabolic disease, Y. J. Parl et al; 2021, JBC Reviews, Volume 297, Issue 5, P897-911.
<https://doi.org/10.1016/j.jbc.2021.101296> Copyright C 2021 by the Authors. Published by Elsevier Inc on the behalf of American Society for Biochemistry and Molecular Biology. Creative Commons license CC BY 4.0.

1.2.1. CHROMOSOME TERRITORIES

In eukaryotes, chromosomes occupy specific compartments, called chromosome territories ¹⁴². Techniques such as Hi-C or chromosome painting have unraveled chromosomes organization in the nucleus ¹⁴². In differentiated cells, active genes are typically located between chromosome territories while repressed regions tend to be located inside territories or at the nuclear periphery ¹⁴³. Chromosomes with similar features tend to be located next to each other and in specific part of the nucleus¹⁴⁴. For example, chromosomes containing a few numbers of genes are more located towards the periphery, at the nuclear lamina while gene-rich chromosomes are more centrally located in the nucleus. While the position of a specific chromosome is not defined across cell of the same population¹⁴⁵, their position is quite stable during the same cell cycle ¹⁴⁶ (**Figure 8A**).

1.2.2. HETEROCHROMATIN AND EUCHROMATIN

The concepts of heterochromatin and euchromatin covers multiple similar aspects to the A and B compartments. The terms themselves come from the visual description of nuclear structures imaged in 1928 by electron microscopy ¹⁴⁷. Some “heterochromosomes” were highly packed during the whole cell cycle, in contrast with the other part of the chromatin ¹⁴⁷. Works in drosophila have later demonstrated that different chromosomes were bearing different number of heterochromatin regions, systematically or in specific to certain cell types, giving rise to the term heterochromatin facultative (cell-type specific) or heterochromatin constitutive (always present in all cell types) ¹⁴⁸, and that the relocation of genes from euchromatin to heterochromatin could lead to functional gene silencing ¹⁴⁹. Later more cytogenetic studies have correlated electron-dense regions present on chromosomes and gene silencing, deepening slow our understanding of such heterochromatin and

euchromatin, with for example the heterochromatinization of the X chromosome in cat neuronal cells, the Barr body, and the hypothesis of dosage compensation ¹⁵⁰. Nowadays, with the development of molecular biology techniques like ChIP-seq ^{33,39} and the proposition of a “histone code”¹²¹, our understanding of the chromatin substructure and its relation to function have greatly improved. The historical description of euchromatin and heterochromatin facultative or constitutive are associated with specific histone marks (as described in Chapter 1.1.3) and recognized by effector proteins (chromatin writers, erasers, and readers) (**Figure 7**) that can change the local chromatin structure and compaction, leading to control of gene expression levels locally. For example, constitutive heterochromatin is typically associated with the Heterochromatin Protein 1 (HP1), and the histone marks H3K9me2/3 or H4K20me2/3 ¹⁵¹. These chromatin regions contain mostly repetitive sequences and concentrate at the centromeres and telomeres and is proposed to be related to genome stability ^{152–154}, by suppressing recombination and transcription of repetitive sequences ¹⁵⁵. In the contrary, facultative heterochromatin is associated with the PcG protein complexes and the histone mark H3K27me3 in the repressed state, but can switch to an active chromatin state, becoming euchromatin in specific conditions ¹⁵¹. Recently, a third heterochromatin type was described as “black” heterochromatin ^{152,156}, which does not present any enrichment for certain histone marks, contains low-expressed genes and are highly associated with the nuclear lamina ^{152,156}.

1.2.3. A/B COMPARTMENTS

Euchromatin and heterochromatin are mutually exclusive in the nucleus and by this mean, organize the genome into separated regulatory domains in 3D. This organization in the 3D genome can be recapitulated by the concept of A and B compartments ^{157,158}. In the nucleus, chromatin is typically separated in two compartments: active (A) and inactive (B) ¹⁵⁹ (**Figure 8A, B**). Such compartments have been brought to light by the first Hi-C studies of the human genome ^{157,160}. The chromatin regions from the A compartment tend to interact more with chromatin regions from the same compartment, then with chromatin regions associated with B compartment ¹⁵⁷. In general, genomic regions from the A compartment tends to be more gene-rich, display a higher GC-content, bear histone marks for open chromatin

and active transcription and be more centrally localized in the nucleus. By opposition, genomic regions from the B compartment are the opposite, with low density of genes, histone marks related to compacted chromatin state and gene silencing, and more displaced towards the periphery of the nucleus, mainly consisted of LADs ¹⁵⁷.

1.2.4. TOPOLOGICALLY ASSOCIATED DOMAINS

At a smaller scale than A/B compartments, the genome is further finely organized at the levels of chromosomal regions. Topologically Associated Domains (TADs) are large 3D genome structures, typically ranging from tens to hundreds of kilobases (**Figure 8A**) ^{161,162}, where the DNA interacts more frequently than any other DNA outside the TADs. The boundaries of TADs are enriched for insulator proteins, like CTCF, that keep the TADs to interact with one another and the cohesion complex (**Figure 8A, B**) ^{162,163}, as removal of these proteins lead to the disruption of the structure of the TADs ^{164,165}. To form such TADs, the DNA is extruded by the cohesion complex through a DNA loop extrusion mechanism, stopped by the encounter with the TAD boundary elements, such as CTCF ¹⁶⁶. Genomic DNA is organized in a such way to facilitate promoter-enhancer interaction through DNA looping ¹⁶⁷. Indeed, an enhancer located in one TAD interacts preferentially with the promoters inside the same TADs ¹⁶⁸. Moreover, genes that are co-regulated are often found in the same TAD as well ¹⁶⁹.

Interestingly, while removal of TAD boundaries causes developmental problems in mouse, certainly due to the wrong enhancer-promoter interactions ^{170,171}, global depletion of CTCF or cohesion change widely the TAD structure in vertebrates but does not have a strong effect on gene expression, except on certain genes ^{172,173}. These studies suggest that TADs are only important for the expression of a small set of genes. CTCF and cohesion are ubiquitously expressed and more involved in constitutive interactions between enhancers and promoters ¹⁷⁴. Some cell-specific interactions can be performed by other proteins that CTCF, along with the cohesion complex. For example, the expression of Sox2 in embryonic stem cells is dependent on the promoter-enhancer interaction mediated by cohesion-Mediator complexes ¹⁷⁴. Other examples show complete independence with the cohesion complex with enhancer-promoter loops mediated by tissue-specific TFs, such as GATA1 at the b-globin locus ¹⁷⁵.

1.2.5. SUPER-ENHANCERS AS A WAY TO REGULATE MASTER GENES

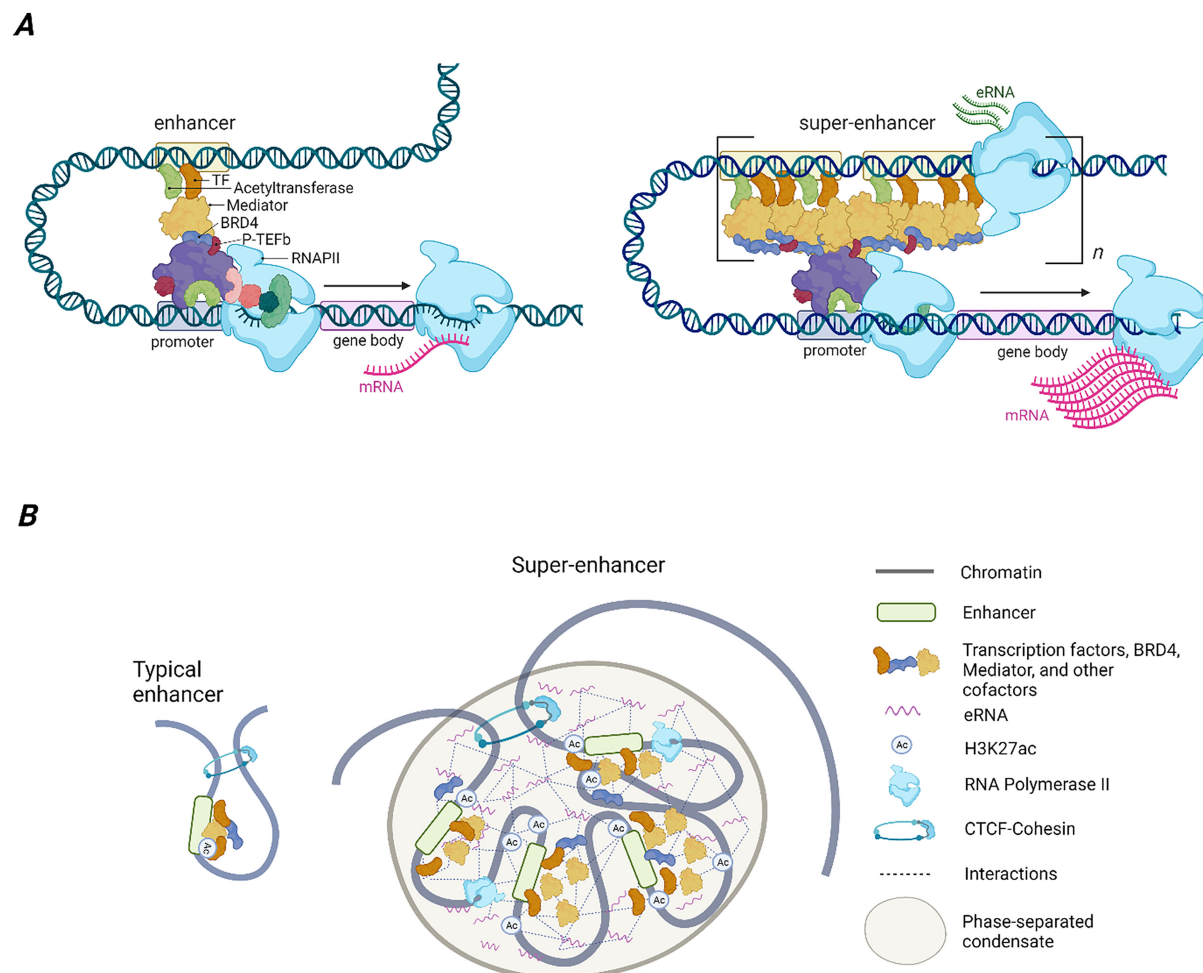


Figure 9: Structure and components of a conventional enhancer *versus* a super-enhancer

From Sonia Dębek, Przemysław Juszczynski, Super enhancers as master gene regulators in the pathogenesis of hematologic malignancies, *Biochimica et Biophysica Acta (BBA) – Reviews on Cancer*, Volume 1877, Issue 2, 2022, 188697, ISSN 0304-419X, <https://doi.org/10.1016/j.bbcan.2022.188697>. Copyright C 2022 by Elsevier. Creative Commons license CC BY 4.0. Authorization of reuse obtained by Third Party Permission and Reprints, RightsLink.

At the intermediate between TADs and DNA loops are found a recent model proposed to regulation transcription, called super-enhancers (SEs). They have first been described in the context of pluripotent gene programs in embryonic stem cells (ESCs) and cell identity in more differentiated cells. Master transcription factors like Oct4, Sox2 and Nanog, and the co-activator Mediator are forming high local

enrichment at the enhancers of genes that control the pluripotent state (**Figure 9A**)¹⁷⁶. These enhancers are also special by their larger size, their high transcription factor density and variability of content, as well as their ability to regulate transcription, in comparison with regular enhancers (**Figure 9A**)¹⁷⁶. Nowadays, the definition of a super-enhancer is still empirical and based on two criteria: 1) a region containing a lot of enhancers (up to 12.5 kb, 2), detected usually by H3K27Ac ChIP-seq peaks and 2) regions with high binding of master regulators that are important in specific cell subtype (like Oct4, Sox2 and Nanog in ESCs) and presence of activators like the Mediator complex and p300¹⁷⁷. Interestingly, some master TF active at SEs participate to their own transcription regulation by SEs, revealing a positive feedback mechanism¹⁰⁶.

However, the definition of SE is controversial as the criteria to define them are quite subjective and many studies use this concept with different definitions¹⁷⁸. Yet the difference in term of transcription activity between conventional enhancer and SEs is quite striking, especially related to the number of transcription factors and co-activators associated to it (**Figure 9A**). In the case of conventional enhancer, the transcription activity can be linearly related to the transcription activity, while in the case of SEs, the relation is sigmoidal, encouraged by cooperativity between the TFs and co-activators¹⁷⁶. Regulation at SEs was proposed to be regulated by Liquid Liquid Phase Separation (LLPS). Indeed, the master TFs and co-activators present at SEs have usually a high valency (how many interactions can do with other factors) and are present in high number¹⁷⁶. Transcription starts abruptly and range rapidly to a maximum when the TFs and co-activators are in a phase-separation state, in a switch-like manner¹⁷⁹. This model explains why the expression of SE-regulated genes is so much affected by small changes in TFs level and removal of some enhancers^{176,180}. Interestingly, super-enhancers are typically found inside sub-TADs, a loop formed by two interacting CTCF sites, co-occupied by the cohesion complex (**Figure 9B**). Loss of those CTCF boundaries leads to a decrease in transcription of the SE associated genes but conduct to an activation of neighboring genes outside of the sub-TAD¹⁸¹. This study connects the model of LLPS in gene regulation and genome organization at the gene level, as well as the importance of keeping the integrity of sub-TADs at SEs for proper transcription. More recently, a study showed that among all super-enhancers, some of them were non-hierarchical. More precisely, inside each SE, one enhancer could make more contacts with the other enhancers than expected. This

specific “hub-enhancers” are highly enriched for interactions with CTCF and Rad21 while they are usually only found at TAD boundaries. Removal of these hub-enhancers by knock-out resulted in decrease of transcription and change in the local chromatin landscape¹⁸². This study reveals how much CTCF and Rad21 are important for SE function, as well as the role of enhancer is primarily seeding super-enhancers and high levels of TFs and co-activators.

1.2.6. CLUSTERING OF THE TRANSCRIPTIONAL MACHINERY IN THE NUCLEUS

Recent studies have suggested that SEs are related to condensates by cooperative gathering of many factors from the transcriptional machinery^{46,47,49}. However, clustering of the transcriptional machinery has been observed much before the discovery of SEs. The first time that transcription factor clustering was observed was in 1995, where the glucocorticoid receptor (GR) was observed in discrete clusters using fixed imaging in humans’ cells. However, at this time, they concluded that these clusters did not have a direct role in transcription activation because they did not colocalize with RNA Pol II clusters or with the splicing factor SC-35¹⁸³. Since then, clustering has been observed for many TFs and co-activators in many different organisms. Along with a such variety of proteins and model organism, many words have been used to characterize the TF and co-activator clustering: clusters, hubs, condensates, punctate, foci, aggregates, droplets... Some of these words are usually associated with a specific meaning on how those clusters are formed¹⁸⁴. Here I will use the word clustering as I do not make any assumption on the biophysical nature of protein interactions within the clusters, and I refer only to high local accumulation of proteins at the molecular level. I will however mention the conclusion of each individual study if they mention such interactions.

Almost all type of factors involved at any step of transcription can form clusters: TFs, downstream proteins involved in signaling pathways, co-activators, chromatin adapters and elongation factors¹⁸⁵. Especially, one important protein complex that forms clusters is the RNA Pol II. The RNA Pol II has been shown to cluster as early as 2002¹⁸⁶ and then studied in many papers¹⁸⁷⁻¹⁹². A study showed that the CTD of RBP2 can undergo LLPS in vitro. They further showed that this CTD is essential for the

formation of RNA Pol II clusters in human cells. Indeed, by truncation of the CTD, they showed that the CTD length controls the RNA Pol II clustering as well as its diffusion in the human cells ^{193,194}. They finally showed that CTD phosphorylation in Serine 5 leads to the dissolution of the RNA Pol II droplets in vitro ¹⁹⁴. These results have been later taken in a review where they suggest that several kinds of droplets can be used to go from initiation to elongation of transcription ¹⁴¹. Moreover, clusters of RNA Pol II and Mediator localize where transcription happens ¹⁹⁰, suggesting that the RNA Pol II co-clusters with other proteins.

1.3. TRANSCRIPTIONAL MACHINERY CLUSTERING

Models of molecular diffusion and affinity models alone have difficulties to explain how transcription factors can bind to their specific targets so fast. Indeed, the rate of binding is more than 1000-fold faster than what is predicted with diffusion and binding affinity alone ¹⁹⁵. Moreover, more recent studies have suggested more complex interactions than specific and non-specific binding to DNA^{67,68}. Such complex interactions are happening particularly inside confined environment.

1.3.1. FROM STOICHIOMETRIC INTERACTIONS TO LIQUID-LIQUID PHASE-SEPARATION

A new paradigm about how the nuclear is organized has been raised in the last decades. Development of new genomics and microscopy methods has highlighted specialized compartments of DNA, RNA, and proteins in the nucleus, concentrating the molecular players to similar functions in 3D nuclear territories ¹⁹⁶.

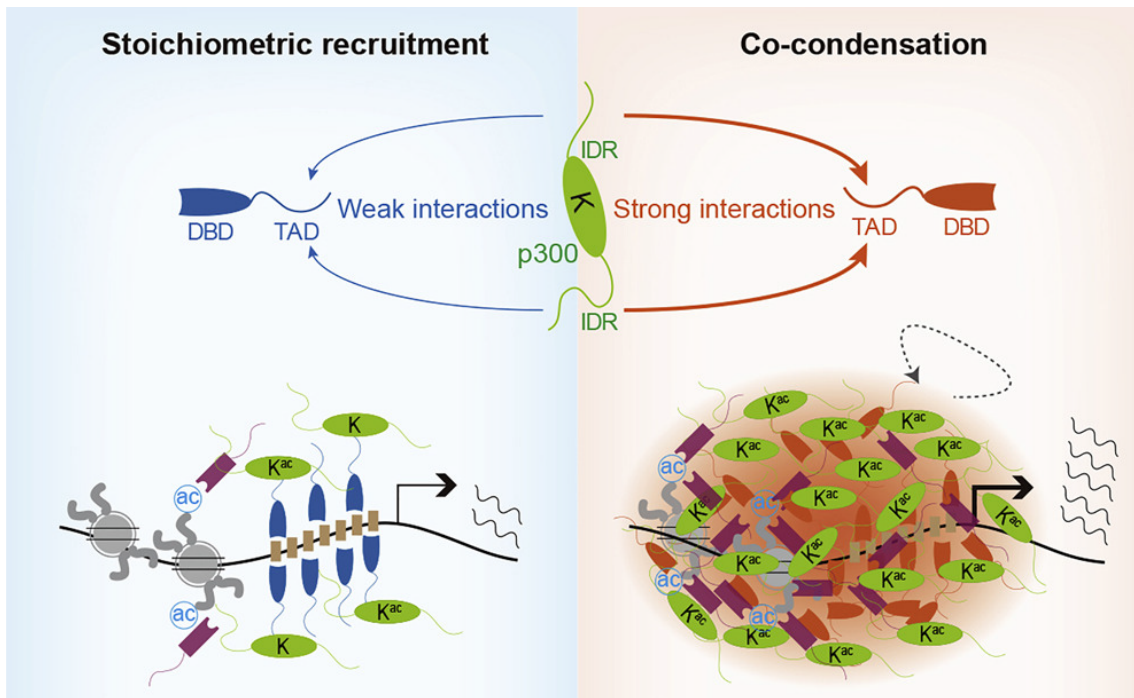


Figure 10: Examples of stoichiometric recruitment and co-condensation through multivalent interactions in the context of transcription

Adapted from Liang Ma, Zeyue Gao, Jiegen Wu, Bijunyao Zhong, Yuchen Xie, Wen Huang, Yihan Lin, Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics, *Molecular Cell*, Volume 81, Issue 8, 2021, Pages 1682-1697.e7, ISSN 1097-2765, <https://doi.org/10.1016/j.molcel.2021.01.031>. Copyright C 2021 by Elsevier BV. Creative Commons license CC BY 4.0. Authorization of reuse obtained by Third Party Permission and Reprints, RightsLink.

Classical view of the macromolecule complex formation: high-affinity interactions, specificity and stoichiometry

We classically imagined that complexes of DNA, RNA and proteins are built through stoichiometric interaction: the number of each component is proportional among themselves. Moreover, the interactions are thought to be conjugated through well-defined structured domains and interactions done through high affinity chemical interactions ¹⁹⁶ (**Figure 10**, blue part). Examples of complexes with fixed stoichiometry are for example the RNA Pol II complex, constituted of 12 proteins interacted with high affinity, along with general transcription factors like TFIIB ¹⁹⁷.

Low-affinity and non-stoichiometric interactions via low complexity domains

In the recent years, a new word came into the field to describe assembly of molecules interacting mostly through low-affinity interactions with IDRs, cooperative and multivalent interactions: “condensate”¹⁹⁸. The word condensate is used for such locally rich assembly of molecules that have a variable stoichiometry¹⁹⁹. For proteins, typically, such interactions are mediated through charged and polar amino acids.²⁰⁰ In the case of TF, it is thought that TFs condensates together by a facilitated interaction through their low complexity domains or intrinsically disordered regions (IDRs): regions that do not fold in a specific structure like usual domains (alpha-helix, beta-sheet...). IDRs contains amino acid with many charges and potentially forming hydrogen bonds, forming weak interactions of diverse chemical nature, but not covalent, called multivalent interactions. However, clustering can also occur outside of IDR interactions^{201,202}. The ADs of transcription factors, typically contains IDRs with a certain enrichment in proline, serine, threonine or glutamine residues, or special secondary structures^{49,203–205}. ADs and other IDRs are suggested to facilitate TF self-interactions, promoting clustering or phase-separation^{199,206}.

Liquid-liquid Phase-Separation

Molecules can undergo phase-separation and they can exist in different physical states: liquid, solid or “gel-like” state. Stronger the interactions between the components, more rigid the structure is. It has been suggested since years now that many compartments in the cell are membrane-less and hold through such a process called liquid-liquid phase separation (LLPS)¹⁸⁴. Many structures in the cell and in the nucleus are considered to follow the rules of LLPS. One of the first discovered was the formation of P granules in *C. elegans*, now used as a basic model to study LLPS²⁰⁷. The first structure related to transcription regulation that has been described is the transcription occurring in the nucleolus²⁰⁸. Many recent studies in the field of transcription have demonstrated that LLPS seems to be a widespread mechanism for transcription regulation, including the RNA Pol II²⁰⁹, co-activators such as Brd4⁴⁹ and the mediator complex^{46,47}, transcription factors like OCT4⁴⁶ and TAF15²¹⁰, splicing factors like SRSF1 and SRSF2⁴⁷, RNA processing factors like PTBP1²¹¹ and even

genome organization, especially chromatin with HP1^{212,213} and the Polycomb proteins^{214,215}.

Recognizing LLPS is based on the ability of condensates to fuse, do fission and specific diffusion pattern²⁰⁷. They also display a spherical surface tension with the surrounding medium²⁰⁷. As part of a liquid state, components are uniformly distributed inside the condensates²¹⁶. However, the rules of detected phase-separation *in vivo* are not clear^{217,218} and studying phase-separation *in vivo* remains very challenging^{216,218}. To form LLPS condensates, the components must be found in a high concentration in the same local space. However, a such high concentration is very difficult to reach and maintain in the crowded nucleus, especially by low-energy interactions. Often, such nuclear compartments need the seeding through higher affinity interactions with constrained structure, for example TFs binding to DNA, histones modifications or the nuclear lamina^{219,220}.

1.3.2. MECHANISM AND REGULATION OF TF CLUSTERING

TF clustering in the nucleus occurs often at specific genomic location as shown by recent studies and depends largely on interactions between IDRs^{46,49,209}. However, such clusters need to be seeded by DNA elements in other to facilitate the formation process. A recent study explored the DNA features that facilitate such clustering²²¹. They showed that the TFBS number, density and specificity are important to drive cluster formation. By comparing these features and the ones found at enhancers, they observed strong correspondences, suggesting that enhancers process the DNA features necessary for TF cluster formation. Moreover, they conclude that both TF-DNA interactions and the IDR-IDR interactions are important for cluster formation²²¹.

However, other studies that have used deletion mutant of TF and co-activators forming clusters in the nucleus have shown discrepancies in these results. A study using 75 IDRs from TF during early development in fly have shown that the IDRs alone are not sufficient to drive clustering in the nucleus²²². Especially, other studies have demonstrated that IDRs are essential for clustering and subsequent function of the protein. For example, the core IDR of tumor suppressor UTX is important for clustering of UTX. Deletion and mutagenesis of this IDR lead to a lack of clustering and dysregulation of the tumor suppression effect of UTX²²³. Another work showed

that the oncogenic transcription factor YY1 transactivation domain is essential for cluster formation, especially a histidine cluster from this AD ²²⁴.

Interestingly, some other studies have focused on the role of the DNA binding domain is clustering and have found that they are essential to make clusters. A point mutant in the endogenous Ubx TF, unabling the DNA binding, did not lead to cluster formation in fly ²²⁵. In another study, they used the TF Sox2 and the co-activator Brd4 to study which protein domain is important for clustering. For both, the point mutation or removal of DNA binding domain led to the decrease in the density of clusters in the nucleus while the removal of the AD did not affect clustering in mESC nuclei ²⁰¹. Using quantitative imaging in living yeast cells and truncation mutants, a study showed that the TF Gal4 needs both DNA binding and IDRs to form cluster in vivo²²⁶. A last study shows that the glucocorticoid receptor forms clusters in humans' cells dependent on phase-separation. They showed that the DBD but not the IDRs, were essential for cluster formation²⁰². All the studies together show that there is no consensus in which domain of the TF is essential for clusters, but that it could depend on each specific TF.

RNA has been shown to enhance the cluster of some proteins, especially related to the LLPS model ²²⁷. Such a model has been also demonstrated in the case of transcription regulation: the transcripts coming from transcription initiation stimulate cluster formation while the mRNA coming from transcription elongation is dissolving TF clusters in vivo²²⁸. Another way to regulate clustering is by the binding of repressor. During the yeast heat shock response, the heat shock factor 1 (Hsf1) make clusters with Mediator and RNA Pol II²²⁹. The binding of the chaperone Hsp70 to a specific site on Hsf1 represses clustering.

1.3.3 – PROPERTIES AND ROLE OF TF CLUSTERING IN TRANSCRIPTION REGULATION

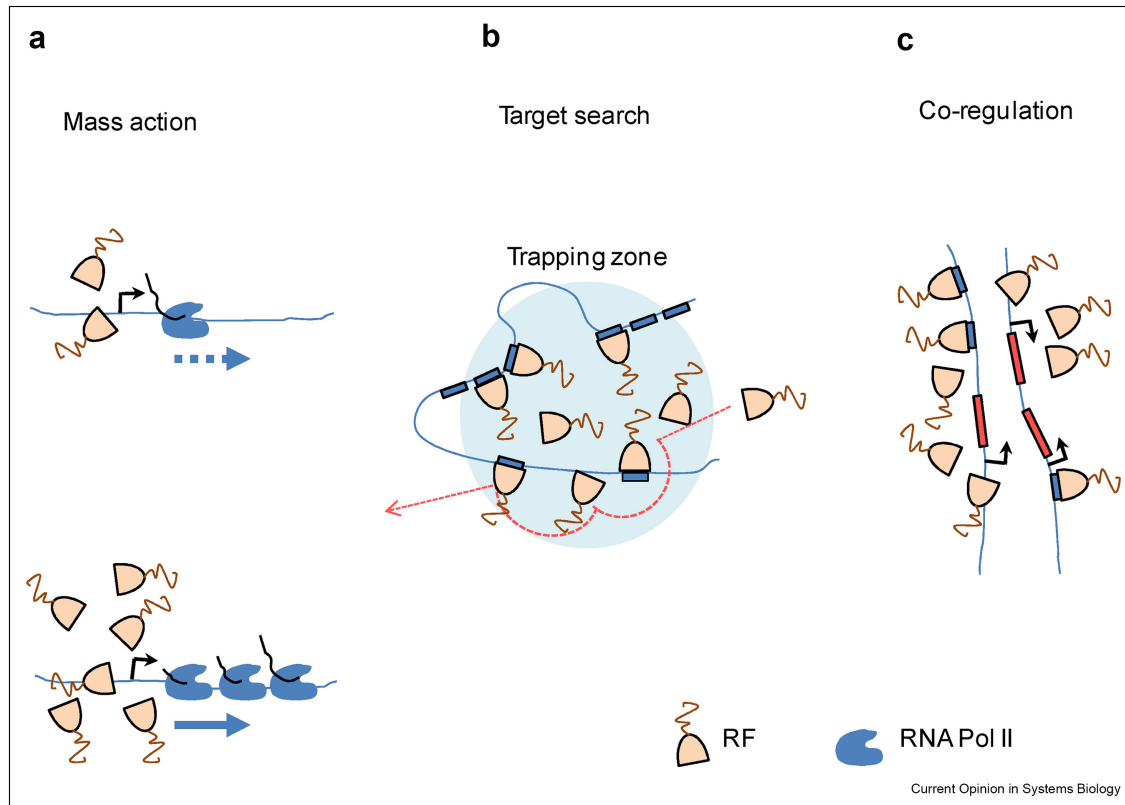


Figure 11: Mechanisms by which TF clustering is regulating transcription

From Jieru Li, Alexandros Pertsinidis, Nanoscale nuclear environments, fine-scale 3D genome organization and transcription regulation, *Current Opinion in Systems Biology*, Volume 31, 2022, 100436, ISSN 2452-3100. <https://doi.org/10.1016/j.coisb.2022.100436>. Copyright C 2023 by Elsevier BV. Creative Commons license CC BY 4.0. Authorization of reuse obtained by Third Party Permission and Reprints, RightsLink.

While transcriptional condensates have been studied for a while, especially their physical properties in line with LLPS, studies about their potential function on transcription only start to emerge. Here, I make a summary of the studies treating about the potential role of TF clustering on transcription, independently of knowing if they are liquid-like.

Bringing enhancers and genes together

The emerging insights from early imaging studies have revealed the spatial proximity between transcription factor (TF) clusters, target promoters, and distal enhancers, as evidenced by nanoscopy imaging techniques ²³⁰. Notably, recent research has demonstrated that Nanog genes from sister chromatids share the same Brd4 cluster, shedding light on the dynamic nature of TF clustering ²³⁰. Further exploring the phenomenon, the current model suggests that TF clustering, particularly within the context of SEs, plays a crucial role in bringing distant DNA segments together, ultimately influencing genomic organization. For instance, E2-responsive enhancers recruit multiple TFs such as GATA3 and ER-alpha, which cluster at MegaTrans enhancers, characterized by a high local concentration of enhancers. These TFs, rich in intrinsically disordered regions (IDRs), exhibit liquid-liquid phase separation (LLPS) properties both in vitro and in vivo, as supported by the disruption of weak interactions using 1,6-hexanediol, resulting in a substantial decrease in ER-alpha cluster number and corresponding binding at MegaTrans enhancers. Notably, MegaTrans enhancers have been found to facilitate the spatial interactions of distant enhancers, as exemplified by the proximity of TFF1 and the DSCAM-AS regions, located in separate topologically associating domains (TADs), induced by E-2 induction ²³¹.

Furthermore, in yeast, the oncogenic factor YY1 not only co-localizes with co-activators and the transcriptional machinery within clusters but also mediates the convergence of enhancer elements and the promoter of its target gene FOXM1, as confirmed by fluorescence in situ hybridization (FISH) and chromosome conformation capture (3C) techniques ²²⁴. Notably, a mutant protein incapable of forming clusters further emphasizes the significance of TF clustering in mediating these genomic interactions. The recruitment of Sox2 and Brd4 into nuclear clusters in mouse embryonic stem cells (mESCs) has been found to rely on the specific molecular recognition of DNA and chromatin binding sites, suggesting that interactions mediated by IDRs might not be indispensable for efficient incorporation into clusters. High-frequency pair-wise genomic interactions have pointed to the presence of a scaffold of enhancer cluster super-enhancers, underlying the formation of Sox2 and Brd4 clusters at active pluripotency genes in mESCs ²⁰¹. Additionally, the role of TF clustering in modulating transcription from low-affinity enhancers has been underscored in drosophila embryos. The co-concentration of the transcription factor

Utrabithorax (Ubx) and its co-factor Homothorax (Hth) within clusters has been associated with the colocalization of active transcription sites and enhancers, supporting the notion that related enhancers from different chromosomes can share the same local microenvironment ²²⁵.

TF clustering could also help in bringing not only enhancers but genes together. In yeast, the protein heat shock factor 1 (Hsf1) has been shown to form inducible condensates with Mediator and RNA Pol II upon stress. The condensates even form interactions among HSR genes from different chromosomes and this condensation is mediated by IDRs. Transcription factor clustering could thus change the genome organization reversibly ²²⁹ .

Recruiting co-activators and the transcriptional machinery

The assembly of enhancers and various DNA elements into clusters appears to promote the binding of other components of the transcriptional machinery within these clusters. Numerous studies have highlighted the enrichment of TF clusters and their role in facilitating the clustering of other factors involved in transcription. In yeast, the oncogenic transcription factor Yin Yang 1 (YY1) forms liquid-like clusters, which exhibit co-localization with EP300, BRD4, MED1, and RNA Pol II ²²⁴. Similarly, the Hsf1 transcription factor in yeast forms clusters associated with Mediator and Pol II ²²⁹. Additionally, research has demonstrated the direct recruitment of the transcriptional machinery by co-activator clustering in vitro. For instance, BRD4 and MED1 form clusters in the nucleus, specifically at super-enhancers (SEs) in live cells. The purified intrinsically disordered regions (IDRs) of MED1 have been shown to concentrate the transcription apparatus from nuclear extracts in vitro, thereby establishing a link between Mediator clusters and potential recruitment of RNA Pol II ⁴⁹. This co-condensation mechanism is proposed to increase histone acetylation and gene target activation, leading to a non-linear transcription response, contrary to the effects observed with single TF recruitment at gene targets ²³².

In summary, the formation of clusters facilitates the recruitment and increased concentration of associated transcription components, surpassing the typical stoichiometric concentrations allowed by DNA binding sites. For instance, the clustering of the tumor suppressor UTX in the nucleus enriches the downstream effector methyltransferase MLL4. Disruption of clustering through the selective

deletion of UTX IDRs results in the reduction of tumor suppressor function and embryonic stem cell differentiation ²²³. These findings collectively underscore the vital role of clustering in mediating intricate interactions within the transcriptional landscape, thereby influencing gene expression.

Affecting binding properties: Decreasing search time, increasing binding rate or stabilize DNA binding

The concept of liquid-liquid phase separation (LLPS) has been proposed to accelerate biochemical reactions ^{199,206,233}. In the context of transcriptional regulation, a widely accepted biophysical model suggests that transcription factors (TFs) locate their targets through facilitated diffusion ²³⁴. This search process involves rapid three-dimensional (3D) diffusion to locate the genomic region of interest, followed by a slower scanning of the same region to find the DNA binding site, where they bind with high affinity ²³⁴. However, the observed search times for a target have been found to be significantly shorter than those expected based on the theory of free diffusion followed by binding, indicating the involvement of additional mechanisms⁵⁹. The directed motion facilitated by IDRs during the genome search may decrease the search time and contribute to cluster formation. Correlated with their roles in target search, IDRs not only specify the genome regions to search but also modify diffusion behavior. For instance, the differential binding and diffusion characteristics of the hypoxia-inducible factors HIF-1alpha and HIF-2alpha, recognizing the same DNA binding motif in vitro (the hypoxia response element), were found to be dependent on the IDRs of the activation domains ²³⁵. In the context of transcriptional regulation, intrinsically disordered regions (IDRs) are implicated in facilitating the binding of specific genomic regions, thereby reducing the time required to search for specific targets. Notably, the transcription factors Msn2 and Yap1, despite recognizing the same DNA binding site, occupy distinct genomic regions, a specificity determined by their IDRs. Removal of the IDRs results in both factors binding to the same sets of binding sites. Although this study did not directly demonstrate that TF clustering reduces search time, it highlighted the contribution of IDRs to target search ⁵⁵.

Recent advances propose that TF clustering could decrease target search time by enhancing intersegment transfer, thereby accelerating the localization of specific DNA targets. In *Saccharomyces cerevisiae*, the glucose-sensing repressor Mig1 forms clusters in the cytoplasm that translocate into the nucleus upon glucose stimulation in

the external media to regulate specific gene targets. Single-molecule fluorescence imaging of Mig1 revealed the formation of clusters comprising 6-9 proteins on gene targets. Investigations into the diffusion movements of Mig1 *in vitro* using a molecular crowding agent indicated that Mig1 clustering facilitated target search by increasing intersegment transfer, thereby reducing the search time for a target ²³⁶. Additionally, clustering has been suggested to create a confined environment, leading to a decrease in search time by limiting 3D diffusion. Live-cell single-molecule tracking (SMT) experiments demonstrated that the Polycomb repressor complex unit CBX-PRC1 forms clusters through DNA interactions. The search for specific binding sites of CBX2 was speed up within these condensates by limiting the free diffusion, thus reducing the number of non-specific sites explored ²³⁷. Moreover, clustering induces confined motion, leading to a decrease in 3D diffusion. Short SMT trajectories of the glucocorticoid receptor (GR) revealed that the GR exhibits confined motion within clusters, mediated by intrinsically disordered regions (IDRs), that might amplify transcription⁶⁹. These findings collectively highlight the critical role of clustering in accelerating target search and modulating the dynamics of transcriptional regulation.

Additionally, TF clustering has been implicated in stabilizing DNA binding. Live-cell single-molecule imaging studies have demonstrated that TFs form clusters at both synthetic and endogenous genomic loci, stabilizing DNA binding, recruiting RNA Polymerase II, and activating transcription. These interactions are dependent on IDRs but not on LLPS²⁰⁴. Transitioning to the role of transcription factor CTCF, it is observed to be trapped in specific nuclear compartments likely constituting CTCF clusters. This trapping mechanism, mediated by an RNA binding domain, increases the efficiency of target search by 2.5-fold, proposing an alternative mode of target search termed anisotropic diffusion⁶⁸.

To conclude, IDRs facilitate interactions with DNA or other IDRs, thereby enhancing the efficiency of the TF search process. IDRs may also increase binding frequency by facilitating faster exchange of TFs within the same cluster, consequently increasing the TF on-rate. Finally, IDRs may stabilize DNA binding through DNA interactions or IDR-IDR interactions. How are TF clustering influencing the transcriptional bursting is still unclear ²³⁸ and could potentially rely on the amount of TF within clusters.

Transcription activation by high local amount of regulators: accelerates rate-limiting steps

Numerous studies have provided evidence that artificially increasing TF clustering at a promoter can enhance the transcription of specific genes. In a recent study, an artificial TF, TerR, as a DNA binding domain, coupled with VP16 as a transactivation domain, with or without the addition of FUS IDRs, was used to study the impact of clustering on transcription. They demonstrated that clusters formed at the tetO system, whether transfected in cells or integrated into the cell genome, increased the transcription of the target gene *Seap*. However, the study distinguished between clustering through protein-protein interactions, as seen in the Cry2 system, and clustering mediated by both protein-protein interactions and low-energy interactions facilitated by FUSn IDRs. The strength of activation was observed to vary depending on the mode of clustering. Furthermore, the combination of both types of clustering resulted in an increased number of clusters, although the impact on the concentration of transcription factors within these clusters at the target gene promoter was not explicitly stated ²³⁹.

In an optogenetic study, it was demonstrated that TFs from the FET-family can form phase-separated clusters in the nucleus of living cells, driving transcription. The blue light-mediated clustering of TAF15 was found to be positively correlated with the fluorescence of the RNA Pol II CTD at the TAF15 cluster. Moreover, the clustering of IDRs from FET-family TFs revealed that the intensity of TAF15 was associated with transcription levels, as observed through live-imaging. These findings collectively suggest that TAF15 clustering amplifies gene expression ²⁴⁰.

Using an artificial DNA array based on UAS elements to control the clustering of a GAL4-VP16 protein construct in fly embryos, a study demonstrated the formation of GAL-VP16 fusion protein clusters at the UAS elements. The average fluorescence within the cluster was positively correlated with transcription at the reporter gene, indicating the temporal coordination of TF clustering with transcriptional bursting ²⁴¹.

Quantitative analysis using nanoscopy imaging revealed that the size of the Cdk9 cluster at the *Pou5f1* locus resulted in the tunable initiation of transcription. Similarly, at the *Nanog* gene, the size of the Brd4 cluster was correlated with the bursting kinetics of the *Nanog* transcripts ²⁴². These findings collectively underscore the significant role of TF clustering in modulating gene expression dynamics and highlight the importance of spatial organization in transcriptional regulation.

Inhibition of transcription

While TF clustering facilitates the recruitment of TFs to target genes, the presence of non-DNA-bound TFs can potentially impede transcription activation. Quantitative microscopy studies conducted on the yeast TF Gal4 within its natural context unveiled that Gal4's self-interactions within clusters play a vital role in the recruitment of TFs to target genes. It was observed that non-DNA-bound TFs exert inhibitory effects on the activation of transcription, indicating a delicate equilibrium in the impact of TF clustering between target search and transcription activation ²²⁶. Building upon this notion, the idea of liquid-like TF droplets suggests a neutral or inhibitory effect on transcription activation. For instance, in a specific study, real-time single-cell fluorescence microscopy combined with multiple synthetic TFs at a reporter gene array demonstrated that while phase separation has the potential to bolster activation strength on the AD, the actual formation of clusters assumes a neutral or even inhibitory role in transcription activation ²⁴³. Similarly, a separate investigation simulating reconstituted mitochondrial transcription *in vitro* revealed that the formation of clusters had a detrimental impact on the transcription output compared to the same condition in bulk solutions ²⁴⁴. In another domain, X chromosome inactivation (XCI) is the process through which one of the X chromosomes is silenced in the cells of female mammals. The non-coding RNA Xist is transcribed from the X chromosome to be silenced and spreads across the entire chromosome, initiating a cascade through the binding of the complex SHARP-SMRT-HDAC3. However, the concentration of SHARP on the X chromosome significantly surpasses the quantity of Xist. It was demonstrated that the recruitment of SHARP to the inactive X chromosome occurs in a non-stoichiometric manner, facilitated by the large intrinsically disordered region (IDR) of SHARP. This transition enables SHARP to form multivalent interactions with itself rather than with Xist, ultimately amplifying the transcriptional silencing induced by Xist ²⁴⁵.

1.4. THE ZYGOTIC GENOME ACTIVATION IN ZEBRAFISH AS A MODEL TO STUDY TRANSCRIPTION REGULATION

1.4.1. GENERAL PRINCIPLES UNDERLYING ZGA

The maternal-to-zygotic transition

The maternal-to-zygotic transition (MZT) is the period when the developmental control is transferred from maternal proteins and RNA into proteins ²⁴⁶. We called the start of the transcription in the embryo the zygotic genome activation (ZGA). It takes count of multiple concomitant events that go from fertilization until fully active zygotic genome and maternal protein and mRNAs clearance ²⁴⁷. ZGA is essential for embryos to continue to develop as a blockage of zygotic transcription leads to developmental arrest prior to gastrulation in many organisms ^{248,249}.

Transcription landscape during early embryonic development

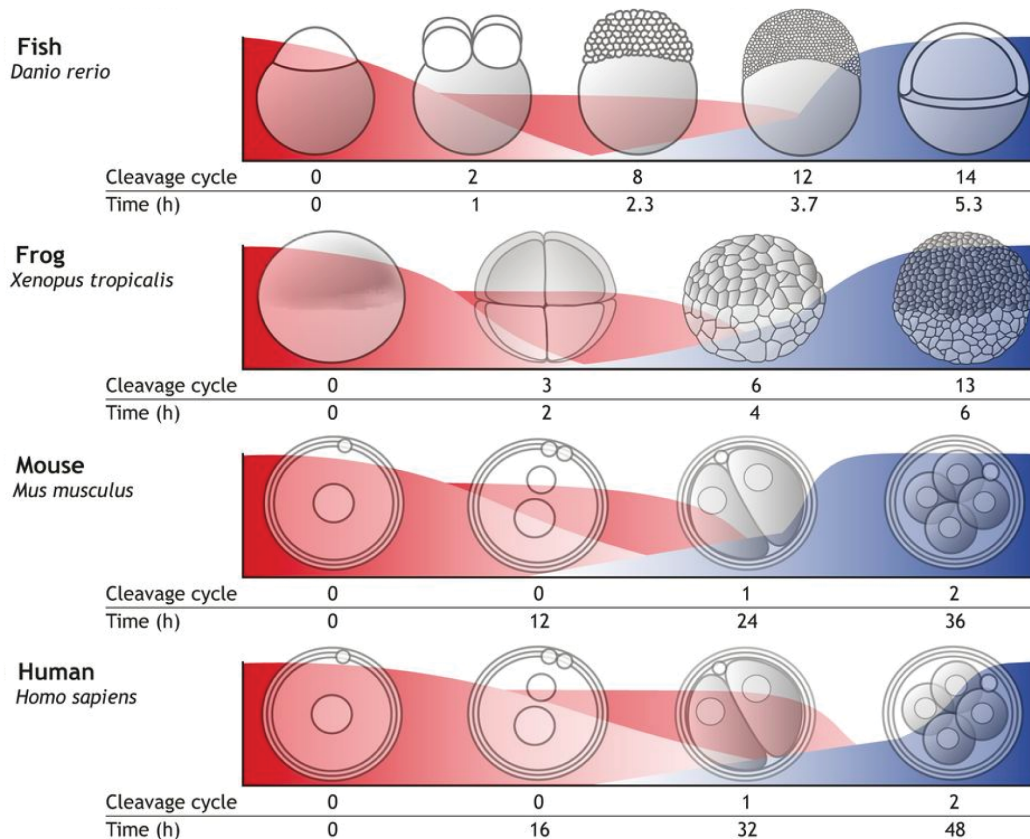


Figure 12: The maternal-to-zygotic transition in fish, frog, mouse and human

Adapted from Nadine L. Vastenhouw, Wen Xi Cao, Howard D. Lipshitz; The maternal-to-zygotic transition revisited. *Development* 1 June 2019; 146 (11): dev161471. doi: <https://doi.org/10.1242/dev.161471> Copyright C 2023 by Elsevier BV. Creative Commons license CC BY 4.0. Authorization of reuse obtained by Third Party Permission and Reprints, RightsLink.

Differences can be distinguishing between fast-developing embryos (generally anamniotes) and slow developing embryos (generally amniotes) ²⁵⁰. Fast developing embryos usually display rapid and synchronous cell divisions alternating rapidly S and M phases until mid-blastula transition (MBT). In zebrafish, rapid cellular division occur until the 10th cell cycle (3 hpf at 28C) before becoming asynchronous^{251,252}. In *Xenopus laevis*, the cell cycle slows down at the 13th cell cycle²⁵². In both species, the increase in length of the cell cycle coincides with a sharp progress of zygotic transcription. The first transcribed zygotic genes are transcribed from 8-cell stage in *Xenopus* and include the miRNA pri-mir427²⁵³. However, what is considered that first phase of ZGA start at 128-cell stage and include gene transcripts involved in gene regulation and development^{254,255}. In zebrafish, zygotic transcripts have been detected from 64-cell stage on and mainly include at this stage the miRNA miR430, involved in maternal transcript decay^{256–258}. At such early stage, the detected transcripts are short, and it was suggested that the transcription of longer genes was stopped at mitosis because of short cell cycles²⁵⁷. From 1000-cell stage (1k-cell stage), a high increase in zygotic transcription occurs and coincide with MBT. It is called the major wave^{259,260}.

In slow-developing embryos, cell division is much slower. The first cell division occurs between 12-36 hours after fertilization and next divisions every 12 to 24 hours²⁶¹ (look for other references for other organisms). In mouse, the first division occurs at 20 hpf even if zygotic transcripts can already detect before in the male pronucleus^{262,263}. A first phase of transcription is detected at the 2-cell stage and corresponds to the minor wave. A second transcriptional phase starts at the 4- to 8-cell stage transition and coincides with blastocyst formation ^{264,265}. Interestingly, many early transcripts include transposons and repetitive elements²⁶⁶. Many of retrotransposons in mammals are derepressed during early development before being repressed again later in ZGA^{267,268}.

1.4.2. THE TIMING OF ZGA AND PROPOSED MODELS

The onset of ZGA varies across species. Two general models have been proposed to explain when in developmental the onset of ZGA is determined: a model based of a specific timer that determinates when transcription initiates, a model based on the number of cells in the embryos and a model based on the size of the cell.

The first model proposes that the embryo can “sense” the developmental time by accumulating transcription activators of ZGA. In zebrafish, the translation of maternal mRNAs encoding for the general activators of ZGA, Nanog, Pou5f3 and Sox19b starts after fertilization ²⁶⁹. They are the most translated TFs in the embryo before ZGA and their knock-down or knock-out lead to developmental arrest ^{269–271}. The co-activator and chromatin read Brd4 has also been shown to be important for ZGA timing ²⁷². We and others have shown that overexpression of such transcription activators and co-activators lead to premature start of ZGA^{272,273}.

The cell size model proposes that ZGA is starting when cells achieve a certain size or a certain DNA to cytoplasm ratio. Increasing or decreasing artificially the amount of DNA in the nucleus has led to premature or delay in ZGA in *Danio rerio*, *Xenopus leavis* or *Drosophila Melanogaster* ^{274–276}. The cell size model has been proposed to be related to the presence of a transcriptional inhibitor in the nucleus which level is titrated as embryo divide and the nuclear volume decreases. While the amount of DNA remains the same, the level of such inhibitor decreases as the embryo divides. We and others have shown that the histone content could play the role of such repressor at early stages, in *Danio rerio* ²⁷³ and *Xenopus leavis* ²⁷⁷.

1.4.3. CHROMATIN LANDSCAPE AND GENOME ARCHITECTURE DURING ZGA IN ZEBRAFISH

After fertilization, the pronuclei containing highly specialized chromatin landscape needs to reorganize to give rise to a totipotent cell state. First, in all vertebrates, protamines from the male pronuclei are replaced with maternally loaded histones ²⁷⁸. However, some species have a partial coverage of the paternal DNA with paternal nucleosomes and up to 100% of the genome in the case of zebrafish²⁷⁹. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish

sperm²⁷⁹. After fertilization, the zebrafish genome is progressively changing during the subsequent cell cycles. In the contrary of mammals, zebrafish embryos do not undergo a full demethylation of the genome^{280,281}. The H3K27Ac is the earliest histone mark detected during zebrafish development, concomitant with transcription activation²⁸². Loss of the histone of the histone acetyltransferases P300 or in the contrary increase presence of p300 and the acetyltransferases Brd4, respectively led to reduction and premature zygotic transcription²⁸³. Acetylation of the H3K27 is thus very important for transcription during ZGA. H3K4me3 and H3K27me3 histone marks, respectively associated with transcription activation and repression, are not present after fertilization²⁸⁴. Both marks are slowly reintroduced in the genome and some genes are positive for both marks while not transcribing yet, remaining poised until later transcription^{284,285}. Interestingly, the H3K27me3 repression mark is not as important during zebrafish development as mutants for this specific methyltransferase is leading to drastic changes in gene expression at early stages and embryos development normally²⁸⁶. Finally, the H3K9me3 mark is surprisingly absent at early stage and only established after ZGA, while proposed to be essential for genome stability^{287,288}.

1.4.4. ZGA IN ZEBRAFISH AS A MODEL TO STUDY TRANSCRIPTION REGULATION

The zebrafish model organism presents a good opportunity to investigate the regulation of transcription, particularly in the context of Zygotic Genome Activation (ZGA). Preceding ZGA, the nucleus is quiescent, as no transcriptional activity can be detected^{257,289}, excepted for low level stochastic transcription²⁹⁰. The dynamic process of Zygotic Genome Activation (ZGA) in zebrafish is marked by a significant reorganization of the chromatin landscape²⁹¹. Notably, the absence of stable heterochromatin formations²⁹¹ and intricate 3D genome structures when transcription is initiated²⁹², such TADs, presents a unique advantage for investigating the intricate interplay between transcriptional activation, 3D genome architecture, and the establishment of crucial chromatin marks during early embryonic development. Additionally, the transparency of zebrafish embryos provides a distinct advantage, facilitating precise and real-time imaging studies, enabling researchers to observe the progressive activation of transcription across different phases of the cell

cycle^{275,293}. The zebrafish, being a vertebrate specie, exhibits prolific reproductive capabilities with external fertilization, coupled with rapid and reproducible developmental processes^{275,293}. With its evolutionary conservation of numerous genes shared with humans, findings derived from zebrafish studies hold the potential to offer valuable insights applicable to a broader spectrum of species²⁹⁴. Moreover, the feasibility of genetic manipulation through techniques such as injection further enhances the versatility of the zebrafish model for investigating the regulation of transcription and its implications for developmental biology²⁹⁴.

