Roles of endogenous retroviruses in transcriptional condensates

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by
Vahid Asimi

2023

Research work for this dissertation was performed under the supervision of Dr. Denes Hnisz at the Max Planck Institute for Molecular Genetics in Berlin, Germany from October 2018 to May 2023. This dissertation was submitted to the Department of Biology, Chemistry, Pharmacy of the Freie Universität Berlin in June 2023.

1st reviewer: Dr. Denes Hnisz
Department of Genome Regulation
Max Planck Institute for Molecular Genetics, Berlin, Germany

2nd reviewer: Prof. Dr. Florian Heyd
Institute of Chemistry and Biochemistry, Laboratory of RNA Biochemistry
Freie Universität Berlin

Date of defense: 17.10.2023

Acknowledgements

I would like to thank Dr. Denes Hnisz for giving me the opportunity to conduct my PhD research in an outstanding research environment. Denes is a true visionary, and I am beyond excited to follow his future scientific career. First principles reasoning, critical thinking and open collaboration are some of the lessons I learned from you that I will cherish for the rest of my life.

I thank my brilliant co-authors, Abhishek Sampath Kumar, Henri Niskanen, Christina Riemenschneider, Sara Hetzel and Julian Naderi, for their dedication, perseverance and creative approaches to experimentation. I learned the true meaning of teamwork from you. I thank Amèlia Aragonés Hernández for positive energy, friendship and valuable advice. I extend my gratitude to all other Hnisz lab and Meissner department members and all other collaborators who contributed to the publication of our work.

I thank Prof. Dr. Florian Heyd for valuable input and active participation in my thesis advisory committee. I thank Prof Dr. Alexander Meissner for support and for creating a unique and thriving scientific atmosphere.

I thank Maxi for unconditional friendship and long philosophical discussions that motivated me during this journey. I already consider you a lifelong friend. I thank Raha for meaningful conversations that made me see the world in new ways. You are the most amazing person I know.

I thank Henry Paige, Lennart Hesse and Saman Azizi for their support and for being part of my inner circle for over 10 years now. I wish you every success and hope our friendship lasts forever.

I am most grateful to my parents for their continuous encouragement and selfless commitment to my success. I thank my father for cultivating my intellectual spirit from an early age.

I thank my brother for his generosity and support during most difficult times. I am proud of you and everything you have accomplished.

I dedicate this work to my mother for always leading by example and being my true hero.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

Date: 23.06.2023 Signature:

Table of Contents

Acknowledgements ..... *V*

Declaration of Independence ..... *VII*

List of figures *1*

List of tables *4*

List of abbreviations *5*

Summary *7*

Zusammenfassung *8*

1 Introduction *11*

1.1 Membraneless organelles 11

1.2 Liquid-liquid phase separation 12

1.3 Nuclear organization 16

1.4 Heterochromatin 17

1.5 Retrotransposons 18

1.6 RNA in formation and regulation of transcriptional condensates 23

1.7 Thesis rationale and aims 25

2 Methods *26*

2.1 Cell culture 26

2.2 Generation of the TRIM28-FKBP mESC line 27

2.3 Generation of the TRIM28-FKBP iPSC line 28

2.4 Inactivation of NANOG::GFP in the TRIM28-FKBP iPSC line 29

2.5 Generation of shRNA IAPEz/MMERVK10C/ MMERVK9C/MMETn knockdown mESC lines 30

2.6 Integration of PiggyBac transposon encoding Dox-inducible ERVs 32

2.7 Deletion of ERVs in the TRIM28-FKBP mESC line 33

2.8 TRIM28 degradation 34

2.9 Western blot 34

2.10 Proteomics sample preparation and liquid chromatography with tandem mass spectrometry (LC-MS/MS) 34

2.11 RNA isolation and quantitative Real-Time PCR (qRT-PCR) 35

2.12 RNA-FISH combined with Immunofluorescence (IF) 36

2.13 TrueSeq Stranded mRNA-seq 39

2.14 TT-SLAM-Seq 40
2.15 H3K27Ac and H3K9me3 ChIP-Seq 41
2.16 RNAPII and MED23 ChIP-Seq 42
2.17 Average image and radial distribution analysis 43
2.18 Bioinformatics 44
2.19 RNA-Seq processing 44
2.20 TT-SLAM-Seq processing 45
2.21 Public ChIP-Seq data 45
2.22 ChIP-Seq processing 46
2.23 Enhancer and super-enhancer annotation 47
2.24 Detection of eRNA expression and preparation of TT-SLAM-Seq histograms 47
2.25 Differential gene expression analysis 49
2.26 Retrotransposon element definition 49
2.27 ChIP-Seq enrichment analysis 50
2.28 Motif enrichment 51
2.29 Retrotransposon expression quantification 51
2.30 Statistics and reproducibility 52
3 Results 53
3.1 Rapid and selective degradation of TRIM28 in mESCs 53
3.2 Reduced SE transcription in TRIM28-degraded mESCs 55
3.3 Reduced SE-condensate association in TRIM28-degraded mESCs 62
3.4 Derepressed ERVs form nuclear foci that overlap with RNAPII condensates 67
3.5 SE-enriched TFs rescue transcriptional condensate localization 80
3.6 IAP RNA contributes to transcriptional condensate localization in TRIM28-degraded mESCs 82
4 Discussion 86
4.1 Significance of the transcriptional hijacking model 86
4.2 Future perspective: RNA in transcriptional condensates 87
4.3 Future perspective: Competition for transcriptional machinery in vivo 92
4.4 Future perspective: TRIM28 condensates 95
4.5 Implications for disease and open questions 97
5 Bibliography 102

List of figures

Figure 1.1. Membraneless organelles in eukaryotic cells 12

Figure 1.2. Principles of transcriptional condensate formation 15

Figure 1.3. Classes of retrotransposons in the mouse 20

Figure 1.4. Epigenetic regulators of ERV repression and transcription 21

Figure 1.5. ERVs undergo waves of highly controlled epigenetic derepression during early development 22

Figure 1.6. Major types of RNA molecules produced by RNA polymerase II 24

Figure 3.1. TRIM28-FKBP knock-in strategy A 53

Figure 3.2. Acute and reversible degradation of TRIM28 in mESCs. 54

Figure 3.3. Selective and acute TRIM28 degradation does not significantly impact levels of pluripotency factors in mESCs 55

Figure 3.4. TT-SLAM-Seq detects nascent transcription 56

Figure 3.5. TRIM28 degradation leads to reduction of SE transcription 57

Figure 3.6. TRIM28 degradation leads to genome-wide changes in nascent transcription 58

Figure 3.7. TRIM28 degradation leads to reduction of nascent SE transcription and IAP derepression. 59

Figure 3.8. TT-SLAM-seq validates the effects of TRIM28 degradation on nascent transcription at the miR290-95 SE 60

Figure 3.9. qRT-PCR validation of the TT-SLAM-seq data at the miR290-295 and Klf4 loci 61

Figure 3.10. Global analysis of TRIM28 depletion effects on nascent transcription at intergenic SEs, intergenic typical enhancers and de-repressed LTRs 62

Figure 3.11. Reduced SE-condensate association in TRIM28-degraded mESCs 63

Figure 3.12. Additional quantifications of RNAPII colocalization at the miR290-295 and Fgf4 locus 65
Figure 3.13. PALM microscopy cell line setup 66
Figure 3.14. TRIM28 degradation induces derepression of IAPs 68
Figure 3.15. De-repressed IAPs form nuclear foci that associate with RNAPII condensates. 69
Figure 3.16. Further quantification of RNAPII and MED1 colocalization with de-repressed IAPs 70
Figure 3.17. Retrotransposon de-repression in TRIM28-degraded ESCs 71
Figure 3.18. TRIM28-degradation leads to reduced H3K9me3 and increased MED23/RNAPII occupancy at ERVs a 72
Figure 3.19. ERV-associated transcriptional condensates incorporate neighboring genes 73
Figure 3.20. ERV-neighboring genes associate with transcriptional condensates 74
Figure 3.21. RNAPII condensates contribute to Cthrc1 upregulation 75
Figure 3.22. CRISPR/Cas9-mediated deletion of ERVs at the Cthrc1 locus 76
Figure 3.23. Gene expression at the Cthrc1 locus in ERV deletion cell line 77
Figure 3.24. IAP sequences are enriched for TF binding motifs 78
Figure 3.25. Characterization of NFY and NRF1 co-localization with Cthrc1 RNA 79
Figure 3.26. Pluripotency TF overexpression system in TRIM28-FKBP iPSCs 81
Figure 3.27. Pluripotency TFs rescue transcriptional condensate localization in TRIM28-degraded iPSCs 82
Figure 3.28. ERV knockdown rescues transcriptional effects of TRIM28 depletion 83
Figure 3.29. ERV knockdown counteracts ERV derepression in TRIM28-depleted cells 84

Figure 3.30. ERV knockdown rescues transcriptional condensate localization in TRIM28-degraded mESCs. 85
Figure 4.1. Transcriptional hijacking model. 86

List of tables

Table 1. Oligonucleotides required for TRIM28-FKBP mESC line generation. 28