In the early zebrafish embryo, bulk activation of transcription starts at around 3hpf (1k-cell stage)^{251,257,259,289,295,296}. However, zygotic transcripts have been detected as early as 64-cell stage^{257,295,297}. First events of transcription can be visually observed in the nucleus by the presence of two S2P transcription bodies^{270,272,297-301}. While many transcription bodies are observed at this stage²⁷⁰, only two go into elongation states and thus produce actively transcripts^{270,272,297-301}. The two S2P-positive transcription bodies are positive for mir430 transcription^{270,297-299,301} and are seeded by the mir430 locus transcription itself as demonstrated by the deletion of the locus^{270,297,299,301}. Brd4 has also been shown to be essential for ZGA²⁷².

Nanog, Sox19B and Pou5f3 are three important TFs activating^{269,270,302} ZGA in zebrafish and the mir430 locus transcription^{269,270}. Knock-out of the *nanog* gene abolishes mir430 transcription²⁷⁰, while knock-down of Sox19 and Pou5f3²⁶⁹, or knock-out of Sox19b have only slight effects on mir430 transcription²⁷⁰. Nanog forms multiple clusters in the nucleus at early stage in the zebrafish embryo²⁷⁰. Two of them only colocalize with active transcription, more specifically the transcription bodies seeded by *mir430* transcription²⁷⁰.

1.5. AIM OF THE THESIS

Recent studies have shown that TF clustering could increase TF efficiency to activate transcription by acting on its capacity to bind DNA (by decreasing search time, increase or stabilize DNA binding at DNA targets)^{269,270,302}, by bringing together enhancers or genes^{269,270,302} and recruit other TFs and co-activators^{269,270,302}. All these elements suggest that the clustering of TFs might facilitate transcription. However, other recent studies have highlighted the potential negative role of

clustering in transcription initiation, contradicting previous suggestions. Moreover, the increase in local concentration of the TF has been reported to correlate with transcription bursting and transcription levels^{201,241,303–305}. However, these few studies that have approached this question mostly used methods based on artificially induced clustering or artificial DNA arrays to seed clustering. Here I take advantage of the zygotic genome activation in zebrafish to study how the clustering of Nanog into clusters at the *mir430* locus is correlated with transcription activation. I take the advantage of being able to label live TF clusters, as well as target DNA and native transcripts in high temporal resolution to study how the *mir430* Nanog cluster formation correlates quantitatively and spatially with transcription at the *mir430* locus.

2. Materials and Methods

2.1 FISH WORK

2.1.1. FISH MAINTENANCE

Zebrafish were maintained under standard conditions, as previously described. In this work, only nanog $-/-$ were used, as to have no Nanog protein background. Embryos were raised at 28 C in Danieau's 0.3X or blue water.

2.1.2. NANOG FISH LINE MANAGING

Nanog $-/-$ fish are a kind gift from Onitchtchouk lab. Adult fish for nanog $-/-$ were fin-clip and genotyped as previously described (Veil et al, 2018). Fish were outcrossed every three generations with Wild-Type ABTL.

2.2. PREPARATION OF EMBRYOS FOR IMAGING

2.2.1 – PREPARATION OF THE COMPONENTS FOR VISUALIZATION

Cloning of Nanog-HaloTag and NLS/NIS-MCP-mNeonGreen

HaloTag sequence was obtained from the MPI protein facility and amplified by PCR from the vector using HaloTag specific primer. HaloTag was subsequently cloned into an empty pCS2+ vector using Gibson Assembly® (E2611, NEB), adding in 5' FseI and AscI sites for easy gene integration as well as the DNA sequence of a protein linker. Nanog coding sequence was amplified from a pCS2+_Nanog-mNeonGreen plasmid with hangovers for Fse I and Asc I sites, and finally integrated into the pCS2_HaloTag plasmid using T4 DNA ligase (NEB, Mo202S). Nanog-mNG, Sox19B-mNG and Pou5f3-mNG were amplified and cloned as previously described (Kuznetsova et al., 2023). The tdMCP gene was amplified from plasmid pME-NLStdMCP-tagRFP (AddGene #86244) and cloned into pCS2+ plasmid with zebrafish-optimised

mNeonGreen tag at its C-terminus. To facilitate optimum levels of nuclear background, one Nuclear Localization Signal (NLS) sequence and one copy of Nucleus Export Signal (NES) were incorporated at the N-terminus. mRNA was synthesized using the Ambion mMESSAGE mMACHINE SP6 Transcription Kit (AM1430; ThermoFisher Scientific, Waltham, MA), quantified, aliquoted to single-use aliquots and stored at -80°C .

In vitro transcription

Nanog-mNeonGreen, Nanog-HaloTag, Sox19B, Pou5f3 mRNA and NLS/NIS_tMCP_mNeonGreen were *in vitro* transcribed using the mMESSAGE mMACHINE™ SP6 Transcription Kit (AM1340, Invitrogen™) from NotI-linearized plasmid for 2h at 37°C , followed by digestion of the plasmid DNA template using TURBO DNase (AM1340, Invitrogen™) for 15 min at 37°C . *In vitro* transcription reactions were cleaned up using the RNeasy MinElute Cleanup Kit (QIAGEN, 7404) and resuspended in RNase-free water. Obtained mRNAs were run on 1% agarose gel aside with RiboRuler High Range RNA Ladder (SM1823, Thermo Scientific™) to check for correct size and integrity, quantified using the NanoDrop (NanoPhotometer® NP80) and Qubit (Qubit fluorometer®, invitrogen) and finally diluted using RNase-free water at the desire concentration prior to conditioning at -80°C .

Mir430 DNA labelling

Recombinant dCas9 expression and purification. The gene encoding a catalytically inactive *Streptococcus pyogenes* Cas9 (D10A/H840A) (dCas9) was cloned in a T7 express vector with a His-maltose binding protein (MBP) tag at the N-terminus. The dCas9 sequence also contained two copies of Nuclear Localization Signal (NLS) sequence at the N-terminus and one copy of NLS at the C-terminus, to facilitate nuclear import. The *S. pyogenes* Cas9 D10A/H840A mutant was expressed in T7 express strain (NEB) containing the pRARE plasmid (Novagen) and cultured at 37°C in terrific broth medium supplemented with chloramphenicol ($17\ \mu\text{g mL}^{-1}$), kanamycin ($100\ \mu\text{g mL}^{-1}$) to $\text{OD}_{600} = 0.5$. Cultures were then shifted to 18°C and induced with $0.2\ \text{mM}$ IPTG overnight.

Cells were lysed in lysis buffer ($50\ \text{mM}$ Tris pH 8.0, $1\ \text{M}$ NaCl, $1\ \text{mM}$ DTT), supplemented with protease inhibitor cocktail (Roche). To remove any nucleic acid

contaminants, polyethylenimine (PEI) was added to the clarified lysate (0.25% w/v) and the sample was clarified by high-speed centrifugation after 10 min incubation on ice. Clarified lysate was filtered through an 0.45- μ m filter and loaded on a MBP Trap column. The column was washed with lysis buffer without DTT and cleavage buffer (20 mM HEPES, 250mM KCl, 10% glycerol, 1 mM DTT). Protein was eluted with elution buffer (20 mM HEPES, 250mM KCl, 10% glycerol, 1 mM DTT and 10mM Maltose) and cleaved with PreScission protease overnight to remove the His-MBP affinity tag. After cleavage, the protein was separated from MBP using cation-exchange chromatography with a 5 ml SP Sepharose HiTrap column (GE Life Sciences). Fractions containing dCas9 protein were pooled and the protein was concentrated with spin concentrators (Amicon Ultra 15, MWCO 30 k; Millipore), diluted to final concentration of 2.5mg/mL using storage buffer (20mM HEPES, 250mM KCl pH 7.25), flash-frozen in liquid nitrogen and stored at -80°C .

Preparation of In vitro transcribed gRNAs. To target miR430 locus we used two sgRNAs with published seed sequences (Chan et al., 2019) with eight MS2 stem loops inserted in the tetraloop of each guide RNA. DNA templates for T7 transcription were prepared by PCR reactions against plasmid pPUR-hU6-sgRNA-Sirius-8XMS2 (Addgene #121942) using a common reverse primer (AAAAAAGCACCGACTCGGTGCC) and unique forward primers (containing the T7 promoter, the seed sequence and overlap sequence with plasmid template):

sgmiR4301_F:

taatacgactcactataGAGGGTACCGATAGAGACAA gtttgagagctactgccatgagga

sgmiR4302_F:

taatacgactcactataGGCTGAGTGTTAACGACTGgtttgagagctactgccatgagga

The PCR products were purified using QIAquick PCR purification kit (Qiagen). The purified product was used as a template for T7 in vitro transcription (HiScribe T7 High Yield RNA Synthesis Kit (NEB)). In vitro transcribed sgRNAs were DNase-treated, purified by phenol:chloroform: isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The sgRNA pellets were dissolved in 20 mM HEPES (pH 7.5) and 300 mM KCl. To refold purified sgRNAs, the sgRNAs were incubated at 70°C for 5 min and slowly cooled down to room temperature. MgCl_2 was then added to 1 mM final concentration and the sgRNA samples were incubated at 50°C for 5 min and slowly

cooled down to room temperature. The sgRNAs were quantified, aliquoted to single-use aliquots and stored at -80°C .

Visualize miR430 locus. To visualize miR430 locus, 1- μL aliquot of untag-dCas9 protein (2.5mg/mL) was diluted in 9 μL of 20 mM HEPES, 300 mM KCl solution to a final concentration of 0.25mg/mL. An amount of 50 ng of each sgmiR4301-8xMS2 and sgmiR4302-8xMS2 targeting the miR430 locus were mixed with the untag-dCas9 protein solution and incubated at 37°C for 10 minutes. After incubation, the assembled dCas9-sgRNA RNP complex was stored on ice until injections. To visualize the locus, 1 nL of assembled RNP complex and 25pg of NLS-NES-tdMCP-mNG mRNA was injected in 1-cell stage embryo.

miR430 RNA labelling

MiR430 RNA were visualized using the MORpholino Visualisation of Expression (MoVIE) coupled in 3' end with the red-emitting Lissamine tag³⁰⁷. Briefly, it is a morpholino-based array binding on the 5' ends of miR430 transcripts, coupled with a fluorophore and allows detection of pre-mir-miR430 RNA (before maturation). It was ordered from GeneTools (<https://www.gene-tools.com/>).

RNA Pol II S5P with Fab

RNA Polymerase II was visualized using Fab-based live endogenous modification labelling (Fab) conjugated with a Cy5 dye ³⁰⁸⁻³¹⁰. Fluorescently labeled Fabs specific to RNA Pol II Ser5P was prepared from monoclonal antibodies specific to Pol II Ser5 phosphorylation. Monoclonal antibodies were digested with Ficin or Papain (ThermoFisher Scientific), and Fabs were purified through protein A-Sepharose columns (GE Healthcare) to remove Fc and undigested IgG. After passing through desalting columns (PD MiniTrap G25; GE Healthcare) to substitute the buffer with PBS, Fabs were concentrated up to >1 mg/ml using 10 k cut-off filters (Amicon Ultra-0.5 10 k; Merck). Fabs were conjugated with Cy5 (N-hydroxysuccinimide ester monoreactive dye; GE Healthcare) to yield 1.8 dye/protein ratio. After the buffer substitution with PBS, the concentration was adjusted to 1.2 mg/ml.

Transcription inhibition

For inhibiting transcription, α -amanitin (A2263, Sigma) was diluted at a concentration of 50 mg/mL and mixed with Nanog-mNeonGreen mRNA and the Fab S5P-Cy5 in a common injection mix. A total amount of 0.25 ng was injected in 1-cell stage embryos. The inhibition was confirmed by a developmental arrest prior to gastrulation ²⁷⁰.

2.2.2 – DECHORONATION, INJECTION, HEALTH CHECKING AND SOAKING

Injection plates were used to block the embryos during injection. Injection plates were made using 1.5% agarose dissolved in Danieau's 0.3X or blue water and from specifically made molds (AdaptativeScienceTools.com). Dechoronation was achieved either before injection using 2 mg/mL of Pronase in E3 medium (Roche, 165921), or after injection using forceps. Injection was performed with glass needles (Thin-Wall Capillary, 4", 1.0 mm, TW100F-4, World Precision Instruments), pulled using a needle puller (SUTTER INSTRUMENT CO. Model P-97) and installed on an air injector (Pneumatic PicoPump, PV830, World Precision Instruments). After injection, embryos were kept in an incubator (memmert) at 28° C until desired stage is reached and frequently observed under a stereoscope (Motic SM2-171) to remove potential dead embryos. For Nanog-HaloTag labelling, embryos were subsequently soaked in 5 μ M of JFX650-HaloTag dye (CS315109, Promega) diluted in Danieau's 0.3X for 20 min followed by direct mounting on imaging plate.

2.2.3 – MOUNTING

Mounting was performed using a 0.8% low melting agarose solution (UltraPure Low Melting Point Agarose, 16520050, ThermoFisher) diluted in Danieau's 0.3X. The agarose solution was warm up at 70C to dissolve and then kept at 37C during embryo mounting. Between 10 and 15 embryos were transferred into an glass Eppendorf filled with the agarose solution and then moved at the surface of an imaging plate (u-Dish 35 mm, high, Ibidi). After a waiting time of 10 minutes to allow agarose polymerization, the plate containing the embryos was then brought to the microscope.

2.2. IMAGING

2.2.1. MICROSCOPE

Imaging for time-lapse (Chapter 3.1, 3.2, 3.3, 3.4 and 3.5):

Imaging was performed on an inverted Nikon Ti2 microscope associated with a Yokogawa CSU-W1 spinning-disk confocal unit using a Nikon 100X Oil CFI Plan Achromat Microscope Objective. Images were acquired using a simultaneous identical duo cameras system, Photometrics Prime 95B and excited using one of the four available laser lines: 405, 488, 561 and 638. Embryos were always maintained at a temperature of 28 °C using a fully enclosed temperature-controlled chamber.

Imaging for FRAP experiments (Chaper 3.6):

Data were acquired using a Real-Time Confocal Microscope Andor Eclipse Ti, inverted stand and motorized objective revolver mounted with a FRAPPA unit. The spinning-disk unit is a Yokogawa CSU-X1. The objective was a water-immersion 60x Plan Apochromat VC DIC from Nikon with a numerical aperture of 1.2. The system is equipped with a dual camera system, both Andor iXON 897 Monochrome EMCCD, with a dixel size of 16 μm . The 405 laser line was used to bleached Nanog clusters or the nucleoplasm and the images were acquired using the 488 laser line with an exposure time equal to 100 ms. 15 frames were acquired before bleaching and 125 frames were acquired after bleaching to follow the recovery of the bleached Nanog clusters, all on a single plane. Single time-lapses were acquired on a period covering approximatively 15 seconds. To check the transcriptional status of the Nanog cluster, the first and last frame of the movie were also acquired simultaneously to the green channel in the red channel using the 561-laser line. For bleaching, the single-point bleaching tool was selected and an area of around 1 μm was subsequently bleached.

2.2.2. IMAGING CONDITIONS :

The following injection mix and image conditions were used for the different imaging experiments in Chapter 3.1, 3.2, 3.3, 3.4 and 3.5. The elements were injected in nanog $-/-$ embryos.

Tableau 1: Injection mix and imaging conditions for all imaging experiments in this thesis

Chapter	Time step	Nanog	RNA Pol II	miR430 transcripts	Mir430 DNA	α -amanitin
3.1	2 min	mNeonGreen (180 pg)	Fab Cy5	/	/	/
3.2 3.3.2	15 s	mNeonGreen (180 pg)	Fab Cy5	/	/	/
3.2.2 3.2.4	15 s	mNeonGreen (180 pg)	Fab Cy5	/	/	25 pg
3.3.3	15 s	mNeonGreen (180 pg)	/	MOVIE lissamine (25 nM)	/	/
3.3.4 3.4	15 s	HaloTag (JFX650) (210 pg)	/	MOVIE lissamine (25 nM)	MCP-mNeonGreen (25 ng)	/
3.5	15 s	Unlabelled (120 pg)	/	MOVIE lissamine (25 nM)	MCP-mNeonGreen (25 ng)	25 pg
3.5	15 s	/	/	MOVIE lissamine (25 nM)	MCP-mNeonGreen (25 ng)	/

Most of the imaging was made using simultaneous acquisition using a duo camera system. Only for acquisition of Nanog, *mir430* DNA and transcripts at the same time, Nanog and *mir430* DNA were acquired first followed by miR430 transcripts. By consequence, there are around 7 seconds of difference between those acquisitions. It can be seen in the images and will be treated during image analysis.

The following injection mix and image conditions were used for the different imaging experiments in Chapter 3.6. These elements were injected in ABTL WT embryos.

Visualized elements	Injected elements
Nanog and RNA Pol S2P	Nanog mNG mRNA (100 pg), Fab recognizing S2P RNA Pol II (1.3 ng)
Sox19B and RNA Pol S2P	Sox19B mNG mRNA (200 pg), Fab recognizing S2P RNA Pol II (1.3 ng)

Pou5f3 and RNA Pol S2P	Pou5f3 mNG mRNA (250 pg), Fab recognizing S2P RNA Pol II (1.3 ng)
------------------------	--

2.3. IMAGE ANALYSIS

2.3.1. PRE-PROCESSING :

Time-lapse max projection and nucleus mapping

Each time-lapse obtained after imaging was projection on the Z axis, to obtain a 2D max projection in the Z plane. In the same time, all .nd2 time-lapses were converted into .ims files using the Imaris file converter (ref). If the time-lapse contains several cell stages, they were first isolated before nucleus mapping. Each individual nucleus of the same stage was assigned a unique ID and reported on a "map". Thus, any nucleus used for analysis can be traced back to the raw data³¹¹.

Cell-stage determination

Cell stage was determined using distances between the center of neighboring nuclei in 3D as described before ²⁷⁰.

Correction of channel registration

Sometimes, data can have a slight registration shift between the channels. To correct for the channel registration, nuclei have been segmented in both channel and X, Y centered positions of the nucleus have been retrieved. Using Imaris, registration was corrected by aligning the nucleus center in X and Y positions.

Isolation of nuclei

Single nuclei were cropped from the whole field of view using either Fiji or the 3D crop function in Imaris (RRID:SCR_007370, Bitplane).

2.3.2. SEGMENTATION OF THE SIGNAL IN 3D :

General parameters for segmentation :

	Nucleus	Nanog	miR430 RNA signal	S5P-Cy5	Mir430 DNA signal	Mir430 DNA signal
Algorithm	Surface	Spot	Surface	Surface	Surface	Spot
Channel	Nanog/ <i>mir430</i> DNA	Green/Far red	Red	Far red	Green	Green
Initial parameters	Object-Object Statistics	Object-Object Statistics	Object-Object Statistics	Object-Object Statistics	Object-Object Statistics	Object-Object Statistics
Diameter	9 μm	XY : 0.325 μm Z : 0.650 μm	0.5 μm	0.325 μm	0.5 μm	0.325 μm
Smoothing	0.5 μm					
Thresholding	Based on histogram	1st: based on curve 2nd: based on curve	Based on curves	Based on curves	Manual	Manual
Final filtering	Remove other nuclei or vesicles	Remove extranuclear objects	Remove extranuclear objects	Remove extranuclear objects	Remove extranuclear objects	Remove extranuclear objects
Tracking	None	Manual	Manual or autoregression model	Manual or autoregression model	Manual or autoregression model	Manual

Specificities for each signal :

Nucleus : The nucleus was always segmented using fluorescence from Nanog. If Nanog was not present or unlabelled, the fluorescence of MCP-mNeonGreen was used. Segmentation was performed based on the histogram of intensities in the 3D volume. Clear distinction of voxels belonging to background or nucleus could be made and consistent segmentation of the nucleus could be achieved (cite Figure and Table).

Nanog mNeonGreen/Nanog-HaloTag (JFX650) : Once the nucleus was segmented, spots were detected and size was determined using different thresholds based on the nuclear mean intensity. For Nanog clusters detected with mNeonGreen, we used 21 nuclei of different stages and background fluorescence to make a calibration curve that we will use to segment all Nanog clusters (**Tableau 5**). The first threshold allow to filter among spots detected on the image from a given diameter (**Tableau 5, Figure Supplemental 2**). The second threshold give the size of the spot by expanding from the center of the pre-detected spots (**Tableau 5**,

Figure Supplemental 3). The same principle was used for Nanog-HaloTag (JFX650) using 22 nuclei (

Figure Supplemental 4, Tableau 6).

mir430 DNA signal: Unfortunately, the method used to detect Nanog, and other structures could not be used for the mir430 DNA. No correlation curve with a good correlation with to the mean intensity could be obtained. Either using the Spot or Surface algorithm, thresholds were determined using visual inspection. All thresholds were reported on a table (data not shown).

RNA Pol II S5P: Signal was segmented using the Surface algorithm of Imaris. A calibration curve was building using 21 nucleus of different stages and background fluorescence (**Figure Supplemental 5, Tableau 7**). To segment RNA Pol II S5P in analyzed nuclei, the nucleus mean intensity of the correct channel was used to determine which threshold to use.

MiR430 RNA : Signal was segmented using the Surface algorithm of Imaris. A calibration curve was building using 22 nucleus of different stages and background fluorescence (**Tableau 8**,

Figure Supplemental 6). To segment RNA Pol II S5P in analyzed nuclei, the nucleus mean intensity of the correct channel was used to determine which threshold to use.

2.3.3. TRACKING :

Manual tracking

For Nanog and *mir430* DNA signal (spot), after segmentation, the different signals were always tracked during time manually as the automatic tracking from Imaris was not able to correctly linked clusters over time, linking clusters that were not the same. Other objects, except the Nucleus, were sometimes tracked manually as well because the same problem.

If a cluster or a transcription body was dividing in two or more parts, the different parts were linked together to the same track, as part of the same object over time.

Connection of Nanog clusters:

After segmentation, Nanog clusters were tracked manually using the following method: Nanog transcribed clusters were detected using the S5P RNA Pol II transcription body signal. All Nanog clusters that were colocalizing with the transcription body were considered as part of the same transcribed Nanog cluster. For time points before transcription, tracking was performed going back in time, frame by frame, from the transcription start. All Nanog clusters and subclusters that clearly connected to the next time point were manually associated together. If an association between clusters was not clear, the specific transcribed Nanog cluster was removed from the dataset.

Automatic tracking

Most of time, *mir430* RNA signal, S5P-Cy5 and *mir430* DNA signal were tracked automatically using the tracking plugin of Imaris. The algorithm used was autoregression model, with a maximum connecting distance of 4 μm , and a missing number of time points depending on the sample.

2.3.4. COLOCALIZATION ANALYSIS IN 3D :

Nanog and the *mir430* DNA were segmented as described above. Once segmented, a mask of the signal was performed, and each allele was isolated in a specific channel. Using the tools Coloc2 (<https://imagej.net/plugins/coloc-2>; Susanne Bolte and Fabrice Cordelières, 2006) in Fiji, percentage of overlap between the two masks was calculated in 3D. To calculate the Pearson's correlation score between the two signals, *mir430* DNA mask was applied on both raw channels to delimitate the area of calculation using JaCoP (<https://imagej.net/plugins/jacop>) in Fiji. Then, correlation score was calculated taking in account all the voxels delimited in this volume in 3D.

2.3.5. SEGMENTATION AND IMAGE ANALYSIS IN 2D :

To describe the *mir430* DNA shape in 2D, Imaris file containing the 3D *mir430* DNA segmentation mask was max projected in Fiji. Then the mask was binarize and shape were tracked over time using TrackMate ³¹². After importing the tracked shapes into

the ROI manager, diverse metrics were measured using the Measurements function. If the mask of the same *mir430* DNA allele was separated in two or more shapes, there were manually merged as one unique object using the ROI manager before measuring features.

2.4. DATA PROCESSING AND PLOTTING

2.4.1. ASSOCIATION BETWEEN NANOG CLUSTERS AND TRANSCRIPTION STATUS :

Association of Nanog-mNG clusters and RNA Pol II transcription:

Statistics were extracted for both types of segmentation. To associate Nanog clusters with transcription status at each time point, manually calculated distance between the gravity center position of the segmented signals and distance calculated by the Imaris software were considered (*Shortest distance to Surfaces*). If the calculated distance between the center of gravity of both segmented signals was inferior to 1 μm and the Imaris software distance inferior to 0.5 μm , then both signals were linked together at each time point.

Association of Nanog-HaloTag (JFX650) and MoVIE:

The same method as applied as for Nanog-mNG and RNA Pol II transcription, except that the calculated distance between the center of gravity of both segmented signals was inferior to 1.5 μm and the Imaris software inferior 1 μm . The threshold distances were increased to take in account for the time lapse between Nanog/*mir430* DNA acquisition and miR430 DNA acquisition.

2.4.1. DETERMINE THE CATEGORY OF NANOG TRACKS: GROWING OR MERGING.

Categorization of Nanog clusters into phenotypes:

For **Figure 16**, phenotyping was performed by visual inspection.

For **Figure 17**, phenotyping was performed on R using the Nanog cluster tracks.

Single cluster: Always a single cluster detected before and at transcription activation.

More clusters: In the last 10 time points before transcription, at least two spots were detected at one time point.

Two or more clusters (short): We can detect two or more clusters prior to transcription for less than three frames in a row.

Two or more clusters (long): We can detect two or more clusters prior to transcription for three or more frames in a row.

Three clusters: At least three clusters were detected at one time prior to transcription start.

Clusters never merging: There are always two or more clusters detected at each time point from 3 frames before transcription start until 3 frames after transcription start. For clusters that start transcription while two or more were detected, if only one spot was detected at the maximum 3 frames before or after, they are not considered by either as “short” or “long” merging events.

Nanog clusters and transcription start:

If only one Nanog cluster was detected at transcription start, the cluster was considered as « merged » at transcription start. If two or more clusters were detected, it was not « merged » at transcription start. Among the « non merged » clusters at transcription start, if only one cluster was detected in the last five frames prior to transcription start, it was considered as fused before transcription start. If one cluster was detected in the last next three frames prior to transcription start, it was considered as fused after transcription. For other cases, it was considered as « never merging ».

Determine merging time:

If only one single cluster was always detected prior to transcription, it was considered as non-merging. In the case of Two or more clusters (short) and Two or more clusters (long) phenotypes, the merging time was determined as the time point when only one Nanog cluster is detected prior to transcription start. If several time points fill these criteria, then the latest time point is considered as the merging time. If merging was not happening prior to transcription start and it was occurring in the last three time points after transcription then, it was considered as the merging time. If no time point with a single cluster was detected in a range of -3 to 3 frames of transcription start, it was considered as not merging and no fusion time was calculated.

2.4.2. ANALYSIS OF BOUND AND UNBOUND STATE FOR *MIR430* DNA SHAPE

Time-serie analysis of *miR430* loci distances

We used movies of labeled *mir430* loci in presence and absence of Nanog to understand how Nanog influences their dynamics. The *mir430* locus was segmented

as described above using the spot algorithm in Imaris. Specifically, we looked at how the distance between all detected loci evolves over time. Whenever at a given time point only one spot was detected, we considered the distance to be zero. For all other time points, we calculated the geometric mean between all detected spots and then calculated their distance as the average distance between all loci and the geometric center. If more than two spots were detected, we calculated the distance the geometric distance as well as the maximum distance between the most distanced clusters. The resulting trajectories of distances between miR430 loci exhibit an oscillatory behavior between the state in which only one spot is visible (compacted state) and the state in which loci are at some distance between each other's (decompacted state). We define an oscillation as the consecutive time points when more than one time point is detected.

Hypothesis testing

We investigated how Nanog changes the dynamics of distances between miR430 loci performing hypothesis testing. From the miR430 loci trajectories we measured several parameters (see below) and compared the distributions of these parameters in presence and absence of Nanog. We considered the dynamics in the absence of Nanog as a null model for the role of Nanog. We first tested whether Nanog has a role in bringing the loci together (motor hypothesis). We reasoned that if Nanog has a role in bringing loci together then the speed at which trajectories come together should be faster in presence of Nanog. Then, we tested whether Nanog has a role in keeping loci together once they are attached (glue hypothesis). We reasoned that if nanog has a role in keeping loci together then the time interval during which loci stay attached should be longer in presence of Nanog.

We performed one-sided Mann-Whitney tests on the distributions of speeds, times and distances in presence and absence of Nanog to test if they were significantly different.

Definitions

Seed to come together: From max: maximum distance value in oscillation divided by the time it takes to detect again only one spot. From first detached: first distance value in oscillation divided by time it takes to detect again only one spot. For each oscillation in each trajectory, we measured one speed value. We then pooled

all speed values and plotted a histogram for all trajectories in presence and absence of Nanog, respectively.

Time to come together:

From max: for each oscillation, the time it takes to detect one spot only starting from the time at which the distance is maximum.

From first detached: for each oscillation, the time it takes to detect one spot only starting from the time at which the loci detached.

For each oscillation in each trajectory, we measured one time value. We then pooled all time values and calculated a histogram for all trajectories in presence and absence of Nanog, respectively.

Distance in decompacted state: For each trajectory we measured all distance values in each oscillation. We then pooled all distance values and calculated a histogram for all trajectories in presence and absence of Nanog, respectively.

Time compacted state: For each trajectory, we measured the duration of all time intervals between two successive oscillations. We then pooled all time values and calculated a histogram for all trajectories in presence and absence of Nanog, respectively.

2.4.3. DATA NORMALIZATION BEFORE PLOTTING

Normalization: All normalization were performed on a range of time point between -10- and 10-time frames between transcription start, to avoid the bias of very bright Nanog clusters at mitosis. Normalization was performed by dividing the values to the maximum value of the track. If one or more spot/shape was detected at one time point, their volume or total intensity were summed up before normalization. Their mean intensity was averaged before normalization.

2.4.4. PLOTTING AND GRAPH CONSTRUCTION

To make graphs, statistics were important into R-Studio (Integrated Development for R. Rstudio, PBC, Boston, <http://www.rstudio.com/>). Data were pre-processed and then plotted using ggplot2³¹³.

2.4.5. IMAGE ANALYSIS, NORMALIZATION AND FITTING FOR FRAP DATA

Tracking and mean intensity measurement:

Spots:

Nanog clusters for which incomplete bleaching or escaping the field of view during acquisition were discarded. TrackMate, installed in Fiji, (Tinevez et al. 2017) was used to follow the bleached spot and the recovering fluorescence. The estimated blob size is 1 μm and the threshold is manually determined to obtain two to three fold more detected spots than frame numbers. The simple LAP tracker was used as the tracking algorithm with the following parameters:

- Linking max distance = 1 μm
- Gap-closing max distance = 1 μm
- Gap-closing max frame gap = 15

Gaps in the track are closed using the option "Close gaps by introduction new spots". This option allows to fill the gaps when the bleached spots are not detectable. Added spots in the track are 1 micro in diameter and the fluorescence is the one of the last positions of detectable spots. The mean intensity as a function of the time were exported as a table and loaded into Rstudio.

Nucleoplasm: The same approach was used for data in which the nucleoplasm was bleached, except that the bleached region was not followed by TrackMate. A circle ROI of the same size as the bleached zone was made. This ROI is determined for the first frame after bleaching. Only bleached region of size less than 1 μm was selected for this study. Mean intensity was measured for this ROI along the whole time-lapse.

Analysis using Rstudio:

Fluorescence normalization: A region of the nucleoplasm, where no spots and no specific roughness in the fluorescence pattern, was measured. This region is considered as a reference zone for fluorescence: presence of fluorescent molecules but not bleached. This region allows to correct later for photobleaching and decrease in fluorescent molecule quantity, the nucleus being considered as a close space. The normalization process consisted in two fluorescence measurements:

- Mean intensity of a 1 μm diameter region covering the tracked spot.
- Mean intensity of a region of the nucleoplasm, where the fluorescence is smooth and no other spot is found, to correct for loss of fluorescence in the nucleus by bleaching

(bleaching of the spot and photobleaching of the molecules over time). To normalize, the different fluorescence measurements are calculated as shown in Equation 1³¹⁴. Normalization for the bleaching of the nucleoplasm is done the same way.

Equation 1:

Normalized fluorescence

$$= \frac{\text{Mean fluorescence spot} * \text{Mean fluorescence reference prebleaching}}{\text{Mean fluorescence spot pre - bleaching} * \text{Mean fluorescence reference}}$$

Fitting:

Normalized fluorescence was associated with corresponding time points from the metadata. Normalized fluorescence for spots and nucleoplasm was fitted single exponential equation using a R function³¹⁵⁻³¹⁷.

Equation 2: $Y = a * (1 - e^{-b*X}) + d$

Half-time of recovery was calculated with the following formula^{314,317}:

Equation 3: $\text{Half - time} = \frac{\log_2(\frac{1}{2})}{-b}$

Percentage of immobile fraction was calculated with the following formula^{314,317}

Equation 4: $\text{Immobile fraction} = (1 - (a + d)) * 100$

3. Results :

As discussed in the Introduction, every element related to transcription is compartmentalized in the nucleus. Transcription factors do not derogate to this observation, forming highly concentrated clusters in the nucleus in many organisms^{185,318}. However, while this is a wide studied topic nowadays, it is still unclear how TF clusters are related to transcription initiation¹⁸⁵.

We previously showed that in early zebrafish embryo, there are multiple local accumulation of initiating RNA Pol II in the nucleus, that we termed transcription body. While they are many S5P+ transcription body at early stage in nuclei of zebrafish

embryos, defined as transcription bodies positive for S5P CTD, only two of those bodies are also S2P+ transcription bodies, defined as positive for S2P CTD, and thus actively transcribing²⁷⁰. These two large S2P transcription bodies are the result of the *mir430* locus transcription, as shown by their absence in *mir430* -/- embryos³¹⁹. To differentiate between the two types of bodies, S5P+ transcription bodies that do not go into elongation and the S5P+/S2P+ transcription bodies, seeded by the *mir430* locus transcription, I consider the S5P+ transcription bodies that do not engage into elongation as non-*mir430* transcription bodies and the elongating transcription bodies seeded by *mir430* transcription as *mir430* transcription bodies.

The TF Nanog, essential for zebrafish development and ZGA^{269-271,320}, forms multiple clusters in the nucleus²⁷⁰. We have shown previously that only two of those Nanog clusters precede and colocalize with the *mir430* transcription body (**Figure 12C**). While we studied the order of events of the transcription body assembly²⁷⁰, we still do not know how Nanog clusters at the transcription body are formed and how the temporal dynamics of Nanog clustering to *mir430* transcription initiation.

3.1. NANOG CLUSTERS POSITIVE FOR TRANSCRIPTION INITIATION CONTAIN MORE NANOG THAN ALL OTHER CLUSTERS

While TF have been shown to form clusters in other systems³¹⁸, it is unclear if they are all functional and what makes the difference between TF clusters colocalizing with transcription bodies and the others. Here, I take the advantage of multiple Nanog TF clusters in the early zebrafish embryos, of which only two are colocalizing with active transcription. I followed up our work published recently by asking what makes the Nanog clusters positive for transcription initiation distinct from the others²⁷⁰.

3.1.1 - NANOG MAKES MULTIPLE CLUSTERS IN THE NUCLEUS, TWO OF WHICH COLOCALIZE WITH TRANSCRIPTION INITIATION

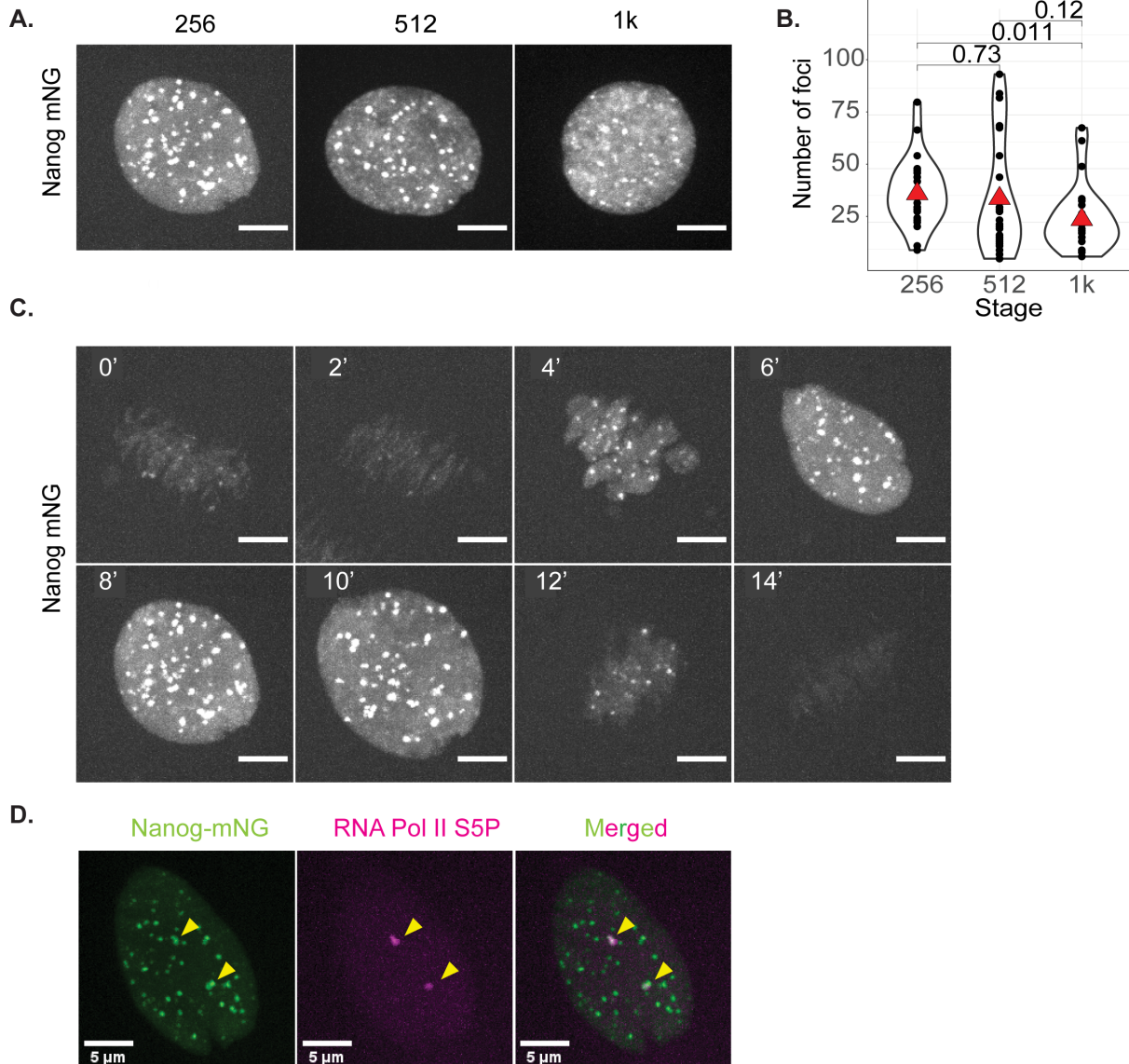


Figure 12: Nanog clusters are present at different stages, during the whole cell cycle, but only two are positive for transcription.

A. Snapshots of Nanog-mNG injected embryos at 256-, 512-, and 1k-cell stage at the moment of transcription activation. **B.** Number of Nanog clusters in the nucleus at 256-, 512-, and 1k-cell stage upon transcription initiation. The red triangle represents the median. Each black dot represents a nucleus. For calculating these values, the following numbers of nuclei/embryos were used :256-cell stage : N1=4, N2=8, n=23, 512-cell stage : N1=5, N2=12, n=24, 1k-cell stage : N1=6, N2=12, n=23, where N1 the number of biological replicates, N2 the number of

embryos and n the number of nuclei analyzed. **C.** Snapshots of embryos injected with Nanog-mNG every two minutes at the 256-cell stage from metaphase to metaphase. **D.** Images of Nanog-mNG, RNA Pol II S5P and merged at 256-cell stage at the frame when transcription was observed for the first time. All scale bars in this pannel represent 5 μ m length.

We have previously shown that the multiple Nanog clusters are present in the nucleus at early zebrafish development^{270,299}. I was interested in understanding better the temporal and spatial dynamics of all Nanog clusters during the cell cycle and across development. To do so, I injected Nanog-mNG as a mRNA, as well as the Fab recognizing RNA Pol II S5P in 1-cell stage nanog^{-/-} embryos³⁰⁸⁻³¹⁰. To visualize transcription, I used the Fab recognizing initiating RNA Pol II (S5P) instead of the elongation RNA Pol II (S2P) as I was interested in detected the first event of transcription, even prior to transcription elongation. After injection, I let embryos develop until desired stage and imaged them at 256-, 512- and 1k-cell stage using a spinning-disk imaging approach at a 2-minute time resolution.

To understand better the dynamics of Nanog clusters in the early zebrafish embryo during development, I was interested in counting the number of clusters over development. To do so, I counted the number of clusters using the licensed software Imaris, following the imaging analysis pipeline described previously (Methods, Chapter 2.3.2). I can observe that Nanog makes multiple clusters in the nucleus at the three studied stages (**Figure 12A**). At the moment of transcription start (first detection of S5P transcription bodies), there are on average 38 Nanog clusters at 256-, 37 at 512- and 27 at 1k-cell stage (**Figure 12B**). The minimum number of Nanog clusters observed in one nucleus are 12 at 256-, 8 at 512- and 9 at 1k-cell stage while their maximum reaches respectively 81, 94 and 69. The number of clusters are thus highly variable between nuclei among the same cell stage (**Figure 12B**). The number of Nanog clusters is stable between 256 and 512-cell stage but decrease significantly between 256- and 1k-cell stage ($p=0.011$, Student t-test). These results suggest that the number of Nanog clusters is quite variable between nuclei and decreases over development.

During the cell cycle, Nanog clusters can be observed on condensed chromosomes during mitosis and all along the cell cycle (**Figure 12C**). The clusters are thus not only restricted to the moment when transcription happens but are constantly present. Their presence during mitosis could reveal a role in bookmarking to prepare for the next cell cycle^{321,322}.

By labelling S5P transcription bodies, I explored the spatial relationship between Nanog clusters and initiation of transcription. As I am not labelling the miR430 transcripts in this experiment but only labeling the S5P transcription bodies, I define the Nanog clusters colocalizing with S5P transcription bodies as S5P+ Nanog clusters and the Nanog clusters not colocalizing with S5P transcription bodies as S5P- Nanog clusters. As previously observed ²⁷⁰, among all Nanog clusters, only two S5P+ Nanog clusters are visually colocalizing with S5P transcription body (**Figure 12D**).

To conclude, I showed that multiple Nanog clusters are present in the nucleus during three consecutive cell cycles, and they are also present during the different step of the cell cycle. However only two of these Nanog clusters are colocalizing with S5P transcription bodies.

3.1.2 – WHAT MAKES THE S5P+ NANOG CLUSTERS SPECIAL COMPARED TO THE S5P- NANOG CLUSTERS?

I have shown that Nanog makes multiple clusters in the nucleus, two of which only are visually colocalizing with S5P transcription bodies ²⁷⁰ (**Figure 12D**). I aim to understand what makes the two S5P+ Nanog clusters distinct from the S5P- Nanog clusters. To do so, I used the same data as previously (Chapter 3.1.1, **Figure 12**) and selected only one time point in the cell cycle, the moment when S5P transcription are detected for the first time. Then I compared the two types of Nanog clusters for several features. To avoid any bias related to the developmental stage, I selected only the nuclei at 1k-cell stage.

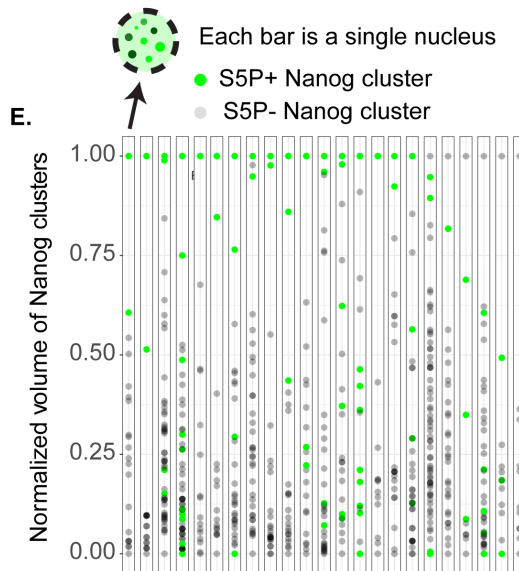
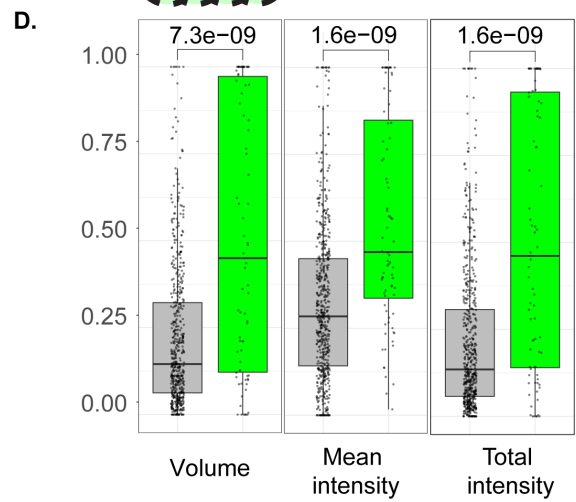
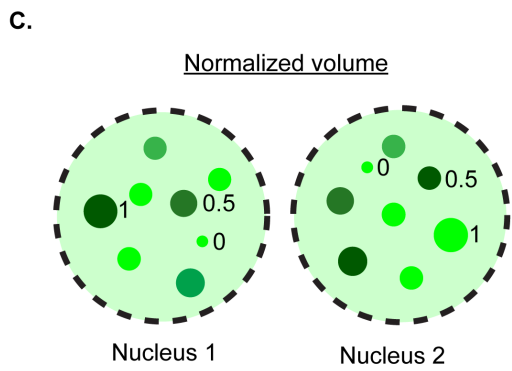
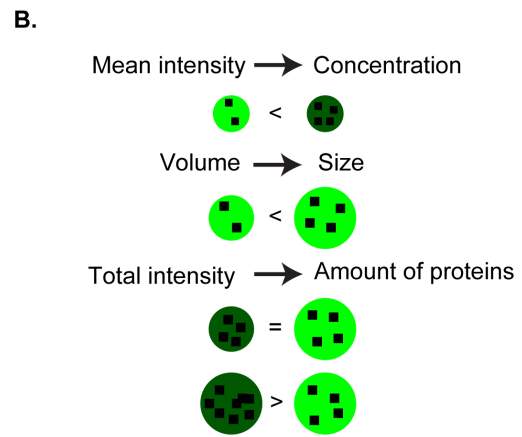
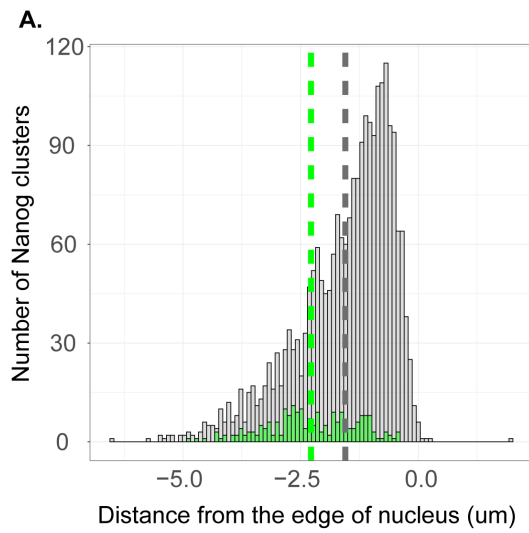


Figure 13: The S5P+ Nanog clusters show distinct features compared to the S5P- Nanog clusters

A. Histogram showing the distribution of the distance between single Nanog cluster and the nuclear surface. Green bar represents S5P+ Nanog clusters, grey bars represent S5P- Nanog clusters. Dotted lines represent the average distance to nuclear for both categories. Difference between the two distributions : $p=2.2E10-16$, Student t.test. **B.** Schematics explaining the selected features to study Nanog clusters. Mean intensity is used to represent the concentration of Nanog proteins in the nucleus, volume is used for the size of the Nanog clusters and total intensity for the total amount of Nanog proteins inside the nucleus. **C.** Schematics explaining how the values for mean intensity, volume and total intensity have been normalized in D and E. Each of the cited features is normalized for values inside the same nucleus. For example, the smallest volume value is normalized to 0 and the highest volume value is normalized to 1. **D.** Boxplots showing normalized mean intensity, volume and total intensity (average) for all the nuclei pulled together. P-values are obtained from Student t-test. **E.** Graphs showing the normalized volume at the level of individual nuclei at 1k-cell stage. Each vertical rectangle represent a nucleus and each spot represents a Nanog cluster. In green : S5P+ Nanog clusters. In grey : S5P- Nanog clusters. 1k-cell stage : $N1=6$, $N2=12$, $n=23$. where $N1$ the number of biological replicates, $N2$ the number of embryos and n the number of nuclei analyzed. Mean intensity, volume and total intensity were normalized for values inside the same nucleus.

First, I was interested in studying the localization of S5P+ and S5P- Nanog clusters in the nucleus. I compared the distance from the nuclear periphery for both types of Nanog clusters and I observed that the S5P+ Nanog clusters adopt a more central position than the S5P- Nanog clusters, respectively distant from the nuclear periphery in average of 2.28 μm and 1.54 μm (**Figure 13A**, $p=2.2E-16$, Student t.test). This is consistent with the earlier observations that transcribing genes have the tendency to be localized more often in the center of the nucleus, compared to the inactive genes, which tend to localize towards the nuclear periphery^{323,324}.

Then, I inquired how the Nanog amount inside Nanog clusters correlates with transcription initiation. To do so, I compared the levels of fluorescence between the S5P+ Nanog clusters and the S5P - ones. I selected three parameters related to the fluorescence or size of Nanog clusters to obtain an indirect read-out of cluster features: 1) mean intensity for concentration of Nanog proteins, 2) volume for size and 3) total intensity for the total amount of Nanog protein present in the cluster (**Figure 13B**). I measured these three parameters for all Nanog clusters at the image frame when I detected S5P transcription bodies for the first time and plotted these three parameters split for the S5P+ or S5P- Nanog clusters. In order be able to compare the three chosen

features between different nuclei, values were normalized based on the minimum and maximum value of each feature inside one given nucleus (**Figure 13C**). For example, for the normalization of the volume, the smallest Nanog clusters in the nucleus was associated with a value of 0 and the largest with a value of 1.

Comparing normalized values of volume, mean intensity and total intensity of S5P+ and S5P- Nanog clusters by pulling all values obtained from all the nuclei from the dataset, I observed that, on average, S5P+ Nanog clusters are larger ($p=7.3e-09$, Student t-test), have a higher Nanog concentration ($p=1.6e-09$, Student t-test), resulting in more Nanog proteins ($p=1.6e-09$, Student t-test) than S5P- Nanog clusters. This behavior was observed at 1k-cell stage (**Figure 13D**), and true at 256- and 512-cell stage as well (data not shown). Interestingly, significance is higher for the volume measurement, suggesting that the size is factor reflecting a larger difference between the S5P+ and S5P- Nanog clusters, in the contrary of the concentration. As the total intensity is a combination of mean intensity and volume, the total intensity measurements are therefore more significant as well.

Next, I was wondering if the S5P+ Nanog clusters are also the largest, the most concentrated in Nanog and containing the most Nanog, at the level of individual nuclei. To do so, I isolated each individual nucleus at 1k-cell stage, ordered the Nanog clusters based on their normalized values for mean intensity, volume or total intensity, and split them depending on if they are S5P+ or S5P- Nanog clusters (**Figure 13E**). An example for the volume at 1k-cell stage, I could observe that the S5P+ clusters are the largest in only 50% of the nuclei. I observed the same with the mean intensity or the total intensity, and at different stages as well (data not shown). This result indicates that even if in average, S5P+ Nanog clusters are larger, more concentrated and contain more Nanog than S5P- Nanog clusters, (**Figure 13D**), this is not true at the level of individual nuclei (**Figure 13**). Moreover, from previous results (**Figure 12D**) and published work ²⁷⁰, we have observed that only two Nanog clusters are colocalizing with S5P transcription bodies at this developmental stage. However, I observed here that in some individual nuclei, more than two Nanog clusters were S5P+ Nanog clusters per nucleus (**Figure 13E**, green spots), suggesting that more than two Nanog clusters can colocalize with S5P transcription bodies in the nucleus.

To conclude, I observed that S5P+ Nanog clusters are more centrally located in the nucleus, are larger, more concentrated for Nanog, and containing more Nanog proteins than S5P- Nanog clusters on average (when all nuclei are pooled

together) but not in individual nuclei. Moreover, I observed that more there are more than two S5P+ Nanog clusters, contradicting previous observations and published work. I will now focus in understanding why I detect more than S5P+ Nanog clusters in the nucleus.

3.1.3 – S5P+ NANOG CLUSTERS DISSOCIATE IN SMALLER CLUSTERS UPON THE GROWTH OF TRANSCRIPTION BODY

S5P+ Nanog clusters are not always the largest, the most concentrated, or containing more Nanog proteins if we look at the nuclei individually (Chapter 3.1.2, **Figure 13**). Moreover, I showed that there are more than two S5P+ Nanog clusters, inconsistent with previous work ²⁷⁰. As described previously, in the early zebrafish embryos, we can observe two types of S5P transcription bodies in the nucleus: 1) the two S5P+/S2P+ transcription bodies that are seeded by *mir430* transcription, that I defined here as the mir430 transcription body, and 2) S5P transcription bodies that are positive for elongation of transcription, that I defined here as the non-mir430 transcription bodies. I propose two hypotheses to explain why I quantified more than two S5P+ Nanog clusters in each nucleus: 1) the S5P+ Nanog clusters colocalizing with the mir430 transcription body are detected as several Nanog clusters, while only visually looking like one (**Error! Reference source not found.A**, hypothesis #1) or there are some S5P+ Nanog clusters colocalizing with some non-mir430 transcription bodies in the nucleus (**Error! Reference source not found.A**, hypothesis #2). Both hypotheses are not mutually exclusive.

To test if one of these hypotheses explain why I observed more than two S5P+ Nanog clusters in the nucleus, I need to differentiate between the mir430 and non-mir430 transcription bodies. We have shown previously that the mir430 transcription body appear first during the interphase, followed later by the non-mir430 transcription bodies. While the mir430 transcription grow during the whole cell cycle, the non-mir430 transcription body size remains constant ²⁷⁰. The mir430 transcription body are thus the largest of the transcription bodies at any moment of interphase. Based on this information, I will use the following criteria to differentiate between the two types of transcription bodies: if one or two transcription bodies are present in the nucleus, I will consider them are mir430 transcription bodies. If three

or more transcription bodies are present in the nucleus, the two largest ones are considered as the mir430 transcription body, and all the other are considered as non-

mir430 transcription bodies. Moreover, to differentiate between S5P+ Nanog clusters colocalizing with the mir430 transcription bodies from the S5P+ Nanog clusters

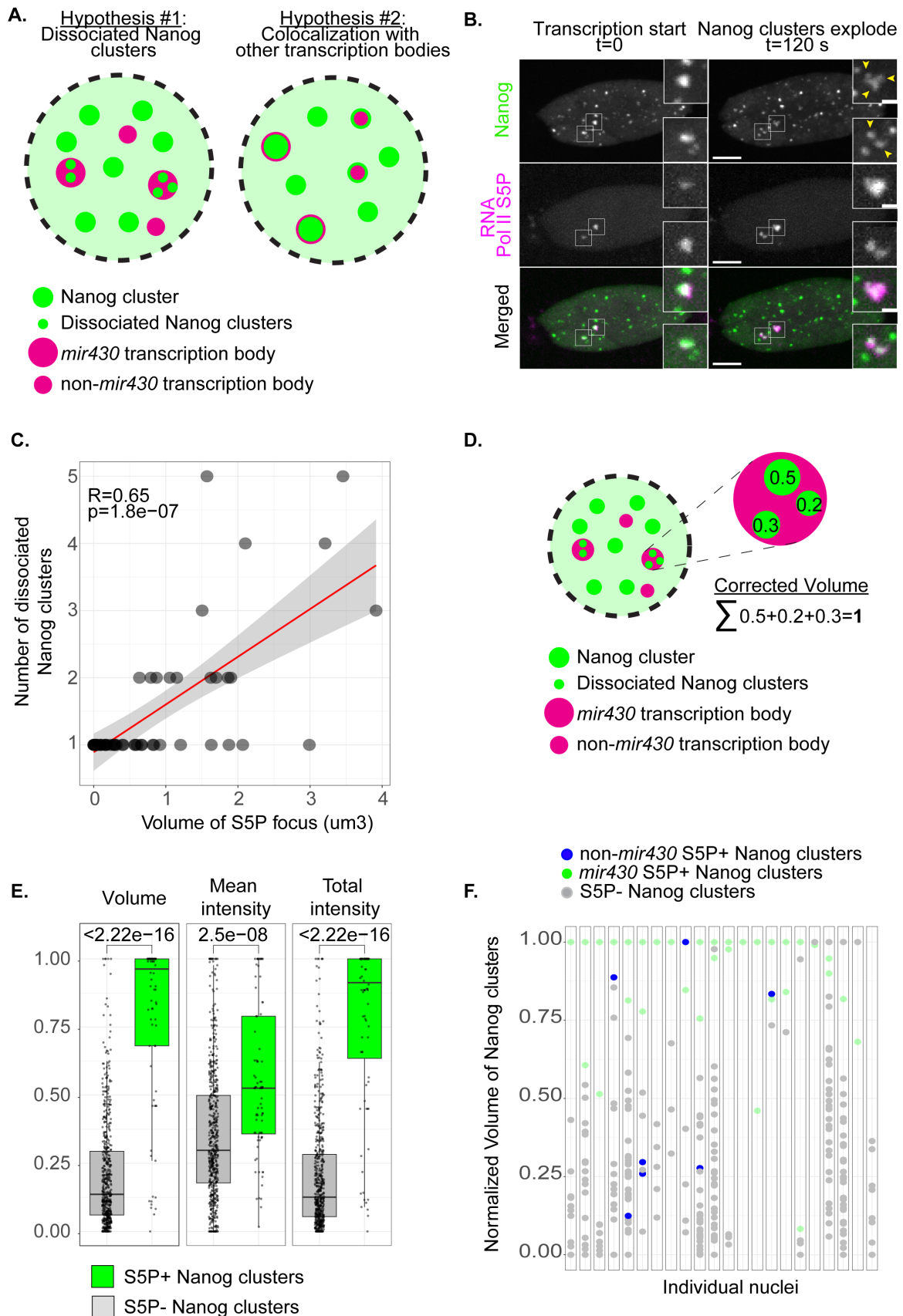


Figure 14: The mir430 S5P+ Nanog clusters dissociate in smaller clusters upon mir430 transcription body growth

A. Schematics explaining the two proposed hypotheses explaining the numbers of S5P+ Nanog clusters observed in individual nuclei. 1. The Nanog signal colocalizing with mir430 transcription bodies consist of several Nanog clusters. 2. Some S5P+ Nanog clusters are colocalizing with non-mir430 transcription body. **B.** Snapshots of Nanog-mNG, RNA Pol II S5P, and merged signals at 1k-cell stage as transcription starts (t=0 second) and as transcription progresses (t=120 seconds). Small squares give a zoom of Nanog S5P+ Nanog clusters. Yellow arrows at t=120s indicates smaller Nanog clusters resulting from a larger cluster. Larger scale bar = 5 μm . Smaller scale bar = 1 μm . **C.** Number of Nanog clusters colocalizing with a single S5P transcription body as a function of the volume (μm^3) of the body at 1k-cell stage. Each black dot represents a transcription body. The red line indicates the linear correlation. R coefficients and p-values are obtained from Spearman's rank correlation coefficient tests. **D.** Schematics representing how the volume has been corrected for the dissociation of Nanog clusters. The volumes of dissociated smaller clusters colocalizing with the same transcription body were sum up and considered as one Nanog cluster. **E.** Boxplots showing normalized mean intensity, total intensity and volume (average, all nuclei pooled together) at 1k-cell stage corrected for the number of Nanog clusters inside the S5P transcription body. P-values are obtained from Student t-test. **F.** Graphs showing the normalized volume (for each nucleus) at 1k-cell stage corrected for the number of Nanog clusters inside the RNA Pol II S5P body. In green : S5P+ Nanog clusters colocalizing with the mir430 transcription body. In blue : S5P+ Nanog clusters colocalizing with non-mir430 transcription bodies. In grey : S5P- Nanog clusters. For calculating the metrics, the following numbers of nuclei/embryos were used : 1k-cell stage : N1=6, N2=12, n=23, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed. Mean intensity, volume and total intensity were normalized for values inside the same nucleus.

colocalizing with non-mir430 transcription bodies, I define respectively mir430 S5P+ Nanog cluster and non-mir430 S5P+ Nanog clusters (**Error! Reference source not found.A**).

Our group and others have observed that as transcription progresses, *mir430* transcription bodies grow during the cell cycle ^{270,283,297,299,307}. Moreover, we and others observed that the mir430 S5P+ Nanog clusters dissociate, from one large cluster to many smaller clusters, as the mir430 transcription body grow ^{270,299}. I made the same observation with my data: Nanog clusters colocalizing with the mir430 transcription bodies dissociate into smaller clusters (**Error! Reference source not found.B, C**). I was wondering whether the number of detected mir430 S5P+ Nanog clusters was related to the growth of the mir430 transcription body. To do so, I plotted the number of S5P+ Nanog clusters colocalizing with the same mir430 transcription,

as a function of the volume of this *mir430* transcription body (**Error! Reference source not found.C**). I observed a significant positive correlation of 0.65 at 1k-cell stage ($p=1.8e-08$, Spearman's rank correlation coefficient tests). (**Error! Reference source not found.C**). This result indicates that the larger the *mir430* transcription body is, the more S5P+ Nanog clusters are found in this body, suggesting that more the transcription process pursue, more the *mir430* S5P+ Nanog cluster dissociates into smaller clusters. The first proposed hypothesis is then true, the number of S5P+ Nanog clusters is influenced by the growth of the *mir430* transcription body. I can observe that if all the S5P+ Nanog clusters colocalizing with the same *mir430* transcription body are considered as one unique cluster, the number of S5P+ Nanog clusters is lower in each individual nucleus (**Error! Reference source not found.F**). To avoid any confusion, the Nanog clusters coming from the dissociation of the same *mir430* S5P+ transcription body are defined as the dissociated Nanog clusters.

Such a dissociation process could have an impact on the measured mean intensity, the volume, and the total at the S5P+ Nanog clusters. Indeed, I used a two-minute time step for the imaging of the data presented in this section. I chose to quantify the features of S5P+ Nanog clusters at the moment when I detect any S5P transcription body for the first time in the cell cycle. However, transcription could start at any moment within this two-minute window and already lead to the dissociation of the S5P+ Nanog cluster into smaller ones. The number and features of S5P+ Nanog clusters could be then impacted: we detect more S5P+ Nanog clusters, but of smaller sizes (**Error! Reference source not found.A**). To correct for the measurements of the dissociation process, I summed up the volume and total intensity of the dissociated Nanog clusters, colocalizing with the same *mir430* transcription body and considered all the dissociated Nanog clusters as one unique cluster (**Error! Reference source not found.D**). For the mean intensity, I averaged the intensity of dissociated Nanog clusters. After correction, when looking at the S5P+ Nanog clusters from all nuclei pooled together, the significance of the differences between S5P+ and S5P- Nanog clusters increased for the volume and the total intensity at 1k-cell stage ($p<2.22e-16$, Student t-test). However, the significance of the mean intensity did not change, suggesting that the dissociation impacts more the volume of the *mir430* S5P+ Nanog clusters than their concentration in Nanog (**Error! Reference source not found.E**), in comparison with the data that were not corrected (**Figure 13D**).

Focusing on individual nuclei, I observed that more S5P+ Nanog clusters were the largest in each nucleus, compared to before correction. Accordingly, after correction, the S5P+ Nanog clusters are now the largest in 74% of all individual nuclei at 1k (Error! Reference source not found.F, one vertical rectangle represents one nucleus). However, after correction, 14% of the nuclei at 1k-cell stage still displayed more than S5P+ Nanog clusters. Potentially, as proposed previously (Error! Reference source not found.A, hypothesis #2), some S5P+ Nanog clusters are colocalizing with non-mir430 transcription bodies as well. To confirm this hypothesis, I selected all the S5P+ Nanog clusters and colored them depending on which type of S5P transcription body they are colocalizing with. Mir430 transcription and non-mir430 transcription bodies were differentiated based on their moment of appearance during interphase and their size, as described above (Error! Reference source not found.F, respectively green and blue). Interestingly, I can observe that, after correction for the dissociated Nanog clusters and further categorization of the S5P+ Nanog clusters, only two S5P+ Nanog clusters are colocalizing with the mir430 transcription bodies in each nucleus. All the other S5P+ Nanog clusters are associated with non-mir430 transcription bodies. However, their number is low as they represent only 17.4% of all the S5P+ Nanog clusters at 1k-cell stage (all nuclei combined). Moreover, they are not present in every nucleus (Error! Reference source not found.F). In contradiction with previous results, they are thus more than two S5P+ Nanog clusters in the nucleus, as some S5P+ Nanog clusters colocalize with non-mir430 transcription bodies as well. However, their frequency is low, and we might have missed them visually previously, while an exhaustive and unbiased quantitative method could detect them.

To summarize, mir430 S5P+ Nanog clusters dissociate into smaller Nanog clusters upon the growth of the mir430 transcription body. When correcting for this dissociation, the link between the amount of Nanog inside S5P+ Nanog clusters and transcription become even stronger. Some S5P+ Nanog clusters can colocalize with non-mir430 transcription bodies, but their number is low, and they are not present in each nucleus across the population. However, it is unclear how such high amount of Nanog are reached by the S5P+ Nanog clusters. As the mir430 S5P+ Nanog clusters are constantly present in the nucleus, I will focus on those to explore this question. More precisely, I want to investigate the formation process of the mir430 S5P+ Nanog clusters before transcription, to understand better how they are reaching high amounts of Nanog

upon transcription initiation, and how such amounts correlate temporally with transcription activation.

3. 2. THE MERGING OF SMALLER NANOG CLUSTERS INCREASE TF AMOUNT LOCALLY AND CORRELATES WITH TRANSCRIPTION INITIATION

A few studies have demonstrated correlation between the TF intensity in clusters and transcription factor level, or bursting. However, they mostly used artificial clustering or clustering occurring on an artificial DNA array. Moreover, none of them studied how the TF clusters forms, and how the formation process is correlated with TF amount and transcription initiation. I intend to study this in our system, taking advantage of the natural clustering of S5P+ mir430 Nanog in the nucleus, and our ability to measure the levels of fluorescence in a high temporal resolution as the same time as imaging the transcription of its target gene, the mir430 locus.

3.2.1 – THE MIR430 S5P+ NANOG CLUSTER ACCUMULATE NANOG PROTEINS BY GROWTH OR BY MERGING OF SMALLER NANOG CLUSTERS

Previous work in our lab and others have shown that transcription starts at the mir430 locus rapidly after the end of mitosis ^{270,298}. Moreover, we have shown that a Nanog cluster precedes and colocalize with mir430 transcription ²⁷⁰, indicating that the formation of the mir430 S5P+ Nanog clusters happen rapidly before transcription. I have shown previously that S5P+ Nanog clusters are enriched for Nanog proteins compared to S5P- Nanog clusters (**Figure 13, Error! Reference source not found.**). To understand how the mir430 S5P+ Nanog clusters reach such high amount of TF and what this could reflect on the mir430 transcription initiation, I injected as previously Nanog-mNG as a mRNA, as well as the Fab recognizing RNA Pol II S5P in 1-cell stage. However, I decided to image the formation process of the mir430 S5P+ Nanog clusters prior to transcription initiation with a time interval of 15 seconds, compared with the previous time interval of 2 minutes to catch the mir430 Nanog cluster formation with a higher temporal resolution.

To describe the formation process of the mir430 S5P+ Nanog clusters, I selected the first frame when I can detect the mir430 transcription body, selected the colocalizing Nanog cluster with this mir430 transcription body and followed it back in time (**Figure 15C**).

Interestingly, I observed that the mir430 S5P+ Nanog cluster can be assembled in two mechanisms:

- By growth: one small cluster appearing after mitosis, that grows (48.1% of cases) (**Figure 15A**, lower zoom and B, upper zoom, **Figure 15D**),
- By merging: Two or more small Nanog clusters that can merge together, colocalizing later with the mir430 same transcription body (51.9% of cases) (**Figure 15B**, upper zoom, **Figure 15E**).

These two mechanisms, growing and merging, are not mutually exclusive, as the mir430 Nanog subclusters can grow as well (**Figure 15E**). If during the mir430 Nanog cluster formation, at least two smaller clusters were observed, coming from the same merged mir430 Nanog cluster, they were classified in the merging category. To differentiate the mir430 S5P+ Nanog clusters forming by growth only or by merging, I define the first category as the growing category, involving the growing mir430 Nanog clusters (**Figure 15D**) and the second category as the merging category, involving the merging mir430 Nanog clusters (**Figure 15E**). Moreover, in the case of the mir430 merging Nanog clusters, to precisely describe which clusters I am referring to, I differentiate before and after merging. Before merging, two or more small Nanog clusters are present and are defined as the mir430 Nanog subclusters (**Figure 15E, light grey**). The result of the merging of the mir430 Nanog subclusters is defined as the merged mir430 Nanog cluster (**Figure 15E, dark grey**).

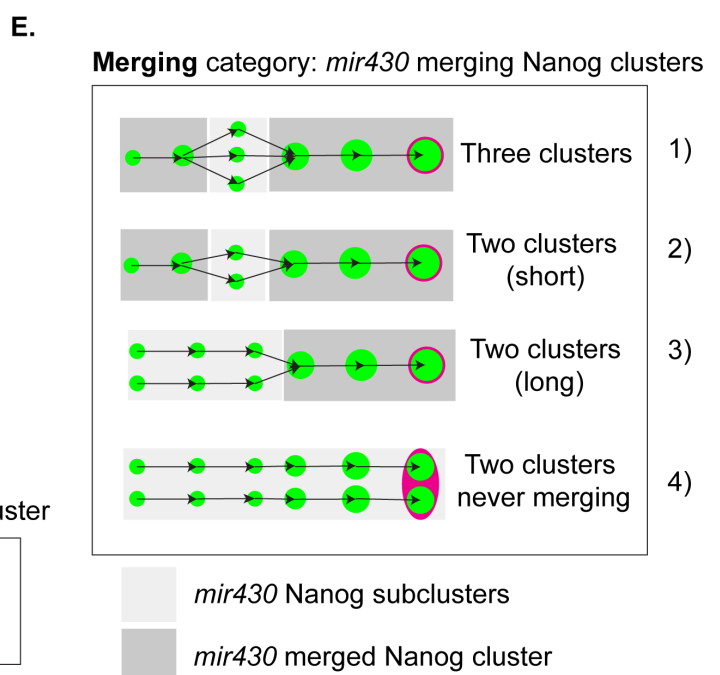
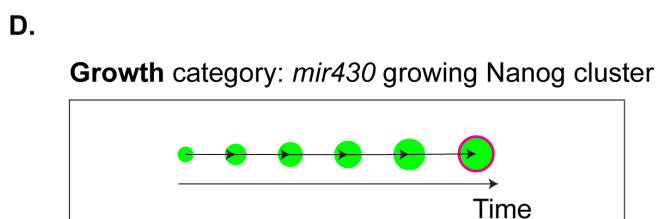
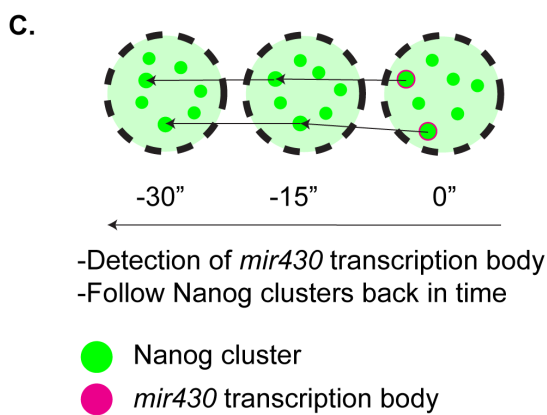
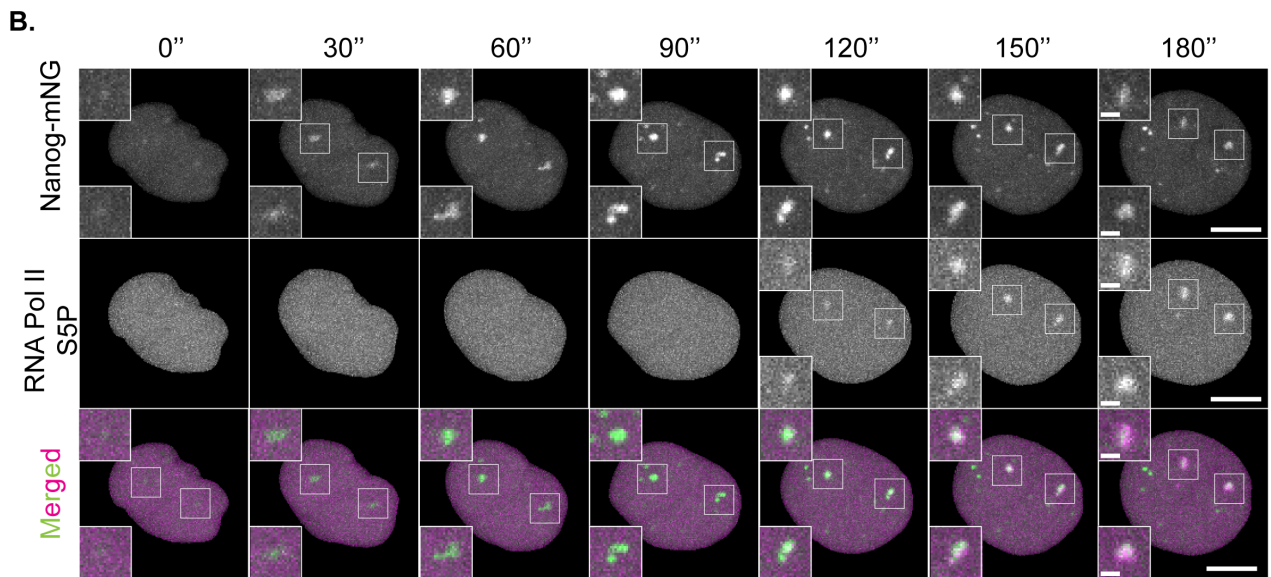
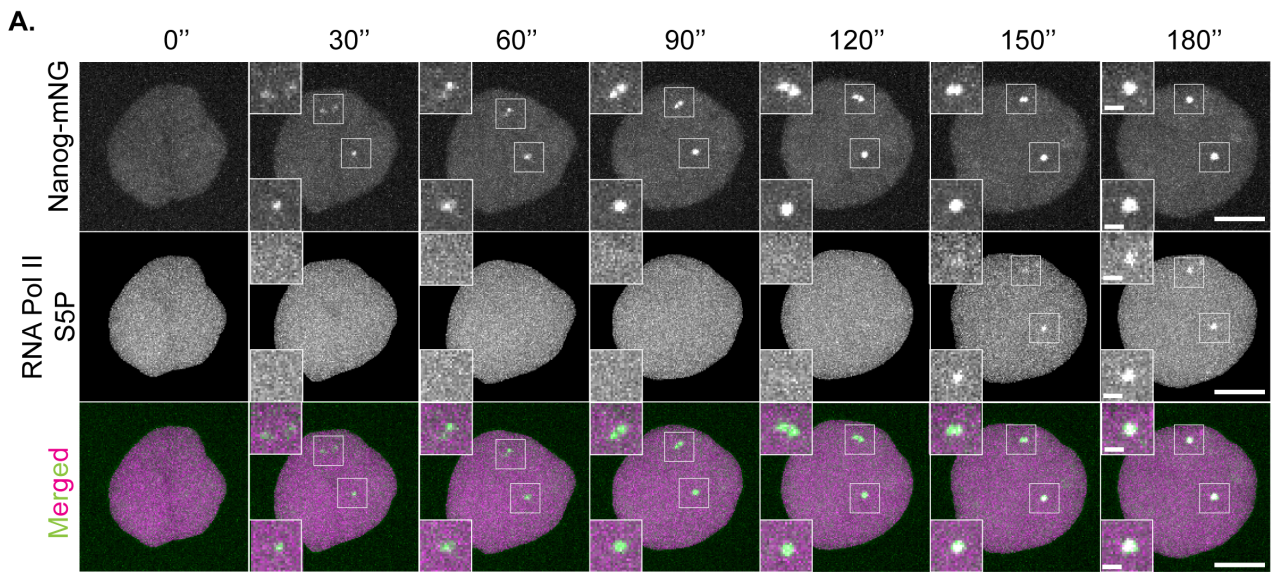


Figure 15: The *mir430* S5P+ Nanog clusters can form by growth or by merging of smaller clusters.

A. Time-series of the same *nucleus* at 256-cell stage every 30 seconds showing the formation of two *mir430* Nanog clusters, *mir430* transcription bodies and merged. Upper zooms show a cluster formed by growth and merging of smaller clusters. Lower zooms show a cluster formed by growth only. **B.** Time-series of the same nuclei at 512-cell stage every 30 seconds showing the formation of two *mir430* Nanog clusters, *mir430* transcription bodies and merged. Upper zooms show a cluster formed by growth only. Lower zooms show two smaller clusters never merging, both colocalizing with *mir430* transcription body. Bigger scale bars = 5 μm . Smaller scale bars = 1 μm . **C.** Schematics representing the method to study the *mir430* Nanog cluster formation. The same *mir430* Nanog cluster is followed back in time. **D.** Schematics representing the growth category for the *mir430* Nanog cluster. A single *mir430* Nanog cluster grow until transcription start. **E.** Schematics representing the merging category for *mir430* Nanog cluster, and the four associated subcategories. At least two *mir430* Nanog subclusters are merging and splitting.

I have observed that in the case of the *mir430* merging Nanog clusters, the Nanog cluster formation process is quite variable in term of how many *mir430* Nanog subclusters I can visualize and for how long they can be observed during consecutive time frames before merging (**Figure 15E**). To be able to study further the potential functional of merging, especially related to the time it takes for *mir430* Nanog subclusters to merge, I decided to categorize the merging category further into four subcategories based on, 1) what is the maximum number of *mir430* Nanog subclusters that can be observed during the *mir430* Nanog formation process, 2) for how long I can observe the same *mir430* Nanog subclusters before they merge into a larger as the merged *mir430* Nanog cluster (**Figure 15D**).

- First subcategory consists of all Nanog cluster formation process where I can observe three distinct *mir430* Nanog subclusters, cluster during at least one frame, then merging all together into a larger merged *mir430* Nanog cluster. This subcategory is referred as “Three clusters”
- Second subcategory consists of two *mir430* Nanog subclusters, observed during one or two consecutive frames, before merging into a larger merged *mir430* Nanog cluster. This subcategory is referred as “Two clusters (short)”.

- Third subcategory consists of two mir430 Nanog subclusters, observed during at least three consecutive frames, before merging into a larger merged mir430 Nanog cluster. This subcategory is referred as “Two clusters (long)”.
- Fourth subcategory consists of two mir430 Nanog subclusters, that are never merging in the last five frame before or three frames after transcription start. This subcategory is referred as “Two clusters never merging”.

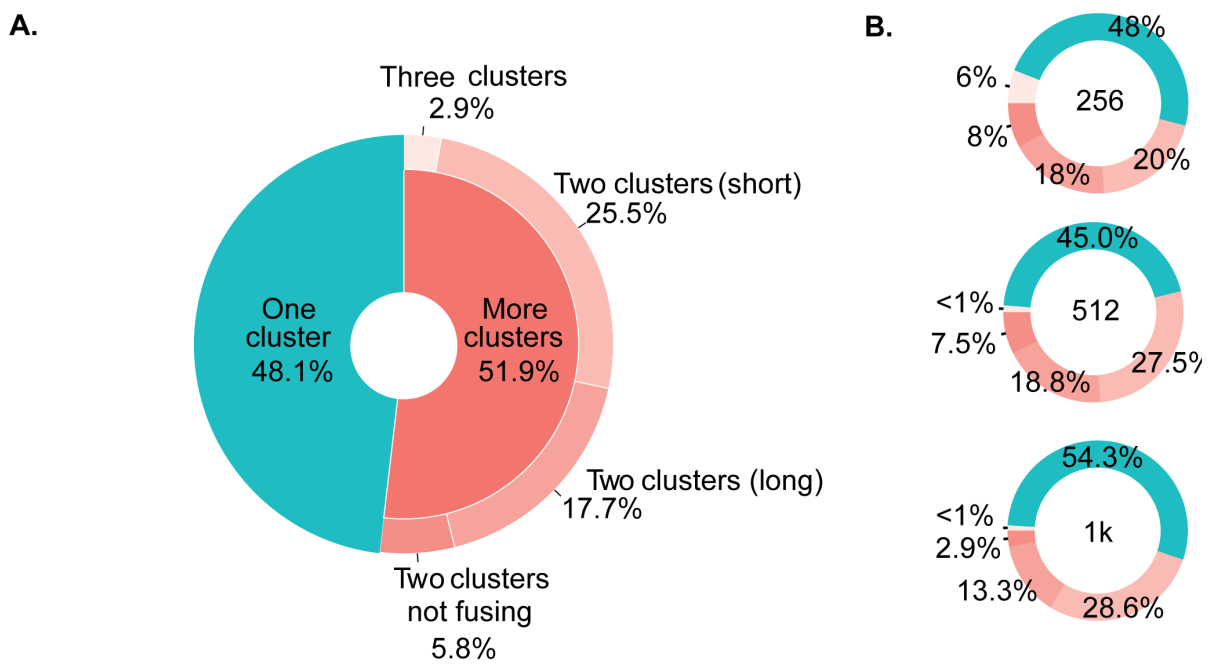


Figure 16: Frequency of the different mir430 Nanog cluster formation mechanisms.

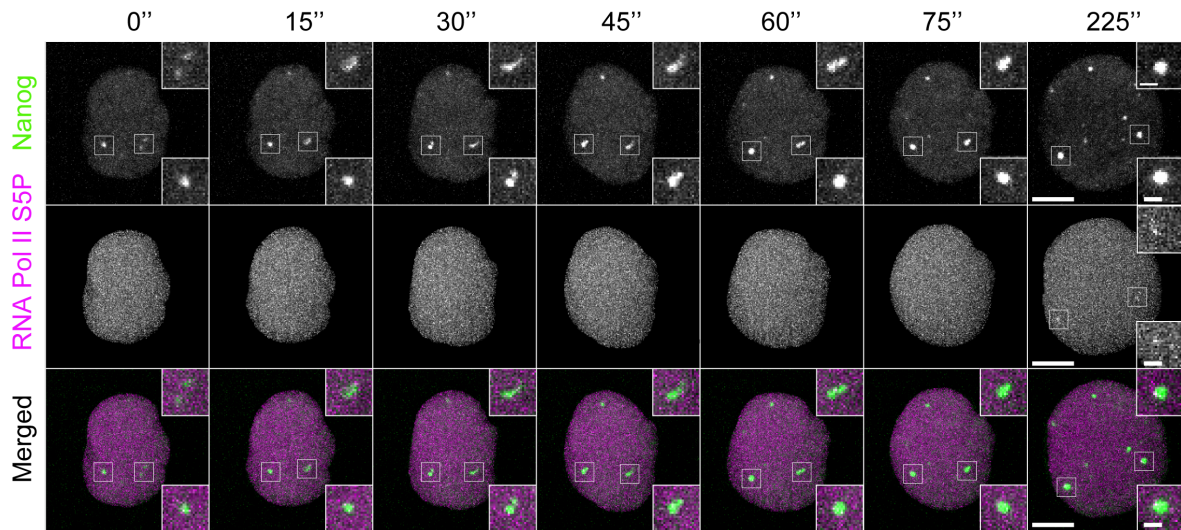
A. Pie-chart showing how often the growing and merging categories, and subcategories, can be observed. All stages mixed. Phenotypes were determined by visual inspection (see Methods). **B.** Pie-chart showing the frequency of the different mechanisms occurring at for 256-, 512- and 1k-cell stage. Categories and subcategorization were determined by visual inspection (see Methods). For calculating the number, the following numbers of nuclei/embryos were used: 256-cell stage: N1=5, N2=8, n=31, 512-cell stage : N1=5, N2=8, n=55, 1k-cell stage : N1=5, N2=7, n=68, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed.

To study the frequency the two mechanisms of mir430 Nanog cluster formation, as well as the frequency for the subcategories concerning the merging category, I counted the occurrences for each category and subcategory. First, I observed that the growing and the merging occurs at the same frequency (48.1% of the growing category and 51.9% for the merging category). Concerning the subcategory of the merging mechanism, I observed almost exclusively two mir430 Nanog subclusters (94.4% of the cases, subcategory 2, 3 and 4 all together) and never more than three mir430 Nanog subclusters (5.6% of the cases, first subcategory) (**Figure 16A**). The merging process is quite fast as in 49.2% of the cases, the mir430 Nanog subclusters merged in less than three frames after their first detection (subcategory 2), that being less than 45 seconds. I can observe that mir430 Nanog subclusters took more than 45 seconds to merge in 34.1% of the cases (subcategory 3). These results suggest that the time it takes for mir430 Nanog subclusters to merge is quite variable. Interestingly, in 11.1% of the cases, I could also observe two mir430 Nanog subclusters that were never merging but both colocalizing with the same mir430 transcription body (**Figure 15B**, lower zoom, subcategory 4). The distribution of the different subcategories from 256-cell stage until 1k-cell stage is very stable (**Figure 16B**), suggesting robust processes underlying mir430 Nanog cluster formation that is independent of the developmental stage.

To conclude, I observed that the formation process of the mir430 Nanog cluster can be classified in two categories: the growing category and the merging category. These clusters can form either by growth only or by growth and concomitant merging of mir430 Nanog subclusters. Both mechanisms occur at equal frequency. The merging time of smaller merging Nanog clusters is variable, and they are mostly two mir430 Nanog subclusters, rarely three.

3.2.2 – TRANSCRIPTION IS NOT RESPONSIBLE FOR THE MERGING OF NANOG CLUSTERS

A.



B.

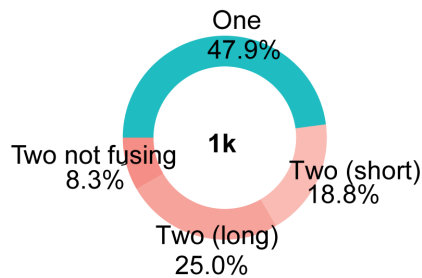


Figure 17: Transcription is not responsible for the merging of Nanog clusters.

A. Time-series of the same nuclei at 512-cell stage every 30 seconds showing the formation of two *mir430* Nanog clusters, *mir430* transcription bodies labeled with the Fab recognizing the RNA Pol II S5P, and merged. Upper zooms show a cluster formed by growth and merging of smaller clusters. Lower zooms show a cluster formed by growth only. Bigger scale bar = 5 μm . smaller scale bar = 1 μm . **B.** Pie-chart showing how often the different phenotypes can be observed at 1k-cell stage. All stages mixed. Phenotypes were determined based on the tracking (see Methods). For calculating the number, the following numbers of nuclei/embryos were used: 1k-cell stage : N1=3, N2=3, n=26, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed.

RNA has been involved in the formation ^{27,110,227,318} as well as the dissociation of TF clustering²²⁸, especially eRNA at SEs ^{325,326}. Moreover, we have shown previously that transcription is responsible for dissociation of Nanog clusters in smaller clusters

after transcription starts (**Error! Reference source not found.**)²⁷⁰. These studies and results underlie the importance of RNA in TF cluster formation. Considering the role of RNA in TF clustering formation and dissociation, I wondered if transcription was impacting the *mir430* Nanog cluster formation.

To assess this dependence, I injected as previously Nanog-mNG as a mRNA, as well as the Fab recognizing RNA Pol II S5P and the transcription elongation inhibitor α -amanitin in 1-cell stage embryo. The α -amanitin has been reported to lead to the degradation of the RNA Pol II during the initiation step of transcription, but not preventing from the phosphorylation of the 5th serines on the CTD³²⁷. By consequence, even if no mature mRNA is produced, I can still distinguish the S5P+ and S5P- Nanog cluster. I then image using a time interval of 15 seconds to study the formation of the *mir430* Nanog cluster.

In the absence of transcription, I could still observe *mir430* Nanog subclusters merging to give rise to the merged *mir430* Nanog cluster (

Figure 17A). Moreover, the proportion of the different mechanisms described previously (**Figure 15D, E**) did not change in a drastic way (

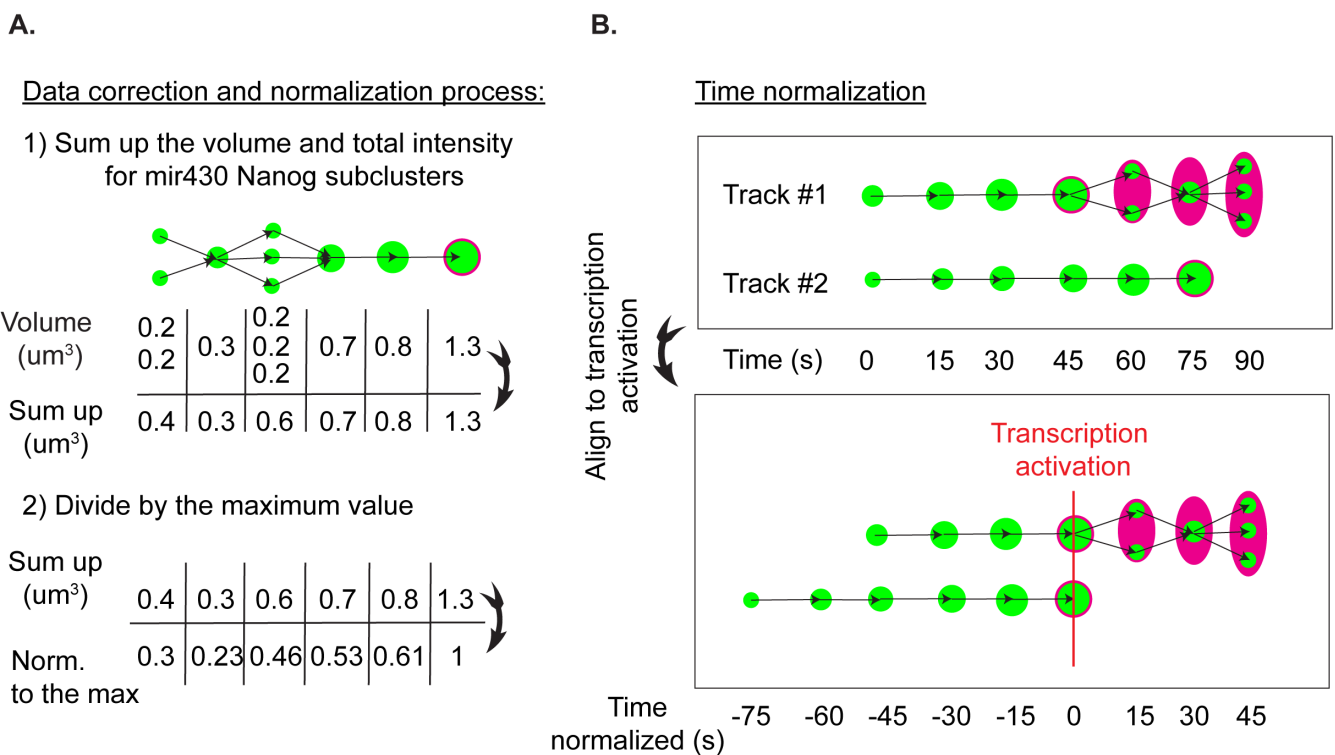
Figure 17B). Indeed, in the case which transcription was inhibited, 47.9% of the *mir430* Nanog clusters were classified in the growing category while 52.1 % of the Nanog clusters were classified in the merging category, while in the presence of transcription, the percentages were respectively of 48.1% and 51.9%. In the case of the merging category, I further classified the different observed in four subcategories to explore better the dynamics of the merging mechanism (**Figure 15E**). While in the case of transcription inhibition, I did not observe cases where three *mir430* Nanog subclusters were present (subcategory 1,

Figure 17B), the frequency for two *mir430* Nanog subclusters for a short (subcategory 2, less than three frames) or a longer time (subcategory 3, more than three frames) were of respectively 18.8% and 25% while they were of 25% and 17.7% in the absence of inhibition. These results suggest that the transcription is not influencing the time it takes for *mir430* Nanog subclusters to merge. Finally, 8.3% of the *mir430* Nanog subclusters were never merging in the inhibition dataset while it was 5.7% in the absence of inhibition.

Altogether, these results suggest that *de novo* transcription is not influencing in a significant way the dynamics of Nanog cluster formation related to the growth and merging of *mir430* Nanog subclusters.

3.2.3 – HIGH LEVELS OF NANOG CORRELATES WITH ACTIVATION OF TRANSCRIPTION

As discussed in chapter 3.2.1, the S5P+ Nanog clusters are the largest, the most concentrated in Nanog and the clusters containing the most Nanog in the nucleus when transcription starts. I wondered what the temporal relationship between transcription activation and Nanog amount was at the mir430 Nanog cluster. To be able to compare the results without the bias of the stage, all the results in this chapter are for the 1k-cell stage.



● *mir430* Nanog cluster

Figure 18: Mir430 Nanog fluorescence levels and time line are normalized

A. Schematic explaining how the mir430 Nanog fluorescence level are normalized in the case of several mir430 Nanog subclusters at the same time point. If two or more mir430 Nanog subclusters are present at the same time point, the volume and total intensity of these subclusters is summed up. Their mean intensity is averaged. After this operations, the values for the same mir430 Nanog cluster over time are further normalized to be able to compare all Nanog mir430 clusters from different nuclei. Each value for the volume, total intensity or mean intensity is divided by their respective max values that the mir430 Nanog is reaching in the track. Such normalization allows to better reflect when the maximal values are reached for each mir430 Nanog clusters. **B.** Schematic explaining how the tracks for mir430 Nanog clusters are aligned in time. Transcription initiation is considered as the 0 in the time line, and all other time points are subtracting for the time point when transcription start, all of mir430 Nanog clusters.

To answer this question, I quantified the mean intensity, the volume and the total intensity of the mir430 Nanog clusters over time as described previously (**Figure 13B**). I measured the mean intensity, the volume, and the total intensity of mir430 Nanog clusters using a time interval of 15 seconds using the same data as previously (**Figure 15**). As I do not start imaging at the same time during the cell cycle for each nucleus, I used the frame when the mir430 locus starts to be transcribed to normalize the time and be able to compare the data by using a unique time reference (here, *mir430* transcription initiation, **Figure 18B**). Furthermore, I have shown that the mir430 Nanog clusters can form by the merging of mir430 Nanog subclusters giving rise to a larger merged mir430 Nanog cluster. As the mir430 Nanog subclusters are detected individually, their respective volumes and total intensities need to be sum up in order to take in account of the total amount of Nanog that is colocalizing with the mir430 same transcription body (**Figure 18A**). Then, when two or more mir430 Nanog subclusters colocalizing with the same mir430 transcription body, were present, their volumes and total intensities were summed up, their mean intensities averaged, allowing to obtain only one value for each measurement at the same time point. Moreover, because I am interested in comparing the evolution of the mean intensity, volume, and total intensity of the mir430 Nanog signal colocalizing with one mir430 transcription body, I decided to normalize the values to the maximum reached over time (**Figure 18A**). For example, if the same mir430 Nanog cluster over time is reaching its maximum volume two time points before transcription activation, all the values of volume for this mir430 Nanog cluster over time will be normalized to this maximum.

When looking at the mean intensity, an indirect read-out for concentration, I observed that 150 seconds before transcription starts, mir430 Nanog clusters were already reaching in average at 83% of their maximum concentration. (**Figure 19A**). The concentration increases and peaks 15 seconds after transcription start, then decreasing slowly. Taking in consideration the lowest levels, the concentration of Nanog proteins inside the mir430 Nanog clusters only increased by 13%. However, on average volume increases from 25% at 105 seconds before transcription starts to a peak of 83% of the maximum at transcription activation, plateaus for 60 seconds before decreasing to reach 50% 150 seconds after transcription started. A similar behavior can be observed for total intensity. The results suggest that while the mir430 Nanog clusters are already concentrated in Nanog proteins, their size varies the most, influencing the total amount of Nanog as well. Moreover, the decrease observed in volume, mean intensity and total intensity after transcription start might be related to the dissociation of mir430 Nanog clusters into dissociated Nanog

clusters, upon the mir430 transcription body growth (**Error! Reference source not found.B**).

Then, I looked at individual example of mir430 Nanog clusters. I could observe a similar trend: the mean intensity is increasing, while the volume and the total intensity increased the most. (**Figure 19B**). As for the average, this individual mir430 Nanog cluster reaches a maximum of volume and total intensity at transcription start. However, for this single example, peak of mean intensity is reached at 60 seconds after transcription start while it is at 30 seconds in average (**Figure 19A**), showing variability between single examples and the average distribution. Moreover, average curves showed that the transcription starts at 83% of the maximum volume, not 100%. It suggests that individual mir430 Nanog clusters do not need to reach a maximum of mean intensity, volume or total intensity to start transcription.

To check this dependance, I plotted the time at which the maximum peak for each feature is reached compared to transcription start (**Figure 19C**). I can observe that for most of the mir430 Nanog clusters, they reached their maximum peak after transcription has started, but the moment when the maximum is reached is variable, ranging between 45 seconds before until 60 seconds after transcription started. Only 25% of the transcribing Nanog clusters reached their peak of fluorescence and volumes prior to transcription start (**Figure 19D**). These results suggest that mir430 Nanog clusters do not need to reach a peak of concentration, volume, or total amount of proteins, for mir430 to start transcribing, but a certain proportion of their maximum, suggesting a threshold.

As it seems that the peak of fluorescence is not the most important predictor of transcription activation, I checked at what percentage of their maximum transcription starts. I observed that on average, transcription starts when Nanog clusters reach 97% of their maximum mean intensity, 83% of their maximum volume and 85% of their maximum total intensity (**Figure 19E**). Thus, most of the Nanog cluster do not start transcription when they reach their maximum level of fluorescence but a proportion of it. It is consistent with the plateau observed for the average volume and total intensity around transcription start (**Figure 19A**). It suggests that reaching high levels of Nanog close the maximum peak is sufficient to lead to activation.

To conclude, levels of fluorescence at mir430 Nanog clusters are highly correlated with transcription initiation. While the mean intensity does not varies much, the volume and the total intensity increase in average by more than three-fold. Moreover, mir430

transcription does not initiate when a peak of concentration, volume, or total amount of Nanog proteins is reached at the mir430 at the mir430 Nanog cluster, but only a proportion close from these maxima. These results suggest that reaching a peak is not the most important but rather that there is a threshold to reach, to lead to transcription activation.

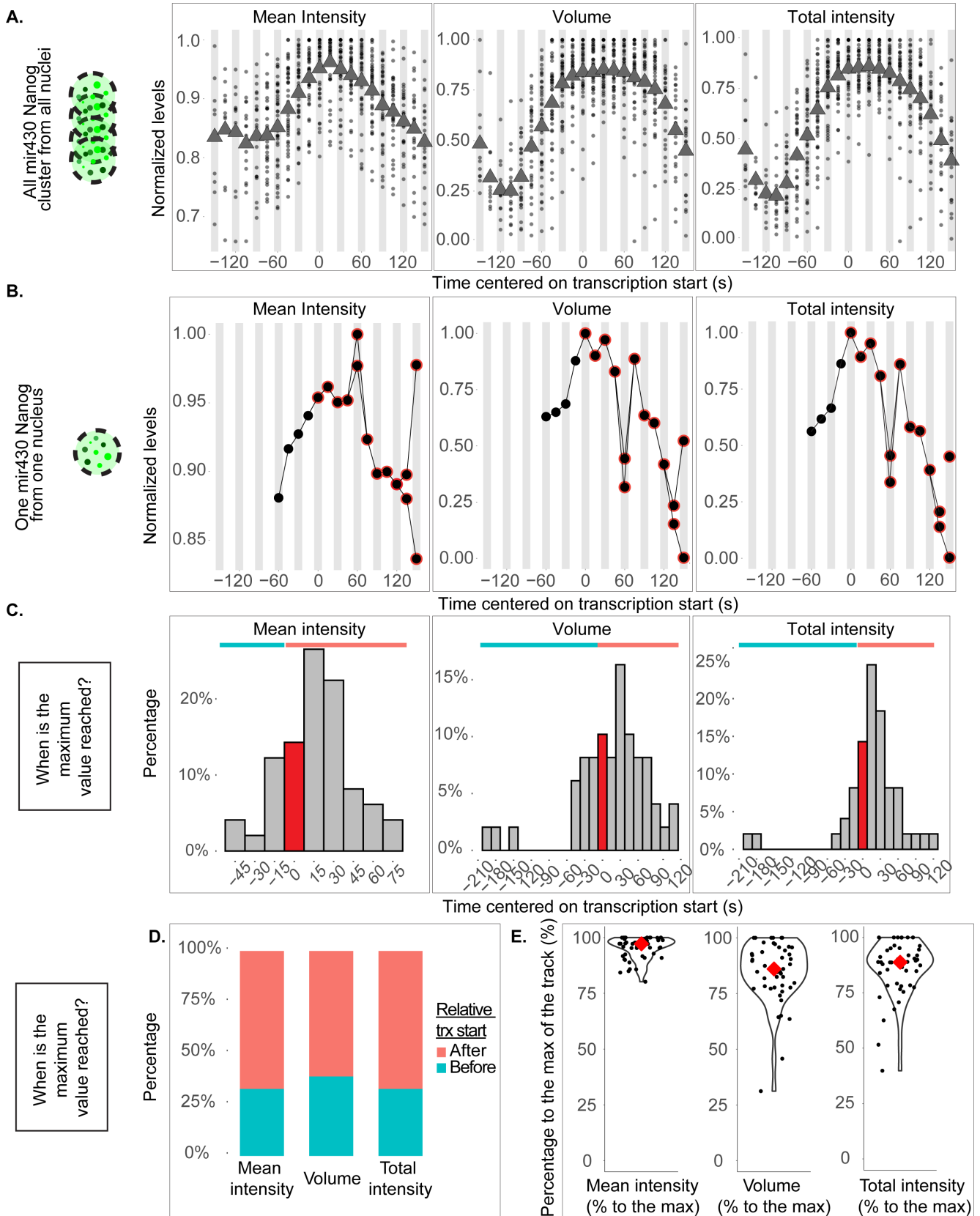


Figure 19: *Mir430* transcription is activated when *mir430* Nanog clusters reach high amount of Nanog.

A. Normalized mean intensity, volume, and total intensity of mir430 Nanog clusters during one cell cycle centered on transcription initiation mixing at 1k-cell stage. **B.** Example of a single track from panel A. Normalized mean intensity, volume, and total intensity for an individual example of mir430 Nanog cluster centered on transcription initiation. Red strokes indicate time points when the mir430 transcription body is present. **C.** Histograms representing at which time point around transcription start Nanog clusters reach their maximum for mean intensity, volume, and total intensity. Red bar indicates the transcription start. Negative times indicates that the maximum peak is reached before transcription initiation, positive times after. **D.** Histogram showing if the mir430 transcription body starts transcription before or after maximum peak for mean intensity, volume, and total intensity for the associated mir430 Nanog clusters. **E.** Percentage of the max value for which Nanog clusters start transcription for mean intensity, volume, and total intensity. Red diamond represents the mean. For this figure, the following numbers of nuclei/embryos were used :256-cell stage: $N1=5$, $N2=9$, $n=19$, 512-cell stage: $N1=5$, $N2=7$, $n=21$, 1k-cell stage: $N1=5$, $N2=6$, $n=27$, where $N1$ the number of biological replicates, $N2$ the number of embryos and n the number of nuclei analyzed. Mean intensity, volume and total intensity are normalized for the maximum in time of each Nanog cluster.

3.2.4 –MIR430 NANOG CLUSTERS HAVE A MAXIMUM TF LOADING CAPACITY

Thus far, I showed that mir430 transcription starts when Nanog clusters reach high concentration for Nanog, volume and total amount of Nanog proteins, but not necessarily a peak, suggesting the presence of a threshold (**Figure 19C, D, E**). However, to define a threshold, I need to know what the maximum possible reached for the concentration, the size and the total amount of mir430 Nanog clusters. Indeed, as transcription leads to the dissolution of Nanog clusters (**Error! Reference source not found.A, Figure 19A**), it is possible that levels of Nanog would keep increasing if transcription did not start.

To test this, I used the same data as in Chapter 3.2.2 and again measured the mean intensity, volume and total intensity of mir430 Nanog clusters when transcription initiation and elongation is inhibited with alpha-amanitin. As described in chapter 3.2.2, the use of alpha-amanitin block the RNA Polymerase II during transcription initiation but does not stop the phosphorylation of the 5th serines of CTD, allowing me to still detect when the transcription process is engaged ³²⁸. To normalize the intensity levels and the volume, I used the same method as previously described (**Figure 18A, B**).

By plotting the mean intensity, volume and total intensity centered the first appearance of RNA Pol II S5P signal, I observed that mir430 Nanog levels do not drop anymore but stay rather stable for all the three parameters (**Figure 20A, B, C**). This result suggests that there are no more Nanog proteins recruited in the clusters after reaching peak levels and suggests that there is a maximal loading capacity underlying those Nanog clusters. Potentially, the underlying DNA seeding the mir430 Nanog clusters cannot bind more TFs, when the mir430 Nanog clusters reach a peak of fluorescence. The existence of such a maximal loading capacity is important as it gives consistency to the idea that there is a threshold of Nanog to reach in order to active mir430 transcription. Interestingly, such a maximum loading capacity has already been reported for Sox2 and Brd4 clusters in mESC cells²³⁰.

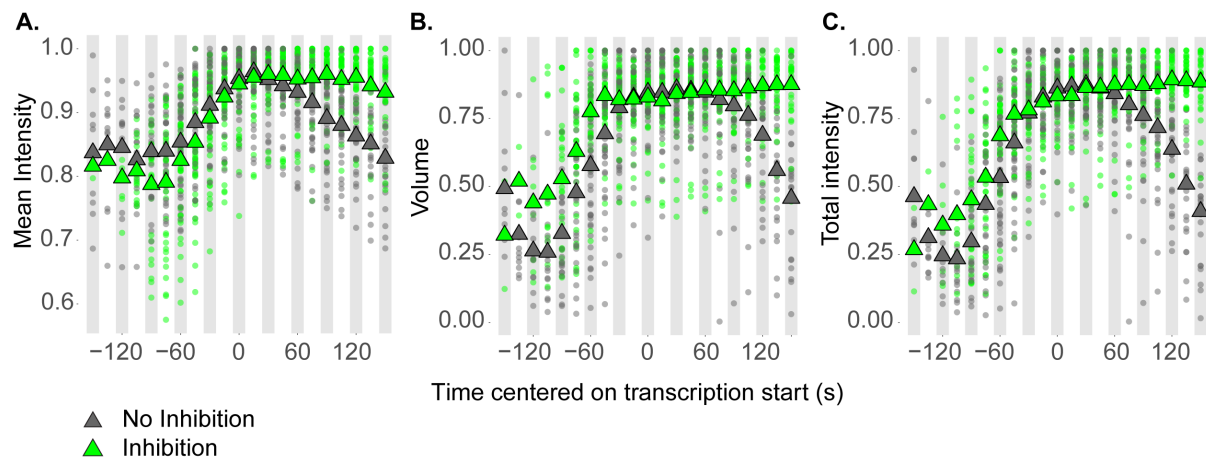


Figure 20: mir430 Nanog clusters do not grow further in transcription-inhibited conditions.

A. Mean intensity of mir430 Nanog clusters during one cell cycle centered on transcription initiation at 1k-cell stage. **B.** Volume of mir430 Nanog clusters during one cell cycle centered on transcription start at 1k-cell stage. **C.** Volume of mir430 Nanog clusters during one cell cycle centered on transcription start at 1k-cell stage. **Green:** Transcription inhibited data. **Grey:** WT situation. For calculating the number, the following numbers of nuclei/embryos were used: WT data: 1k-cell stage: $N_1=5$, $N_2=6$, $n=27$. Transcription-inhibited data: $N_1=3$, $N_2=3$, $n=26$, where N_1 the number of biological replicates, N_2 the number of embryos and n the number of nuclei analyzed. Mean intensity, volume and total intensity are normalized for the maximum in time of each Nanog cluster (**Figure 18A**).

3.2.5 – GROWING OR MERGING: NO DIFFERENCE IN LEVEL OF NANOG AND TIME OF ACTIVATION

I have previously shown that the mir430 Nanog clusters can form by two ways: by growth only, or by merging of mir430 Nanog subclusters that independently grow as well (**Figure 15E**). Following this observation, I have categorized the mir430 Nanog clusters in two categories based on these mechanisms: 1) Growing: Clusters for which only one cluster can be observed at all time points before transcription initiation, and 2) merging: clusters for which I can observe two or more mir430 Nanog subclusters during at least one time point before transcription initiation. I wondered if these two observed mechanisms exhibit a different dynamic growth and whether these two mechanisms relate to difference in transcription activation timing. To do so, I used the same data as in Chapter 3.2.3.

First, I checked whether the mir430 Nanog clusters reached the same volume at transcription initiation, depending on how they form, by growth or merging. The volume here is not normalized as previously because I am interested in comparing the raw volume values between all the nuclei. The difference in volume is not statistically significant between the two categories (**Figure 21A**, Wilcoxon test, $p=0.48$). Previously, I further classified the merging category where we observed at least two smaller merging Nanog clusters (**Figure 18C**). In the case of merging of clusters, if more than two mir430 Nanog subclusters were observed during one or two consecutive time points, I defined it as “Two clusters (short)”, while if more than two mir430 Nanog subclusters could be observed for three or more consecutive time points, I defined it as “Two clusters (long)”. Checking for the volume of mir430 Nanog subclusters in breaking down further for these three categories, I did not observe any difference either (**Figure 21A**, Wilcoxon test, $p=0.41$, $p=0.47$ and $p=0.98$). These results suggest that the mir430 Nanog clusters do not reach a different size at the transcription initiation, depending on how they form.

I hypothesized then that the different observed mechanisms could lead to a difference in activation time. I start imaging at different moment of the cell cycle, I used previously the transcription activation time to normalize the time lime and be able to compare the different data together. Here, I do not have any other time referential as I want to compare the transcription start time itself. I thus used the two

mir430 Nanog clusters from the same nucleus to correlate mechanism of formation and transcription start time as the two mir430 Nanog clusters from the same nucleus are imaged in the same time, and are thus at the same moment of the cell cycle (**Figure 21B**). I observed that in nucleus which the mir430 Nanog clusters were formed by the same mechanism, both mir430 Nanog clusters start to be positive for transcription on average with one time point difference (**Figure 21B**, left boxplot). There was no difference between the activation times when the mir430 Nanog clusters were formed by different mechanism either (**Figure 21B**, right boxplot). The differences in time of activation are not significant between the two comparisons (**Figure 21B** Wilcoxon test, $p=0.32$). However, it is interesting to observe that transcribing Nanog clusters can have large variation in their transcription activation time, even if they are found in the same nucleus. Finally, I was interested in comparing the mean intensity, volume, and total intensity over time as (as for **Figure 19A** and **Figure 20**) for the two categories. Mean intensity, volume and total intensity were normalized as previously described (**Figure 18A**). I did not observe a strong difference between the distribution associated to a specific category at any of the studied parameters (**Figure 21C**), suggesting that they are not differences in mir430 Nanog cluster growth depending on the mechanism of formation.

To conclude, the mechanism of formation of the mir430 Nanog cluster colocalizing with the mir430 transcription body does not influence the size of the mir430 Nanog cluster at the moment of transcription initiation nor when transcription is activated. Moreover, both mechanisms lead to similar growth pattern, suggesting that the Nanog clusters reach the same amount of Nanog at the same speed, independently of how they are formed.

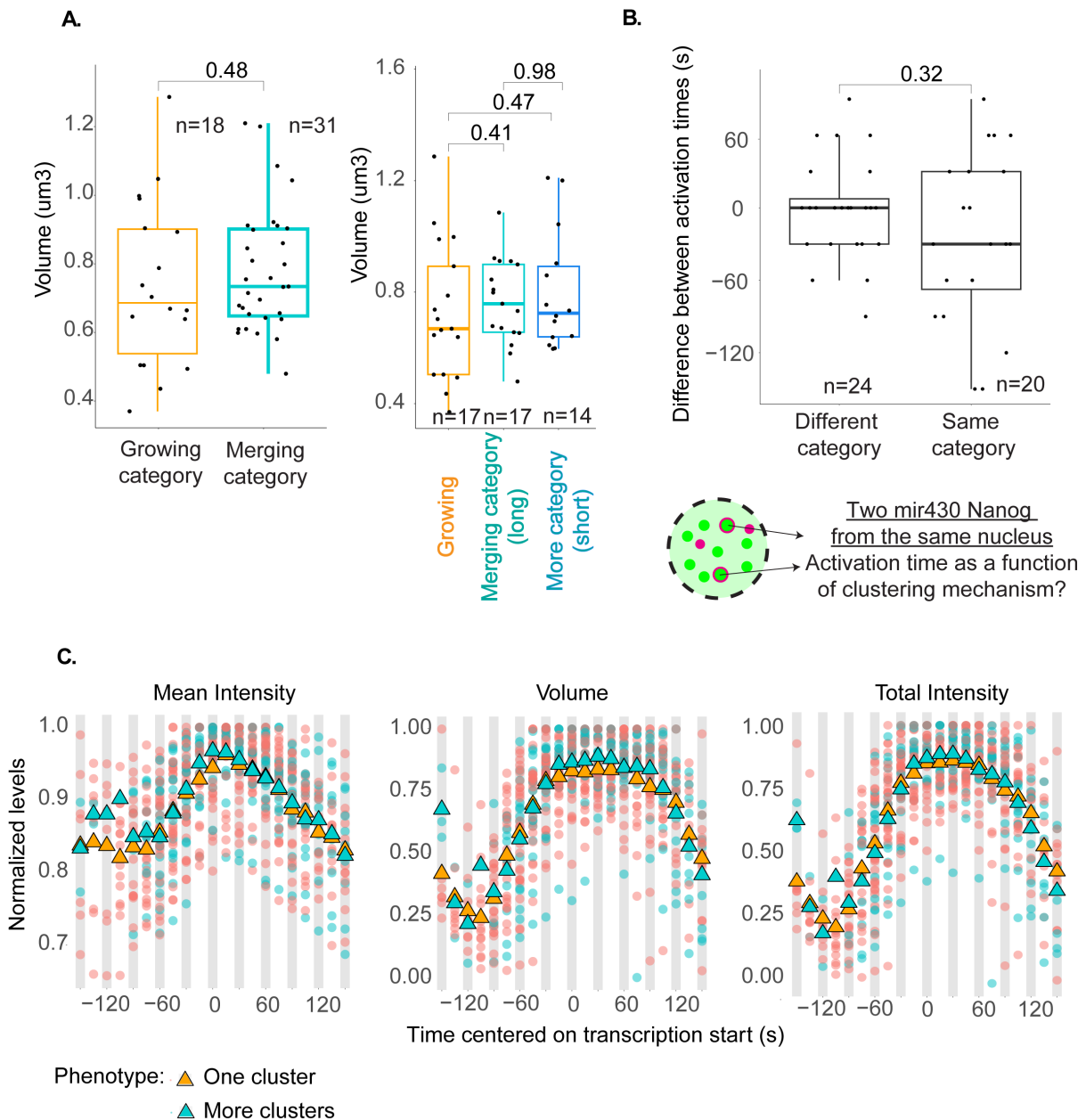


Figure 21: The mir430 Nanog clusters do not grow differently based on their mechanism of formation.

A. Boxplots showing the volume of Nanog clusters at activation with different way of forming at 1k. (left) Comparing one cluster versus more cluster. (right) Comparing one cluster versus a breakdown of more cluster possibilities (defined in **Figure 15E**). Nomenclature: s: short, l: long. P-values from Wilcoxon test. **B.** Difference in time of transcription activation between clusters of displaying the same mechanism of formation (identical phenotype) versus different mechanisms of formation (different phenotype), within the same nucleus. Wilcoxon test. **C.** Distribution of average mean intensity, volume, and total intensity for both type of formation mechanism over time (s). Time is centered on transcription activation. For calculating the number, the following numbers of nuclei/embryos were used: WT data: 1k-cell stage: N1=5, N2=6, n=27, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed. **Orange:** Always one cluster. **Blue:** At least two clusters during one frame. Mean intensity, volume and total intensity are normalized for the maximum in time of each Nanog cluster.

3.2.6 - MERGING OF SMALLER NANOG CLUSTERS IS IMPORTANT TO REACH HIGH ENOUGH LEVELS OF NANOG AND START TRANSCRIPTION.

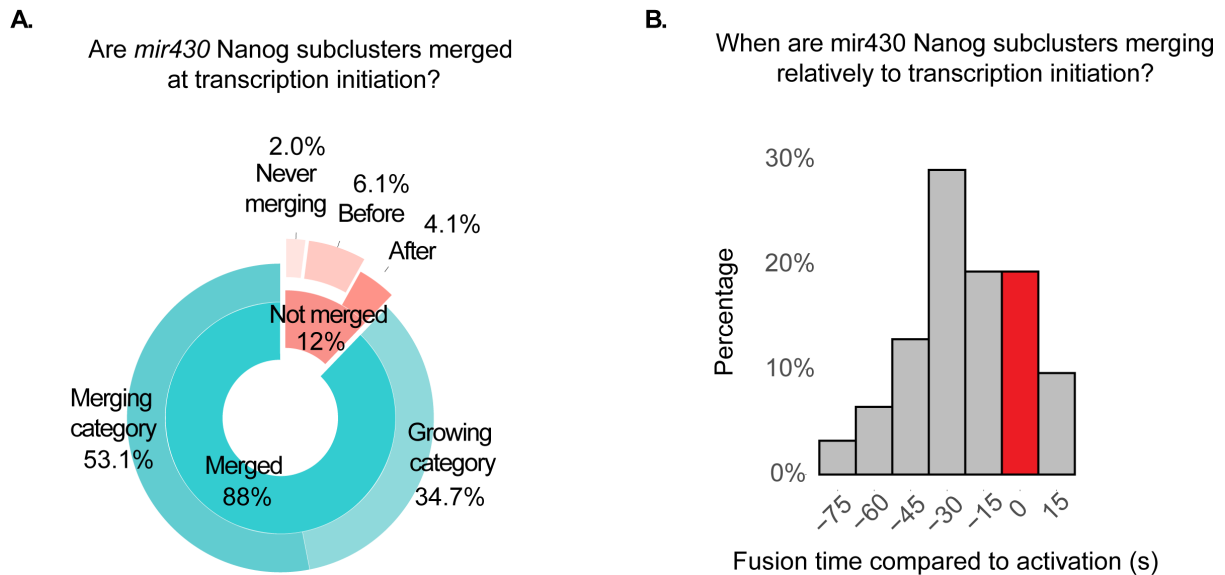


Figure 22: *Mir430* Nanog clusters can start transcription without merging.

A. Pie chart showing the proportion of *mir430* Nanog subclusters merged or non-merged at transcription activation at 1k-cell stage. If they are not merged, do they ever merge before or after transcription initiation **Blue**: merged at transcription start. **Salmon**: not merged at transcription start. **B.** Histogram showing the merging time compared to transcription start. **Red** bar indicate when transcription start. Negative times indicates than merging happened before transcription start, positive times after transcription start. For calculating the number, the following numbers of nuclei/embryos were used : WT data : 1k-cell stage : $N_1=5$, $N_2=6$, $n=27$, where N_1 the number of biological replicates, N_2 the number of embryos and n the number of nuclei analyzed.

I have previously shown that *mir430* Nanog clusters colocalizing with the *mir430* transcription body can be formed by two mechanisms: growing or merging (

Figure 17A). As I have found that the two different mechanisms are not related to a difference of volume of the *mir430* Nanog cluster at the moment of activation, a difference in time of activation nor a difference in growth rate (**Figure 21**), I was wondering how important is the merging process, and especially what is the correlation between the merging of smaller merging Nanog clusters and transcription activation. I have also showed that the *mir430* Nanog clusters could reach a threshold

of Nanog amount to start transcription (**Figure 19**). The merging of mir430 Nanog subclusters could potentially lead to an increase in concentration or volume to support transcription start.

I first asked whether transcription always starts when one mir430 Nanog cluster is observed, meaning in the case of the merging category, that the mir430 Nanog subclusters are merged together when transcription initiates. Among all mir430 Nanog clusters, 88% of them are visualized as one single cluster when transcription starts (**Figure 22A**). Among the remaining 12% for which at least two mir430 Nanog subclusters can be visualized when transcription start, 6.1% have merged at least once prior to the onset of transcription, 4% merged at least once within the next 45 seconds after transcription starts and 2% never merged (always two or more mir430 Nanog subclusters are detected). These results suggest that most of the mir430 Nanog clusters are detected as one mir430 merged Nanog cluster when transcription starts, suggesting that merging is important. Focusing only on the merging category, when I look at the time of merging relative to transcription start, 67% of the smaller clusters merged within 30 seconds prior to the onset of transcription (**Figure 22B**). This result suggests that the exact moment of merging is not correlated with the transcription initiation but the merging of the mir430 Nanog subclusters has to occur prior to transcription activation.

Then, I was wondering how the merging of mir430 Nanog subclusters was contributing to Nanog cluster growth. Merging two or more mir430 Nanog subclusters into a larger mir430 merged Nanog cluster could increase Nanog levels in different manners: increase the concentration but keep the volume constant, increasing the volume but not the concentration, or a mix of those two possibilities.

To answer this question, I selected only the mir430 merging clusters and I investigated the levels of mean intensity, volume and total intensity centered on the frame when the mir430 Nanog subclusters merged (**Figure 23A**). I observed that while the mean intensity is not changing before and after merging, volume is greatly by two-fold, leading to a big change in total intensity as well. It suggests that the merging of Nanog clusters leads to an increase in volume of a mir430 merged Nanog cluster but do not increase the concentration. It suggests that the two mir430 Nanog subclusters allow the local increase of Nanog proteins without changing Nanog concentration. It also suggests that the concentration is then only increasing by growth, potentially by the binding of Nanog proteins locally.

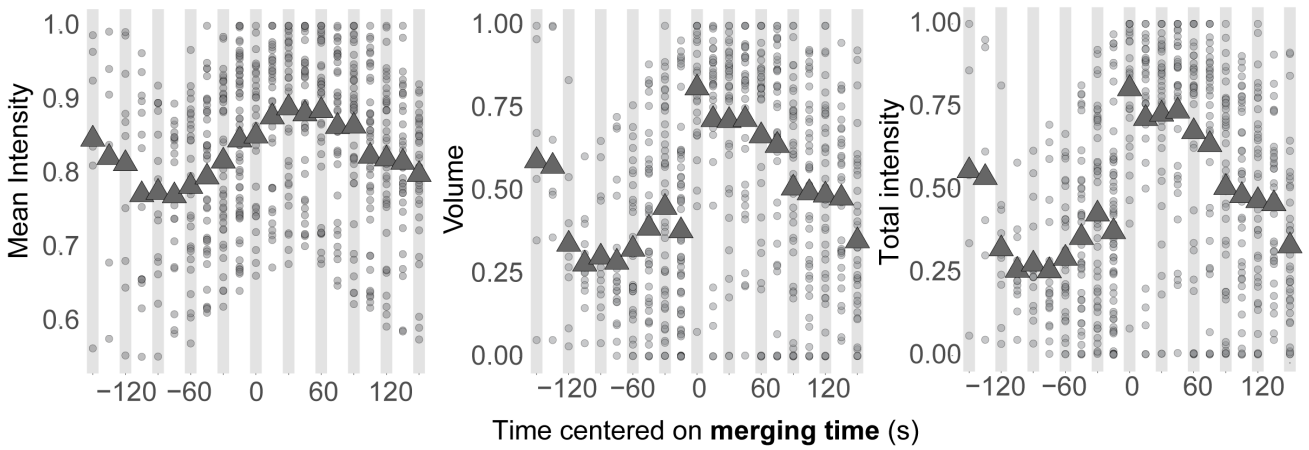
I have shown previously that the mir430 Nanog clusters can load a maximum amount of Nanog before transcription start (**Figure 20**). One can imagine that if two mir430 Nanog subclusters are two halves of the merged mir430 Nanog clusters, then they can only contain half the maximum amount of Nanog each. But I observed that transcription starts on average at 85% of the maximum total amount of Nanog that a mir430 Nanog cluster can load (**Figure 20**). Merging of the two mir430 Nanog subclusters would be then essential to reach a threshold amount. To test this hypothesis, I decided to focus on the mir430 Nanog subclusters that are separated for at least three frames before merging. This allows to follow the volume of the mir430 Nanog subclusters on consecutive time points and see how it evolves. I plotted the volume of mir430 Nanog subclusters (yellow) and compared it to the sum of volume of both mir430 Nanog subclusters (red), as well as the volume of mir430 Nanog growing clusters only formed by growth (black) (**Figure 23B**). We can observe that on average, the total intensity for each mir430 Nanog subclusters increased slowly while the sum of both mir430 Nanog subclusters is equivalent to the mir430 Nanog growing clusters. These results suggest that the merging of both mir430 Nanog subclusters make the total amount of the Nanog proteins increased, above the observed threshold described previously (**Figure 19**). Merging is thus important for reaching high levels of Nanog.

Observing how merging of mir430 Nanog subclusters seem to be important for reaching high levels of Nanog, there are still 12% of mir430 Nanog clusters that start transcription when mir430 Nanog subclusters have not merged, represented here in total 6 Nanog mir430 clusters. I hypothesize that mir430 Nanog subclusters clusters can be of different size and that if one of the mir430 Nanog subclusters is large enough, it can contribute to sufficient amounts of Nanog to reach the threshold and lead to transcription. To confirm this hypothesis, I followed the mir430 Nanog clusters for which transcription starts when mir430 Nanog subclusters are not merged. I considered the sum of their volume at the time transcription starts as 100% and observed what proportion each of the mir430 Nanog subclusters make up. I observed in all cases, one of mir430 Nanog subclusters is larger than the other one (**Figure 23C, E**). For two of the examples, the difference was extreme (**Figure 23E, 2**) and **3**). Indeed, the two largest clusters of each pair was forming 98% of the volume of the sum of both mir430 Nanog subclusters. Interestingly, when I looked at the distance at which the mir430 Nanog subclusters were from the mir430 transcription body at transcription initiation, I observed that the larger cluster of each pair was exactly coinciding with the mir430

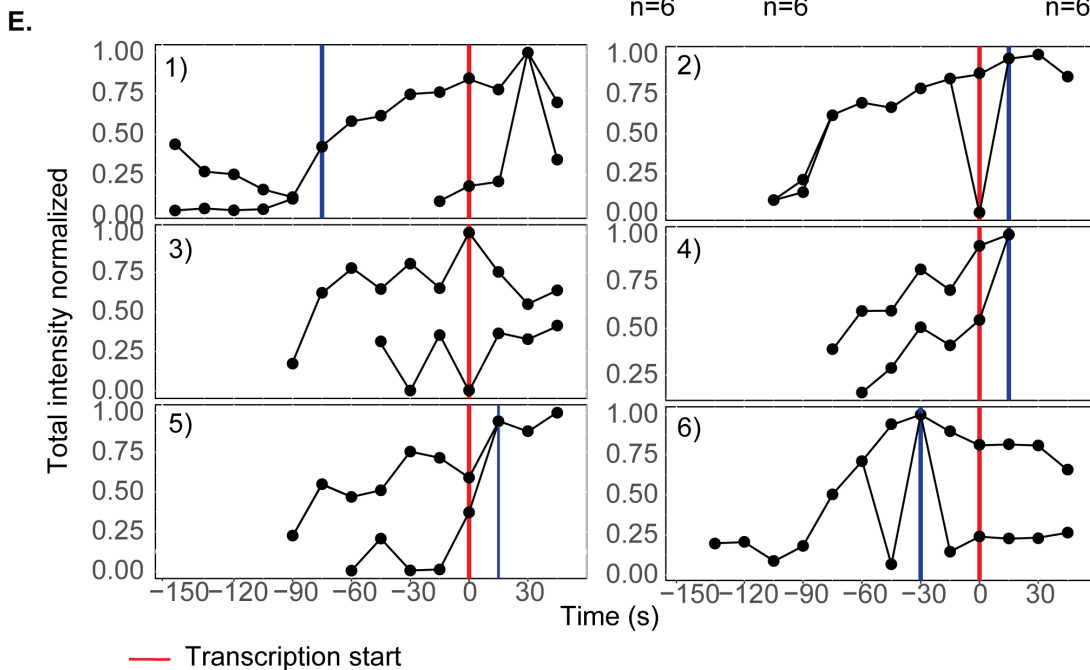
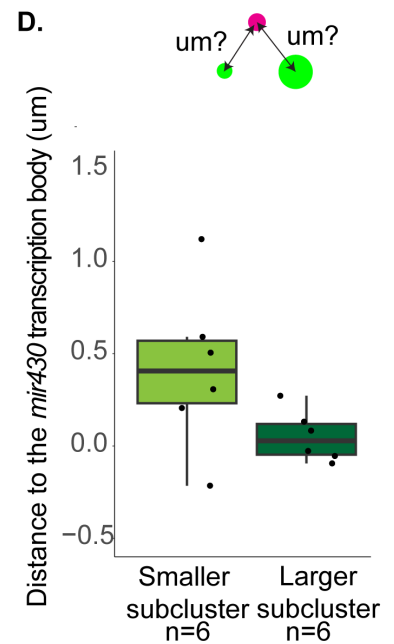
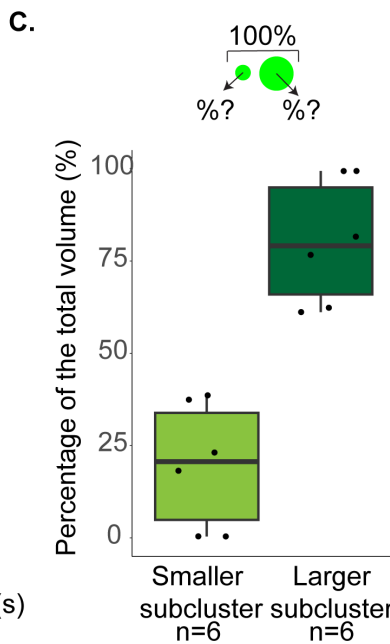
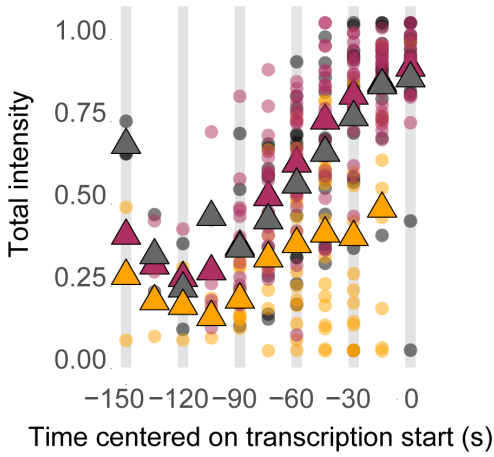
transcription body (distance close to 0 or negative), while the smaller clusters were further apart (**Figure 23D**). Taken together, my results suggest that even if the mir430 Nanog subclusters are not merged when transcription starts, there is a difference in volume between these two subclusters and the largest one is large enough to start transcription. These larger clusters are also the ones that are the closest from with the mi430 transcription body when transcription starts. These results confirm even further that there is a threshold amount of Nanog to reach at the mir430 locus to start transcription, while in that case, the merging of mir430 Nanog subclusters is not necessary.

To conclude, I have shown that the merging of mir430 Nanog subclusters leads to an increase in volume, important to have enough Nanog proteins, reach a threshold and leads to transcription. Transcription can still start if the smaller subclusters did not merge but only on the largest one.

A. What is impact of merging on *mir430* Nanog levels?



- ▲ Growing category
- ▲ Merging category: sum
- ▲ Merging category: no sum



— Transcription start

Figure 23: Merging is essential to reach high levels of Nanog.

A. Distribution of the mean intensity, volume, and total intensity of mir430 Nanog merging clusters over time (s). Time is centered on merging time. **B.** Distribution of the total intensity of mir430 Nanog clusters with only growth (grey), individual mir430 Nanog subclusters before merging (brown) or sum of both mir430 Nanog subclusters (yellow) over time (s). Time is centered on transcription start. **C.** Percentage representing the total volume for both clusters when summing up both volumes at the moment of transcription start. **Light green:** smaller cluster; **Dark green:** larger cluster. **D.** Distance to the transcription body for the smallest or the largest cluster at the moment of transcription start. **Light green:** smaller cluster; **Dark green:** larger cluster. **E.** Proportion of the total intensity of individual mir430 Nanog subclusters for which transcription does not start when clusters are merged over time (s). Time is relative to transcription start. **Red line** indicates transcription start. **Blue line** indicates merging time. For calculating the number, the following numbers of nuclei/embryos were used: A: WT data: 1k-cell stage: N1=5, N2=6, n=27. B: N1=3, N2=4, n=12. C, D: N1=4, N2=6, n=6, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed. Mean intensity, volume and total intensity are normalized for the maximum in time of each Nanog cluster.

3.3. MIR430 NANOG SUBCLUSTERS ARE BOTH SEEDED BY THE *MIR430* LOCUS

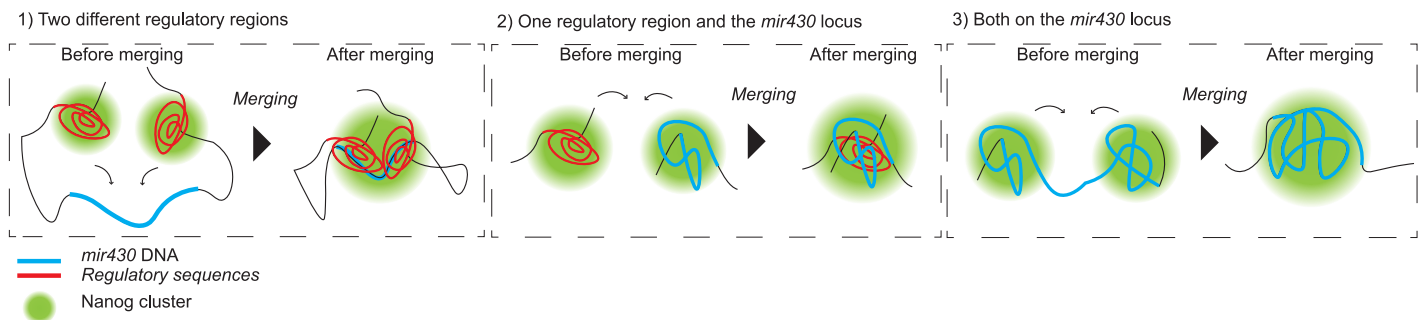


Figure 24: Schematics representing different hypotheses about mir430 Nanog cluster seeding regions

Upon mir430 transcription start, only one *mir430* Nanog clusters is detected in 88% of the cases, suggested a role of merging in *mir430* transcription activation (**Figure 22**). My results suggest that merging is important to reach enough Nanog proteins locally and correlate with transcription start (**Figure 23**). A recent study using a DBD-deletion form of Nanog showed that Nanog does not form clusters in the nucleus, suggesting that the Nanog clusters are DNA-seeded ²⁹⁹. To understand better what is the role of merging and how this process regulates transcription, we need to understand what the nature of the DNA underlying the Nanog clusters is.

As I have shown previously, two *mir430* Nanog clusters colocalize with the *mir430* transcription. Based on these observations, several hypotheses can be made about the underlying sequences of the *mir430* Nanog subclusters (**Figure 24**):

- 1) Both the *mir430* Nanog subclusters are unknown DNA regions, potentially enhancers, and get in close contact with the *mir430* locus before activation.
- 2) One of the the *mir430* Nanog subcluster is an unknow DNA sequence and the other one the *mir430* locus, getting in close contact prior to *mir430* activation.
- 3) Both the *mir430* Nanog subclusters are seeded by the *mir430* locus.

Adding these different hypotheses, a recent study revealed that TF clustering could concentrate regulatory regions from different chromosomes²²⁹. The transcribing Nanog subclusters could also be seeded by other DNA regions on different chromosomes. In this section, I will test these different hypothesizes using live imaging.

3.3.1 - NANOG MERGING CLUSTERS ARE SEEDED BY DNA REGIONS LOCATED ON THE SAME CHROMOSOME

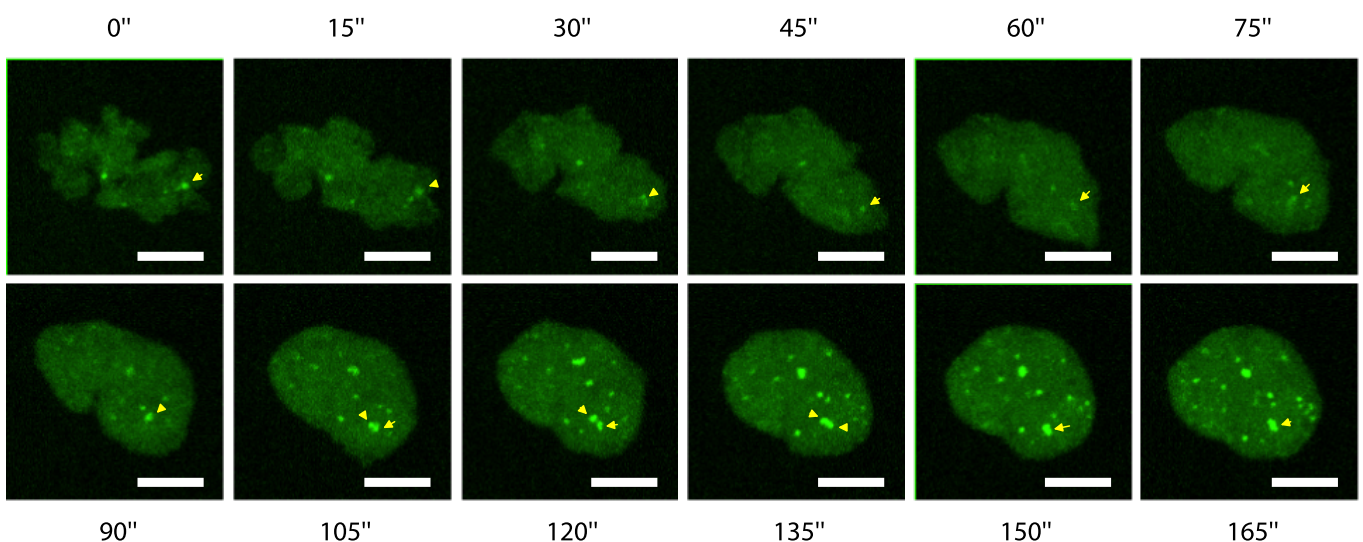


Figure 25: Nanog merging clusters are seeded by DNA sequences on the same chromosome.

Snapshots of **Nanog-mNG clusters**, as well as nuclei, over time (s) during the disappearance of Nanog signal at the mitosis-interphase transition. Scale bar indicates 5 μm . The same mir430 Nanog cluster over time is indicated with yellow arrows.

First, to test if the merging Nanog clusters are seeding by DNA regions on the same chromosome, I was first interested in observing what is the dynamics of the mir430 Nanog clusters from the end of mitosis until transcription initiation. To do so, I imaged again Nanog-mNG and mir430 transcription body labeled with the Fab recognizing the mir430 5'P transcription body during the transition from mitosis to interphase. Nanog clusters are also found on mitotic chromosome as previously shown (**Figure 12C**,

Figure 25 at $t=0''$). The mir430 Nanog at the mitotic form unique and large cluster. The mir430 Nanog clusters are splitting from a larger cluster into the mir430 Nanog subclusters (

Figure 25, $t=105''$ and $t=120''$) that merge again (

Figure 25, from $t=135''$ until $t=150''$) before transcription starts. This observation suggests that the merging Nanog clusters could be seeded by DNA sequences physically linked, e.g., on the same chromosome. Moreover, those sequences should be close to each other on the chromosome as the two smaller clusters splitting from the larger cluster stay in proximity.

Interestingly, I can observe that the first Nanog clusters to form after mitosis are the mir430 Nanog clusters (

Figure 25, $t=75''$), followed later by the appearance of non-mir430 Nanog clusters (

Figure 25, $t=90''$), suggesting that Nanog binds first at the mir430 locus after mitosis followed by other DNA regions.

Taken together, my results suggest that the merging Nanog clusters could be seeded by DNA regions on the same chromosome.

3.3.2 - BOTH MERGING NANOG CLUSTERS COLOCALIZE WITH MIR430 MICRORNA

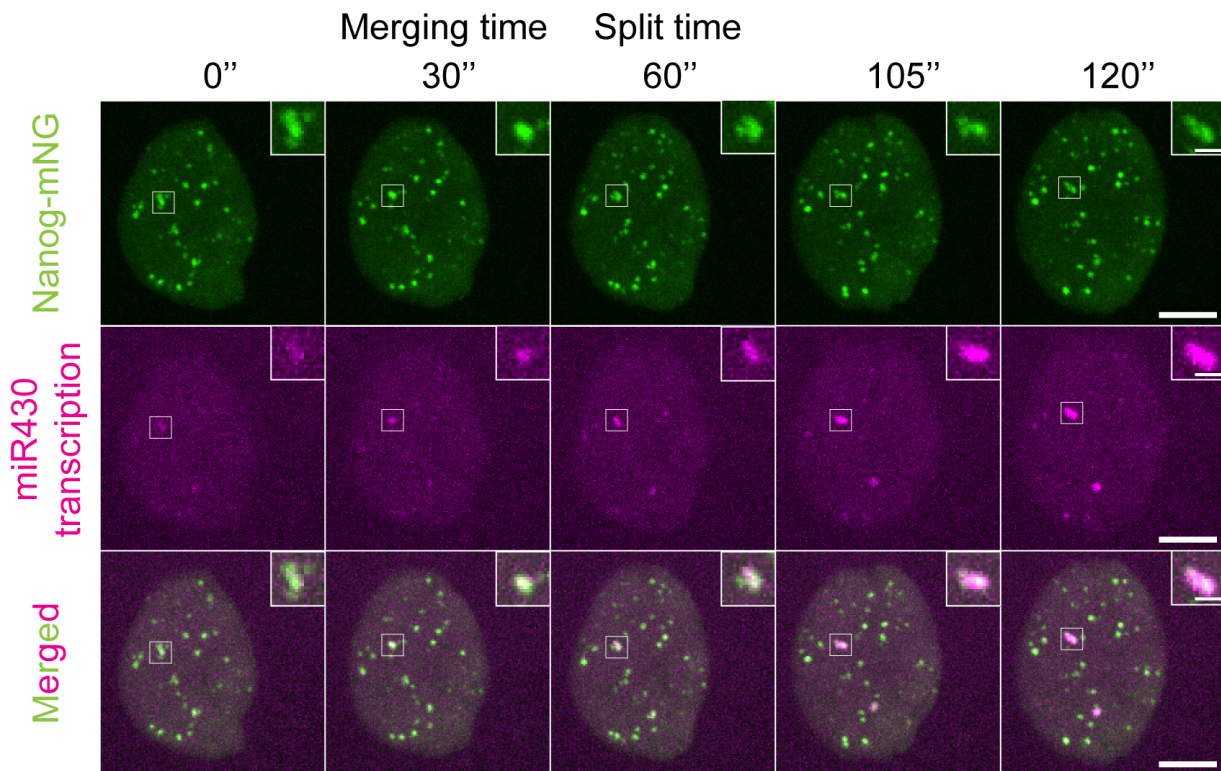


Figure 26: Both Nanog merging clusters are positive for miR430 transcription.

Snapshots of *Nanog*-NG clusters, *miR430* transcription (MoVIE-lissamine) and merge over time (s). Merging time represents the moment when smaller *Nanog* clusters merged. Split time represents the moment when the transcribing *Nanog* cluster split up again. Bigger scale bar = 5 μm , smaller scale bar = 1 μm .

To investigate where is the *mir430* locus in comparison to the *mir430* *Nanog* subclusters, I next look at the position of *miR430* native transcripts in comparison to the *mir430* *Nanog* subclusters. I have shown that in majority of cases (88%), transcription start on merged *Nanog* clusters (**Figure 22A**). However, in 12% of the cases, transcription can also start when the *mir430* *Nanog* subclusters are not merged, with the largest cluster of the pair being on average the closest to the *miR430* transcription body (**Figure 23D**). However, in this experiment, I labeled the RNA Pol

II S5P and not miR430 transcripts. I will test now using the miR430 transcripts how the two mir430 Nanog subclusters are positioned compared to mir430 transcription.

To assess whether both, one or none of the mir430 Nanog subclusters were positive for *mir430* transcription, I used an antisense oligonucleotide (miR430 MoVIE) to label for the presence of miR430 nascent transcripts ²⁹⁸. I observed that both the mir430 Nanog subclusters could be positive for *mir430* transcription prior to merging (

Figure 26, t=0”). Interestingly, when the mir430 Nanog merged split again into smaller clusters after transcription initiation (t=60 s), I observed that both were still positive for *mir430* native transcripts. The results suggest that both the mir430 Nanog subclusters are seeded by the *mir430* locus.

As I am labelling *mir430* nascent transcripts, I expect that the underlying transcribed mir430 locus is also located at the same place or in close proximity. This result that the unmerged Nanog clusters could be both seeded by the *mir430* locus. However, I cannot exclude that the transcripts are colocalizing on both unmerged Nanog clusters without being produced there.

3.3.3 - BOTH MERGING NANOG CLUSTERS ARE SEEDED BY THE *MIR430* LOCUS

This chapter and all the next ones where the mir430 DNA labelling is involved have been in collaboration with Ramya Purkanti. She developed the mir430 locus live-labelling. Using this technology, I imaged Nanog, mir430 DNA and RNA at the same time to study how both Nanog clusters and the mir430 locus colocalize.

To check where the *mir430* locus localizes with respect to the mir430 Nanog subclusters, I need to be able to visualize the *mir430* locus live. Ramya Purkanti from my lab developed a tool to visualize the *mir430* locus live (Figure Supplemental 8). Briefly, the dCas9 protein recognizes the *mir430* locus in a sequence-specific manner through two sgRNAs (**Figure Supplemental 8A, B**). The sgRNAs are designed such that it contains eight stem loops (**Figure Supplemental 8A**), which are recognized and bound by MCP-mNG proteins (**Figure Supplemental 8A, C**), permitting the detection of the *mir430* locus by tracking MCP-mNG (**Figure Supplemental 8C**).

We have shown that using this visualization tool, mir430 labelling is specific, as only two MCP-mNG signal clusters are observed in the nucleus during the cell cycle, colocalizing with miR430 native transcripts (**Figure Supplemental 9A, B**). We could reproduce mir430 DNA expansion, following the transcription body growth and miR430 RNA accumulation (**Figure Supplemental 9**), as previously published³⁰¹. Finally, transcription-inhibition using alpha-amanitin demonstrated that the expansion of the mir430 locus was not occurring in absence of miR430 transcripts, further confirming the specificity of mir430 labelling using this visualization tool (**Figure Supplemental 9C, D**).

To visualize where are localized the mir430 Nanog subclusters, the mir430 DNA and RNA in the nucleus prior to mir430 activation, I imaged Nanog using HaloTag conjugated with the JFX650 dye, MCP-mNG to visualize *mir430* DNA and MoVIE-lissamine to visualize miR430 RNA. I then imaged at the time interval of 15 seconds prior to mir430 transcription initiation. I inquired first if both of the merging clusters were seeded by the mir430 locus. Visual inspection of the images revealed that when the mir430 Nanog subclusters appeared, both colocalize with the MCP-mNG signal (**Figure 27A, t=0''**), suggesting that both mir430 Nanog subclusters can be seeded. To check for this, I quantified the number of cases where either none, one or the mir430 Nanog subclusters are colocalizing with *mir430* DNA signal prior to merging (**Figure 27B**). In 86% of the cases, both the mir430 Nanog subclusters colocalized with *mir430* DNA, strongly suggested that the mir430 Nanog subclusters are seeded by the *mir430* locus. However, there are still 13.7% of the cases where only one mir430 Nanog subcluster, or none colocalizes with the *mir430* locus (**Figure 27B**). To explain this result, I can draw two hypotheses: 1) the *mir430* locus is not always seeding both the mir430 Nanog subclusters, or 2) the absence of colocalization could be a technical reason due to the imaging experiment. Indeed, I observed previously that the Nanog signal is weakening during the mitosis to interphase transition (**Figure 17A,**

Figure 25).

To be certain that some mir430 Nanog subcluster are not colocalizing with *mir430* DNA because of technical issues, I wanted to check how often the *mir430* locus and mir430 Nanog clusters colocalize by at least one pixel over time. In the case of merging clusters, I consider the merging clusters as belonging to the same mir430 merged Nanog cluster. Plotting the percentage of Nanog clusters colocalizing with the

mir430 DNA, I observed that 150 seconds before activation, 100% of Nanog clusters overlap with the *mir430* locus but this percentage decreases to 75% in the next 15 seconds, before gradually increasing again to 100% at 60 seconds prior to activation (**Figure 27C**). This result suggests that some *mir430* Nanog clusters and the *mir430* DNA are not overlapping at all during 5 consecutive time points, that is 75 seconds. Next, I looked at the percentage of overlap between the Nanog cluster and the *mir430* DNA signals. I observed that they show a similar trend: percentage of colocalization between Nanog and *mir430* DNA decreased (-150s to -135s) before sharing a larger percentage of overlap again in the next time points (**Figure 27D**). These quantifications suggest that there is a decrease in colocalization between *mir430* DNA signal and *mir430* Nanog clusters. However, this decrease is related to a specific time window. The complete absence of any overlap suggests that the signal for *mir430* DNA or *mir430* Nanog clusters is absent during this specific time window. This sudden decrease in colocalization between both signals could explain why I observe for some cases none or only one of the merging Nanog clusters to colocalize with the *mir430* locus.

All together, these results suggest that both merging Nanog clusters are seeded by the *mir430* locus itself. However, some discrepancies are observed for some early time points, such as a decrease in the percentage of overlapping merging clusters with *mir430*DNA.

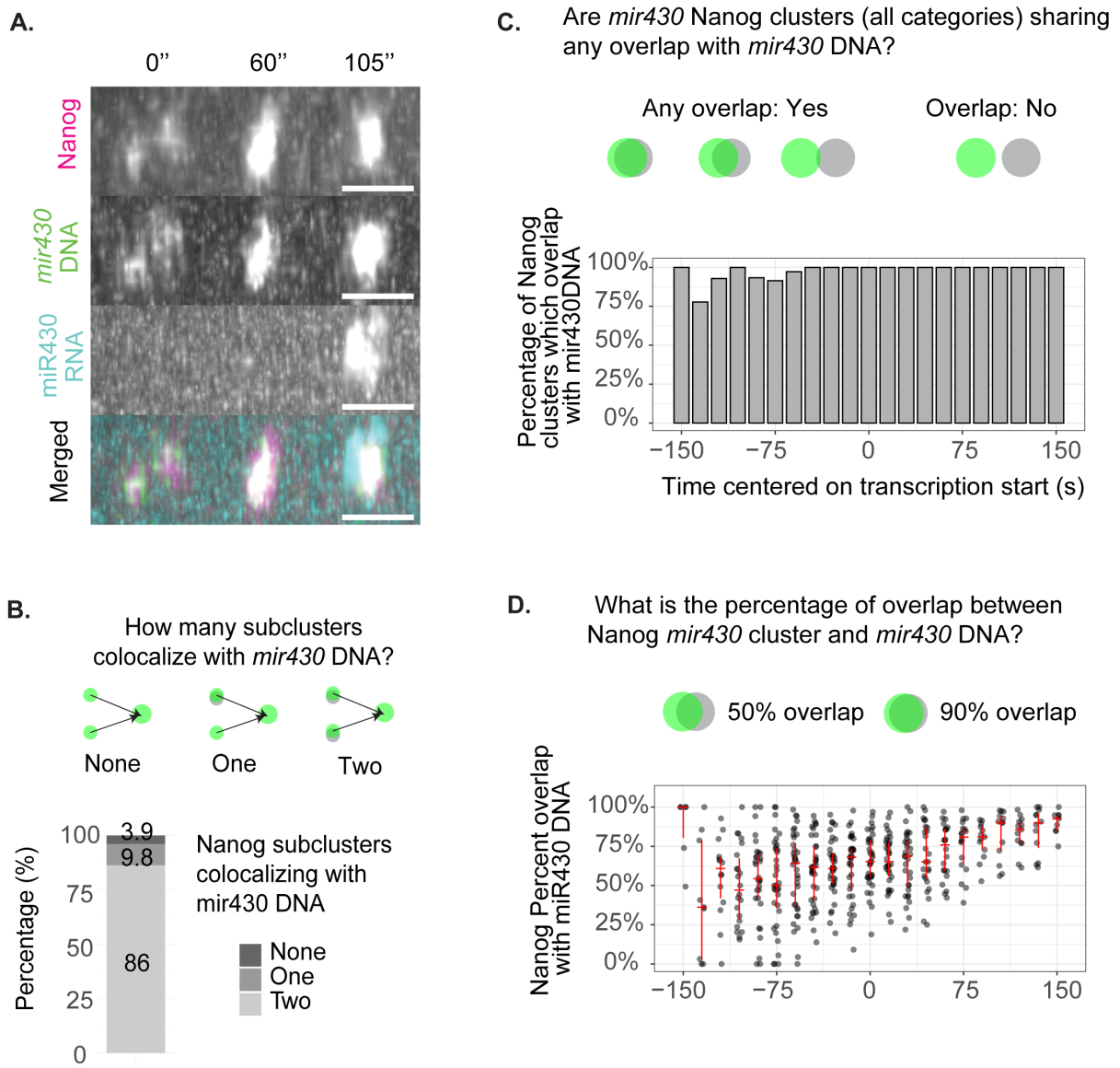


Figure 27: Both Nanog merging clusters are seeded by the *mir430* locus DNA.

A. Snapshots showing **Nanog** (HaloTag-JFX650), ***mir430* DNA** (MCP-mNG) **miR430 RNA** (MoVIE lissamine) and merged. Three time points are shown: before merging (0”), at the time of merging (60”) and transcription start (105”). Scale bars represent 1 μ m. **B.** Percentage of occurrences where none, one or both Nanog merging clusters are colocalizing with *mir430* DNA signal. **C.** Distribution of the percentage of Nanog signal that overlap with *mir430* DNA at each time point (s). Time is centered on transcription start. **D.** Distribution of the percentage of Nanog clusters colocalizing with *mir430* DNA over time (s). The horizontal **red** bar represents the mean, the top and bottom of vertical **red** bars represent respectively 25 and 75% of the distribution. Time is centered on transcription start. $N_1=4$, $N_2=7$, $n=37$, $a=51$, where N_1 the number of biological replicates, N_2 the number of embryos, n the number of nuclei analyzed and a the number of alleles.

Pannels B and C were plotted by Ramya Purkanti.

3. 4. THE *MIR430* LOCUS COMPACTS AS THE *MIR430* NANOG CLUSTERS ARE MERGING

3.4.1. NANOG AND *MIR430* DNA SIGNAL DISAPPEAR UPON DNA RELAXATION DURING TELOPHASE

In the previous section, I have shown that the *mir430* Nanog subclusters colocalize with the *mir430* locus, suggesting that they are seeded by the *mir430* locus (**Figure 27A, B**). Moreover, I also showed in section 3.2 that the *mir430* Nanog subclusters merge prior to transcription (**Figure 15A, E**) to reach up high total amount of Nanog proteins close to their maximal loading capacity, reaching a threshold that correlates with transcription activation (**Figure 23A**). Based on these results, I wondered whether the *mir430* had the same dynamics before transcription initiation: the *mir430* DNA locus concentrate locally prior to transcription. For simplicity, I define here this process as the *mir430* locus compaction.

To decipher if the *mir430* locus is compacting prior to transcription, I need to quantify precisely the *mir430* locus DNA signal over time. However, my previous results suggested that the colocalization between *mir430* Nanog clusters and *mir430* DNA signals is decreasing during a few time points because one of the signals is disappearing (**Figure 27C, D**). I proposed in chapter 3.3.3 that the decreasing level of colocalization indicates a scenario in which their signal is temporarily absent before recovering. In this case, signal disappearance can have an impact on the *mir430* DNA signal and therefore on further quantification analysis. Indeed, the merging of the *mir430* Nanog subclusters and the decrease in overlap of both signals occur at the same time of the cell cycle. If the *mir430* signal is disappearing while I want to study the potential compaction of the *mir430* locus, it can have an impact on the analysis.

Therefore, taking these factors into account, I decided to the reason of the decrease in overlap of *mir430* and Nanog signals to then properly study if the *mir430* locus is compacting as the same time as the *mir430* Nanog clusters are merging.

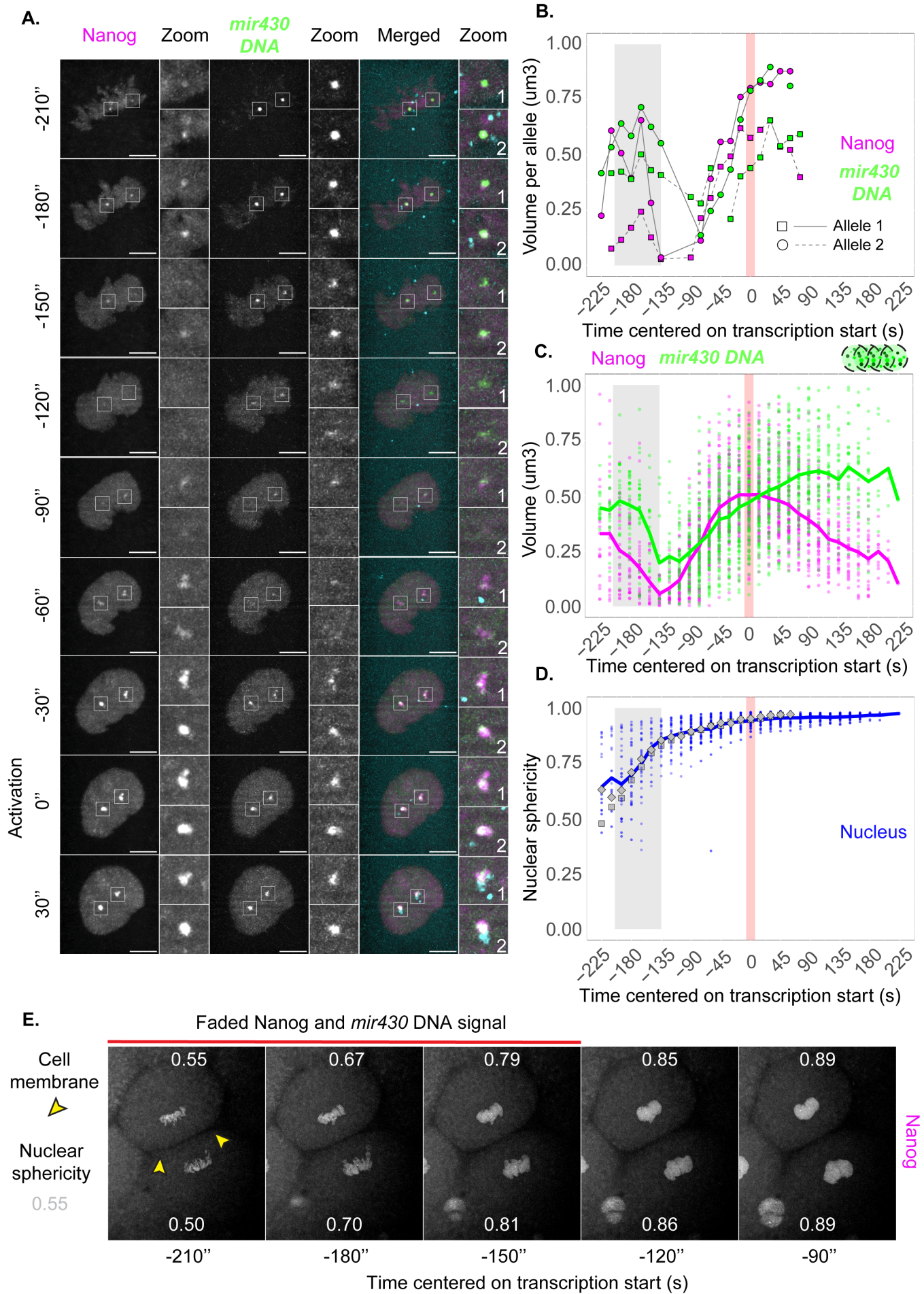


Figure 28: Nanog and mir430 DNA signal are fading off upon DNA relaxation.

A. Snapshots of **Nanog** (HaloTag-JFX650), **mir430 DNA** (MCP-mNG) and **RNA** (MoVIE-lissamine) for the nucleus and both **mir430** alleles as a function of time. Time range from -210 seconds to 30 seconds after transcription start, at 1k-cell stage. MiR430 RNA is only shown for the merged. Transcription activation is marked down at the $t=0$ seconds. Scale bar = 5 μm . **B.** Volume of **Nanog** (magenta) and **mir430 DNA** (green) segmented signal for both **mir430** alleles shown in A. Left allele is represented with plain line and rounds. Right allele is represented with dotted line and squares. Time ranged from -225 seconds until 225 seconds with time centered on transcription start at $t=0$ second (red line). Grey window indicates time period in which both volumes decrease largely. **C.** Volume of **Nanog** (magenta) and **mir430 DNA** (green) segmented signals for all mir340 Nanog clusters present in the dataset. Volume is not normalized and indicated in μm^3 . Average is shown as a bold line. Time range a, grey window and red line as explained in B. **D.** Nuclear sphericity as a function of time. **Bold** line represent the average. **Square** represents the upper nucleus displayed in E. **Diamond** represents the lower nucleus displayed in E. Time range a, grey window and red line as explained in B. **E.** Snapshots of two daughter cells and nuclei during telophase, as labeled by Nanog-HaloTag-JFX650 bound on mitotic chromosomes. Upper nucleus is the one represented in A. Time range from -210 to -90 seconds before transcription start. **Yellow arrows** represents the presence of the cell membrane between the two daughter cells. Sphericity for both nuclei are indicated as comparison to D. Nucleus 1 and nuclear 2 are indicated by the same number over time. Scale bar = 10 μm . For B, C and D. $N_1=4$, $N_2=7$, $n=45$, $a=88$, where N_1 the number of biological replicates, N_2 the number of embryos and n the number of nuclei analyzed and a the number of alleles.

To do so, I used the same dataset as in Chapter 3.3.3, where Nanog is labeled with Nanog HaloTag (JFX650), mir430 DNA with MCP-mNG and miR430 transcripts with MoVIE. First, I examined individual nuclei at the moment of the cell cycle when overlap between Nanog and **mir430** DNA is decreasing (**Figure 28A**). At the first time points (-210s to -180s), I noticed that Nanog and **mir430** DNA signals are on mitotic chromosomes and colocalize (-210s to -180s). For Nanog, the signal begins to decrease at -150s and has disappeared (or nearly) by -120s. In contrast, for **mir430** DNA, the signal is still detectable but begins to decrease at -150s and begins to disappear by -90s. Both are not distinguishable from the background fluorescence respectively at -120 s and -60 s relative to transcription start and are coming back progressively within the next seconds before transcription. These results confirm that rather than having both signals at different places at this moment of the cell cycle, they fade and disappear for a short period. These observed patterns for these two examples of the **mir430** alleles are also reflected by the quantified volume of segmented Nanog and **mir430** DNA signal (**Figure 28B**). Both are decreasing (-210s to -150s), sometimes not being segmentable anymore (at $t=-120$ s and $t=105$ s), before increasing again (from $t=-90$ s) before transcription start.

Pulling all the *mir430* alleles from the dataset and averaging the volume of Nanog and *mir430* DNA signal, I can observe on average a large decrease of segmented signal volumes (from $t = -195$ s to $t = -135$ s) before transcription start (**Figure 28C**). Both signals are reappearing progressively again (from $t = 105$ s). These results confirm that the *mir430* DNA signal as well as Nanog signal are disappearing from the DNA for a certain time, explaining the results from the previous section (**Figure 27**).

Next, I investigated the potential reason of signal disappearance during the mitosis to interphase transition. As the chromatin change dynamically at this moment to prepare the nucleus for the new cell cycle, I wondered if it was not due to such changes, especially DNA relaxation during telophase. Potentially, either both Nanog and the dCas9 could be released as DNA relaxes, or the DNA is so relaxed that the signal coming from the bound proteins is completely diluted and below detection limit. In both cases, I can approximatively determine at which moment of the cell cycle the signals disappear based on the presence of the cell membrane, position of nuclei and nuclear membrane shape and determine if it occurs at the same time as DNA relaxation. To be able to relate these different criteria cited above and the moment of the signal disappearance, I can use the nuclear sphericity as an indirect time measurement to indicate the stage of the cell cycle (**Figure 28D, E**). I noticed that Nanog and *mir430* DNA signals start to disappear when the average nuclear sphericity is 0.72 (**Figure 28E**). At this moment of the cell cycle, chromosome sets are already separated from the metaphase plate, indicating that the disappearance of both signals occurs from anaphase onwards (**Figure 28**, second image). Moreover, I can already distinguish the cell membrane separating both daughter cells (as indicated by two yellow arrows) indicating that the cytokinesis is almost completed or already over (**Figure 28E**). As cytokinesis is happening in late telophase, as well as DNA relaxation, time of signal disappearance is coherent with the DNA relaxation timing.

To further correlate disappearance of both *mir430* DNA and Nanog signal, I decided to look at the chromosome structure during signal disappearance. Nanog is a DNA-binding protein and binds DNA also during mitosis (**Figure 12C**). The signal of Nanog can thus indirectly delimitates the shape of the chromosomes during mitosis, indirectly informing about the state of DNA relaxation. Using Nanog signal, one can observe that the two sets of DNA packages from the two daughter cells resemble chromosome-like structures (-210 s, **Figure 28E**, first image). These structures become smoother and edgeless as cell cycle progress, suggesting that DNA, initially

compacted into visible chromosomes from prophase until anaphase, relaxed into a more chromatin diffuse state along with the nuclear membrane reformation. This change in shape is concomitant with the time of signal disappearance as correlated by the nuclear sphericity of Nanog clusters (**Figure 28D, E**). My observations suggest that DNA relaxation is indeed concomitant with the time of signal disappearance. Further investigations are necessary to understand what is happening at this sensitive moment of the cell cycle and functionally relates DNA relaxation to signal disappearance.

To conclude, both Nanog and *mir430* signals are becoming weaker during telophase. My data suggest that this decrease in signal coincides with DNA relaxation. As we want to study the potential compaction of the *mir430* locus prior to transcription activation, I need to take this in consideration in my image analysis approach.

3.4.2. THE *MIR430* LOCUS COMPACTS WITH A SIMILAR DYNAMICS TO MERGING NANOG CLUSTERS PRIOR TO TRANSCRIPTION

As the *mir430* DNA signal is shortly disappearing prior to transcription start before coming back slowly (**Figure 28A, C**), the intensity itself is not a reflective indicator of the *mir430* DNA compaction state over time. To study *mir430* compaction prior to transcription, I must adapt my analysis method to this technical issue. As the amount of fluorescence labelling the *mir430* locus itself is changing over time, I cannot use parameters depending on the fluorescence intensity or even on the size, as it calculated directly from the intensity values. To quantify for compaction over time, I instead decided to use a metrics independent of intensity and based on shape description. Indeed, even if the fluorescence intensity decreases, I expect the shape not to be impacted by such a decrease in fluorescence. To be able to characterize the degree compaction of the *mir430* locus based on the AR, I want to define a threshold value for which I assume that the *mir430* locus is completely compacted (as much as imaging allows me to detect, as the signal is influenced by the PSF). Moreover, I decided to carry those measurements on max projected images. Indeed, the *mir430*

DNA signal is affected by the point spread function with abnormalities stronger in the Z direction and potentially biasing the shape measurements. (**Figure 29A**).

Given the variable shape of the *mir430* DNA mask and the possibility to have two *mir430* DNA densities for one allele, I decided to use the aspect ratio which is the ratio of the major axis over the minor axis of the best fitted ellipse on the segmented mask (**Figure 29A**). In an ideal case, I expect the total compaction of the *mir430* locus as a perfect circle corresponding to an aspect ratio of 1. To confirm this hypothesis on real data, I selected all the aspect ratio values for all *mir430* DNA signal that were measured at a moment of the cell cycle prior to DNA relaxation, that I defined when the nuclear sphericity was inferior to 0.6 and found an average value for the aspect ratio of 1.3. As DNA is very compacted during mitosis, I consider the *mir430* locus to be in a strong state of compaction when reaching a value of 1.3 or below. In the contrary, more elongated is the object, longer will be the major axis, higher the aspect ratio is.

Then, I studied the compaction state of the *mir430* locus over time (**Figure 18B**). as previously, as I do not start to image at the same time during the cell cycle, I aligned the tracks at the *mir430* transcription initiation to have a normalized time scale.

First, centering on the aspect ratio values for the entire dataset on transcription initiation ($t=0$), I observed that the aspect ratio is increasing on average prior to transcription start (-240 s to -105 s, **Figure 29C**, grey line). Thus, this increase in aspect ratio could reflect DNA going from a more condensed to relaxed state. It is indeed coinciding with the moment when *Nanog* and *mir430* DNA signal are disappearing, itself coinciding with DNA relaxation timing (**Figure 29C**, grey window). The AR for the *mir430* DNA can still be calculated while the signal is fading off because it is still detectable for some *mir430* alleles in the dataset. Moreover, it is interesting to observe a such increase in AR as the same time as DNA relaxation is occurring. It suggests that the AR is reflecting the DNA relaxation independently from the intensity levels, only relying on the shape. The aspect ratio is then decreasing on average from -105 seconds until reaching a minimum 30 seconds after transcription start. I can interpret this decrease as the compaction of *mir430* DNA locus prior to transcription. Following this decrease, the AR is increasing after transcription initiation, suggesting DNA expansion. We and others have shown that upon the growth of the *mir430* transcription body, the *mir430* DNA also expands (**Figure Supplemental 9A. B**). Moreover, it also show that the AR is sensitive to the

compaction state of the *mir430* locus. Indeed, the increase in AR after transcription coincide with an increase in volume of the *mir430* locus detected signal (**Figure Supplemental 9B**). As the DNA amount itself is not changing, the AR is indirectly indicating how locally concentrated the *mir430* DNA is.

Altogether, these results suggest the *mir430* locus is compacting at the same time as the *mir430* Nanog clusters are merging.

Next, to further correlate *mir430* DNA compaction and merging of the *mir430* Nanog clusters, I was wondering if the compaction of the *mir430* locus will be different depending on the number of detected Nanog clusters prior to transcription, to imitate the category described previously: one cluster, the growing mechanism, *versus* more clusters, the merging mechanism (see **Figure 15D, E**). Individual curves show that there are different types of compaction dynamics (**Figure 29B**). First, some of allele do not show a big change in compaction, as the AR stay stable over time or only slightly decrease before the initiation of transcription. For the examples 1,2,3 and 4, the aspect ratio is quite low already since the beginning, suggesting that compaction might already have occurred or that the *mir430* locus was already compacted after mitosis (**Figure 29**). Most of those examples are actually single clusters one even if we can also observe them in the multiple-cluster category (**Figure Supplemental 7**). Some other single-cluster category has a slight gradual decrease (**Figure 29B.1**) or a drastic decrease (**Figure 29B.3**) followed by no variation until transcription start. The multiple-cluster category is however much more dynamics in term of aspect ratio even if most of the examples reach a value under the average one for mitosis, set to 1.3. (**Figure 29B. 4-6, Figure Supplemental 7**). Even if some are mostly flat, suggesting that they are already compacted (**Figure 29B.4**). Then, others show a very large increase, suggesting decompaction and then decrease of aspect ratio, suggesting compaction. Such dynamics could result from DNA relaxation (increased aspect ratio, followed by *mir430* compaction prior to transcription initiation (**Figure 29B.5**). Finally, I observed examples where the aspect ratio does not decrease at the transcription start (**Figure 29B.6**). All of these suggest that the *mir430* compaction is quite variable for each allele. However, the single-category alleles have a lower dynamic change of AR ratio compared to the multiple-category clusters, suggesting that the variation of aspect ratio could be correlated with the number of Nanog clusters. Interestingly, not only the multiple-cluster category DNA *mir430* alleles are compacting but the single-cluster category as well, even if at a lower degree. Finally,

most of the mir430 DNA alleles reach an AR value close to the average one in mitosis, suggesting that most of them have reached the maximum level of compaction that I can detect, prior to transcription activation.

Plotting the average aspect ratio for both categories over time confirm further the previous observations that there are some differences in their dynamics (**Figure 29C**). First, both show that the AR decrease at the moment of DNA relaxation, suggesting relaxation of the mir430 locus DNA after mitosis (-240 s to -105 s). However, while the single cluster alleles are compacting again rapidly to reach their average minimum value 45 seconds before transcription start, the multiple-cluster alleles do not show a decrease in aspect ratio before 60 seconds prior to transcription start and reach a minimum exactly when transcription start. These results suggest that single-cluster alleles compact faster than the several-cluster alleles (**Figure 29C**). Interestingly, the single-cluster category is reaching higher peaks of AR than the multiple-cluster category, suggesting that the single-cluster mir430 DNA allele is more relaxed before transcription initiation. However, looking at the individual examples of at each mir430 DNA allele, the single-cluster category does not seem to be more relaxed before transcription initiation. To ensure that this peak is not due to individual high AR values at a certain time point and further confirm that the multi-cluster category compact more than the single-cluster category, I compared the maximum AR value reached before transcription initiation between both single-cluster and multiple-cluster category (**Figure 29**). I can observe that the multiple-cluster category has on average much higher maximum aspect ratio values than the single-cluster category ($p=0.00029$, Student t-test). This result is in contradiction with the average AR curve measured for the single-category and suggest that the peak we observe at $t=-135$ s might be due to individual higher values that push the average up. This also suggest that the single-cluster category is less expanded after DNA relation compared to the multiple-cluster category, as the maximum AR reached before transcription initiation is lower than in the multiple-cluster category. Potentially, the single-cluster category mir430 DNA does not need to compact as much as for the multiple-category one because the mir430 locus is less expanded in the 3D space and already concentrated. To test of the mir430 locus is compacting less, I measured how much the mir430 DNA alleles have compacted prior to transcription, by calculating the ratio of the maximum AR before transcription to the one at transcription start and compared both categories. This measurement gives an indirect measure of how much the mir430 locus has

compacted prior to transcription, defined as the degree of compaction. In the case of single-cluster alleles, the degree of compaction was lower than for the multiple-cluster category, suggesting that the single-cluster mir430 DNA alleles are compacting less ($p=0.00047$, Student t-test, **Figure 29E**).

To conclude, my findings show that the *mir430* locus is compacting, at the same time as Nanog clusters are merging before transcription initiation. Interestingly, the previous categorization of mir430 Nanog clusters into one or multiple-clusters is correlating with the degree of mir430 DNA compaction prior to transcription: single-cluster category shows a weak degree of compaction, while multiple-cluster category are compacting more. My results also suggest that the single-cluster alleles compact faster than multiple-cluster allele, potentially because they relax less after DNA relaxation. This measurement gives an indirect read-out of how much the mir430 locus has compacted prior to transcription, defined as the degree of compaction. However, to confirm this statement, I would need to be able to label the mir430 locus all the time and check is the volume occupied by the mir430 locus after DNA relaxation. Another possibility is to find by which process the mir430 compacts and inhibit the process.

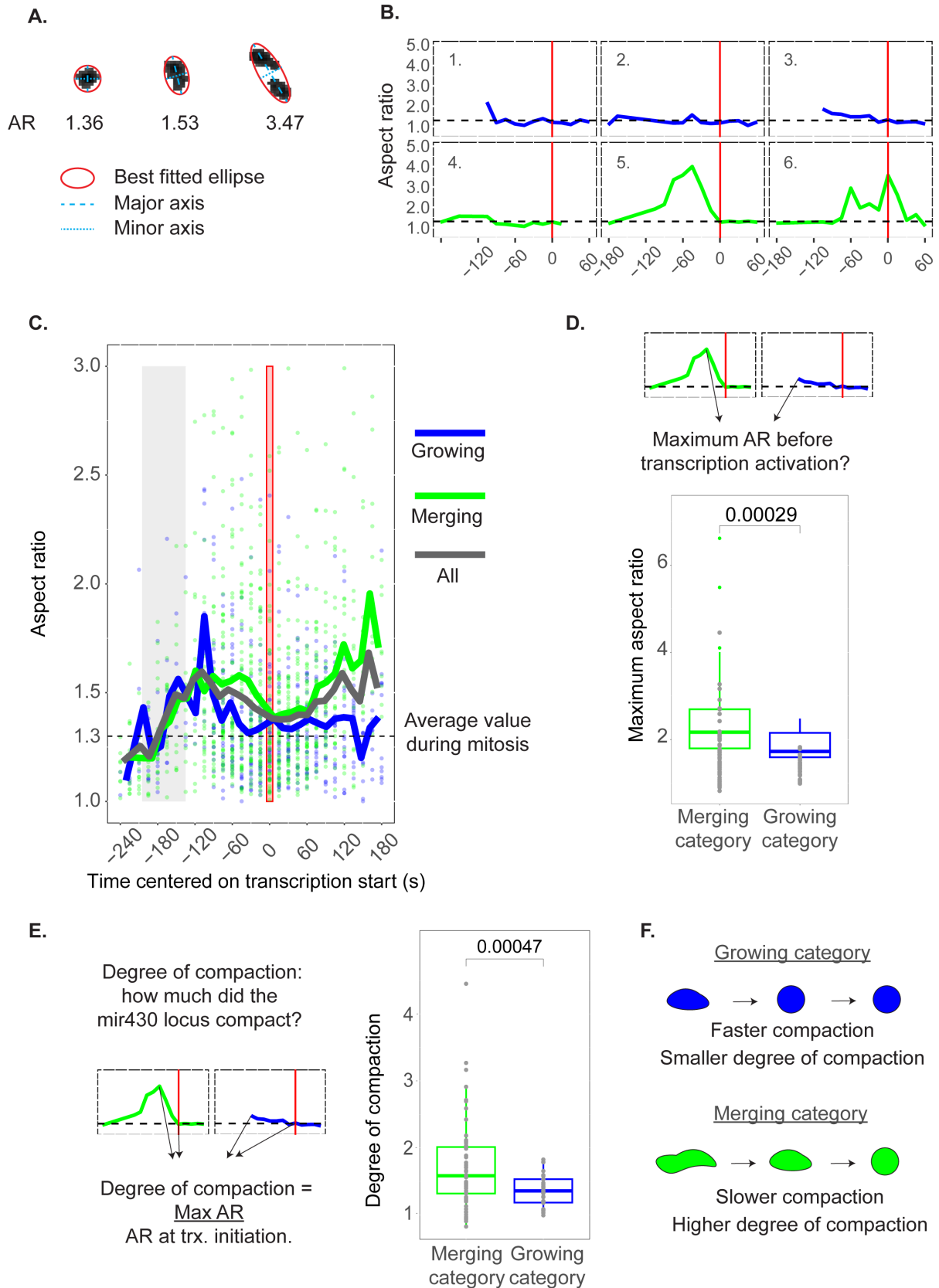


Figure 29: Analysis of the mir430 locus compaction before transcription initiation.

A. Image analysis explanation to segment and characterize *mir430* DNA compaction. Three real 2D segmented shapes and their associated aspect ratio values are given. Red ellipse indicates best fitted ellipse to the segmented mask, dotted line the major axis and plain line the minor axis of the same best fitted ellipse. Aspect ratio is calculated as the major axis over the minor axis. **B.** Evolution of the aspect ratio over time for individual *mir430* alleles. 1-3: **Single cluster category**. 4-6: **Multiple-cluster category**. ‘One cluster’ indicates that only one Nanog cluster was always detected prior to transcription start. ‘More clusters’ indicates that at least two clusters were detected for at least one time point prior to transcription start. Time is relative to transcription start and range from -180 to 60 seconds. Full dataset in **Figure Supplemental 7**. **C.** Aspect ratio of *mir430* DNA mask as a function of time. Complete dataset is represented here as dots. **Grey bold** line represents average. **One cluster** (blue) or **More clusters** (green) Time is relative to transcription start at t=0 and indicated from -240 to 180 seconds. **Red** line represent the transcription start. Grey window represents the moment when Nanog and *mir430* DNA are decreasing drastically on average (see **Figure Supplemental 7**). **Dotted line** represents average aspect ratio for *mir430* segmented mask for which nuclear sphericity is inferior to 0.6. **D.** Maximum aspect ratio for each track for both categories as explained in C. **One cluster** (blue) or **more clusters** (green). Student t-test. P-value=0.00029. **E.** Aspect ratio change for individual track for both categories as explained in C. **One cluster** (blue) or **more clusters** (green). Aspect ratio change is calculated as the maximum value for aspect ratio in a given track to the aspect ratio value at transcription start and represent an indirect measurement for the level of compaction of one given *mir430* allele. Student t-test. P-value=0.00047. For C, D, E and F : N1=4, N2=7, n=45, a=88, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed *and* a the number of alleles. **F.** Schematics representing the *mir430* locus compaction as a function of the mechanism of *mir430* Nanog cluster assembly.

3.4.3. THE AMOUNT OF *MIR430* DNA CORRELATES WITH NANOG LEVELS

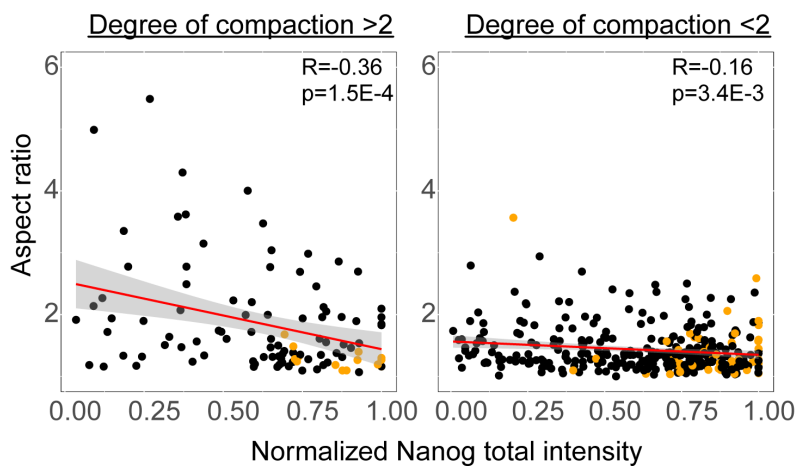
In the previous section, I showed that the single-category cluster are on average compacting less than the multiple-cluster category, suggesting that both Nanog clusters and *mir430* locus follow the same trend (**Figure 29C, F**). I also previously showed that merging of smaller Nanog clusters is important for bringing enough Nanog locally to start transcription (**Figure 23**). Putting these findings together, I next wondered whether the *mir430* locus, by a compaction process, could enriched *mir430* bound Nanog proteins locally. To test this hypothesis, I split the *mir430* DNA alleles in two categories based on aspect ratio: 1) low degree of compaction (degree of compaction inferior to 2) and 2) high level of compaction (degree of compaction superior to 2). The degree of compaction indicates there that the AR decreased more

than two-fold before transcription initiation. Then, following this categorization, I checked how much the aspect ratio is correlated with the normalized Nanog total intensity before transcription (**Figure 30A**). As I have described previously, the total intensity represents an indirect measurement for the total amount of Nanog proteins in a cluster. In the case of the *mir430* Nanog subclusters, I have shown that merging increases the volume rather than concentrating the proteins. If the *mir430* locus compaction increase the total amount of Nanog proteins locally, then the total intensity will be correlated with the degree of compaction.

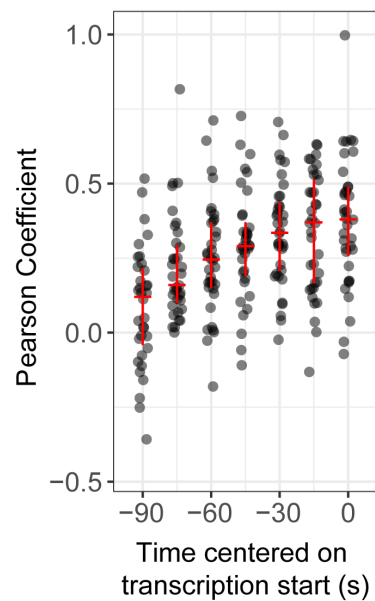
Strikingly, *mir430* alleles for which the degree of compaction is superior to 2 have a stronger correlation to Nanog total intensity compared to *mir430* alleles with lower degree of compaction ($R=-0.36$, $p=1.5E-4$ versus $R=-0.16$, $p=3.4E-3$). This result suggests that, as the *mir430* locus compacts and the AR value decrease, the amount of Nanog proteins increase locally. For the cases where the degree of compaction is low, the AR does not vary so much and is correlated less with the growing amount of Nanog at the *mir430* locus. The degree of compaction is correlated to Nanog total intensity.

The degree of compaction is correlated with higher Nanog levels, suggesting that compaction is concentrating Nanog locally. As DNA compaction also concentrates the amount of DNA locally, I wondered if the amount of Nanog is correlated with the DNA concentration. To test this, I calculated the Pearson Correlation score between the Nanog and *mir430* DNA intensities prior to transcription initiation (**Figure 30B**). I observed that the correlation is in average of 0.4 at transcription start, suggesting that where more DNA is present, more Nanog proteins are also present.

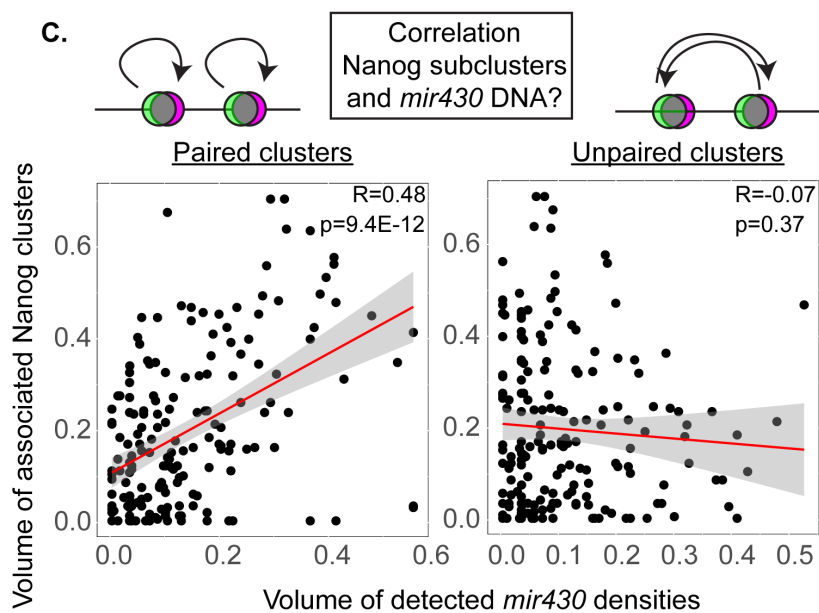
A.



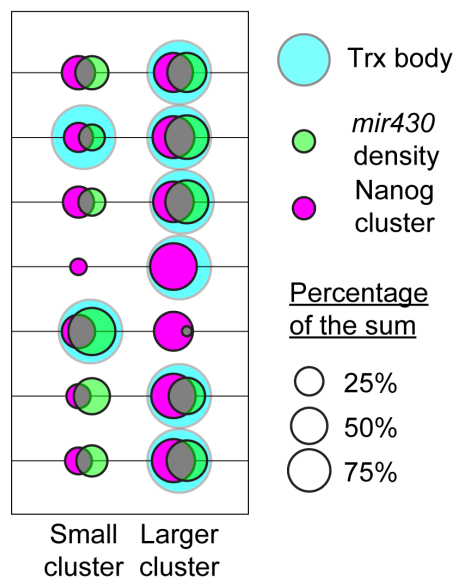
B.



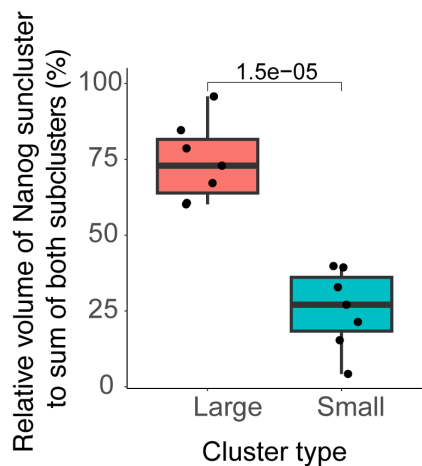
C.



D.



E.



F.

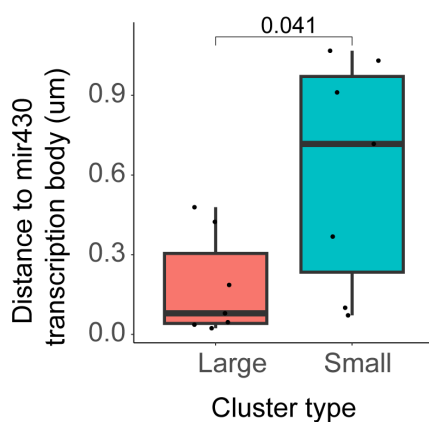


Figure 30: The amount of mir430 DNA correlates with Nanog levels.

A. Correlation between the degree of compaction and the total Nanog intensity depends on the initial level of maximum aspect ratio prior to transcription start (-105s to 0s). Left graph is for aspect ratio superior to 2 and right graph is for aspect ratio lower to 2. **Orange** dots indicates shapes at transcription starts. Nanog signal is normalized to the max of the track. Pearson's correlation and Pearson's correlation ranking test. **B.** Pearson coefficient of Nanog and mir430 DNA intensities prior to transcription. Time is centered on transcription start. **Red** horizontal before indicate the median, top of vertical bar 75% of the distribution and bottom 25% **C.** Correlation between the volume of colocalizing Nanog clusters and MCP densities during average compaction time (105 seconds to 0 seconds) in the case of colocalizing clusters (left) compared to non-colocalizing clusters (right). Pearson's correlation and Pearson's correlation ranking test. **D.** Schematics displaying the two Nanog or mir430 densities from the same mir430 allele at transcription start for aspect ratio superior to 1.7. Nanog spot (**magenta**), mir430 DNA densities (**green**) or miR430 transcription body (**cyan**). Size is proportional to the percentage of the sum of all the two spots composing Nanog or mir430 signal. The position of the transcription body indicates if it is closer from the smallest or the biggest cluster. Smallest cluster on the left and larger cluster on the right. **E.** Volume of both Nanog clusters on the same mir430 locus compared to the sum of both clusters (in %). Student t-test. **F.** Shortest distance of the two Nanog clusters to the mir430 transcription body at transcription start indicated in um. Student t-test. For A : $N1=4, N2=7, n=45, a=88$. B: $N1=3, N2=5, n=21, a=42$. C: $N1=4, N2=7, n=,26 a=34$. D,E,F: $N1=3, N2=4, n=7, a=7$, where $N1$ the number of biological replicates, $N2$ the number of embryos and n the number of nuclei analyzed and a the number of alleles. Nanog total intensity is normalized for the maximum in time of each Nanog cluster.

I have previously shown that the mir430 locus DNA is compacting prior to transcription initiation, reaching the most compacted state around transcription start (**Figure 29A**). However, I have also shown that some of the mir430 DNA alleles could start transcription in a more relax state as well (**Figure 29B.6**, **Figure Supplemental 7**). Moreover, I have proposed previously that high levels of Nanog correlate with transcription activation (**Figure 19A**), either visualized as one large mir430 Nanog cluster or two mir430 Nanog subclusters with unequal sizes, the largest recruiting transcription first. (**Figure 23C**, **E**). Based on these previous results, I wondered if transcription initiation for mir430 DNA alleles for which AR was high at transcription initiation, meaning the alleles that did not show a high level of compaction at the moment of transcription start, is not due to the presence of two mir430 Nanog subclusters, for which one is larger than the other one and potentially recruit transcription. To test this, I decided to segment the mir430 locus with the same algorithm used for the Nanog clusters. This allows me to detect the mir430 DNA as several densities where local maximum peaks are detected.

First, I selected the *mir430* alleles for which the *mir430* Nanog clusters was forming by merging, meaning that I can detect two or more *mir430* Nanog subclusters. I correlated the volume of the individual merging *mir430* Nanog subclusters to the detected *mir430* DNA densities, whatever the degree of compaction of the *mir430* alleles. Interestingly, the *mir430* Nanog subclusters that were associated with one of two *mir430* detected DNA densities, were positively correlated for their fluorescence ($R=0.48$, $p=9.4E-11$, the same pair of Nanog subcluster and *mir430* DNA density). As a control, I also correlated the fluorescence between the *mir430* Nanog subclusters and the *mir430* DNA densities that were not colocalizing, but still belonging to the same *mir430* alleles. The correlation was null and not significant ($R=-0.06$, $p=0.3687$). (**Figure 30C**). These results suggest that even when the *mir430* DNA is seeding two Nanog subclusters, the amount of DNA colocalizing with each Nanog cluster correlate strongly with their volume.

To relate these findings to the case when some *mir430* alleles are not compacting upon transcription start, I selected the *mir430* alleles for which the aspect ratio was greater than 1.7 at transcription start and check if they are, 1) seeding two *mir430* Nanog subclusters at the moment of transcription activation and, 2) if the two Nanog *mir430* subclusters are of different sizes. Only 12 *mir430* DNA alleles were showing this behavior, 7 of them seeding two Nanog subclusters on the *mir430* DNA locus at the moment of transcription initiation (**Figure 30D**). I decided to segment the *mir430* signal the same way I do for Nanog by detecting spots. Interestingly, as described previously, when transcription starts at unmerged *mir430* Nanog clusters, one of them is larger and the *mir430* transcription body is detected closer to the larger subcluster rather than the smaller cluster (**Figure 23D**). Interestingly, the largest Nanog cluster was seeded by the largest *mir430* DNA density as well, conforming further that the amount of *mir430* DNA and Nanog is correlated (**Figure 30D**). As previously observed, the largest Nanog subcluster is the closest from the *mir430* transcription body at transcription start (**Figure 30D, F**).

All together, these results suggest that the *mir430* DNA amount is correlated with underlying Nanog levels. These results suggest that not only the level of Nanog is important locally but the level of *mir430* DNA as well. As Nanog clusters have a maximum binding capacity on the *mir430* locus (**Figure 20**), the only way to increase even further the Nanog concentration locally is to the compaction of the *mir430* locus.

3.5. NANOG IS NOT RESPONSIBLE FOR *MIR430* COMPACTION BUT HELP TO KEEP THE LOCUS TOGETHER

This part was conducted in collaboration with Alessia del Panta Ridolfi (PhD student and physicist from Dr. Alma Dalco's lab (2022) and current in the group of Dr. Sara Mitri) and Ramya Purkanti (post-doc, member of the Vastenhouw lab). Ramya helped in preparing and injecting the embryos before. I did the imaging and most of image analysis. Alessia did the modelling and data analysis.

3.5.1 – SETTING UP THE METHOD TO STUDY THE ROLE OF NANOG IN *MIR430* DNA COMPACTION

Studies suggest that TFs and co-activators can concentrate DNA locally, by bring enhancers or genes together into clusters^{201,225,231,329} and even locally reorganizing the genome²²⁹. This ability of TFs and co-activators to form local enrichment of DNA is thought to be mediated especially by their IDRs^{201,231,329}. In the previous sections, I have showed that the *mir430* locus is compacting at the same time as Nanog clusters are merging (**Figure 22B**, **Figure 29C**) and that the volume of Nanog clusters is correlated with the volume of *mir430* DNA densities (**Figure 30B**, **C**), suggesting that the Nanog levels in a cluster correlates with the amount of DNA. These results led us to wonder whether Nanog itself is responsible for the *mir430* locus compaction. Indeed, it could be interesting to think that for *mir430* activation, Nanog is compacting the locus itself to increase its own local concentration and then leading to activation.

To answer this question, we decided to study the *mir430* locus dynamics using live imaging in presence and absence of Nanog. In a first condition, we used embryos knock-down for the *nanog* gene (*nanog*^{-/-}), in which we injected what all the components to label the *mir430* DNA locus (dCas9, sgRNAs recognizing the *mir430* locus and MCP-mNG, Figure Supplemental 8) and miR430 transcripts (MoVIE-lissamine) (**Figure 26A**). In the *nanog*^{-/-} embryo, the *mir430* locus will not produce any transcripts as Nanog is a TF essential for its activation^{269,270}. We still injected MoVIE to label miR430 transcripts to ensure the non-transcription state of the

mir430 locus. Using these experiment conditions, we can study the *mir430* locus dynamics over time and how it is influenced by the absence of Nanog. In a second condition, we again used the *nanog* *-/-* fish to inject the necessary components to visualize *mir430* DNA and miR430 transcripts, but adding an unlabeled version of Nanog, as a mRNA and the transcription inhibitor alpha-amanitin (**Figure 31A**). In this second experimental condition, we seek to study the dynamics of the *mir430* locus in the presence of Nanog, as a control for the first depletion condition. However, in this second experimental condition, as the *nanog* knock-out is rescued in the embryo by injected *nanog* mRNA, the *mir430* locus is producing transcripts. We have shown previously in the lab that *mir430* transcription is the seed for the *mir430* transcription body ²⁷⁰. But the growth of the transcription body is leading to the expansion of the *mir430* locus ³⁰¹ (Figure Supplemental 9), changing by this way the dynamics of the *mir430* locus due the transcription body growth, that is not occurring in the first experimental condition. Thus, we decided to inject in the second condition the inhibitor of transcription elongation alpha-amanitin. This way no transcripts are produced at the *mir430* locus and the *mir430* DNA dynamics can be compared between both experimental conditions.

To study the *mir430* locus compaction in absence of Nanog, we acquired images of the *mir430* locus compaction every 15 seconds, and we used an approach based on spot detection to detect the *mir430* locus DNA signal (**Figure 31B**). As we are interested in understanding if the *mir430* locus compaction change with or without Nanog, we wanted to calculate the speed of compaction. We then needed to calculate the distance representing the *mir430* locus elongation in 3D. Briefly, maximum intensity peaks are detected locally in 3D on the image. The minimum between two peaks is defined by the diameter of the objects, manually selected, here 0.325 μm . After maximum peak detection, the fluorescence around the peak is considered as the same object, until a certain fluorescence threshold, determined manually. Once the objects are segmented, they are considered in 3D as ellipsoid. If only one ellipsoid is detected at the place where the *mir430* DNA signal is, the *mir430* locus is considered as compacted and if more than one spot is detected, the *mir430* locus is considered as not compacted. In this second case, as more than two ellipsoids are present, a distance between the ellipsoids can be calculated (**Figure 31B**). We consider the distance as an indirect measure of the *mir430* locus compaction state. If only one spot is detected, the distance is considered as 0.

The distance for individual tracks is later reported over time for each *mir430* allele track (**Figure 31C**). We defined the compacted time as all the time points for which the distance is equal to 0 (only one ellipsoid detected) (**Figure 31C**, blue zones). On the other hand, we define the decompacted time as all the time points for which the distance is greater than 0 (several ellipsoids are detected) (**Figure 31C**, orange zones).

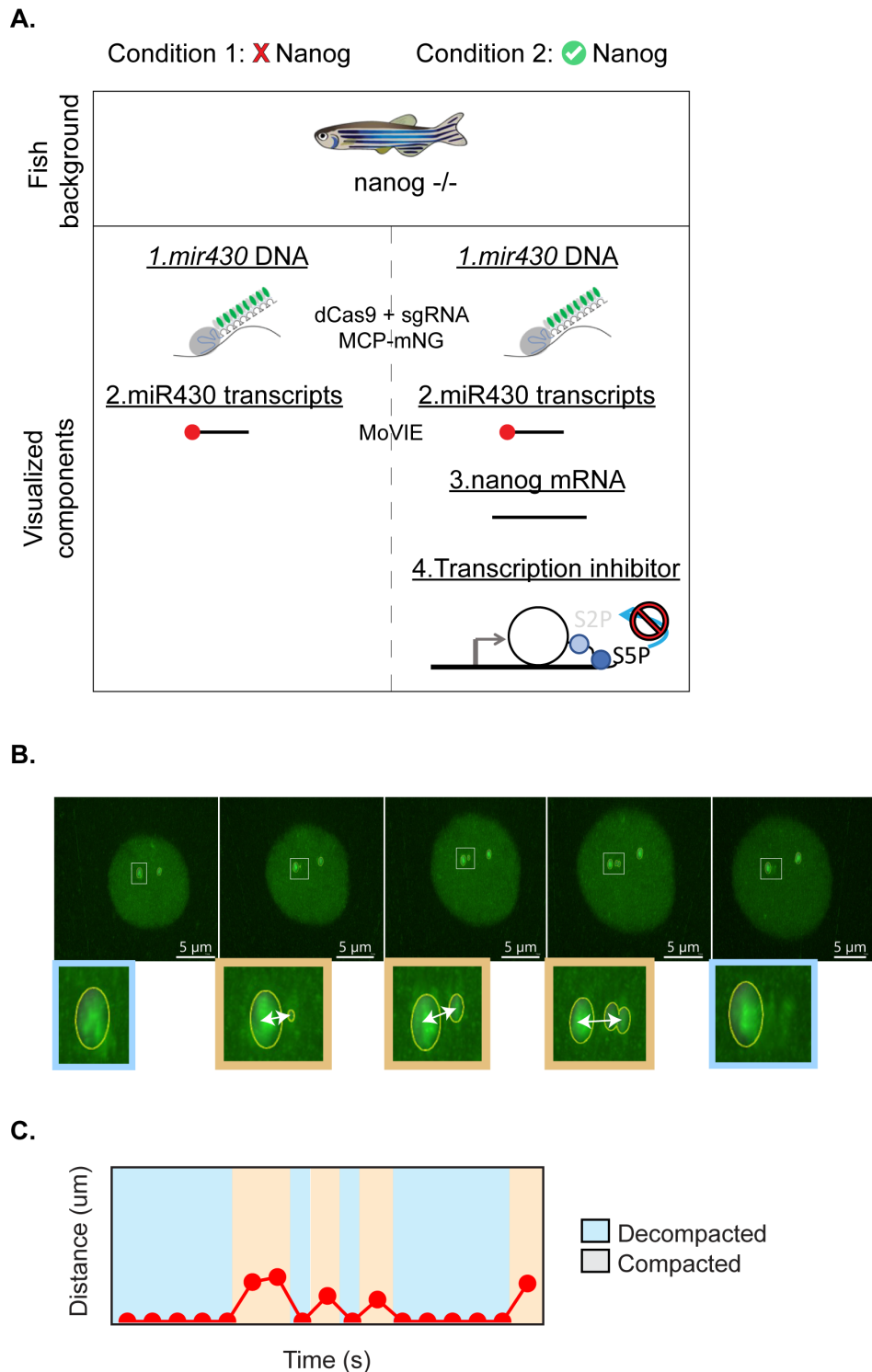


Figure 31: Is Nanog implicated in the mir430 locus compaction? Hypothesis and Methods.

A. Schematics explaining the experimental conditions. B. Schematics explaining the image analysis method. Spots are detected on the *mir430* DNA signal. Detection of one spot is considered as a null distance. If more than two spots are detected, the distance between two further apart is considered as the distance for analysis. C. Example of *mir430* allele distance analysis. If the distance is null, we considered the *mir430* locus as completely compacted and defined it as compacted state. In the contrary, if two or more densities were detected, we considered to be in decompacted state.

3.5.2 –NANOG ITSELF IS NOT RESPONSIBLE FOR *MIR430* COMPACTION

In the cases where Nanog has a role in compacting the *mir430* locus, we would expect that the absence of Nanog has a consequence how fast the *mir430* locus is compacting, that we can quantify using the speed at which the *mir430* densities are moving (**Figure 31B**). First, comparing the tracks in the presence (magenta) or in absence of Nanog (green), we can observe that in absence of Nanog, the number of oscillations between the compacted and decompacted states is higher, suggesting that the *mir430* locus is more dynamic in absence of Nanog (**Figure 32A**). The visual difference between the two datasets is confirmed by the average number of oscillations per track, between compacted and decompacted state: 1.74 in presence of Nanog, 2.6 in absence of Nanog (**Tableau 3**). The *mir430* is thus more often in a decompacted state in absence of Nanog.

Then, to check if Nanog plays a role in *mir430* compaction process, we looked at the speed of compaction. We defined this parameter as the distance between the detected ellipsoids to the time it takes to be in the compacted time again (**Figure 32B**, blue target). The speed of compaction revealed no significant difference between both conditions, in which Nanog is either present or absent ($p=0.91$) (**Figure 32C**). It suggests that Nanog is not influencing the “speed” at which the *mir430* locus compacts.

In contrast with this result, we observed that average time decompacted state is 15.9 seconds (one time frame) in presence of Nanog while it is 25.4 seconds in absence of Nanog (almost two frames) (**Tableau 2**, $p=0.021$, **Figure 32D**). Moreover, interestingly, we observed that the average distance between the detected ellipsoids is higher in absence of Nanog, 0.789 μm against 0.954 μm (**Tableau 2**). To

summarize, while our findings show that Nanog is not responsible for *mir430* compaction, the absence of Nanog does reveal that there are more oscillations in the tracks (**Tableau 3**), the average distance between the detected ellipsoids is higher (**Tableau 2**) and the *mir430* locus takes more time to compact (**Tableau 2, Figure 32D**), but this is not due to a slower compaction as the speed is not different (**Tableau 2, Figure 32C**). These results suggest that the *mir430* DNA structure is more relaxed and decompacts more often in absence of Nanog. We hypothesized that Nanog plays a role in keeping the *mir430* locus together once it is compacted. We called this hypothesis the glue hypothesis.

Tableau 2: Distance between clusters and speed of compaction (1k-cell stage)

1k-cell stage	Nanog	No Nanog	p-value
Average time in decompacted state (s)	15.9	25.4	0.021 (*)
Average speed of compaction (um/s)	0.0427	0.0470	0.89 (n.s.)
Average distance in decompacted state (um)	0.789	0.954	3.26E-05 (***)

Wilconxon test.

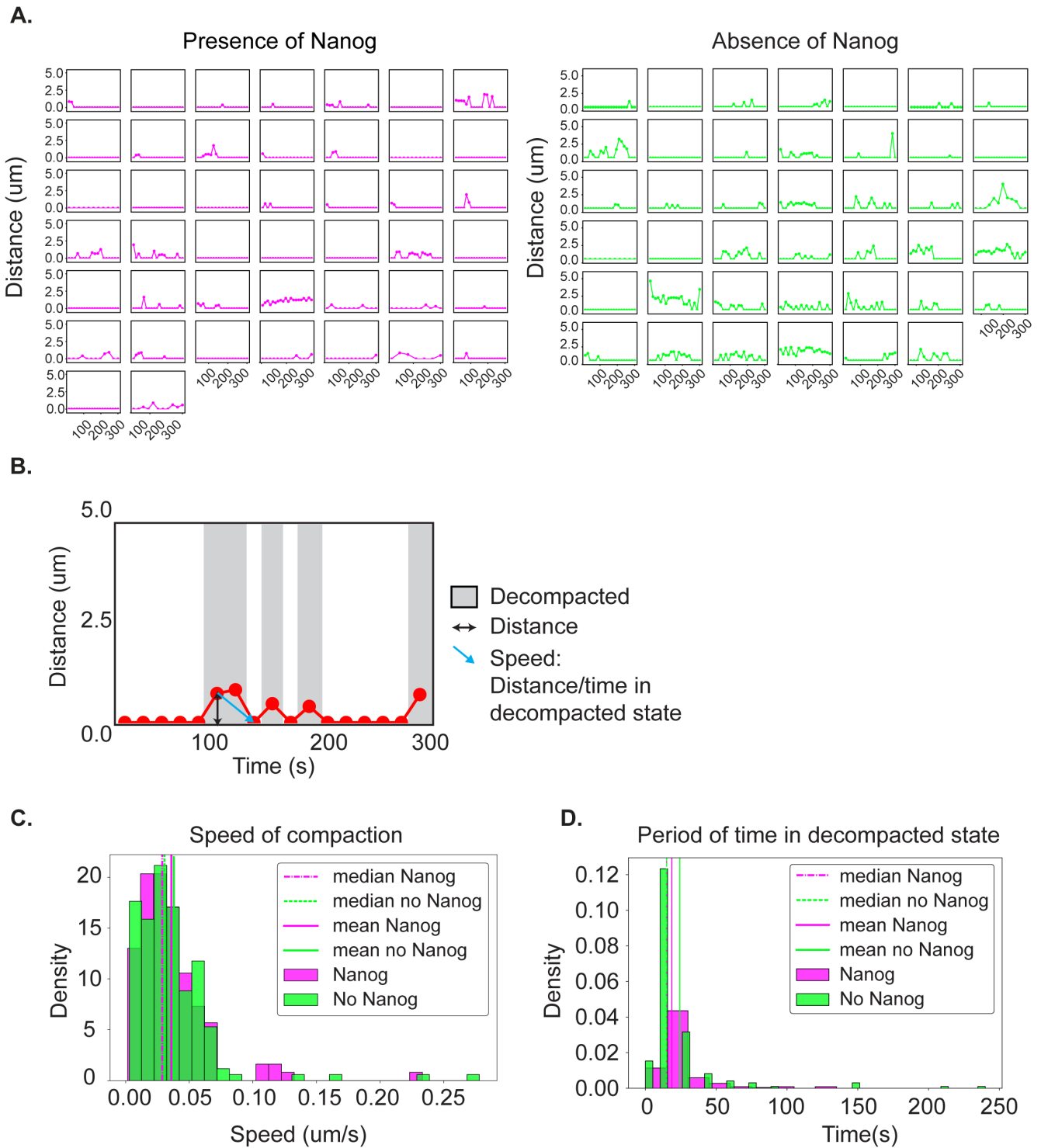


Figure 32: The *mir430* locus compacts as fast in presence and absence of Nanog

A. Distance between each detected spots, over time, for each individual track of the dataset. **B.** Schematics explaining the following parameters: time to come together and speed. **C.** Histogram showing distribution, median and average speed of *mir430* compaction. **D.** Histogram showing distribution, median and average time of *mir430* compaction. In green = absence of Nanog, in magenta = presence of Nanog.

Absence of Nanog: N1=4, N2=13, n=46, a=81. Presence of Nanog: N1=3, N2=17, n=53, a=81, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed and a the number of alleles. All mixed stages: 256-,512-,1k-cell and High stage.

3.6.3 – NANOG ACTS AS A GLUE TO KEEP THE *MIR430* LOCUS TOGETHER

To check if Nanog plays a role in keeping the *mir430* locus together once it is compacted, we decided to focus on the compacted state: when the *mir430* locus is compacted and only one ellipsoid is detected (**Figure 31B**). If Nanog acts as a glue, then the *mir430* locus would decompact less often in presence of Nanog and in absence of Nanog.

Tableau 3: Statistics about decompacted state (1k-cell stage)

1k-cell stage	Nanog	No Nanog	p-value
Tracks always sticking	34.0%	17.0%	/
Tracks always separated	2.27%	2,44%	/
Average number of oscillations	1.068	2.10	/
Average number of oscillations (no flat lines)	1.74	2.60	/
Average time stucked (with flats) (s)	124	71.2	6.09E-05 (***)
Average time stucked (with no flats) (s)			5.45E-05 (***)

First, we observed that in presence of Nanog, 34% of the tracks were never showing any oscillations in presence of Nanog while it is 17% in absence of Nanog (**Tableau 3**). It suggests that in presence of Nanog, the *mir430* locus decompact less often, supporting our model that Nanog could act as a “glue” for keeping the *mir430* locus compacted.

Then, we calculated the average time in compacted state, meaning how long in average the *mir430* locus is compacted before decompacting again, and only one ellipsoid is detected (**Figure 33A**). In presence of Nanog, the average time in compacted state was 123 seconds while it was 71.2 seconds in absence of Nanog ($p=0.0023$), showing a very large difference between the two experimental conditions. However, we have shown mentioned previously that in presence of Nanog, there are more tracks with no oscillations than in absence of Nanog. As it could lead to a bias in the calculation of the average in compacted state, we decided to measure this parameter removing the tracks that do not have any oscillation. We observed the same trend: the *mir430* locus stay compacted on average 89.3 seconds in presence of Nanog and 58.7 seconds in absence of Nanog (**Figure 33B**, $p=0.0011$).

To conclude, in the presence of Nanog, the *mir430* locus is more often compacted than in absent of Nanog. This outcome is not the result of a lower number of oscillations. These results suggest that Nanog is important for keeping the *mir430* locus together, as Nanog could act as a “glue” once the locus is compacted. This could help in activating the *mir430* locus, as we have suggested previously that a certain amount of Nanog is necessary in a small volume (**Figure 19**).

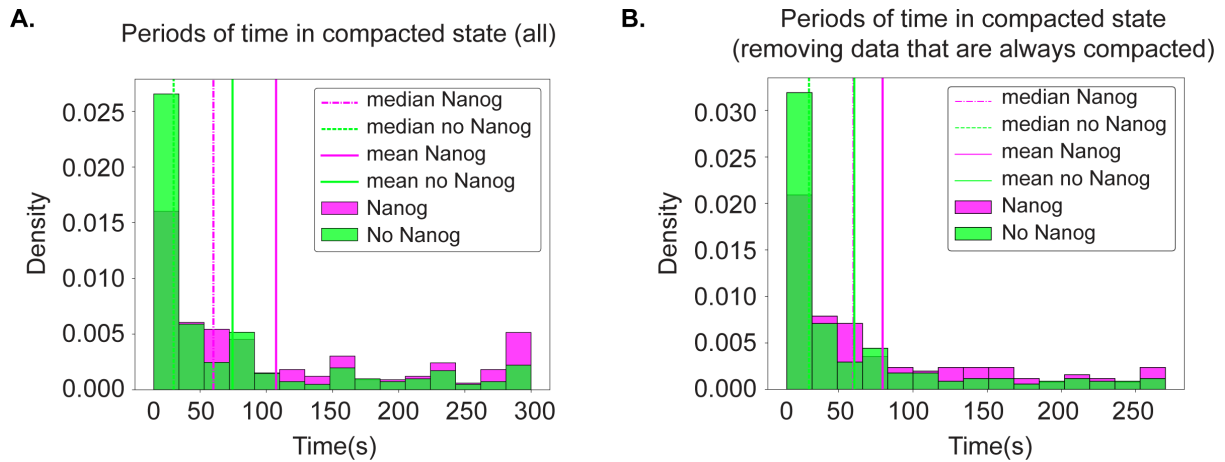


Figure 33: Nanog is important to keep the *mir430* locus together.

A: Histogram showing distribution, median and average time in decompacted state, including tracks with no oscillations. **B:** Histogram showing distribution, median and average time in compacted state, removing tracks with no oscillations. In green = absence of Nanog, in magenta = presence of Nanog. Absence of Nanog:

$N_1=4, N_2=13, n=46, a=81$. Presence of Nanog: $N_1=3, N_2=17, n=53, a=81$, where N_1 the number of biological replicates, N_2 the number of embryos and n the number of nuclei analyzed and a the number of alleles. All mixed stages: 256-, 512-, 1k-cell and High stage.

3.6. DYNAMICS OF NANOG, SOX19B AND POU5F3 TFs IN CLUSTERS DURING ZGA

The following work has been in part published recently ²⁷⁰. The Figure and the legend related to this publication have not been modified compared to the published work and will be indicated.

NANOG

Protein clustering, especially through the LLPS mechanism, have been proposed to accelerate particle movement ^{199,206,233}. Recent studies have shown that TF clustering influences TF dynamics by creating a confined compartment, that in turn accelerates target search, increase the TF binding turnover at TFBS and residence time at TFBS ^{55,68,69,236,237,330,331}. However, such works are recent and the impact of TF clustering on TF dynamics and consequences on transcription is still unclear ¹⁸⁵. To understand better what the dynamics of TF clusters is and how these dynamics change upon transcription, I decided to study the dynamics of Nanog, Sox19b and Pou5f3 TF inside clusters. These three TFs are essential for ZGA ^{269,270,302} and form multiple clusters in the nucleus at early stage in zebrafish embryos^{270,299}. Adding their essential role in development, Nanog forms clusters that precedes and colocalize with the *mir430* DNA locus (**Figure 27**) and *mir430* transcription^{270,299} (**Figure 26**). The *mir430* is the first gene to be largely expressed at early stage in zebrafish, and its transcription can be visualized by *mir430* two transcription body. Such a system allows me to study the dynamics between TF clusters colocalizing with transcription and other clusters in the nucleus.

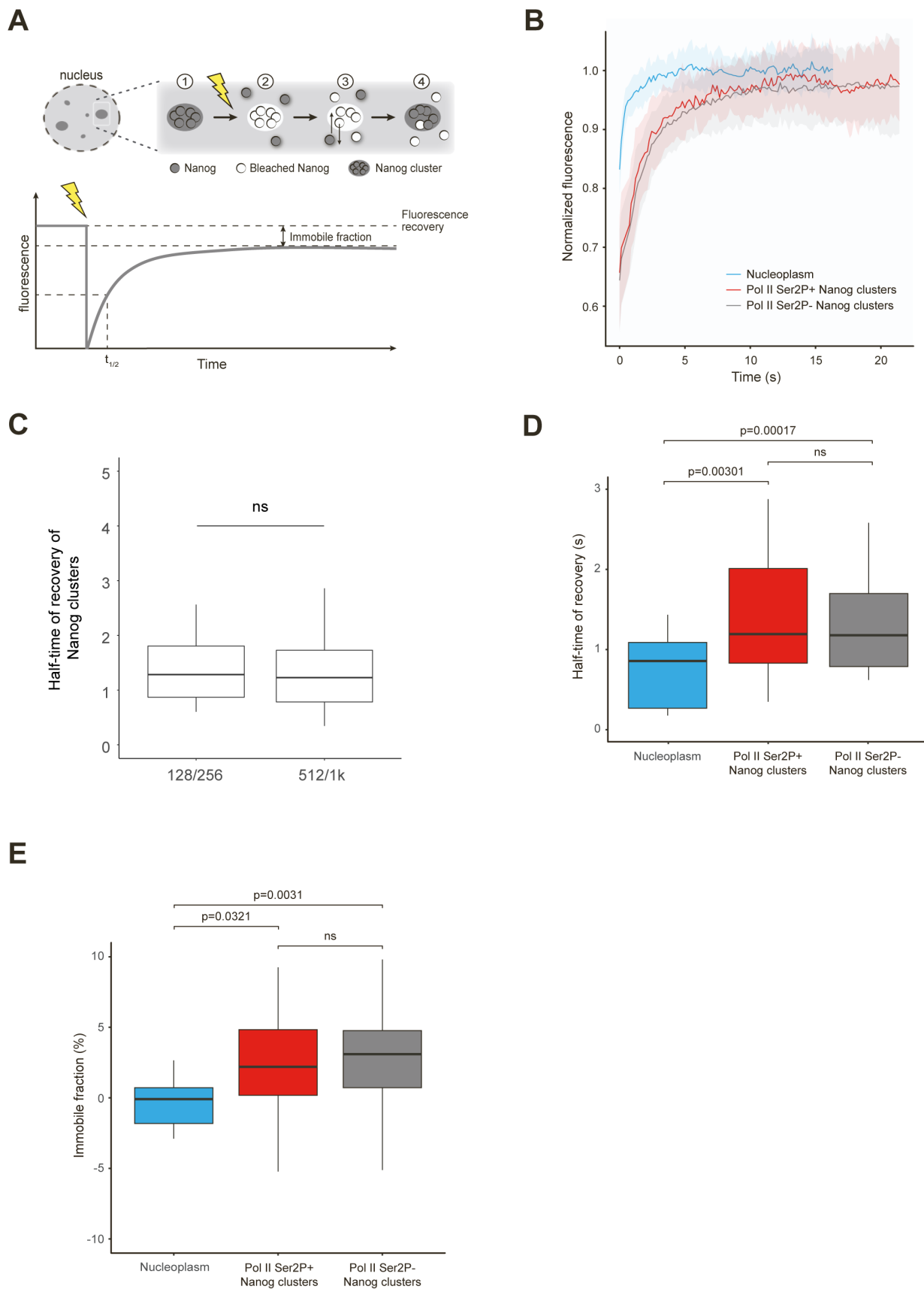


Figure 34: Nanog proteins in clusters recover fast.

A. Schematic of Fluorescence Recovery After Photobleaching (FRAP) assay and analysis. 1 - The mean intensity of fluorescence is determined prior to bleaching. 2 - Fluorescently labelled Nanog proteins in a Nanog cluster are bleached. 3 - If Nanog molecules in the cluster are mobile, bleached Nanog proteins will be replaced with unbleached ones from outside the cluster. 4 - In this example, almost all proteins are exchanged, resulting in an almost complete recovery of Nanog cluster fluorescence. Immobile fraction refers to the difference in recovered fluorescence (after normalizing for photobleaching) compared to the level pre-bleaching. This is a measure for what fraction of Nanog proteins are stably found in the cluster and therefore are not exchanged with unbleached proteins. Half-time of recovery ($t_{1/2}$) refers to the time needed to reach 50% of the maximum fluorescence reached after photobleaching. This number indicates how fast the Nanog proteins exchange between the inside and outside of the cluster and is a measure of mobility. **B.** Recovery curves of Nanog after photobleaching in nucleoplasm (N=1, n=17; blue), Pol II Ser2P-positive clusters (N=3, n=24; red), and Pol II Ser2P-negative clusters (N=3, n=53; gray). Thick lines represent the mean, shaded areas the standard deviation. Pol II Ser2P-positive clusters are those that colocalize with transcriptional activity and thus refer to the Nanog clusters at the mir430 locus. Pol II Ser2P-negative clusters are Nanog clusters that do not colocalize with transcriptional activity. This includes Nanog clusters away from the mir430 locus as well as clusters at the mir430 locus prior to them becoming transcriptionally active. **C.** Half-time of recovery for Nanog clusters at different stages. Early stages represent 256- and 512-cell stage (n=31, N=3), late stages represent 512- and 1k-cell stage (n=46, N=3). **D.** Half-time of recovery for Pol II Ser2P-positive, and Pol II Ser2P-negative Nanog clusters as well as nucleoplasm. **E.** Immobile fraction of Nanog in Pol II Ser2P-positive and negative clusters compared to nucleoplasm as derived from a single exponential fitting from the data shown in B (see Methods).

To study the dynamics of Nanog in the clusters, I decided to perform Fluorescence Recovery After Photobleaching (FRAP) imaging. I had three bleaching targets for Nanog: S2P+ Nanog clusters, S2P- Nanog clusters and the nucleoplasm. I measured the fluorescence of Nanog clusters before, during and after bleaching to study the fluorescence recovery of Nanog clusters. Based on this fluorescence pattern, I used two classical metrics to study fluorescence recovery: the half-time or recovery and the immobile fraction. The half-time of recovery ($t_{1/2}$) refers to the time needed to reach 50% of the maximum fluorescence reached after photobleaching (**Figure 34A**). It gives an interpretation of how fast bleached proteins inside the clusters are exchanged with unbleached proteins outside of the cluster. First, I wondered how is evolving the half-time across stages. It is not significantly changing from 128- to 1k-cell stage (**Figure 34C**). These results suggest that the protein dynamic is globally constant over development, but it cannot be excluded that tiny variations are not captured by FRAP. I then decided to merge the data for all the stages together. the half-time of recovery is short for the S2P+ and S2P- Nanog clusters and for the nucleoplasm. (**Figure 34D**). It shows a very dynamic movement of Nanog proteins

inside and outside the clusters. I assume here that binding of Nanog proteins is almost inexistant in the nucleoplasm (outside of the clusters) and Nanog protein movements is due to free diffusion. The half-time of recovery is significantly slower in Nanog clusters (S2P+ and S2P) compared to the nucleoplasm (**Figure 34D**, respectively $p=0.00301$ and $o=0.00017$, Wilcoxon test), while it is not different between Nanog clusters colocalizing or not with the transcription body (**Figure 34D**). Thus, Nanog proteins are retained longer in the cluster than in the nucleoplasm. These results suggest that Nanog proteins are binding on DNA or other factors in the clusters.

The immobile fraction represents the percentage of Nanog proteins that were not exchanged after photobleaching during the time-lapse (15 seconds) and thus the proportion of retained Nanog proteins in the clusters (either by long-term binding or other interactions) ^{315,317,332}. Consistent with free-diffusing state of Nanog protein in the nucleoplasm, bleaching the nucleoplasm leads to a complete fluorescence recovery with an immobile fraction centered around zero (**Figure 34E**). S2P+ and S2P- Nanog clusters show an immobile fraction significantly different to the nucleoplasm (respectively $p=0.032$ and $p=0.0031$, Wilcoxon test), potentially due to long-term binding of Nanog proteins (**Figure 34B, E**). However, this immobile fraction is very small (less than 5%), confirming the very dynamic state of Nanog proteins in the clusters.

To conclude, Nanog recovers very fast (order of seconds), both in mir430 transcription bodies, as well as in the other Nanog clusters. The recovery of Nanog in clusters, however, is slower than in the nucleoplasm, suggesting that while exchange is fast, it is slower than free diffusion, which is likely caused by (transient) interactions in the clusters (Nanog with DNA and/or other factors). Nanog is expected to freely diffuse in the nucleoplasm, which is in line with a very small immobile fraction in the nucleoplasm. The immobile fraction of Nanog within clusters is only slightly larger, suggesting that the majority of Nanog protein within clusters is not stably bound.

SOX19B

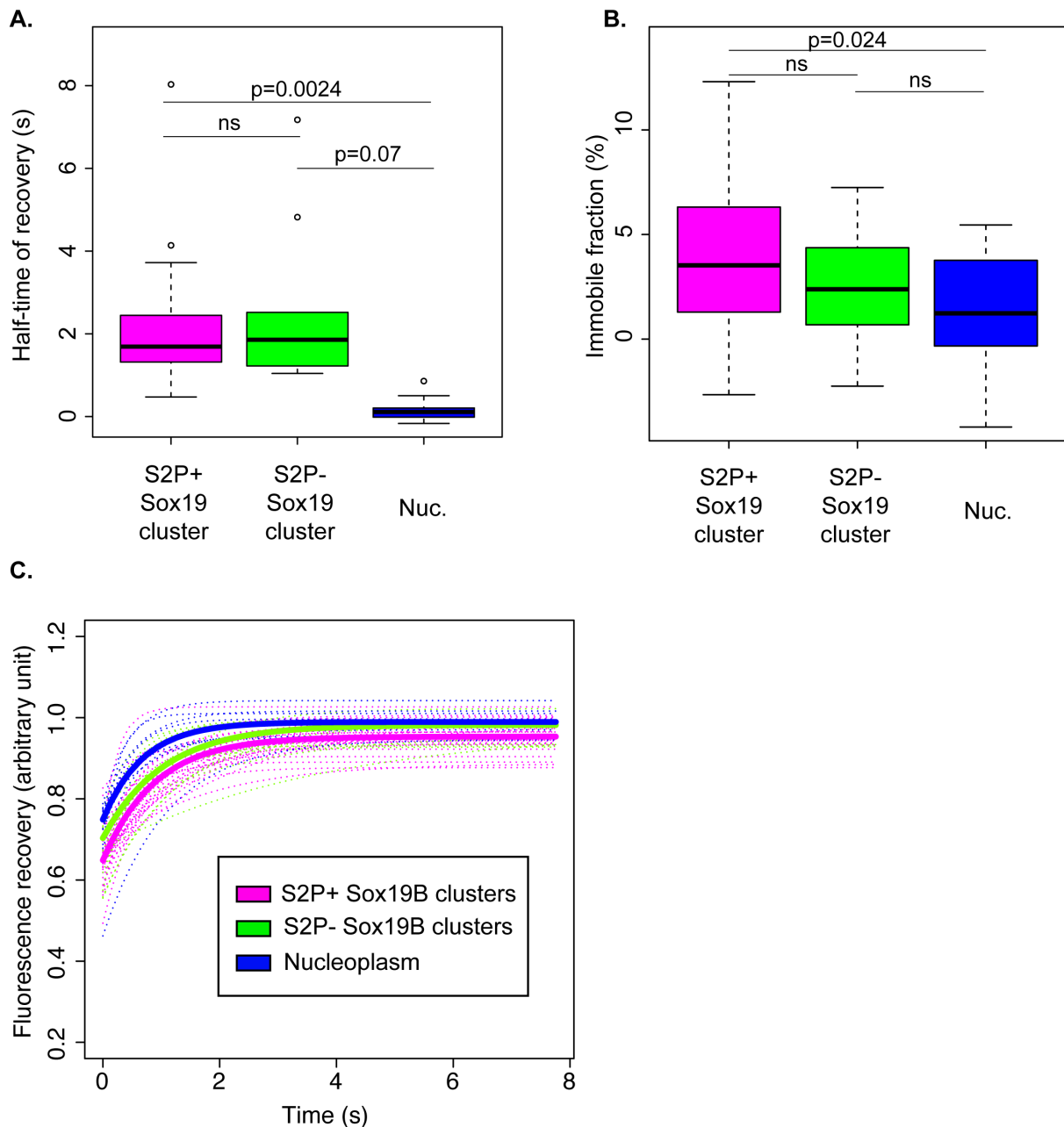


Figure 35: Sox19b proteins in clusters recover fast.

A. Half-time of recovery for Pol II Ser2P-positive, and Pol II Ser2P-negative Sox19b clusters as well as nucleoplasm. (N=1, n=11; blue), Pol II Ser2P-positive clusters (N=2, n=29; magenta), and Pol II Ser2P-negative clusters (N=2, n=10; green). **B.** Immobile fraction of Sox19b in Pol II Ser2P-positive and negative clusters compared to nucleoplasm as derived from a single exponential fitting from the data shown in B (see Methods). **C.** Recovery curves of Sox19b after photobleaching in nucleoplasm. Thick lines represent the mean, shaded areas the standard deviation. Pol II Ser2P-positive clusters are those that colocalize with transcriptional activity and thus refer to the Sox19b clusters at the mir430 locus. Pol II Ser2P-negative clusters are Sox19b clusters that do not colocalize with transcriptional activity. This includes Sox19b clusters away

from the *mir430* locus as well as clusters at the *mir430* locus prior to them becoming transcriptionally active.

In the contrary of *Nanog*, there are only two *Sox19b* clusters in the nucleus at early zebrafish development, preceding and colocalizing with *mir430* transcription²⁷⁰. In this case, I am sure that the *Sox19b* clusters are seeding by the *mir430* locus, as there are not present in the *Sox19b* *-/-* embryos²⁷⁰.

As for *Nanog*, the difference in half-time of recovery between S2P+ *Sox19b* clusters and the nucleoplasm is significant (**Figure 35A, C**, $p=0.0024$), suggesting that the *Sox19b* proteins are not only freely diffusing in the *Sox19b* clusters, but interacting with DNA or other proteins. However, the difference in half-time between S2P- *Sox19b* clusters and the nucleoplasm is not significant (**Figure 35A**, $p=0.07$). Either this result suggests that the *Sox19b* proteins have the same dynamics in clusters as *Sox19b* proteins in the nucleoplasm or the number of observations is too low to be significant. As previously for *Nanog*, the immobile fraction is close to zero for *Sox19b* proteins in the nucleoplasm, consistent with a free-diffusing state in the nucleoplasm. However, the immobile fraction of *Sox19b* proteins in S2P+ *Sox19b* clusters compared to the nucleoplasm is higher (**Figure 35B, C**, $p=0.024$), while it is not significant for S2P- *Sox19b* clusters (**Figure 35B, C**). These results suggest that the *Sox19b* proteins in S2P+ clusters might have a higher residence time than in S2P- *Sox19b* clusters, potentially related to the transcription status.

To conclude, *Sox19b* recovers very fast (order of seconds), with a difference dynamic depending on the transcription status. The recovery of *Sox19b* in clusters, as for *Nanog*, is slower than in the nucleoplasm, suggesting that while exchange is fast, it is slower than free diffusion, which is likely caused by (transient) interactions in the clusters (*Nanog* with DNA and/or other factors). *Sox19b*, as for *Nanog*, is expected to freely diffuse in the nucleoplasm, which is in line with a very small immobile fraction in the nucleoplasm. The immobile fraction of *Sox19b* within clusters is only slightly larger than the nucleoplasm, suggesting that the majority of *Sox19b* protein within clusters is not stably bound, but a small fraction could have a longer residence time in S2P+ *Sox19b* clusters.

POU5F3

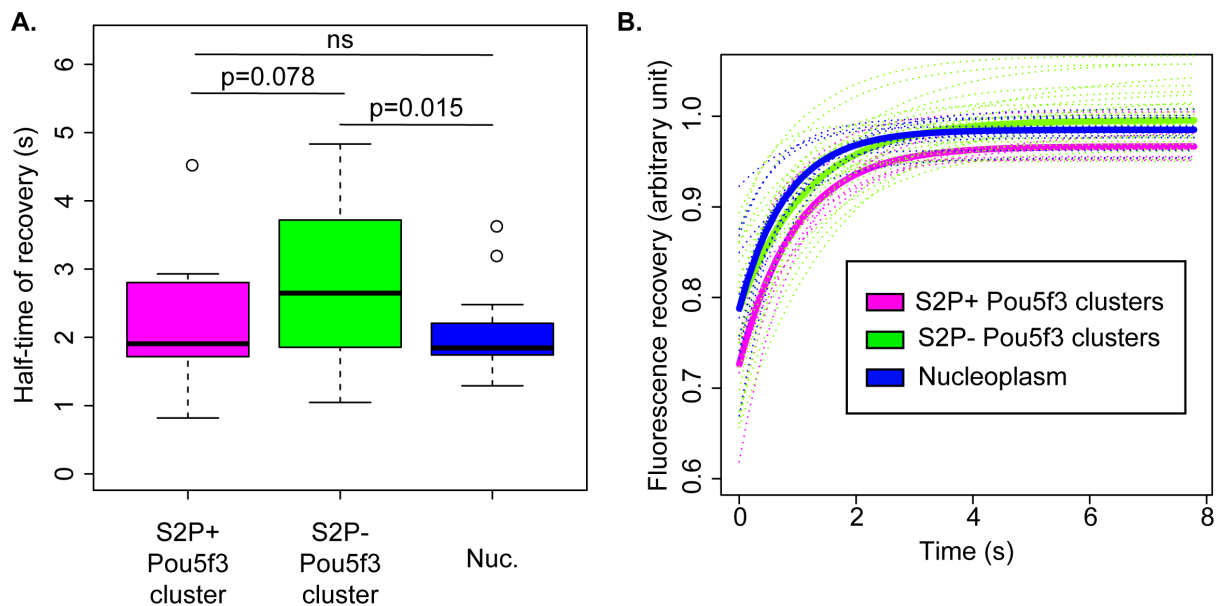


Figure 36: Pou5f3 proteins recovers fast in clusters.

A. Half-time of recovery for Pol II Ser2P-positive, and Pol II Ser2P-negative Pou5f3 clusters as well as nucleoplasm. (N=1, n=8; blue), Pol II Ser2P-positive clusters (N=2, n=15; magenta), and Pol II Ser2P-negative clusters (N=24, n=2; green). **B.** Recovery curves of Sox19b after photobleaching in nucleoplasm. Thick lines represent the mean, dotted lines represent individual observations. S2P+ clusters are those that colocalize with transcriptional activity and thus refer to the Pou5f3 clusters at the *mir430* locus. Pol II Ser2P-negative clusters are Pou5f3 clusters that do not colocalize with transcriptional activity. This includes Pou5f3 clusters away from the *mir430* locus as well as clusters at the *mir430* locus prior to them becoming transcriptionally active.

We have shown previously that Pou5f3 makes multiple clusters in the nucleus²⁷⁰. However, we have shown that none of these clusters are colocalizing with any S5P or S2P transcription body²⁷⁰. In our previous study, we were injecting 150 pg of Pou5f3 mNG in ABTL WT embryos²⁷⁰. In my case, I injected here 250 pg of Pou5f3-mNG in ABTL WT embryos and could observe that some of the Pou5f3 clusters are colocalizing with S2P transcription body, and thus *mir430* transcription. It is possible that injecting more is leading to the clustering of Pou5f3 at the *mir430* locus. The half-time of recovery for S2P- Pou5f3 clusters is different from the nucleoplasm (**Figure 36A**, p=0.015) but not for S2P+ Pou5f3 clusters (**Figure 36A**, p=0.078), suggesting that Pou5f3 proteins diffuse slower than in nucleoplasm in S2P- clusters but as fast as in the nucleoplasm inside S2P+ Pou5f3 clusters. These results are different than for

Nanog and Sox19b, for which proteins are diffusing slower in S2P+ clusters, potentially because of interactions with DNA or other proteins. As previously described for Nanog and Sox19b, I can observe that the immobile fraction for Pou5f3 in the nucleoplasm is close to 0, suggesting free diffusion of the protein. As for Nanog and Sox19b, the immobile fraction of S2P+ Pou5f3 clusters is small, and higher from the nucleoplasm but not significant (**Figure 36B**, $p=0.052$). However, the immobile fraction for S2P- Pou5f3 clusters is significant compared to the nucleoplasm (**Figure 36B**, $p=0.004$) but higher than the nucleoplasm one. As the nucleoplasm immobile is close to 0 because I assume that the Pou5f3 proteins in the nucleoplasm are only diffusing, it is surprising to observe that the S2P+ have even a smaller immobile fraction, some of them being even negative (**Figure 36B**). These results suggest that more Pou5f3 proteins coming back to the Pou5f3 clusters than before bleaching, suggesting a constant flow of new Pou5f3 proteins at the S2P+ clusters, potentially explained by the continuous growth of the Pou5f3 cluster during the image acquisition.

To conclude, Pou5f3 recovers very fast (order of seconds), with a difference dynamic depending on the transcription status. The recovery of Pou5f3 in clusters, in the contrary of Nanog and Sox19b, is as fast as in the nucleoplasm for S2P+ Pou5f3 clusters, suggesting that Pou5f3 proteins might only diffuse inside this clusters. However, S2P- Pou5f3 clusters recover slower, suggesting a while exchange is fast, it is slower than free diffusion, which is likely caused by (transient) interactions in the clusters (Nanog with DNA and/or other factors). The immobile fraction of Pou5f3 within S2P+ clusters is only slightly larger than the nucleoplasm, suggesting that the majority of Pou5f3 proteins within S2P+ clusters is not stably bound, but a small fraction could have a longer residence time in S2P+ Sox19b clusters. However, the immobile fraction for S2P- Pou5f3 clusters is larger than the nucleoplasm one, suggesting a continuous growth of the S2P- Pou5f3 clusters.

All the results together for Nanog, Sox19b and Pou5f3 suggest that the TF are diffusing slower in clusters, maybe due to DNA interactions or proteins interactions through their IDRs^{69,185}. It suggests that if TF clustering increase the chances to active transcription, it is not due a faster diffusion in clusters, but rather by confinement. Interestingly, none of these proteins show a high immobile fraction in clusters, suggesting that they are not binding with a long residence time on DNA, but rather display a fast binding-unbinding on DNA, consistent with previous studies on the dynamics of TF on DNA^{225,333,334}. The difference between the S2P+ and S2P- clusters

is not important for the three TFs, suggesting that the dynamics of these TFs is not changing dramatically depending on the transcription status.

4. Discussion

4.1 ARE THE NANOG CLUSTERS REALLY PRESENT IN A WILD-TYPE BACKGROUND?

I have shown that the Nanog clusters colocalizing with transcription are the largest, the most concentrated and the containing the most Nanog among other Nanog clusters in the nucleus (**Error! Reference source not found.**E, F). However, I have also shown that the number of Nanog clusters in the nucleus is quite variable (**Figure 12B**). The minimum number of Nanog clusters observed among the three stages when transcription start ranges from 8 to 10 depending on the developmental stage, while the maximum reaches 94 at 512-cell stage (**Figure 12B**). A such high variability in the number of clusters raises the question whether these Nanog clusters are present in the nucleus in WT conditions or are resulting from a technical issue, for example related to the experimental conditions.

4.1.1. AN OVEREXPRESSION OF NANOG COULD LEAD TO ARTIFICIALLY INCREASE THE NUMBER OF NANOG CLUSTER IN THE NUCLEUS.

In absence of an endogenously labeled fish line for the *nanog* gene, to visualize Nanog in live-imaging, I injected Nanog as a nanog mRNA coupled with a fluorescent protein, such as mNG or HaloTag (with a dye), in *nanog* double knock-out embryos. I decided to take advantage of injecting in the *nanog* *-/-* to have all Nanog labeled and avoid the pull of unlabeled WT Nanog proteins. For all experiments presented in this thesis, I injected the amount of nanog mRNA that we are using to rescue the development phenotype of *nanog* *-/-* fish, to be as close as possible from the WT conditions. This concentration was previously determined in the lab as 120 pg of *nanog* mRNA per embryo based a previous study and survival assays ^{270,320,335}, equivalent to 180 pg or 210 pg of respectively Nanog-mNG and Nanog-HaloTag when taking the molecular weight of the fluorescent protein in account.

Tableau 4: The different experimental conditions to rescue the developmental phenotype of *nanog* ^{-/-} embryos in three different studies.

Study	Mutant	Injected amount (pg)	Place	Stage	% Rescue
Veil et al.	Mutant 1	50 to 300	Cell	1-cell	35 to 60%
Gagnon et a.	Mutant 2	5	Cell	1-cell	88%
Gagnon et a.	Mutant 2	25	Cell	1-cell	12%
Gagnon et a.	Mutant 2	25	Yolk	4-cell	25%
He et al.	Mutant 3	500	No indication	1-cell	100%
Kuznetsova et al.	Same as Veil et al.	120	Cell	1-cell	Not quantified

However, in the literature, three studies used three different amounts of *nanog* mRNA to rescue the developmental phenotype, injected either in the cell or in the yolk ^{320,336,337}. Adding to the amount injected mRNA, the location where the components are injected, the cell or the yolk (**Figure 37A**), matters. Injecting in the cell allows direct availability of the mRNA to the translation machinery. If the mRNA is injected in the yolk, its availability for translation in the cell will depend on the ooplasmic flow from the yolk to the cell during the first cell cycles of embryo development (**Figure 37B**) ³³⁸. During the first cell stage, the cell grows as the result of actin polymerization wave, resulting in the ooplasm flowing towards the animal pole (**Figure 37B**, two first images). The ooplasmic flow leads to the movement of yolk components into the cell (**Figure 37B**, beads labeled in red) ³³⁸. Both types of injection can have consequences on how much of the injected mRNA will be translated into proteins.

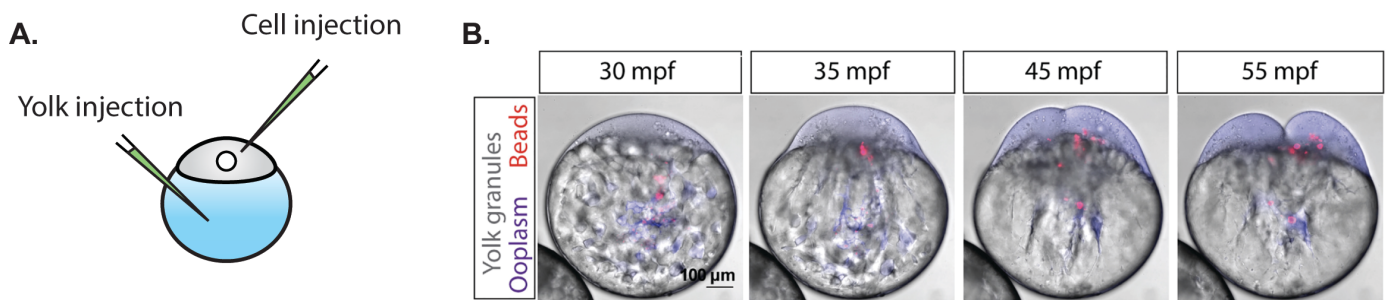


Figure 37: Redistribution of the injected components from the yolk to the cell during the first cell cycles

A. Adapted from Shayan Shamipour, Roland Kardos, Shi-Lei Xue, Björn Hof, Edouard Hannezo, Carl-Philipp Heisenberg, Bulk Actin Dynamics Drive Phase Segregation in Zebrafish Oocytes, *Cell*, Volume 177, Issue 6, 2019, Pages 1463-1479.e18, ISSN 0092-8674, <https://doi.org/10.1016/j.cell.2019.04.030>. Copyright C 2023 by Elsevier B.V. Creative Commons license CC BY 4.0.

In Veil et al, they tried four different amounts of *nanog* mRNA (50, 100, 150 and 300 pg) to check which one was rescuing more embryos after gastrulation, injected in the cell. While no embryo survived past gastrulation in non-injected *nanog* $-/-$ controls, the percentage of rescued embryos for full body axis and head development ranged from around 35% to 60%, uncorrelated with the injected concentrations (**Tableau 4**, first line) ³²⁰. In Gagnon et al, they rescued the *nanog* $-/-$ embryos using either 1) 5 pg or 2) 25 pg of *nanog* mRNA injected in the cell at 1-cell stage, 3) 25 pg of mRNA injected in the yolk at 4-cell stage (**Tableau 4**, second to fourth lines). For embryos injected with 5 pg in the cell at the 1st-cell stage, 88% were fully rescued for full body axis and head development at 24 hpf, while only around 12% of the embryos were rescued injecting 25 pg in the cell at the 1st cell stage. Around 25% of the embryos for 25 pg in the yolk at 4-cell stage. Finally, in He et al, they injected 500 pg of *nanog* mRNA and 100% of the embryos were fully rescued for body axis and head development at 48 hpf (**Tableau 4**, fifth line). For this study, the authors did not mention whether the injection was carried out in the cell or in the yolk. These three studies show how considerably different the rescuing concentrations and the associated percentage of rescue can be. These differences could lay on the variability in the experimental conditions between the three studies. First, they generated independently *nanog* $-/-$ mutants using TALENs, leading to three different mutations in the *nanog* gene. Second, they also did not inject at the same place (cell or yolk), when mentioned and did not use the same method to produce the *nanog* mRNA for rescue. Third, even if not mentioned, we can expect that the way to quantify the *nanog*

mRNA after *in vitro* transcription might vary as well, leading to potential variation in the real amount injected in the embryos. Taken all together, they important variabilities between these experiments that could lead to different results thus in the percentage of rescue of nanog $-/-$ embryos. The study from Gagnon et al, where they used an amount of nanog mRNA as low as 5 pg to rescue the developmental phenotype, suggests that the amount of nanog necessary to rescue the phenotype could be lower than what we used until now. If such a small amount of Nanog is enough to rescue the phenotype, it is possible that we are injecting too much mRNA, leading to an overexpression of Nanog. More Nanog could lead in a greater number of Nanog clusters in the nucleus. Moreover, I have observed a large variation in the number of Nanog clusters between different nuclei (**Figure 12B**). Not only injecting different concentrations could lead to a variation in the number of clusters, but where the product is injected in the cell or in the yolk as well. Indeed, I have observed that injected embryos with Nanog-mNG mRNA for example display a different nuclear background, even within the same embryo (**Figure 38A**). The injected product could stay concentrated locally in the injected area and not diffuse fast enough before translation to reach a uniform distribution in the embryo. By consequence, adding to the potential overexpression by injecting too much of nanog mRNA, the number of Nanog clusters in the nucleus could also be influenced by a heterogenous distribution of the nanog mRNA in the embryo.

To check if the number of Nanog clusters is related to the injected concentration, one could count how many clusters are present in the nucleus as the nanog mRNA injected concentration increases. This experiment is part of an ongoing project in the lab by another PhD student, Shivali Dongre (manuscript in preparation). Another way to study whether the number of clusters in the nucleus is related to Nanog concentration could be to plot the number of Nanog clusters as a function of the nuclear intensity. In this case, we use the nuclear intensity is an indirect read-out for Nanog concentration. Performing such analysis, Shivali Dongre has observed that the number of Nanog clusters is directly correlated to the amount of fluorescence in the nucleus (**Figure 38B**, using my own data). This suggests that the nuclear concentration of Nanog proteins is correlated with the number of Nanog clusters. As I might inject more nanog mRNA than needed for developmental rescue, the presence of the Nanog clusters in the nucleus might not represent what it is happening in WT conditions.

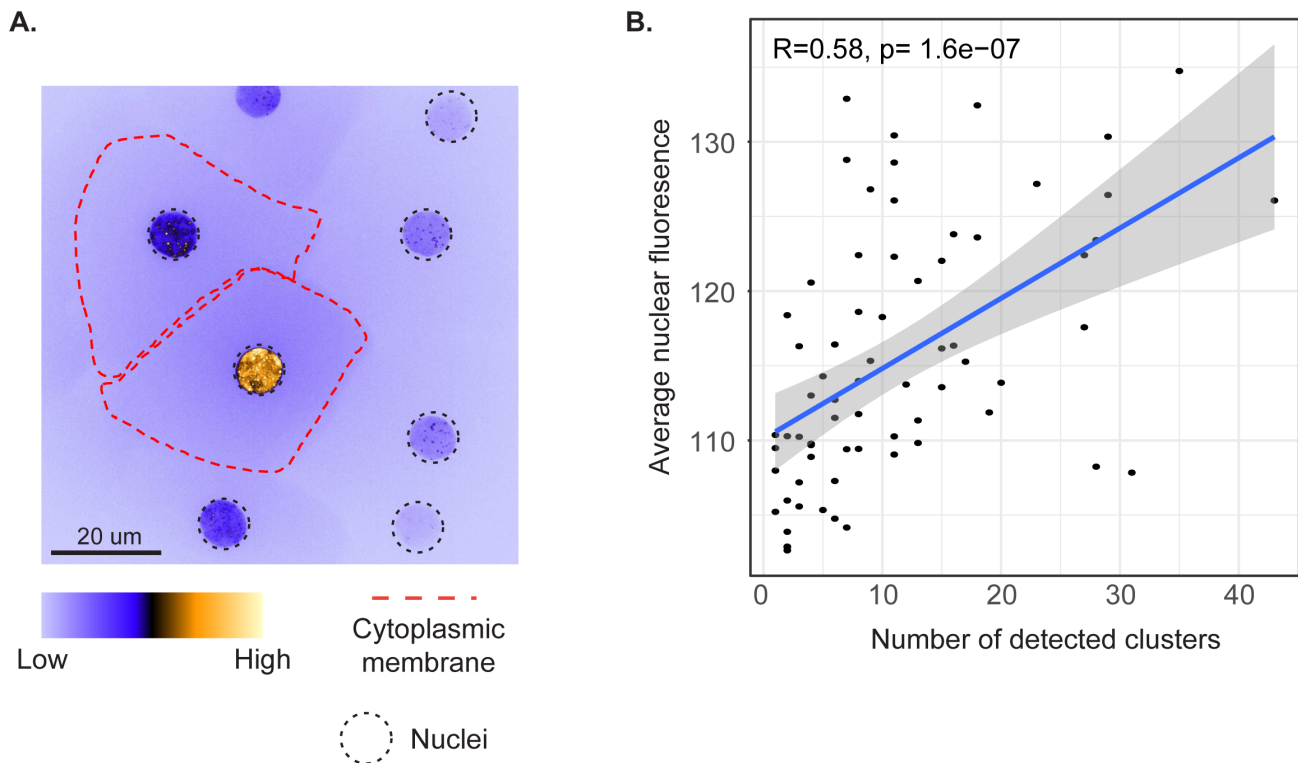


Figure 38: A higher fluorescence nuclear background is correlated with more Nanog clusters.

A. Snapshot of an embryo at 512-cell stage at a magnification of 100x. Each black dotted circle represents a nucleus, each red dotted shape represent a cell. The two nuclei in the two cells surrounded by the red dotted lines have a very different nuclear background, although they are in the same nucleus. **B.** Scatterplot of the number of detected Nanog clusters detected in the nucleus as a function of the average nuclear fluorescence. Each black dot represent a nucleus. The blue line represents the linear correlation between the two metrics. $R=0.58$, $p=1.6e-07$.

4.1.2. HOW TO DETERMINE IF THE NANOG CLUSTERS ARE REAL?

To reveal if they are indeed Nanog clusters in the nucleus and what is the real number in a WT situation, one could use approaches such as fluorescence in situ hybridization and detecting WT Nanog proteins by using a direct antibody against Nanog. However, this method implicates the fixation of the embryos. Such process has been demonstrated to influence the presence and number of protein clusters. Indeed, fixation could lead the appearance or disappearance of protein clusters, as well as changing their number in the nucleus³³⁹. Live imaging to determine if Nanog clusters are present in WT conditions should then be favored. The first and best approach is to

knock-in a fluorescent tag in 5' or 3' of the nanog gene. One could then directly visualize labeled-Nanog proteins without the need of injection. However, it was shown that the addition of a fluorescent protein in 5' or 3' of a protein could change the clustering behavior of the same protein³⁴⁰. Instead of using a fluorescent tag, one could use recognition of a specific epitope by an antibody, or a part of an antibody. For example, Nanobodies or Fabs can be used to recognize specific epitopes of a protein using only the variable part of an antibody. Moreover, their smaller size allows their use in live imaging^{341, 342}. Determine if Nanog clusters are present in the nucleus at early zebrafish development could be further inquired using one of the cited methods.

4.1.3. THE NANOG CLUSTERS AT THE MIR430 LOCUS FORM PREFERENTIALLY COMPARED THE OTHER NANOG CLUSTERS

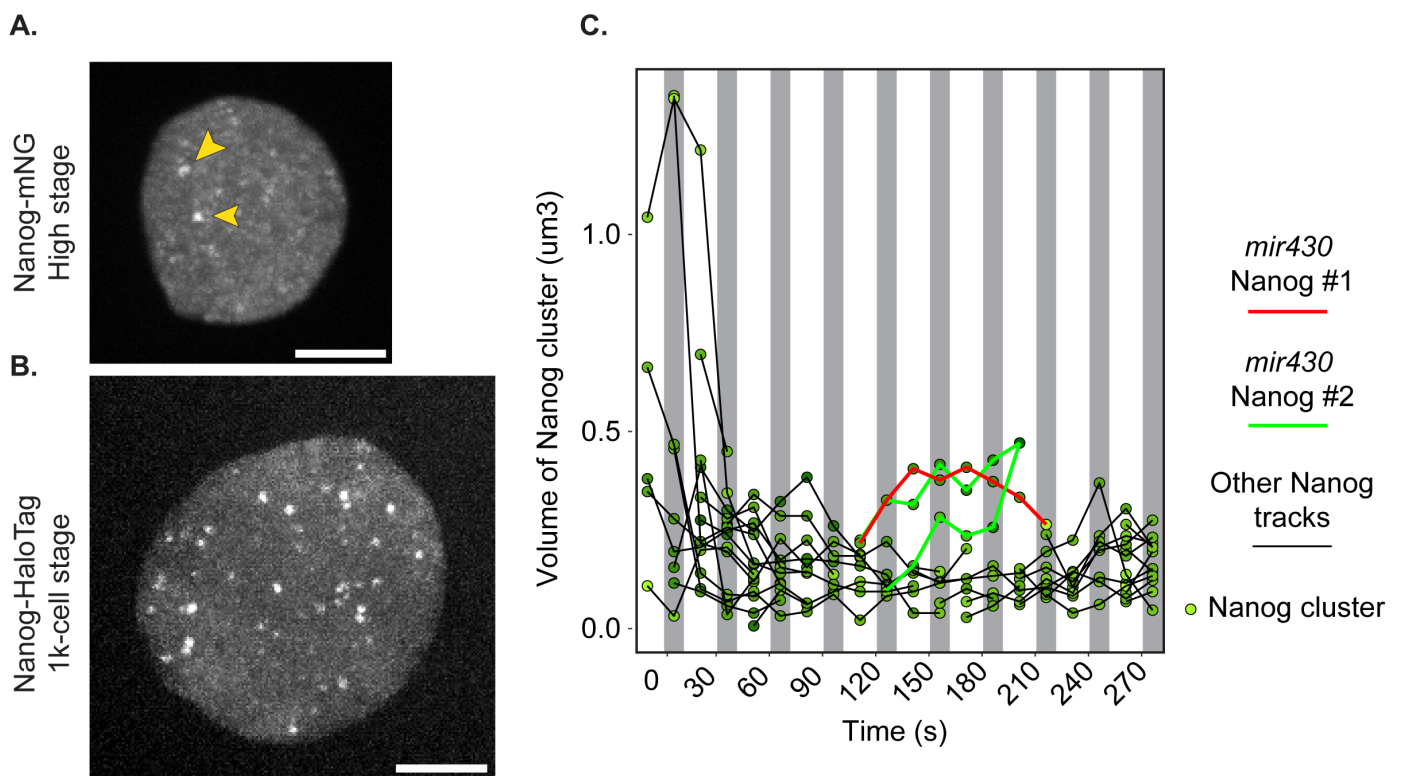


Figure 39: Features of Nanog clusters in different experimental conditions

A. Snapshot of a nucleus at High stage labeled by Nanog-mNG. Only two Nanog clusters are typically visible at this stage, marked here by two yellow arrows. Scale bar is represent 5 um.

B. Snapshot of a nucleus at 1k-cel stage labeled by Nanog-HaloTag (JF646) in a ABTL WT embryo. Nanog-HaloTag (JF646) makes multiple clusters in the nucleus in this experimental condition, while only two Nanog clusters are visible when co-injected with other components in *nanog* *-/-* mutant embryos. Scale bar represents 5 um. **C.** Plot representing the evolution of the volume of several Nanog clusters (um³) from the nucleus as a function of time (s). Each green dot represent a Nanog cluster at a given time. The same cluster over time is related by a line. Back lines represent Nanog clusters that will never be positive for S5P. Green and red lines represent respectively the first and second *mir430* Nanog clusters. For the second *mir430* Nanog cluster, two small Nanog clusters merge to give rise to a larger cluster at t=210 s.

The Nanog clusters in the early zebrafish embryo could be the consequence of an overexpression. However, in some situation, the *mir430* Nanog clusters are preferentially present compared to the other Nanog clusters, suggesting that they are present in WT conditions. Indeed, at the developmental cell stage just after canonical activation of the genome, the High stage, I commonly observe only two Nanog clusters during the cell cycle (**Figure 39A**). These two Nanog clusters are colocalizing with the largest S5P and S2P transcription, positive for *mir430* transcription. Therefore, they are the *mir430* Nanog clusters. Moreover, when *nanog* mRNA is coupled with HaloTag and injected in the WT embryo alone, I can observe multiple Nanog clusters, as with Nanog-mNG (**Figure 39B**), but injecting Nanog-HaloTag with other components in the *nanog* *-/-*, I observe only two Nanog clusters, the ones seeding by the *mir430* locus (**Figure 27A**). These results suggest that while the number of Nanog clusters in the nucleus varies, the *mir430* Nanog clusters are always present. The *mir430* Nanog clusters are also the first one to be reformed in the nucleus after mitosis. Indeed, I have shown that there are multiple Nanog clusters on mitotic chromosomes and in the nucleus (**Figure 12**). However, the Nanog clusters are fading off, sometimes completely disappearing during telophase and reappear gradually as the new cell cycle begins (

Figure 25, Figure 28A, C). Interestingly, the first Nanog clusters to visually come back after mitosis are later becoming S5P positive and thus are the *mir430* Nanog clusters (

Figure 25). After detection and quantification of all the Nanog clusters present in the nucleus during the mitosis to interphase transition, I can observe that the others Nanog, although decreasing in size, are still present during the mitosis to interphase transition (**Figure 39C**, from $t=0$ to $t=105$ s). However, mir430 Nanog clusters grow first (from $t=120$ s, red and green line) while the other Nanog cluster start to grow much later (from $t=210$ s). These results indicate that after mitosis, while Nanog is binding on chromatin to reform clusters, it preferentially forms clusters at the *mir430* locus and later other clusters appear. Such preferential binding of Nanog at the *mir430* locus could be explained by the high number of TFBS at the *mir430* locus. However, while TFBS for Nanog are present on the *mir430* locus, its density is not higher than in other regions of the genome (data do not show, PhD project of Hann Shen Ng). Yet, studies using ChIP-seq and CUT&RUN approaches have shown that Nanog binds at the *mir430* locus^{271,343}, more than in other regions of the genome at such early stages³⁴³ (**Figure 40**), suggesting that Nanog has a high affinity for this region at early stages. All the results together suggest that Nanog binds preferentially at the *mir430* locus and that Nanog clusters might be indeed present at the *mir430* locus in WT conditions.

One can imagine the following scenario: Nanog binds first at the *mir430* locus, forming two clusters there. I have shown that the Nanog clusters seeding by the *mir430* locus have a maximum loading capacity (**Figure 20**). If Nanog is in excess, one can expect that after binding at the *mir430* locus, Nanog TF bind to other regions of the genome for which they have less affinity than the *mir430* locus, forming other Nanog clusters there. If like the *mir430* locus, these other Nanog clusters have a maximum loading capacity, then new Nanog clusters are appearing as the concentration is increasing. This hypothetical scenario is in line with the positive correlation I have observed between the number of Nanog clusters and the average nuclear fluorescence (**Figure 38B**).

4.1.4. ARE NANOG CLUSTERS SEEDING BY THE MIR430 LOCUS REAL?

The *mir430* Nanog clusters might form preferentially compared to other Nanog clusters in the nucleus. However, for the present work in this thesis and relate its

importance in transcription regulation, the most important is to determine if Nanog really form clusters at the *mir430* locus in WT condition.

First, Nanog is essential for *mir430* transcription ^{269,270}, indicating that whatever concentration of Nanog is present in the nucleus, Nanog might at least bind there to regulate *mir430* transcription. Moreover, ChiP-seq and CUT&RUN experiments for Nanog in early zebrafish has shown peaks at the *mir430*, supporting that Nanog indeed bind at the *mir430* locus. Second, I have shown that a certain amount of Nanog is correlated with *mir430* activation (**Figure 19A**, **Figure 20**). However, even when the size of Nanog clusters is only 25% of their maximal size, I can already visualize and segment them (**Figure 19A**, Volume plot). This suggests that even in lower amount of Nanog seeded by *mir430* locus, a cluster is formed. Third, a recent work has used expansion microscopy in early zebrafish embryos to study how Nanog is spatially related to transcription initiation in the nucleus ²⁹⁹. The resolution obtains by such a technique allow them to observe each individual Nanog TF bound on chromatin. They could visualize large Nanog clusters at the *mir430* locus but no other clusters outside of the *mir430* locus. This study suggest that the Nanog clusters seeded by the *mir430* locus are present in the nucleus. However, they did not find the extra clusters of Nanog ²⁹⁹. All these observations together suggest that the Nanog clusters at the *mir430* locus are present in WT conditions. However, the only way to confirm this is to label Nanog endogenously or use labeling methods independent on mRNA translation.

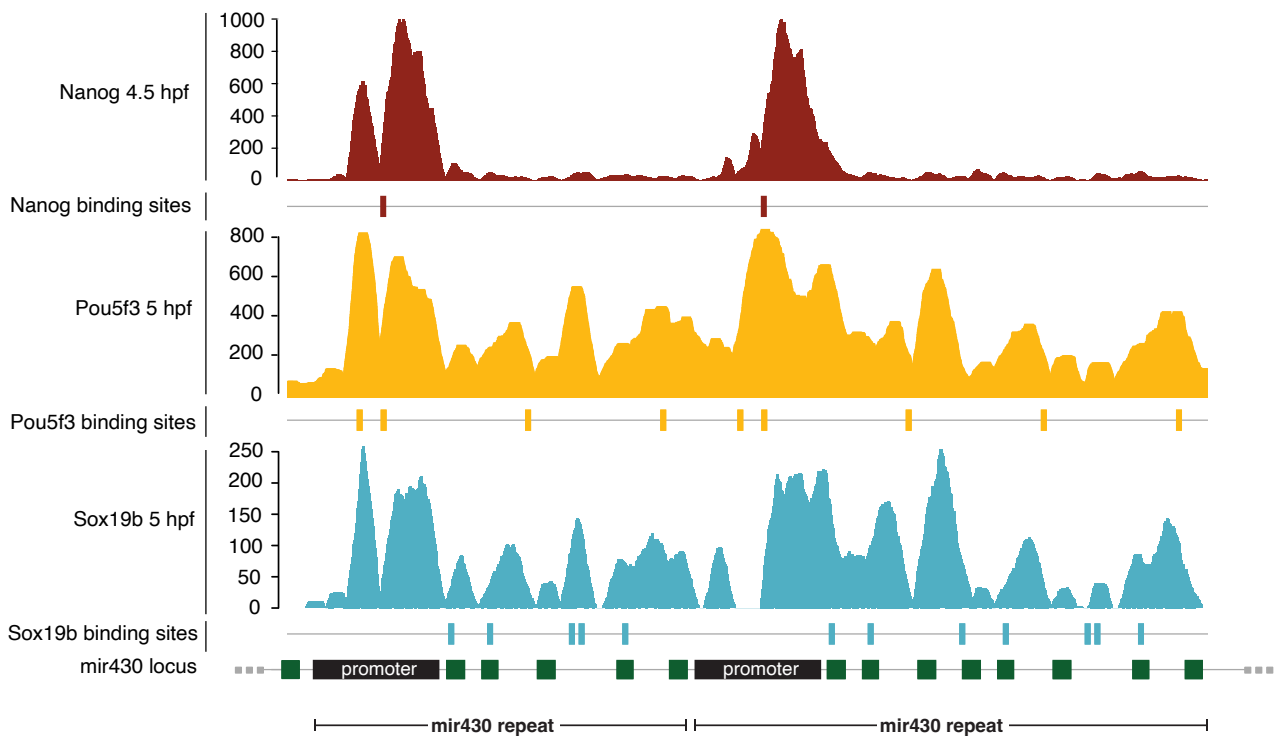


Figure 40: Nanog, Pou5f3 and Sox19b bind at the mir430 locus, as revealed by ChIP-seq experiments
 Schematics representing two mir430 gene repeats and the profile of ChIP-seq peaks for Nanog at 4.5 hpf, Pou5f3 at 5 hpf and Sox19b at 5 hpf, and the TFBS for the three transcription factors. This figure has been created by Hann Ng, PhD student in the Vastenhouw lab.

4.2 A THRESHOLD OF NANOG PROTEINS AT THE MIR430 LOCUS LEAD TO TRANSCRIPTION ACTIVATION

Several studies have shown that the quantity of TFs in cluster are related to gene activation ^{241,303,304,344}. However, these studies are mainly driving conclusions using artificially induced TF clustering or TF clustering seeding by artificial DNA array ^{241,303,304}. Moreover, none of them have suggested that a specific threshold is necessary to activate transcription. Here, we used Nanog clusters which clustering is occurring in the nucleus without the need of artificial system, to show that a certain amount of Nanog needs to be reached at a target gene, the *mir430* locus, for transcription activation. In detail, I showed that there is a high correlation between the amount of

Nanog at the *mir430* locus and transcription initiation (**Figure 19A**). My data also suggest that there is a threshold amount to reach to start transcription, around 85% of the maximum amount of Nanog that can be loaded on the *mir430* locus (**Figure 19A,E, Figure 20**). Similar results were observed labeling Nanog with a different protein (**Figure 28C**).

I quantified the volume in Nanog clusters at the moment of transcription activation, but the values show a large range until a third fold change difference (**Figure 21A****Error! Reference source not found.**). Even inside the same embryo, nuclei can have a different fluorescent background (**Figure 38A**). Such differences in nuclear background can arise from heterogenous distribution of *nanog* mRNA in the embryo or can be related to the imaging itself. To be able to compare the volume of *mir430* Nanog clusters imaged in identical conditions, I have to select *mir430* Nanog clusters that were imaged at the same time in the same nucleus. To check if a specific threshold of Nanog is necessary at the *mir430* locus, I compared the volume of the two *mir430* Nanog clusters at the moment of transcription activation. If the two *mir430* Nanog clusters have the same volume when transcription initiates at the two *mir430* alleles in the same nucleus, the difference of volume should be in theory equal to 0 (**Figure 41A**, intranuclear comparison, blue arrows). To be able to compare the results to a random distribution and test for significance, I also calculated the difference of volume between pairs of *mir430* nanog clusters at the moment of transcription activation, but not belonging to the same nucleus (**Figure 41A**, internuclear comparison, yellow arrows). The difference between the two distributions is not significant, suggesting that the volume of *mir430* Nanog clusters at the moment of activation are not closer than the ones of two *mir430* Nanog clusters from two random nuclei (**Figure 41B**, $p=0.44$).

This result is in contradiction with a specific threshold of Nanog to be reached at the *mir430* locus, to activate transcription.

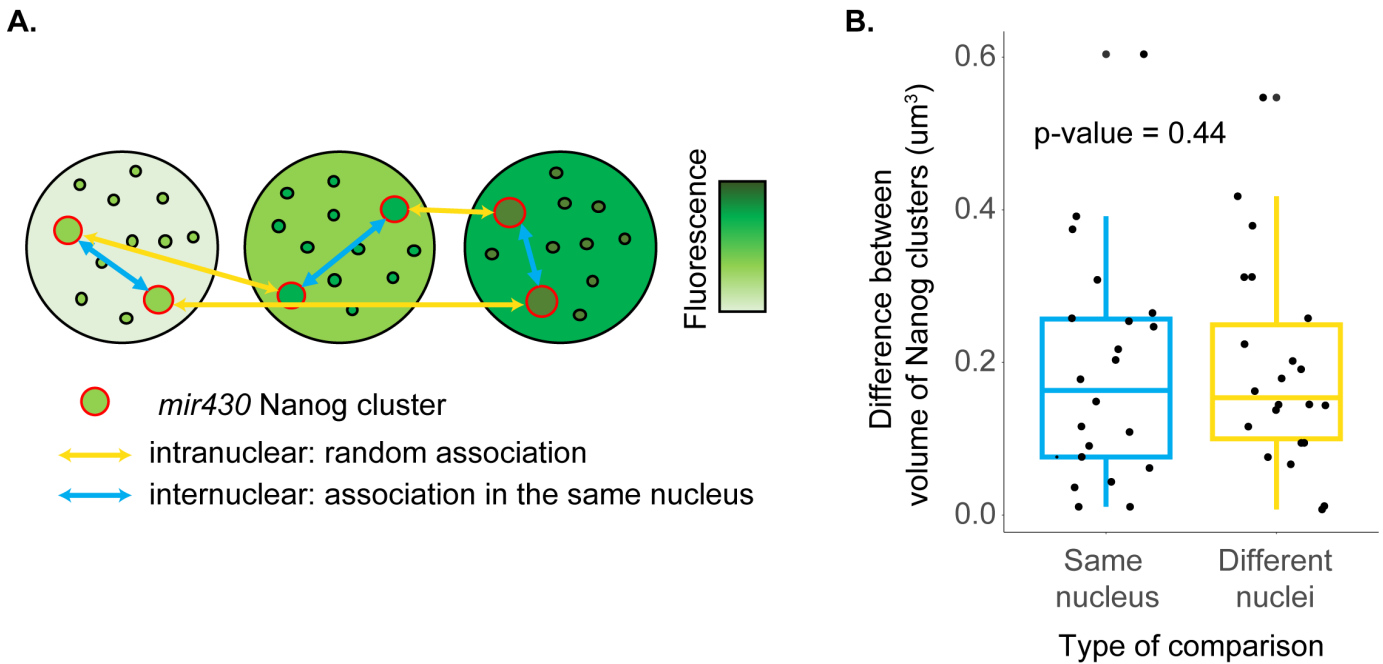


Figure 41: Volume comparison of the two *mir430* Nanog clusters inside the same nucleus versus other nuclei.

A. Schematics explaining the method to compare the volume of *mir430* Nanog clusters inside the nucleus versus other nuclei. Different nuclei have different nuclear backgrounds as represented by the three green shades. The two *mir430* Nanog clusters are represented by the two larger green dots in each nucleus, circled in red. The volumes of the two *mir430* Nanog clusters are subtracted to obtain the difference of volume and reported on the boxplot (blue arrows). The control population is represented by the difference of volume between two *mir430* Nanog clusters randomly paired (yellow arrows) and then reported to the boxplot (B). **B.** Boxplots representing the intranuclear and internuclear volume difference (μm^3) between the *mir430* Nanog clusters from the same nucleus (yellow) or paired randomly between different nuclei (blue). P-value=0.44, Student t.test.

On the other hand, I have shown that the amount of Nanog is correlated with the amount of mir430 DNA (**Figure 30B, C**). However, the underlying structure of the *mir430* locus is not completely known and under debate. Indeed, the *mir430* locus contains the sequence of repeated mir430 genes^{298,301}. As other repeated sequences, the assembly of such long and repetitive part of the genome is challenging³⁴⁵. Several assemblies have been tried, leading to always different results^{346,347} (**Figure 42**). Moreover, for these assemblies, different strains of zebrafish have been used, potentially leading to variation of the mir430 structure (**Figure 42**). We and others have tried to overpass these issues by using long-read sequences to determine the sequence of the *mir430* locus. Among the two published studies, one used PacBio HiFi long-read sequencing and determined that the *mir430* locus is 550 kb long²⁹⁹. The second study could assemble two different contigs of different lengths: one of 577 kb and one of 811 kb, using Nanopore sequencing. They suggest that the two contigs are allelic variants of the mir430 locus³⁰¹. Hann Ng, PhD student in my lab, has determined that the *mir430* locus is around 150 kb long using PacBio HiFi sequencing and composed of at least two different chunks separated by other sequences (unpublished data). Work from Hann Ng also showed that the number of mir430 repeat unit is different in different strains (unpublished data). Altogether, these different studies show that the structure of the mir430 locus is complex, potentially different for different strains and at different alleles.

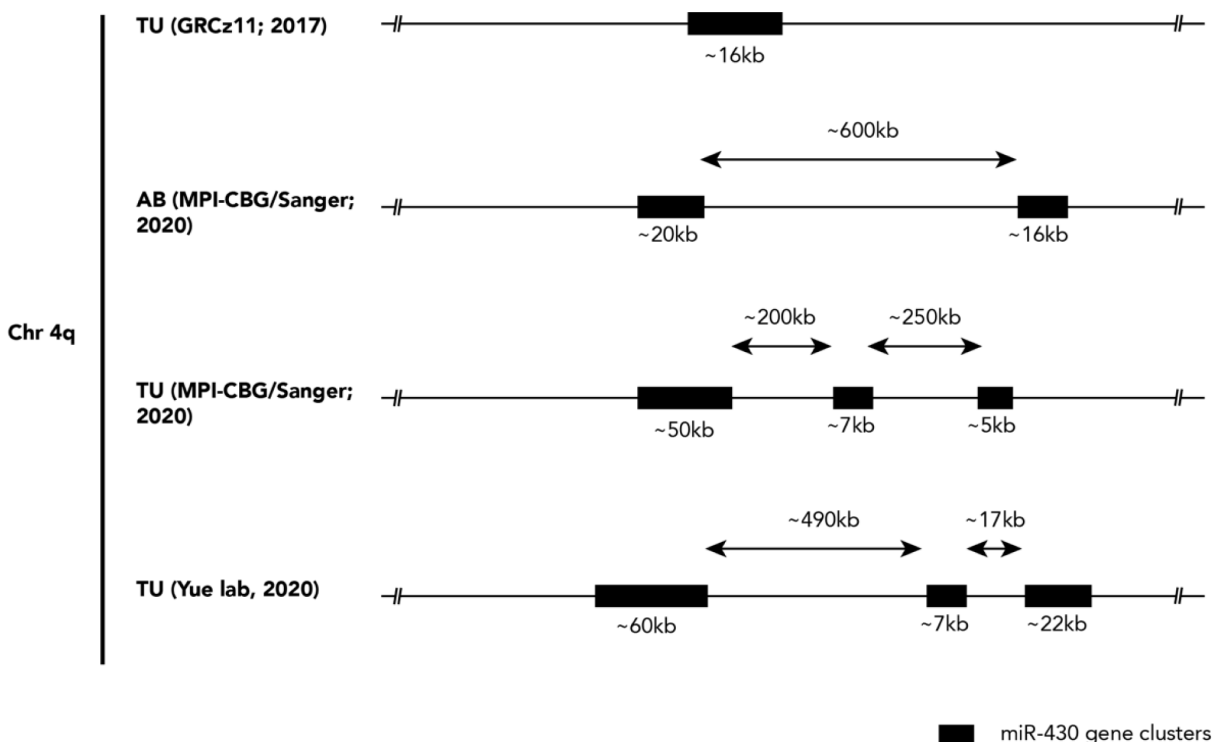


Figure 42: mir430 locus structure across different genome assembly. Four different mir430 locus structures are represented, resulting from four different genome assembly of the long arm of the chromosome 4 in zebrafish. Each assembly was built upon sequencing reads obtained from different strains, as indicated in front of each black line. Each region containing mir430 gene repeats are represented with black boxes. In-between regions not containing mir430 gene repeats are represented with black arrows. The detection of the *mir430* locus in these different genome assemblies and their representations was made by Hann Shen Ng, from the Vastenhouw lab.

Potentially, the size of the underlying *mir430* locus could control how much Nanog can bind there, as suggested by the maximum of Nanog loaded at the mir430 Nanog clusters (**Figure 20**). It could be because the two *mir430* locus in one nucleus have different Nanog thresholds based on their size.

In this thesis, quantification of Nanog clusters has been performed using imaging of nanog $-/-$ ABTL embryos. These fish originated from the crossing of the AB and TL WT strains. Therefore, it is possible that the underlying *mir430* structure is different inside single embryos. Altogether, these experiments and published work suggest that even if I could not quantify a specific volume correlated with *mir430* activation, the presence of a critical threshold cannot be excluded. Indeed, my results suggest that the amount of Nanog is correlated to the amount of *mir430* DNA (**Figure 30B, C**). As the *mir430* locus structure is unclear and could vary from embryo to embryo, it is possible that a specific Nanog threshold exist at *mir430* allele.

Finally, showing that there is a specific threshold is difficult without knowing exactly how many TFs are in the *mir430* Nanog clusters. Some advanced microscopy techniques like Fluorescence Correlative Spectroscopy, Raster image correlation spectroscopy or single-molecule super resolution imaging can allow to count the number of a specific protein in a small volume ^{348,349}.

On other hand, talking about the threshold of a specific factor for gene activation is reductive when we consider how many other factors are important for gene transcription. Indeed, TF can work in a cooperative way ⁷⁰. The amount of a certain TF could then impact greatly the interactions with the other TF and modulate the transcription activation process. In zebrafish, other factors are known to make clusters at the *mir430* locus like Sox19b ²⁷⁰, Brd4v²⁸³, Cdk9 ²⁹⁷, FoxH1 (Shivali Dongre, unpublished data), Pou5f3 (Shivali Dongre, unpublished data). However, three transcription factors are important to activate *mir430* transcription ²⁶⁹, but the loss

Nanog has the largest effect on *mir430* transcription ²⁶⁹. Moreover, the *mir430* transcription is totally abolished in *nanog* $-/-$ ²⁷⁰. These studies suggest that Nanog is the main activator of the *mir430* locus, potentially making its concentration at the *mir430* locus more important than other activators.

4.3 OVERLAPPING IS NOT COLOCALIZATION

I have shown that the *mir430* locus is overlapping by at least one voxel in 100% of the cases at transcription start but that before transcription initiation, these levels drop to a minimum of 75% (**Figure 27C**). This decrease in colocalization is due to the fading off, leading sometimes complete disappearance, of Nanog and *mir430* DNA signals during telophase, potentially related to the unbounding of Nanog and MCP-mNG during DNA relaxation (**Figure 28**). Moreover, Nanog and *mir430* DNA signal show a high correlation score, except during the moment of the cell cycle when the two signal are fading off, sometimes completely disappear (**Figure 27B**). These results altogether suggest that the signals for Nanog and the *mir430* locus DNA colocalize, and that the Nanog clusters colocalizing at the *mir430* locus are seeding by the locus itself.

However, I must take in consideration that *mir430* DNA and Nanog signals are not segmented with the same image analysis algorithm. For the *mir430* DNA, I used the “surface algorithm” available in Imaris, while for Nanog clusters, I used the “spot algorithm”. I chose these algorithms because they gave the most convincing results visually after segmentation. Indeed, the *mir430* DNA signal is labeling a chromatin structure, which shape is often variable. On the other hand, the Nanog clusters are forming a “spotty” pattern in the nucleus and the “spot detection algorithm” was giving better results for such pattern. Moreover, the shape algorithm was displaying unconvincing results while detected the merging Nanog clusters, as they were considered as the same structure by the software. Nevertheless, using the Nanog spot algorithm results is “forcing” the Nanog signal mask to have a spot shape. By consequence, small portions of Nanog clusters might not be including in the mask. This could potentially result in changes of the proportion of colocalizing between Nanog and *mir430* DNA masks.

Another point to take in consideration is the possible spatial shift between the channels, or registration. Both *mir430* and Nanog signals were acquired simultaneously by two different cameras. Even if I regularly used a calibration slide to align the field of view of the camera at the software level, it is possible that some channel shifts remain. As the Nanog clusters are small, close to the theoretical PSF size with green wavelength, such shifts can have big consequences on colocalization analysis. By consequence, the shift correction at the camera level is doubled with a registration correction in the image pre-processing steps, if needed. To correct for registration, I use the nucleus shape in both channels and then calculate the average offset over time in X and Y of both nuclei. Then after registration, the two images are visually inspected to check for proper correction. Even with this careful correction, I cannot exclude that some shifts between the two images in the different channels remain, impacting the colocalization analysis presented in my work.

Finally, I would like to signify that I am using the word ‘colocalization’ in my work for referring to the overlap of both Nanog and *mir430* DNA signals. However, while in publications, the word “colocalization” refers mostly to cases where we compare overlapping between signals in different channels, it is not a real colocalization analysis as intended in image analysis³⁵⁰. Indeed, this term implies that the two structures are exactly at the same place at the molecular level. However, the PSF sizes of Nanog clusters using mNG or HaloTag (JF646) as fluorescent tags are close from the size of the Nanog clusters themselves. A deconvolution analysis could reveal that Nanog and *mir430* DNA signals are localized at different places when correction for light diffraction. Moreover, real colocalization analysis requires extensive correlation analysis, as well as the use of multiple control signals³⁵⁰.

Therefore, to really confirm that Nanog and the *mir430* locus signal are colocalizing, as it is intending in the field³⁵⁰, super-resolution microscopy or FRET could be used. Moreover, showing colocalization between Nanog and *mir430* signals suggest but do not confirm that the merging Nanog clusters are seeding by the *mir430* locus. Experiment using Nanog with mutations defective for DNA binding could show whether Nanog clusters still form at the *mir430* locus. The best approach remains to check if the Nanog clusters are still forming in *mir430* $-/-$ fish. The question then is how to discriminate between the Nanog clusters that were at the *mir430* locus from the others, as I am using RNA Pol II S5P or the *mir430* transcripts to differentiate them from the Nanog clusters. As Nanog HaloTag (+ dye) is only forming two clusters

at the *mir430* locus, when injected what is necessary for labelling mir430 DNA signal, the absence of any Nanog clusters in the *mir430* $-/-$ could confirm that the *mir430* Nanog clusters are indeed seeded by the *mir430* locus.

4.4. IS THE MERGING OF NANOG CLUSTERS ESSENTIAL FOR MIR430 ACTIVATION ?

We observed that in 88% of the cases, the merging Nanog clusters, both seeding by the *mir430* locus, are merged at transcription start (**Figure 22A**). However, in 12% of the cases, I showed that transcription started even if the Nanog clusters were not merged (**Figure 22A**). It raises the question of how important the merging of Nanog clusters for the *mir430* locus activation is.

I showed that transcription start at the *mir430* locus when the *mir430* Nanog clusters reach on average 85% of its maximal size (**Figure 19A**). Unmerged clusters by themselves cannot reach such high level of Nanog (**Figure 23D**), except if one of the subcluster gather most of the Nanog proteins (**Figure 23C**). However, the merging of small Nanog clusters leads to a large increase in Nanog volume (**Figure 23A**), that permit to reach the volume required for activation (**Figure 23C**). These results suggest that the merging of smaller Nanog clusters is important for *mir430* activation.

What is then happening for the cases where transcription starts but the smaller Nanog clusters did not merged? I showed that unmerged Nanog clusters are of different sizes: one smaller and one larger Nanog clusters. The difference in size between the smallest and the largest cluster is quite important (**Figure 23C, E**). Interestingly, the largest of the two unmerged clusters is the closest from the transcription body (**Figure 23D**). These results suggest that, even if the Nanog clusters are not merged, one of them is large enough to lead to transcription activation. The merging of Nanog clusters is then not necessary but is important for increasing locally Nanog levels and drive transcription.

The compaction level of the *mir430* locus is correlated with transcription activation (**Figure 28C**) and happens at the same time as the merging of Nanog clusters (**Figure 28C, Figure 21B**). As Nanog is not important for the *mir430* locus compaction (**Figure 32**), I can hypothesize that the *mir430* locus compaction

movement is the motor driving the merging of Nanog clusters. In the case where the level of compaction prior to transcription start is high (the *mir430* locus start from a more extended conformation to a compacted state), the correlation with Nanog levels is more important than if the level of compaction prior to transcription activation is low (**Figure 30A**). Moreover, I have shown that the Nanog and *mir430* DNA signals positively correlates, either at the level of the larger merged clusters or at unmerged clusters (**Figure 30B, C**). These results show that more the compaction movement is important, more Nanog is enriched at the *mir430* locus, suggesting that *mir430* compaction leads to an increase in Nanog enrichment at the *mir430* locus and in agreement with the hypothesis. The results altogether suggest that *mir430* locus, by seeding the Nanog clusters, is driving their merging by DNA compaction. This indicates that merging of smaller Nanog clusters might not be the most important for *mir430* transcription, but the general increase in local concentration of Nanog TF that the *mir430* locus creates while compacting. To go beyond correlation, I would need to inhibit *mir430* compaction, and to do so, to determine the mechanism of compaction.

4.5. WHAT IS THE MECHANISM FOR *MIR430* LOCUS COMPACTION ?

4.5.1. NANOG IS NOT IMPORTANT FOR *MIR430* LOCUS COMPACTION BUT HELP TO MAINTAIN THE LOCUS COMPACTED

By studying the *mir430* locus compaction in the absence and presence of Nanog, we have shown that the *mir430* locus compaction is different between both conditions (**Figure 32C**). This result suggests that Nanog is not important in the *mir430* compaction process. However, we have shown that in absence of Nanog, the *mir430* locus is more extended, decompacts more often and spend less time in a compacted state (**Figure 32A, D, Tableau 3**). These results suggest that the Nanog has a role in keeping the *mir430* compacted, like a “glue”. Nanog is binding at the *mir430* locus^{271,343} (**Figure 27**) and contains IDRs²⁷⁰. In turn, TFs and co-activators are interacting with through IDRs^{70,318}. In TF or co-activator clusters, it was shown

that the formation of a cluster helps to bring promoters, enhancers, or co-regulated genes in the same local area, facilitating transcription^{201,225,231,329}. Moreover, the effects of clustering TF leads to decrease in diffusion inside the clusters⁶⁹. Potentially, the Nanog proteins clustering at the *mir430* locus, interacting at the same time with the *mir430* DNA, Nanog proteins and other TFs or co-activators, could help in maintaining the locus in a compacted state by IDR-IDR interactions. However, this force is not sufficient to pull *mir430* DNA in a local space (**Figure 32C**).

4.5.2. DNA LOOP EXTRUSION AS A WAY TO COMPACT THE MIR430 LOCUS

One mechanism that leads to DNA compaction is the DNA loops extrusion. DNA loop extrusion is the process that compartmentalize the DNA in the nucleus, using the cohesion complex³⁵¹. Moreover, a study showed that Rad21 and CTCF contribute to ZGA³⁵². Inhibition of the protein Rad21, part of the cohesion complex³⁵¹, has shown to influence genome organization and the distribution of transcription body in the nucleus³⁵². Moreover, they have shown by ChIP-seq that Rad21 and CTCF are both binding on the *mir430* locus during ZGA. A knock-down of CTCF using a morpholino-based approach showed that *mir430* transcription is decreased when less CTCF proteins are present³⁵². However, in zebrafish, the CTCF boundaries might be less important than in other organisms, as TAD boundaries can be found without CTCF bindings³⁵³. Interestingly, the *mir430* locus show a high density of CTCF binding sites in the proximity of the *mir430* locus and Rad21 binds at the *mir430*, as shown by ChIP-seq experiments³⁵². Altogether, it suggests that the DNA loop extrusion could be a mechanism to compact the *mir430* locus. Live-imaging experiments in Rad21 or CTCF-depleted conditions would help us to understand better this hypothesis.

4.6. WHY DO WE SEE SOMETIMES ONE OR TWO CLUSTERS NANOG CLUSTERS ?

I have shown previously that Nanog clusters can merge along with the compaction of the mir430 locus (**Figure 29**). Interestingly, we can observe mostly two and until three mir430 Nanog subclusters prior to transcription start. After transcription start, when the bigger Nanog cluster exploded, we can observe until five Nanog clusters. Why in some cases we can observe only one cluster and in other cases more subclusters?

One hypothesis is the underlying DNA structure of the mir430 locus. As discussed previously, the mir430 locus structure is unknown and might even vary from nucleus to nucleus³¹⁹. However, some genome assembly of the mir430 locus and different results from long read-sequencing experiments have shown that the mir430 locus might not be composed of one straight chunk of mir430 repeats, but several chunks separated by other DNA sequences (**Figure 42**)^{299,319,346}.

In the official current genome assembly, the mir430 locus is displayed as 17 kb long with only one chunk. However, two recent genome assembly from the MPI-CBG/Sanger institute in AB and TU strains have shown that the mir430 locus in the case of the AB strain is split over two clusters of respectively 20 kb and 16 kb, separated by a 600 kb DNA chunk. In the case of the TU strain, we can observe that the mir430 locus is split in three of respectively 50 kb, 7 kb and 5 kb, separated respectively by 200 kb and 250 kb. Another TU assembly have shown that the TU strain is composed of three mir430 blocks, one of 60 kb, one of 7kb and one of 22 kb separated respectively by 490 kb and 17kb (**Figure 42**). These different genome assemblies show that the mir430 locus could be one continuous repeat of mir430 genes or split in several chunks.

Potentially, Nanog could bind on these different mir430 loci, forming distinguishable clusters, that merge to activate transcription. The fish we used for most of the imaging experiments in this thesis are performed using a mix AB/TL strain. The mir430 locus could be different for both strains, leading to a different mir430 structure inside on given embryo. Potentially, the mir430 locus in one of the strains is made of one chunk and the other one of more than one chunk.

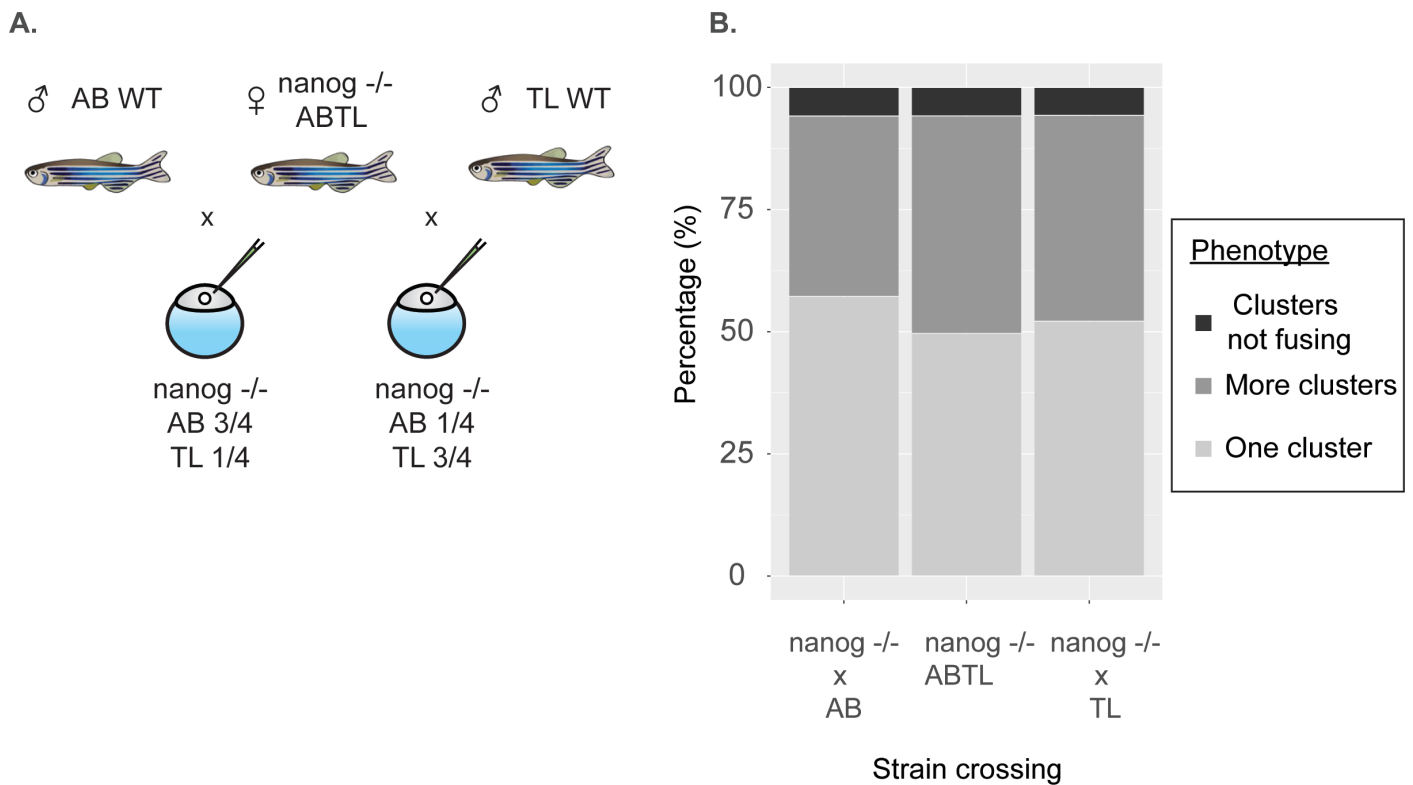


Figure 43: Crossing of different zebrafish strains does not reveal any differences in the *mir430* Nanog assembly.

A. Experimental design of the crossing between two different zebrafish crossing strains to study the *mir430* Nanog cluster assembly. A female nanog -/- with ABTL genetic background is crossed with a AB WT male or a TL WT male. The resulting embryos are depleted for maternally-loaded Nanog and their genetic background is composed of respectively $\frac{3}{4}$ of AB and $\frac{1}{4}$ TL strains for the first crossing and $\frac{1}{4}$ AB and $\frac{3}{4}$ TL strains. **B.** Histogram displaying the distribution of phenotypes of *mir430* Nanog cluster assembly depending on the embryonic genetic background resulting from the different crossing.

To confirm this, I did some preliminary experiment: I used a female nanog -/- fish with AB/TL background and crossed it with an AB or a TL male. Then, I performed the experiment the same way as previously (**Figure 15**) and quantified visually how often I can observe each phenotype. I could not observe any difference between the two types of crossing. However, I performed this experiment only twice for each crossing (**Figure 43A**). Moreover, I could not see a specific association between the phenotype if they were coming from the same embryo or not. Although the number of experiments here is only two, those results suggest that the number of clusters is at least not due to the strain specify.

Another reason for which we could observe one or more clusters is the fact that the mi430 locus 3D genome structure can be variable depending on each mir430 locus. Indeed, the mir430 locus has a lot of CTCF binding sites and Rad21 binds strongly there³⁵². We could imagine that the two clusters are just two different TADs, coming together. Potentially, the length of the mir430 locus is very different in each allele³¹⁹ and several TADs might be necessary to assemble all the mir430 locus together.

4.7. ARE NANOG CLUSTERS PHASE-SEPARATED?

Many studies have used FRAP, among other techniques, to show the dynamics state of such TF clusters and hypothesize liquid-liquid phase-separation (LLPS) as a mechanism for their formation³⁵⁴. The mean half-time of fluorescence recovery for Nanog clusters is 1.8 s (+/- 2.14 s) (**Figure 34**). This value is in the lower range compared to other proteins suggested to go in LLPS state published recently, although values can range from $t^{1/2} = <1$ s until $t^{1/2} = 384$ s²¹⁶. Thus, our work does not contradict the potential LLPS nature of Nanog clusters. However, recovery parameters such as half-time of recovery depends a lot on the experimental approach (type of bleaching, size of the bleached zone, fitting model, *in vivo* or *in vitro*...) and great variations can be observed, even for the same protein^{216,354,355}. Recent studies have warned in using such techniques, including FRAP, for concluding in LLPS as more thorough evidence, such as defining the critical concentration for nucleation of condensates^{216,355}. These experiments are complex *in vivo* and go further the present work.

5. Supplemental material

5.1. SUPPLEMENTAL TABLES

Tableau 5: Nuclei, stage and threshold values for nucleus and Nanog-mNeonGreen

Date	Nucleus	Stage	Chosen time point	Nucleus threshold	Max mean intensity	First threshold	Second threshold
16.06.21	16.06_1k_5_1_7	1k		58.60	230.00	39.5	60.00
16.06.21	16.06_1k_3_2_3	1k		55	206	34	54
16.06.21	16.06_512_4_3_4	512	4	16.80	131.00	11.5	28.50
16.06.21	16.06_512_5_1_5	1k	6	30.00	188.00	24.5	48.70
23.06.21	23.06_512_1_3_1	512		30.00	174.00	23.5	50.90
30.06.21	30.06_512_2_1_1	512	4	4.50	115.00	6	14.20
30.06.21	30.06_512_2_1_2	512	4	6.00	118.00	7	18.70
30.06.21	30.06_1k_3_1_3	1k	4	15.00	133.00	9	30.30
30.06.21	30.06_1k_3_1_5	1k	4	6.00	115.00	6.5	15.60
30.06.21	30.06_1k_3_3_5	1k	6	23.00	153.00	13	31.00
30.06.21	30.06_1k_3_3_6	1k	6	25.00	161.00	16.5	34.90
30.06.21	30.06_1k_5_2_2	1k		8	118	7.5	19.4
31.08.21	31.08_256_1_2_1	256	4	7.00	120.00	6	12.30
31.08.21	31.08_256_1_2_2	256	4	6.00	117.00	6	17.30
31.08.21	31.08_256_2_1_1	256	4	9.00	124.00	7.25	21.30
31.08.21	31.08_1k_3_1_3	1k	3	4.00	109.00	4.75	9.57
05.09.21	05.09_512_1_2_2	512	9	18.00	151.00	12.5	34.00
05.09.21	05.09_512_4_2_6	512		9.09	124	7	20.8
26.10.21	26.10_256_2_1_3	256	4	6.00	112.00	7.5	13.90
26.10.21	26.10_256_4_1_1	256	5	13.00	130.00	12	33.90
26.10.21	26.10_1K_5_2_3	1k	5	19.00	141.00	11.5	33.30

Tableau 6: Nuclei, stage and threshold values for nucleus and Nanog-HaloTag (JFX650)

Date	Stage	File	Nucleus	Mean intensity nucleus red	Nucleus threshold red	First threshold red	Second threshold red
09.01.23		554	1	168	15	8	21.1
			2	189	26.8	11	27.9
			3	129	13	7	25.3
		555	1	147	17	8	25.7
			2	135	15.8	6	21.4
			3	184	23.5	11	31.8
		560	1	184	21.1	10	31.1
			2	182	19.4	9.5	32.1
			3	184	23.5	11	31.8
11.01.23		564	1	165	14.3	10	28.8
			2	150	13.1	11	25.2
			3	170	16.4	8	30
		566	1	170	16.4	8	30
			2	180	21.7	11.5	33
			3	213	40.9	14	39.2
		570	1	162	25.2	11	29.5
			2	206	36.6	15	36.2
			3	213	40.9	14	39.2
12.01.23		571	1	117	7.06	5	23.6
			2	120	8.39	5	22.9
			3	115	6.49	4.5	15
		573	1	132	10.9	7	23.6
			2	135	13.6	7	24.1
			3	115	6.49	4.5	15
		576	1	135	12.2	6	20.2
			2	135	13.6	7	24.1
			3	115	6.49	4.5	15
			2	140	10.7	6.5	21.5

Tableau 7: Nuclei, stage and threshold values for RNA Pol II S5P signal calibration curve

Date	Nucleus	Stage	Max mean intensity	First threshold
16.06.21	16.06_1k_5_1_7	1k	115	2
16.06.21	16.06_1k_3_2_3	1k	115	2.2
16.06.21	16.06_512_4_3_4	512	107	1.45
16.06.21	16.06_512_5_1_5	1k	114	2
23.06.21	23.06_512_1_3_1	512	113	1.8
30.06.21	30.06_512_2_1_1	512	106	1.4
30.06.21	30.06_512_2_1_2	512	108	1.48
30.06.21	30.06_1k_3_1_3	1k	108	1.48
30.06.21	30.06_1k_3_1_5	1k	112	1.75
30.06.21	30.06_1k_3_3_5	1k	112	1.75
30.06.21	30.06_1k_3_3_6	1k	112	1.82
30.06.21	30.06_1k_5_2_2	1k	109	1.6
31.08.21	31.08_256_1_2_1	256	107	1.4
31.08.21	31.08_256_1_2_2	256	106	1.35

31.08.21	31.08_256_2_1_1	256	136	3
31.08.21	31.08_1k_3_1_3	1k	119	2.2
05.09.21	05.09_512_1_2_2	512	124	2.4
05.09.21	05.09_512_4_2_6	512	123	2.5
26.10.21	26.10_256_2_1_3	256	115	2.2
26.10.21	26.10_256_4_1_1	256	118	2.2
26.10.21	26.10_1K_5_2_3	1k	120	2.3

Tableau 8: Nuclei, stage and threshold values for miR430 RNA signal calibration curve

Date	Stage	File	Nucleus	Mean intensity nucleus MOVIE	Shape threshold MOVIE
09.01.23		554	1	132	4.6
			2	132	6.63
			3	132	6.63
		555	1	125	4.299
			2	117	3.94
			3	115	5.29
		560	1	118	4.46
			2	117	4.69
			3	117	4.5
11.01.23		564	1	124	4.83
			2	118	5.23
			3	118	5.23
		566	1	119	5.23
			2	124	5.5
			3	124	5.5
		570	1	120	5.21
			2	130	6.92
			3	131	5.44
12.01.23		571	1	109	2.84
			2	113	3.7
			3	109	3.2
		573	1	126	5.46
			2	124	5.65
			3	124	5.65
		576	1	137	7.72
			2	142	7.2

Tableau 9: Distance between clusters and speed of compaction (all stages mixed)

All stages	Nanog	No Nanog	p-value
Average time to come together	16.7	25.05	0.00412 (**)
Average speed to come together (barycenter, um/s)	0.0465	0.0496	0.91 (n.s.)

Average speed to come together (max, um/s)	0.0436	0.0469	0.92 (n.s.)
Average distance in decompacted state (barycenter, um)	0.833	0.993	1.65E-05 (***)
Average distance in decompacted state from (max, um)	0.800	0.927	3.47E-05 (***)

Tableau 10: Statistics about compacted state (all stages mixed)

All stages	Nanog	No Nanog	p-value
Tracks always sticking	29.8%	18.5%	/
Tracks always separated	2.30%	1.23%	/
Average number of oscillations	1.31	1.89	/
Average number of oscillations (no flat lines)	1.96	2.43	/
Average time sticked (with flats, in seconds)	123	74.4	0.0023 (**)
Average time sticked (with no flats)	89.32	58.7	0.0011 (**)

5.2. SUPPLEMENTAL FIGURES

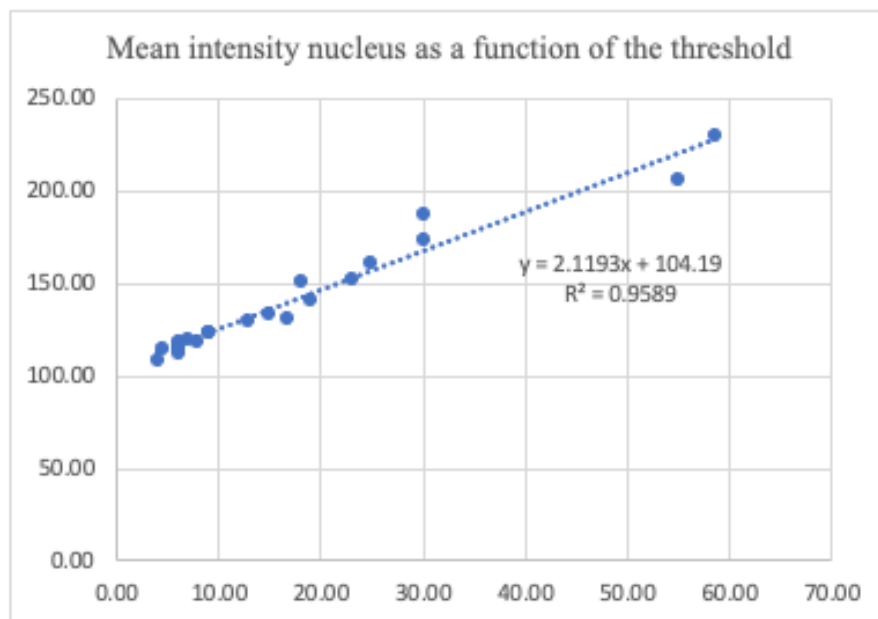


Figure Supplemental 1: Segmentation threshold as a function of mean intensity in the nucleus

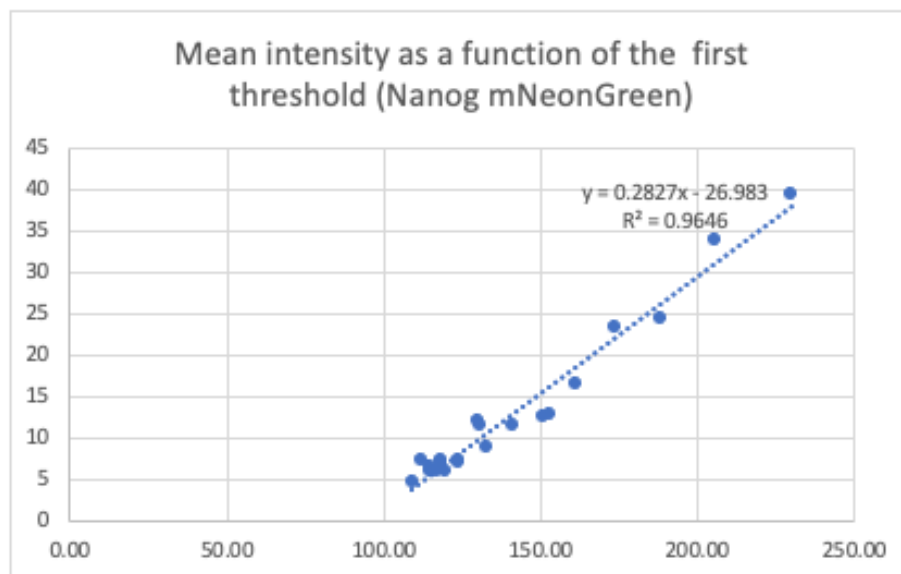


Figure Supplemental 2: First segmentation threshold as a function of nuclear mean intensity (Nanog mNeonGreen)

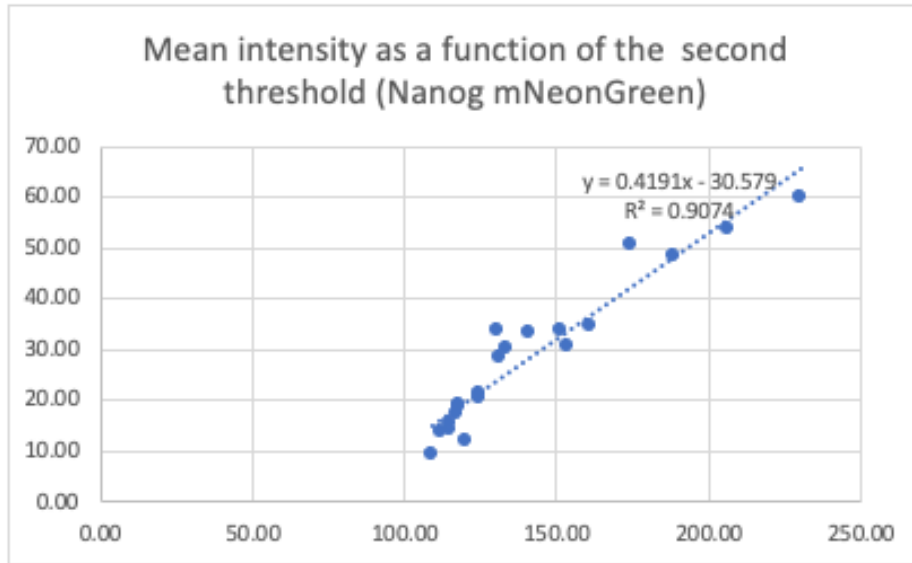


Figure Supplemental 3: Second segmentation threshold as a function of nuclear mean intensity (Nanog mNeonGreen)

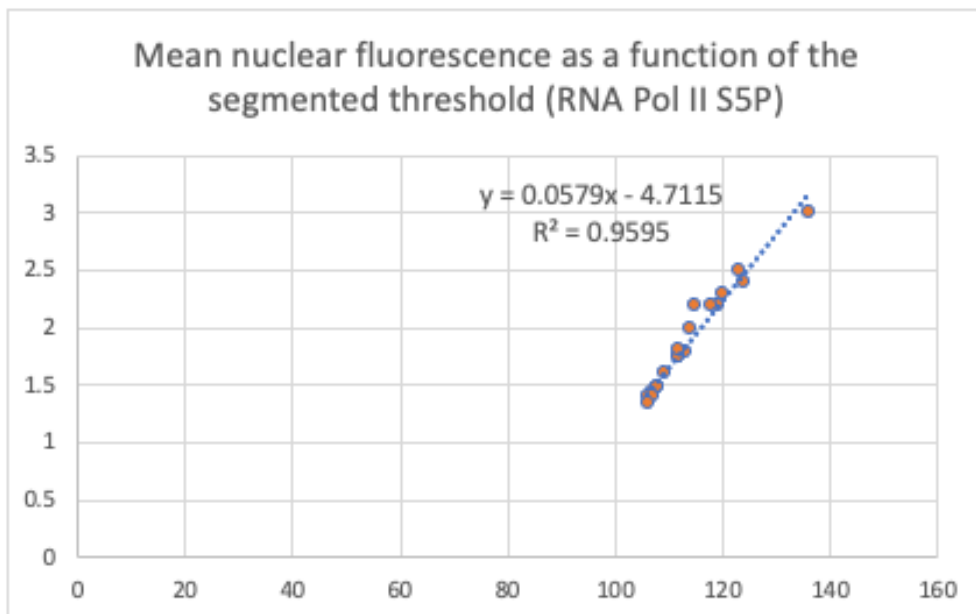


Figure Supplemental 4: Segmentation thresholds for the nucleus and Nanog clusters as a function of nuclear mean intensity (Nanog-HaloTag)

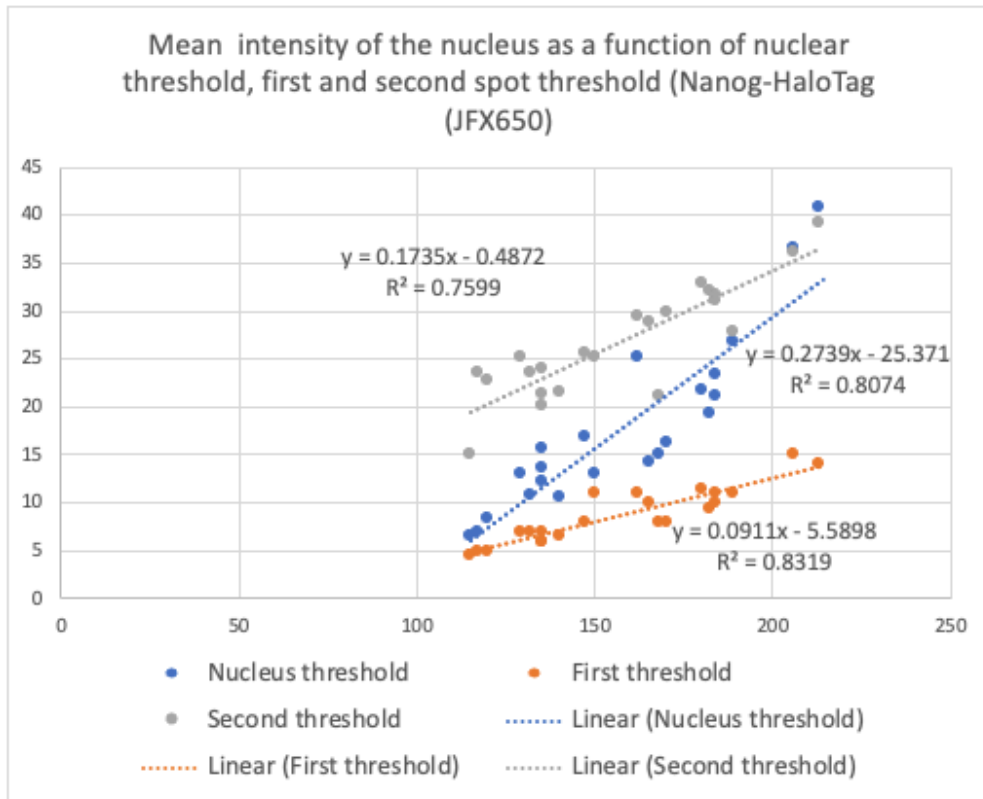


Figure Supplemental 5: Segmentation threshold as a function of nuclear mean intensity (RNA Pol II S5P signal)

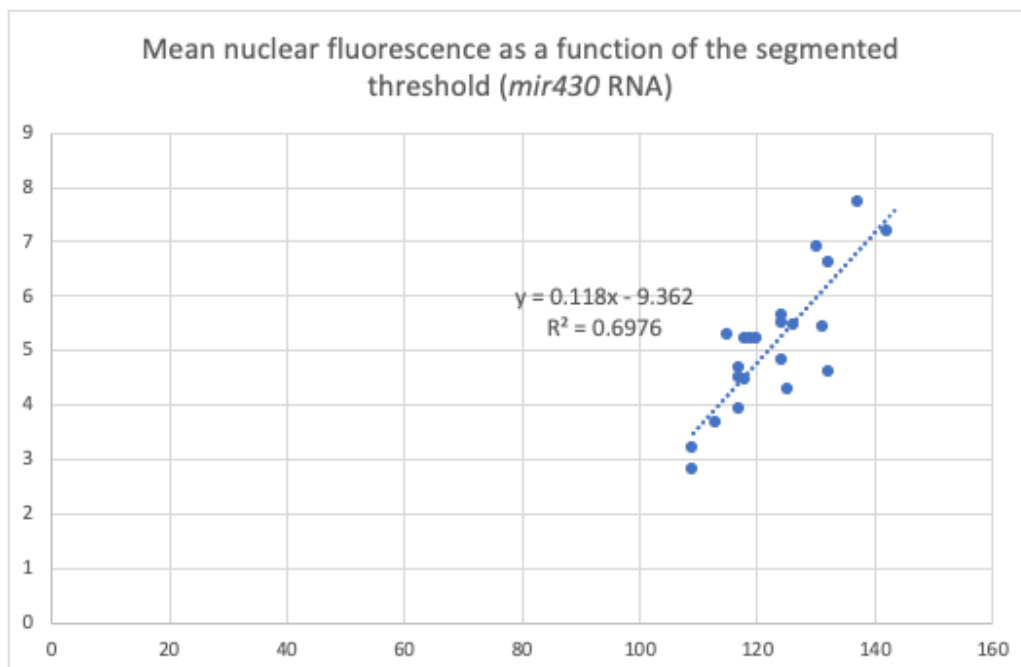


Figure Supplemental 6: Segmentation threshold as a function of nuclear mean intensity (miR430 RNA)

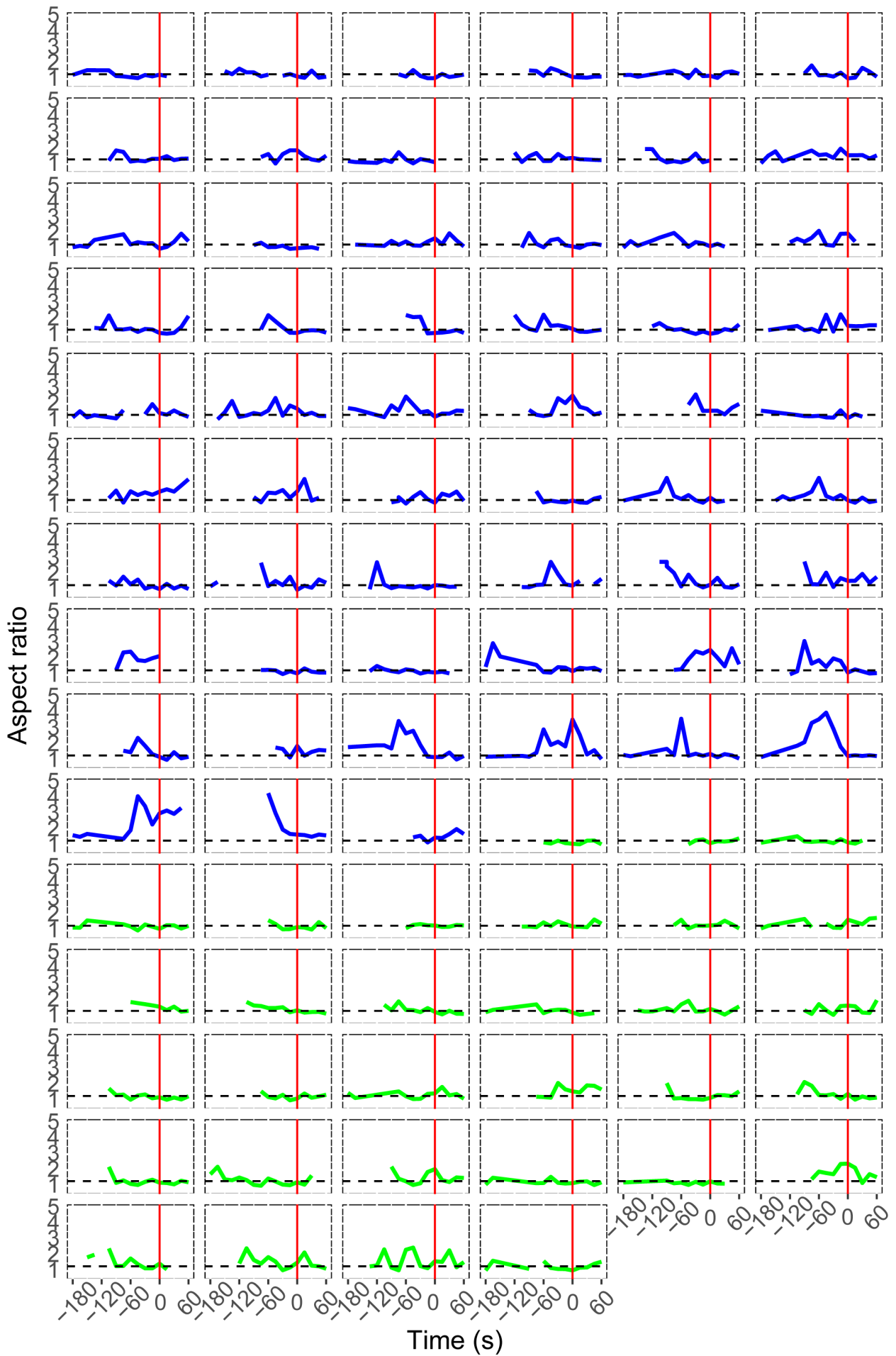


Figure Supplemental 7: Mir430 DNA mask aspect ratio as a function of time for all analyzed alleles

Evolution of the aspect ratio over time for individual *mir430* alleles. **Multiple-cluster** category is indicated in blue; **One-cluster** category is indicated in green. ‘One cluster’ indicates that only one Nanog cluster was always detected prior to transcription start. ‘More clusters’ indicates that at least two clusters were detected for at least one time point prior to transcription start. Time is relative to transcription start and range from -180 to 60 seconds. N1=4, N2=7, n=45, a=88, where N1 the number of biological replicates, N2 the number of embryos and n the number of nuclei analyzed *and* a the number of alleles.

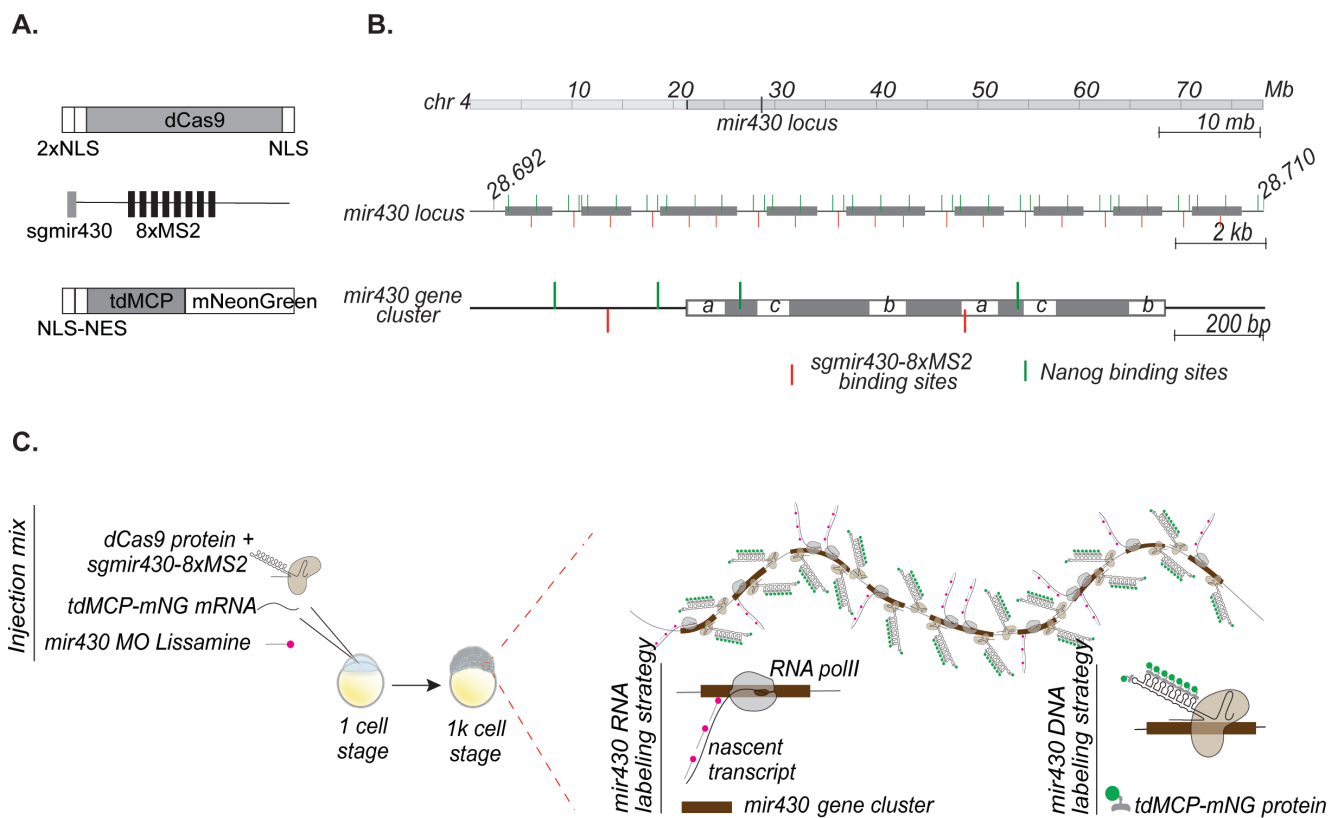


Figure Supplemental 8: Strategy for mir430 DNA visualization.

A. Schematics representing the DNA constructs for the dCas9, the sgRNA and the MCP-mNG. NLS: Nuclear Localisation Signal. MS2: Emesvirus zinderi. tdMCP: tandem dimer MS2 coat protein. NES: Nuclear export signal. B. Schematics representing the *mir430* locus position on the chromosome IV, the structure of the *mir430* locus in the current zebrafish genome assembly GRCz11, and the structure of single *mir430* gene cluster. Red bars indicates where the sgRNAs are binding, green bars indicate where the TFBS for Nanog are present. C. Schematic representing the injection mix permitting *mir430* DNA and RNA visualization live, and how this relates to the binding of the dCas9-sgRNAs on the *mir430* locus.

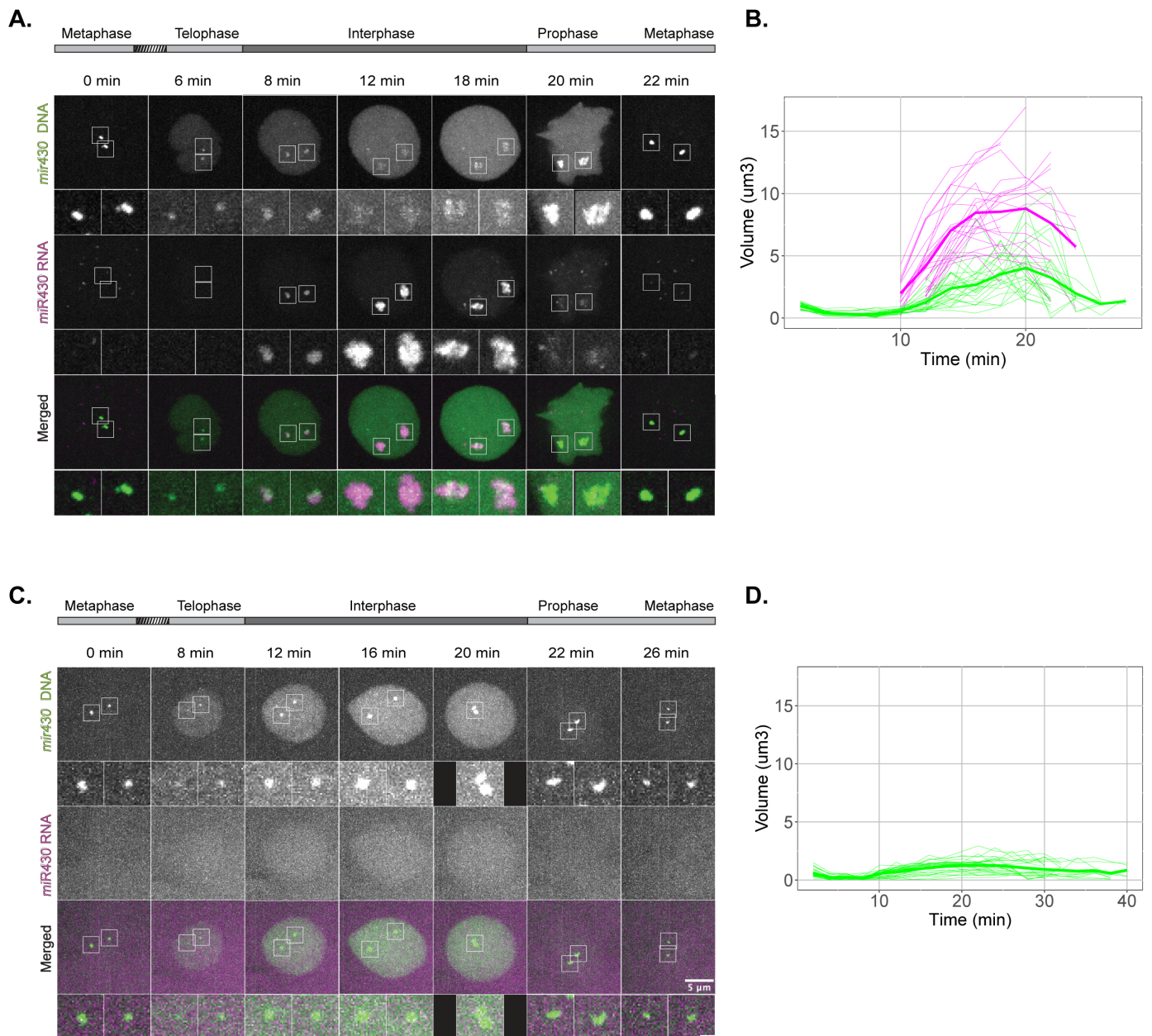


Figure Supplemental 9: Strategy for mir430 DNA visualization.

A. Images of the *mir430* locus and miR430 RNA covering the complete cell cycle at 1K-cell stage. The two *mir430* alleles in each nucleus are shown at higher magnification. Shown are representative images of individual nuclei extracted from spinning disk confocal microscopy time lapse. Scale bar represents 5 μm . **B.** Volume in μm^3 of *mir430* DNA and RNA signal in pink and magenta, respectively, as a function of time at the 1K-cell stage in untreated (left) and a-amanitin injected (right) embryos. Each line is an individual allele, bold lines represent the mean. $N=3$ and $n=13$ (untreated), and $n=14$ (a-amanitin) with N as the number of biological replicates and n the number of individual nuclei. **C.** As B but in a-amanitin injected embryos. Time points shown are different than in B because the cell cycle length is different with and without transcription. **D.** Same as B, but a-amanitin injected embryos. DNA locus is indirectly labelled by MCP-mNeonGreen proteins coupled to dCas9 bound on the *mir430* locus. MiR430 RNA is labelled by MORphilino Visualization Expression (MOVIE). Scale bar represents 5 μm .

6. Bibliography

1. Kornberg, R. D. Eukaryotic transcriptional control. *Trends Cell Biol* **9**, M46–M49 (1999).
2. Bushnell, D. A. & Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: Implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**, 6969–6973 (2003).
3. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119–2137 (2012).
4. Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev* **19**, 1401–1415 (2005).
5. Anthony Weil, P., Luse, D. S., Segall, J. & Roeder, R. G. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* **18**, 469–484 (1979).
6. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
7. Matsuis, T., Segall, J., Weill, P. A. & Roeder, R. G. Multiple Factors Required for Accurate Initiation of Transcription by Purified RNA Polymerase II*. *J Biol Chem* **255**, 11992–11996 (1980).
8. ROEDER, R. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327–335 (1996).
9. Sawadogo, M. & Roeder, R. G. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43**, 165–175 (1985).
10. Roy, R. *et al.* The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* **79**, 1093–1101 (1994).
11. Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**, 1062–1066 (2006).
12. Rougvie, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
13. Strobl, L. J. & Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO Journal* **11**, 3307–3314 (1992).
14. Tome, J. M., Tippens, N. D. & Lis, J. T. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat Genet* **50**, 1533–1541 (2018).
15. Grohmann, D. *et al.* The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol Cell* **43**, 263–274 (2011).
16. Kim, J. B. & Sharp, P. A. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *Journal of Biological Chemistry* **276**, 12317–12323 (2001).
17. Peng, J., Marshall, N. F. & Price, D. H. Identification of a cyclin subunit required for the function of *Drosophila* P-TEFb. *Journal of Biological Chemistry* **273**, 13855–13860 (1998).
18. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics* **2014** *15*:3 **15**, 163–175 (2014).

19. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**, 1616–1625 (2012).
20. Lima, S. A. *et al.* Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* **24**, 1057–1063 (2017).
21. Kühn, U. *et al.* Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. *J Biol Chem* **284**, 22803 (2009).
22. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
23. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110–1122 (2019).
24. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organisation of the genome. *Nat Rev Mol Cell Biol* **20**, 327 (2019).
25. Krivega, I. & Dean, A. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* **22**, 79–85 (2012).
26. Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18**, 956–963 (2011).
27. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 2011 474:7351 **474**, 390–394 (2011).
28. Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**, 26–41 (2018).
29. Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem* **89**, 213–234 (2020).
30. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development* vol. 43 73–81 Preprint at <https://doi.org/10.1016/j.gde.2016.12.007> (2017).
31. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
32. Wurmser, A. & Basu, S. Enhancer-Promoter Communication: It's Not Just About Contact. *Front Mol Biosci* **9**, (2022).
33. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
34. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology* 2016 17:3 **17**, 183–193 (2016).
35. Zabidi, M. A. & Stark, A. Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet* **32**, 801–814 (2016).
36. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: A Database on Transcription Factors and Their DNA Binding Sites. *Nucleic Acids Res* **24**, 238–241 (1996).
37. Johnson, P. F. & McKnight, S. L. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* **58**, 799–839 (1989).
38. Weirauch, M. T. & Hughes, T. R. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell Biochem* **52**, 25–73 (2011).
39. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
40. Nagy, G. & Nagy, L. Motif grammar: The basis of the language of gene expression. *Comput Struct Biotechnol J* **18**, 2026–2032 (2020).
41. Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* **43**, 110–119 (2017).

42. Rodda, D. J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731–24737 (2005).
43. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
44. Dror, I., Golan, T., Levy, C., Rohs, R. & Mandel-Gutfreund, Y. A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res* **25**, 1268–1280 (2015).
45. Brodsky, S., Jana, T. & Barkai, N. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Curr Opin Struct Biol* **71**, 110–115 (2021).
46. Ann Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains In Brief Activation domains from a diverse array of mammalian and yeast transcription factors form phase-separated condensates with Mediator to activate gene expression. *Cell* **175**, 1842-1855.e16 (2018).
47. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 2019 572:7770 **572**, 543–548 (2019).
48. Zamudio, A. V *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes Molecular Cell Article Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol Cell* **76**, 753-766.e6 (2019).
49. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, (2018).
50. Keegan, L., Gill, G. & Ptashne, M. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**, 699–704 (1986).
51. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252–263 (2009).
52. Wang, C., Uversky, V. N. & Kurgan, L. Disordered nucleome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* **16**, 1486–1498 (2016).
53. Burdach, J. *et al.* Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. *Nucleic Acids Res* **42**, 276 (2014).
54. Völkel, S. *et al.* Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3. *PLoS Genet* **11**, (2015).
55. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell* **79**, 459-471.e4 (2020).
56. McNally, J. C., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**, 1262–1265 (2000).
57. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
58. Mahmutovic, A., Berg, O. G. & Elf, J. What matters for lac repressor search in vivo—sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res* **43**, 3454 (2015).
59. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* **32**, 3040 (2004).
60. Gowers, D. M. & Halford, S. E. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO Journal* **22**, 1410–1418 (2003).

61. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
62. Mazzocca, M., Fillot, T., Loffreda, A., Gnani, D. & Mazza, D. The needle and the haystack: single molecule tracking to probe the transcription factor search in eukaryotes. *Biochem Soc Trans* **49**, 1121 (2021).
63. Stracy, M. *et al.* Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. *Mol Cell* **81**, 1499-1514.e6 (2021).
64. Presman, D. M. *et al.* Quantifying transcription factor binding dynamics at the single-molecule level in live cells. *Methods* **123**, 76–88 (2017).
65. Morisaki, T., Müller, W. G., Golob, N., Mazza, D. & McNally, J. G. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nature Communications* *2014 5:1* **5**, 1–8 (2014).
66. Lu, F. & Lionnet, T. Transcription Factor Dynamics. *Cold Spring Harb Perspect Biol* **13**, (2021).
67. Garcia, D. A. *et al.* Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. *Nucleic Acids Res* **49**, 6605–6620 (2021).
68. Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat Chem Biol* **16**, 257 (2020).
69. Garcia, D. A. *et al.* An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Mol Cell* **81**, 1484-1498.e6 (2021).
70. Panne, D. The enhanceosome. *Curr Opin Struct Biol* **18**, 236–242 (2008).
71. Fong, A. P. & Tapscott, S. J. Skeletal muscle programming and re-programming. *Curr Opin Genet Dev* **23**, 568–573 (2013).
72. Afzal, Z. & Krumlauf, R. Transcriptional Regulation and Implications for Controlling Hox Gene Expression. *J Dev Biol* **10**, (2022).
73. Papatsenko, D. Stripe formation in the early fly embryo: principles, models, and networks. *Bioessays* **31**, 1172–1180 (2009).
74. Singh, H., Khan, A. A. & Dinner, A. R. Gene regulatory networks in the immune system. *Trends Immunol* **35**, 211–218 (2014).
75. Sever, R. & Glass, C. K. Signaling by Nuclear Receptors. *Cold Spring Harb Perspect Biol* **5**, (2013).
76. Akerblom, I. E., Slater, E. P., Beato, M., Baxter, J. D. & Mellon, P. L. Negative Regulation by Glucocorticoids Through Interference with a cAMP Responsive Enhancer. *Science (1979)* **241**, 350–353 (1988).
77. Frieze, S. & Farnham, P. J. Transcription factor effector domains. *Subcell Biochem* **52**, 261–277 (2011).
78. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol* **47**, 1–8 (2017).
79. Schmitges, F. W. *et al.* Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res* **26**, 1742–1752 (2016).
80. Kelleher, R. J., Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61**, 1209–1215 (1990).
81. Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).

82. Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci U S A* **93**, 8329–8333 (1996).
83. Stevens, J. L. *et al.* Transcription control by E1A and MAP kinase pathway via Sur2 Mediator subunit. *Science (1979)* **296**, 755–758 (2002).
84. Ito, M., Yuan, C. X., Okano, H. J., Darnell, R. B. & Roeder, R. G. Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. *Mol Cell* **5**, 683–693 (2000).
85. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155–166 (2015).
86. Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**, 235–239 (2005).
87. Nozawa, K., Schneider, T. R. & Cramer, P. Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* **545**, 248–251 (2017).
88. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376–380 (2015).
89. Tsai, K. L. *et al.* Mediator structure and rearrangements required for holoenzyme formation. *Nature 2017 544:7649* **544**, 196–201 (2017).
90. Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M. & Kornberg, R. D. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* **283**, 985–987 (1999).
91. Bernecky, C., Grob, P., Ebmeier, C. C., Nogales, E. & Taatjes, D. J. Molecular Architecture of the Human Mediator–RNA Polymerase II–TFIIF Assembly. *PLoS Biol* **9**, e1000603 (2011).
92. Davis, J. A., Takagi, Y., Kornberg, R. D. & Asturias, F. J. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell* **10**, 409–415 (2002).
93. Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
94. Myers, L. C., Gustafsson, C. M., Hayashibara, K. C., Brown, P. O. & Kornberg, R. D. Mediator protein mutations that selectively abolish activated transcription. *Proc Natl Acad Sci U S A* **96**, 67–72 (1999).
95. Taatjes, D. J., Näär, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science (1979)* **295**, 1058–1062 (2002).
96. Myers, L. C. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev* **12**, 45 (1998).
97. Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361–1375 (1993).
98. Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nat Struct Mol Biol* **17**, 753–760 (2010).
99. Tóth-Petróczy, Á. *et al.* Malleable Machines in Transcription Regulation: The Mediator Complex. *PLoS Comput Biol* **4**, e1000243 (2008).
100. Krivega, I. & Dean, A. Enhancer and promoter interactions — long distance calls. *Curr Opin Genet Dev* **22**, 79 (2012).
101. Dynan, W. S. & Tjian, R. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature 1985 316:6031* **316**, 774–778 (1985).

102. Su, W., Porter, S., Kustu, S. & Echols, H. DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc Natl Acad Sci U S A* **87**, 5504 (1990).
103. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
104. Kadauke, S. & Blobel, G. A. Chromatin loops in gene regulation. *Biochim Biophys Acta* **1789**, 17–25 (2009).
105. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
106. Hnisz, D. *et al.* XSuper-enhancers in the control of cell identity and disease. *Cell* **155**, 934 (2013).
107. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science (1979)* **361**, 1341–1345 (2018).
108. Kim, S. Il, Bultman, S. J., Kiefer, C. M., Dean, A. & Bresnick, E. H. BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. *Proc Natl Acad Sci U S A* **106**, 2259–2264 (2009).
109. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* *2013 494:7438* **494**, 497–501 (2013).
110. Hsieh, C. L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* **111**, 7319–7324 (2014).
111. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358–368 (2016).
112. Uyehara, C. M. & Apostolou, E. 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Rep* **42**, (2023).
113. Gross, D. S., Chowdhary, S., Anandhakumar, J. & Kainth, A. S. Chromatin. *Curr Biol* **25**, R1158–R1163 (2015).
114. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *1997 389:6648* **389**, 251–260 (1997).
115. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the ‘30-nm’ chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A* **103**, 6506–6511 (2006).
116. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**, 175–186 (2011).
117. Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C. & Tremethick, D. J. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nature Structural Biology* *2002 9:3* **9**, 172–176 (2002).
118. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421–434 (2011).
119. Villaseñor, R. & Baubec, T. Regulatory mechanisms governing chromatin organization and function. *Curr Opin Cell Biol* **70**, 10–17 (2021).
120. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem Rev* **115**, 2376–2418 (2015).
121. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
122. Cockerill, P. N. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J* **278**, 2182–2210 (2011).
123. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open

- chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213–1218 (2013).
124. Li, J., Ahn, J. H. & Wang, G. G. Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell Mol Life Sci* **76**, 2899–2916 (2019).
 125. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–318 (2007).
 126. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (1979)* **311**, 844–847 (2006).
 127. Stewart-Morgan, K. R., Petryk, N. & Groth, A. Chromatin replication and epigenetic cell memory. *Nat Cell Biol* **22**, 361–371 (2020).
 128. Choy, J. S. *et al.* DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc* **132**, 1782–1783 (2010).
 129. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
 130. Phillips, T., Education, K. S.-N. & 2008, undefined. Chromatin remodeling in eukaryotes. *qqokbetyes.com*.
 131. Längst, G. & Manlyte, L. Chromatin Remodelers: From Function to Dysfunction. *Genes 2015, Vol. 6, Pages 299-324* **6**, 299–324 (2015).
 132. Gillette, T. G. & Hill, J. A. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ Res* **116**, 1245–1253 (2015).
 133. Mizzen, C. A. *et al.* The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270 (1996).
 134. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature 1996 384:6610* **384**, 641–643 (1996).
 135. Dyson, H. J. & Wright, P. E. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *Journal of Biological Chemistry* **291**, 6714–6722 (2016).
 136. Awad, S. & Hassan, A. H. The Swi2/Snf2 bromodomain is important for the full binding and remodeling activity of the SWI/SNF complex on H3- and H4-acetylated nucleosomes. *Ann N Y Acad Sci* **1138**, 366–375 (2008).
 137. Hock, R., Furusawa, T., Ueda, T. & Bustin, M. HMG chromosomal proteins in development and disease. *Trends Cell Biol* **17**, 72–79 (2007).
 138. Merika, M. & Thanos, D. Enhanceosomes. *Curr Opin Genet Dev* **11**, 205–208 (2001).
 139. Catez, F. *et al.* Network of Dynamic Interactions between Histone H1 and High-Mobility-Group Proteins in Chromatin. *Mol Cell Biol* **24**, 4321 (2004).
 140. Rochman, M., Malicet, C. & Bustin, M. HMG5/NSBP1: a new member of the HMG protein family that affects chromatin structure and function. *Biochim Biophys Acta* **1799**, 86–92 (2010).
 141. Cramer, P. Organization and regulation of gene transcription. *Nature* vol. 573 45–54 Preprint at <https://doi.org/10.1038/s41586-019-1517-4> (2019).
 142. Cremer, T. *et al.* Chromosome territories – a functional nuclear landscape. *Curr Opin Cell Biol* **18**, 307–316 (2006).
 143. Cremer, T. & Cremer, C. *CHROMOSOME TERRITORIES, NUCLEAR ARCHITECTURE AND GENE REGULATION IN MAMMALIAN CELLS*. vol. 292 www.nature.com/reviews/genetics (2001).
 144. Croft, J. A. *et al.* Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. *Journal of Cell Biology* **145**, 1119–1131 (1999).
 145. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, (2010).

146. Walter, J., Schermelleh, L., Cremer, M., Tashiro, S. & Cremer, T. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J Cell Biol* **160**, 685 (2003).
147. Das Heterochromatin der Moose - Emil Heitz - Google Books. https://books.google.ch/books/about/Das_Heterochromatin_der_Moose.html?id=3JwgSQAACAAJ&redir_esc=y.
148. Pimpinelli, S., Gatti, M. & De Marco, A. Evidence for heterogeneity in heterochromatin of *Drosophila melanogaster*. *Nature* 1975 256:5515 **256**, 335–337 (1975).
149. Schultz' And, J. & Dobzhansky, T. H. The Relation of a Dominant Eye Color in *Drosophila Melanogaster* to the Associated Chromosome Rearrangement. *Genetics* **19**, 344 (1934).
150. Barr, M. L. & Bertram, E. G. A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis. *Nature* 1949 163:4148 **163**, 676–677 (1949).
151. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229–244 (2018).
152. Liu, J., Ali, M. & Zhou, Q. Establishment and evolution of heterochromatin. *Ann N Y Acad Sci* **1476**, 59–77 (2020).
153. Fantì, L. & Pimpinelli, S. HP1: a functionally multifaceted protein. *Curr Opin Genet Dev* **18**, 169–174 (2008).
154. Pidoux, A. L. & Allshire, R. C. The role of heterochromatin in centromere function. *Philos Trans R Soc Lond B Biol Sci* **360**, 569–579 (2005).
155. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **34**, 265–288 (2018).
156. Fillion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
157. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
158. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348–1354 (2006).
159. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nature Reviews Molecular Cell Biology* 2015 16:4 **16**, 245–257 (2015).
160. Fortin, J. P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol* **16**, 1–23 (2015).
161. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012 485:7398 **485**, 376–380 (2012).
162. Wang, Q., Sun, Q., Czajkowsky, D. M. & Shao, Z. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nature Communications* 2017 9:1 **9**, 1–8 (2018).
163. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
164. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* **36**, 3573–3599 (2017).
165. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676–684 (2015).

166. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456–E6465 (2015).
167. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics* 2019 20:8 **20**, 437–455 (2019).
168. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24**, 390–400 (2014).
169. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012 485:7398 **485**, 381–385 (2012).
170. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016 538:7624 **538**, 265–269 (2016).
171. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).
172. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
173. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
174. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
175. Krivega, I. & Dean, A. LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Res* **45**, 8255 (2017).
176. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
177. Bruter, V., Rodionova, M. D., Varlamova & Shtil. Super-Enhancers in the Regulation of Gene Transcription: General Aspects and Antitumor Targets. *Acta Naturae* **13**, 4 (2021).
178. Tang, F., Yang, Z., Tan, Y. & Li, Y. Super-enhancer function and its application in cancer targeted therapy. *NPJ Precis Oncol* **4**, (2020).
179. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
180. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
181. Downen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
182. Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nature Communications* 2018 9:1 **9**, 1–12 (2018).
183. Steensel, B. van *et al.* Localization of the glucocorticoid receptor in discrete clusters in the cell nucleus. *J Cell Sci* **108**, 3003–3011 (1995).
184. Wang, B. *et al.* Liquid–liquid phase separation in human health and diseases. *Signal Transduction and Targeted Therapy* 2021 6:1 **6**, 1–16 (2021).
185. Li, J. & Pertsinidis, A. Nanoscale nuclear environments, fine-scale 3D genome organization and transcription regulation. *Curr Opin Syst Biol* **31**, (2022).
186. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & De Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453–1465 (2002).
187. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* **25**, 833–840 (2018).
188. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science (1979)* **341**, 664–667 (2013).

189. Buckley, M. S. & Lis, J. T. Imaging RNA Polymerase II transcription sites in living cells. *Curr Opin Genet Dev* **25**, 126 (2014).
190. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412 (2018).
191. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. doi:10.1126/science.aar2555.
192. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. doi:10.1126/science.aar3958.
193. Sawicka, A. *et al.* Transcription activation depends on the length of the RNA polymerase II C-terminal domain. *EMBO J* **40**, (2021).
194. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* **25**, 833–840 (2018).
195. Mirny, L. *et al.* How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *J Phys A Math Theor* **42**, (2009).
196. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nature Reviews Molecular Cell Biology* vol. 22 653–670 Preprint at <https://doi.org/10.1038/s41580-021-00387-1> (2021).
197. Zhang, Z. *et al.* Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis Transcription initiation by eukaryotic RNA polymerase II (Pol II) requires the coordinated action of Pol II and at least six general transcription factors (GTFs; i.e. (2016) doi:10.1101/gad.285395.
198. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420–435 (2018).
199. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **18**:5 18, 285–298 (2017).
200. Van Der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chemical Reviews* vol. 114 6589–6631 Preprint at <https://doi.org/10.1021/cr400525m> (2014).
201. Li, J. *et al.* Single-gene imaging links genome topology, promoter–enhancer communication and transcription control. *Nat Struct Mol Biol* **27**, 1032–1040 (2020).
202. Stortz, M., Pecci, A., Presman, D. M. & Levi, V. Unraveling the molecular interactions involved in phase separation of glucocorticoid receptor. *BMC Biol* **18**, (2020).
203. Wang, Y. *et al.* A Prion-like Domain in Transcription Factor EBF1 Promotes Phase Separation and Enables B Cell Programming of Progenitor Chromatin. *Immunity* **53**, 1151–1167.e6 (2020).
204. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, (2018).
205. Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **181**, 1062–1079.e30 (2020).
206. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science (1979)* **357**, (2017).
207. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (1979)* **324**, 1729–1732 (2009).
208. Ditlev, J. A., Case, L. B. & Rosen, M. K. Who’s In and Who’s Out—Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666–4684 (2018).
209. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412 (2018).

210. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nature Communications* 2015 6:1 6, 1–12 (2015).
211. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* 2012 483:7389 483, 336–340 (2012).
212. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* 2017 547:7662 547, 241–245 (2017).
213. Larson, A. G. *et al.* Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* 2017 547:7662 547, 236–240 (2017).
214. Seif, E. *et al.* Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nature Communications* 2020 11:1 11, 1–19 (2020).
215. Tatavosian, R. *et al.* Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *Journal of Biological Chemistry* 294, 1451–1463 (2019).
216. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & development* vol. 33 1619–1634 Preprint at <https://doi.org/10.1101/gad.331520.119> (2019).
217. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 176, 419–434 (2019).
218. Irgen-Giorgio, S., Yoshida, S., Walling, V. & Chong, S. Fixation can change the appearance of phase separation in living cells. *Elife* 11, (2022).
219. Padeken, J. & Heun, P. Nucleolus and nuclear periphery: Velcro for heterochromatin. *Current Opinion in Cell Biology* vol. 28 54–60 Preprint at <https://doi.org/10.1016/j.ceb.2014.03.001> (2014).
220. Tzur, Y. B., Wilson, K. L. & Gruenbaum, Y. SUN-domain proteins: ‘Velcro’ that links the nucleoskeleton to the cytoskeleton. *Nature Reviews Molecular Cell Biology* 2006 7:10 7, 782–788 (2006).
221. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates In Brief Transcription-associated proteins form condensates localized at specific DNA elements. *Mol Cell* 75, 549-561.e7 (2019).
222. Hannon, C. E. & Eisen, M. B. Intrinsic protein disorder is insufficient to drive subnuclear clustering in embryonic transcription factors. *Elife* 12, (2023).
223. Shi, B. *et al.* UTX condensation underlies its tumour-suppressive activity. *Nature* 597, 726–731 (2021).
224. Wang, W. *et al.* A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Res* 50, 4917–4937 (2022).
225. Tsai, A. *et al.* Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife* 6, (2017).
226. Meeussen, J. V. W. *et al.* Transcription factor clusters enable target search but do not contribute to target gene activation. *Nucleic Acids Res* 51, 5449–5468 (2023).
227. Lin, Y. *et al.* Narrow equilibrium window for complex coacervation of tau and RNA under cellular conditions. *Elife* 8, (2019).
228. Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* 184, 207-225.e24 (2021).
229. Chowdhary, S., Kainth, A. S., Paracha, S., Gross, D. S. & Pincus, D. Inducible transcriptional condensates drive 3D genome reorganization in the heat shock response. *Mol Cell* 82, 4386-4399.e7 (2022).

230. Li, J. *et al.* Single-gene imaging links genome topology, promoter-enhancer communication and transcription control. *Nat Struct Mol Biol* **27**, 1032 (2020).
231. Nair, S. J. *et al.* Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nature Structural & Molecular Biology* **26**:3, 193–203 (2019).
232. Ma, L. *et al.* Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. *Mol Cell* **81**, 1682-1697.e7 (2021).
233. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39–58 (2014).
234. Bauer, M. & Metzler, R. Generalized Facilitated Diffusion Model for DNA-Binding Proteins with Search and Recognition States. *Biophys J* **102**, 2321 (2012).
235. Chen, Y. *et al.* Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. *Elife* **11**, (2022).
236. Wollman, A. J. M. *et al.* Transcription factor clusters regulate genes in eukaryotic cells. *Elife* **6**, (2017).
237. Kent, S. *et al.* Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. *Cell Rep* **33**, 108248 (2020).
238. de Jonge, W. J., Patel, H. P., Meeussen, J. V. W. & Lenstra, T. L. Following the tracks: How transcription factor binding dynamics control transcription. *Biophys J* **121**, 1583 (2022).
239. Schneider, N. *et al.* Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Sci Adv* **7**, (2021).
240. Wei, M. T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nature Cell Biology* **22**:10, 1187–1196 (2020).
241. Kawasaki, K. & Fukaya, T. Functional coordination between transcription factor clustering and gene activity. *Mol Cell* **83**, 1605-1622.e9 (2023).
242. Li, J. *et al.* Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178**, 491-506.e28 (2019).
243. Trojanowski, J. *et al.* Transcription activation is enhanced by multivalent interactions independent of phase separation. *Mol Cell* **82**, 1878-1893.e10 (2022).
244. Feric, M. *et al.* Mesoscale structure–function relationships in mitochondrial transcriptional condensates. *Proc Natl Acad Sci U S A* **119**, e2207303119 (2022).
245. Markaki, Y. *et al.* Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**, 6174-6192.e32 (2021).
246. Bai, S. *et al.* The maternal-to-zygotic transition revisited. *Development* **146**, (2019).
247. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
248. Golbus, M. S., Calarco, P. G. & Epstein, C. J. The effects of inhibitors of RNA synthesis (α -amanitin and actinomycin D) on preimplantation mouse embryogenesis. *Journal of Experimental Zoology* **186**, 207–216 (1973).
249. Brachet, J., Denis, H. & de Vitry, F. The effects of actinomycin D and puromycin on morphogenesis in amphibian eggs and *Acetabularia mediterranea*. *Dev Biol* **9**, 398–434 (1964).
250. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic genome activation in vertebrates. *Dev Cell* **42**, 316 (2017).
251. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).
252. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696 (1982).

253. Owens, N. D. L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep* **14**, 632–647 (2016).
254. Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. β -catenin Primes Organizer Gene Expression By Recruiting a Histone H3 Arginine 8 Methyltransferase, Prmt2. *Dev Cell* **19**, 220 (2010).
255. Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L. & Klein, P. S. An essential role for transcription before the MBT in *Xenopus laevis*. *Dev Biol* **357**, 478–491 (2011).
256. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
257. Heyn, P. *et al.* The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Rep* **6**, 285–292 (2014).
258. Wei, C., Salichos, L., Wittgrove, C. M., Rokas, A. & Patton, J. G. Transcriptome-wide analysis of small RNA expression in early zebrafish development. *RNA* **18**, 915 (2012).
259. Harvey, S. A. *et al.* Identification of the zebrafish maternal and paternal transcriptomes. *Development* **140**, 2703–2710 (2013).
260. Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**, 1328–1338 (2011).
261. Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. R. Human pre-implantation embryo development. *Development* **139**, 829 (2012).
262. Nothias', J.-Y., Miranda, M. & Depamphilis2', M. L. Uncoupling of transcription and translation during zygotic gene activation in the mouse. *EMBO J* **15**, 5715–5725 (1996).
263. Aoki, F., Worrad, D. M. & Schultz, R. M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181**, 296–307 (1997).
264. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. H. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6**, 117–131 (2004).
265. Wang, J. & Davis, R. E. Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev* **27**, 26–34 (2014).
266. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod* **68**, 651–654 (2003).
267. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–246 (2015).
268. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 2012 487:7405 **487**, 57–63 (2012).
269. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. doi:10.1038/nature12632.
270. Kuznetsova, K. *et al.* Nanog organizes transcription bodies. *Current Biology* **33**, 164–173.e5 (2023).
271. Xu, C. *et al.* Nanog-like Regulates Endoderm Formation through the Mxtx2-Nodal Pathway. *Dev Cell* **22**, 625 (2012).
272. Chan, S. H. *et al.* Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Dev Cell* **49**, 867–881.e8 (2019).
273. Joseph, S. R. *et al.* Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife* **6**, (2017).
274. Jevtić, P. & Levy, D. L. Both Nuclear Size and DNA Amount Contribute to Midblastula Transition Timing in *Xenopus laevis*. *Sci Rep* **7**, (2017).

275. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).
276. Lu, X., Li, J. M., Elemento, O., Tavazoie, S. & Wieschaus, E. F. Coupling of zygotic transcription to mitotic control at the *Drosophila* mid-blastula transition. *Development* **136**, 2101 (2009).
277. Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the *Xenopus* midblastula transition. *Proc Natl Acad Sci U S A* **112**, E1086–E1095 (2015).
278. Braun, R. E. Packaging paternal chromosomes with protamine. *Nature Genetics* **28**:1, 10–12 (2001).
279. Wu, S. F., Zhang, H. & Cairns, B. R. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res* **21**, 578–589 (2011).
280. Ortega-Recalde, O., Day, R. C., Gemmell, N. J. & Hore, T. A. Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nature Communications* **10**:1, 1–10 (2019).
281. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204–220 (2013).
282. Yu, S. H. *et al.* The histone demethylase Jmjd3 regulates zebrafish myeloid development by promoting *sp1* expression. *Biochim Biophys Acta Gene Regul Mech* **1861**, 106–116 (2018).
283. Chan, S. H., Tang, Y., Bazzini, A. A., Moreno-Mateos, M. A. & Giraldez, A. J. Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Dev Cell* **49**, 867–881.e8 (2019).
284. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**:7290, 922–926 (2010).
285. Zhu, W., Xu, X., Wang, X. & Liu, J. Reprogramming histone modification patterns to coordinate gene expression in early zebrafish embryos. *BMC Genomics* **20**, 1–13 (2019).
286. San, B. *et al.* Normal formation of a vertebrate body plan and loss of tissue maintenance in the absence of *ezh2*. *Scientific Reports* **6**:1, 1–16 (2016).
287. Laue, K., Rajshekar, S., Courtney, A. J., Lewis, Z. A. & Goll, M. G. The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nature Communications* **10**:1, 1–10 (2019).
288. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229–244 (2018).
289. Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**, 1328–1338 (2011).
290. Bhat, P. *et al.* SLAMseq resolves the kinetics of maternal and zygotic gene expression during early zebrafish embryogenesis. *Cell Rep* **42**, 112070 (2023).
291. Goll, M. G., Akdogan-Ozdilek, B. & Duval, K. L. Chromatin dynamics at the maternal to zygotic transition: recent advances from the zebrafish model. *F1000Res* **9**, (2020).
292. Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep* **24**, 1–10.e4 (2018).
293. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253–310 (1995).
294. Choi, T. Y., Choi, T. I., Lee, Y. R., Choe, S. K. & Kim, C. H. Zebrafish as an animal model for biomedical research. *Exp Mol Med* **53**, 310 (2021).

295. White, R. J. *et al.* A high-resolution mRNA expression time course of embryonic development in zebrafish. *Elife* **6**, (2017).
296. Pauli, A. *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* **22**, 577–591 (2012).
297. Ugolini, M., Kuznetsova, K., Oda, H. & Kimura, H. Transcription bodies regulate gene expression by sequestering CDK9 1. doi:10.1101/2022.11.21.517317.
298. Hadzhiev, Y. *et al.* A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. *Nature Communications* **2019 10:1** **10**, 1–14 (2019).
299. Pownall, M. E. *et al.* Chromatin expansion microscopy reveals nanoscale organization of transcription and chromatin. *Science* **381**, 92 (2023).
300. Hilbert, L. *et al.* Transcription organizes euchromatin via microphase separation. *Nature Communications* **2021 12:1** **12**, 1–12 (2021).
301. Hadzhiev, Y. *et al.* The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation. *Dev Cell* **58**, 155-170.e8 (2023).
302. Leichsenring, M., Maes, J., Moßsner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science (1979)* **341**, 1005–1009 (2013).
303. Schneider, N. *et al.* Liquid-liquid phase separation of light-inducible transcription factors increases transcription activation in mammalian cells and mice. *Sci Adv* **7**, (2021).
304. Wei, M. T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat Cell Biol* **22**, 1187–1196 (2020).
305. Zuo, L. *et al.* Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nat Commun* **12**, (2021).
306. Kuznetsova, K. *et al.* Nanog organizes transcription bodies. *Current Biology* **33**, 164-173.e5 (2023).
307. Hadzhiev, Y. *et al.* A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. *Nat Commun* **10**, (2019).
308. Kimura, H. & Yamagata, K. Visualization of epigenetic modifications in preimplantation embryos. *Methods in Molecular Biology* **1222**, 127–147 (2015).
309. Stasevich, T. J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **2014 516:7530** **516**, 272–275 (2014).
310. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. doi:10.1093/nar/gkr343.
311. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **2012 9:7** **9**, 676–682 (2012).
312. Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
313. Wickham, H. ggplot2. *ggplot2* (2009) doi:10.1007/978-0-387-98141-3.
314. “What do FRAP curves tell us.
https://wiki.cmci.info/docs/KotaMiura_100311notes.htm.
315. Sprague, B. L. & McNally, J. G. FRAP analysis of binding: proper and fitting. *Trends Cell Biol* **15**, 84–91 (2005).
316. FRAP_Manual. <https://www.embl.de/eamnet/frap/FRAP6.html>.
317. Houtsmuller, A. B. Fluorescence recovery after photobleaching: application to nuclear proteins. *Adv Biochem Eng Biotechnol* **95**, 177–199 (2005).

318. Sabari, B. R., Dall’Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* **45**, 961–977 (2020).
319. Hadzhiev, Y. *et al.* The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation. *Dev Cell* **58**, 155-170.e8 (2023).
320. Veil, M. *et al.* Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* **145**, (2018).
321. Zaidi, S. K. *et al.* Mitotic Gene Bookmarking: An Epigenetic Program to Maintain Normal and Cancer Cell Phenotypes. *Mol Cancer Res* **16**, 1617 (2018).
322. Raccaud, M. *et al.* Mitotic chromosome binding predicts transcription factor properties in interphase. *Nature Communications 2019 10:1* **10**, 1–16 (2019).
323. Kosak, S. T. *et al.* Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science (1979)* **296**, 158–162 (2002).
324. Williams, R. R. E. *et al.* Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* **119**, 132–140 (2006).
325. Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225.e24 (2021).
326. Sharp, P. A., Chakraborty, A. K., Henninger, J. E. & Young, R. A. RNA in formation and regulation of transcriptional condensates. *Rna* **28**, 52–57 (2022).
327. Cheung, A. C. M. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431–1437 (2012).
328. Gong, X. Q., Nedialkov, Y. A. & Burton, Z. F. Alpha-amanitin blocks translocation by human RNA polymerase II. *J Biol Chem* **279**, 27422–27427 (2004).
329. Wang, W. *et al.* A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Res* **50**, 4917–4937 (2022).
330. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (1979)* **361**, (2018).
331. Meeussen, J. V. W. *et al.* Transcription factor clusters enable target search but do not contribute to target gene activation. *Nucleic Acids Res* **51**, (2023).
332. Mueller, F., Stasevich, T. J., Mazza, D. & McNally, J. G. Quantifying transcription factor kinetics: At work or at play? *Crit Rev Biochem Mol Biol* **48**, 492–514 (2013).
333. Tsai, A., Alves, M. R. P. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife* **8**, (2019).
334. Tsai, A., Galupa, R. & Crocker, J. Robust and efficient gene regulation through localized nuclear microenvironments. *Development* **147**, (2020).
335. Pálffy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet* **16**, e1008546 (2020).
336. Gagnon, J. A., Obbad, K. & Schier, A. F. The primary role of zebrafish nanog is in extra-embryonic tissue. *Development* **145**, (2018).
337. He, M. *et al.* Nanog safeguards early embryogenesis against global activation of maternal β -catenin activity by interfering with TCF factors. *PLoS Biol* **18**, e3000561 (2020).
338. Shamipour, S. *et al.* Bulk Actin Dynamics Drive Phase Segregation in Zebrafish Oocytes. *Cell* **177**, 1463-1479.e18 (2019).
339. Miné-Hattab, J. CONDENSATES: When fixation creates fictio. *Elife* **12**, (2023).
340. Uebel, C. J. & Phillips, C. M. Phase-separated protein dynamics are affected by fluorescent tag choice. *MicroPubl Biol* **2019**, (2019).
341. Nelson, A. L. Antibody fragments: Hope and hype. *MAbs* **2**, 77 (2010).

342. Schumacher, D., Helma, J., Schneider, A. F. L., Leonhardt, H. & Hackenberger, C. P. R. Nanobodies: Chemical Functionalization Strategies and Intracellular Applications. *Angew Chem Int Ed Engl* **57**, 2314 (2018).
343. Wang, X. *et al.* Antibody-free profiling of transcription factor occupancy during early embryogenesis by FitCUT&RUN. *Genome Res* **32**, 378–388 (2022).
344. Li, J. *et al.* Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178**, 491–506.e28 (2019).
345. Tørresen, O. K. *et al.* Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res* **47**, 10994 (2019).
346. Yang, H. *et al.* A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* **588**, 337–343 (2020).
347. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
348. Cella Zanacchi, F., Manzo, C., Magrassi, R., Derr, N. D. & Lakadamyali, M. Quantifying Protein Copy Number in Super Resolution Using an Imaging-Invariant Calibration. *Biophys J* **116**, 2195 (2019).
349. Coffman, V. C. & Wu, J. Q. Counting protein molecules using quantitative fluorescence microscopy. *Trends Biochem Sci* **37**, 499 (2012).
350. Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *Am J Physiol Cell Physiol* **300**, C723 (2011).
351. Davidson, I. F. & Peters, J. M. Genome folding through loop extrusion by SMC complexes. *Nature Reviews Molecular Cell Biology* **22**, 445–464 (2021).
352. Meier, M. *et al.* Cohesin facilitates zygotic genome activation in zebrafish. *Development (Cambridge)* **145**, (2018).
353. Pérez-Rico, Y. A., Barillot, E. & Shkumatava, A. Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish. *iScience* **23**, (2020).
354. Taylor, N. O., Wei, M. T., Stone, H. A. & Brangwynne, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys J* **117**, 1285–1300 (2019).
355. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

7. Curriculum Vitae

Noémie CHABOT

Date of birth: 19/11/1993

Nationality: French

Education

09.2011 – 09.2015 – bachelor's in medical studies, Paris Descartes University, France

10.2014 – 07.2015 – Erasmus exchange at La Sapienza University, Rome, Italy

09.2015 – 06.2018 – Master of Genetics, Paris Diderot University, France

Research experience

05.2016 – 08.08.2016 – Student intern in the Chourrout lab, at the SARS International Center for Marine Biology, Bergen, Norway

Project: Evolution of the Hox genes in *Oikopleura dioica*

04.2017 – 08.2017 – Summer intern in the Holt lab, at the Institute for System Genetics, NYU, New-York, USA

Project: Effects of compressive stresses on *Candida albicans* and mammalian cells

01.2018 – 06.2018 – Master intern in the Escudé lab, at the Natural History Museum of Paris, France

Project: Study of the alpha-satellite DNA diversity in the Cercopithecine clade

Publication

Ksenia Kuznetsova, Noémie M. Chabot*, Martino Ugolini*, Edlyn Wu*, Manan Lalit, Haruka Oda, Yuko Sato, Hiroshi Kimura, Florian Jug, Nadine L. Vastenhouw.
Nanog organizes transcription bodies. Current Biology, 2023.

*Equal contribution

Talk and poster presentations

12.2017 – **ASCB**

Poster presentation (related to summer project in the Holt lab)

10.2019 – **Chromosome Territories and Nuclear Architecture**, Mainz, Germany

Poster presentation

12.12.2019 – **Status symposium on “Life?”**, Hannover, Germany

Poster presentation

05.2019 – **Awakening the genome: The maternal-to-zygotic transition**, Prague, Czechia

Poster presentation

03.2021 – **DFG LLPS meeting** (online)

Short talk

06.2022 - **Awakening the genome: The maternal-to-zygotic transition**, Vienna, Austria

Poster presentation

10.2022 – **EMBL conference: Transcription and chromatin**, Heidelberg, 2022

Poster presentation

12.2022 – **Informal meeting “Chromatin at the Lake”** in EPFL

Talk

Student supervision and teaching

10.2021 and 10.2022 – Teacher in practical courses: Introduction to Microscopy (UNIL)

04.2023 – 08.2022 – Supervision of a master student

Project: The role of DNA loop extrusion in the mir430 local compaction

Extracurricular activities

03.2022 – Participation to the competition “Ma thèse en 180 seconds” (UNIL)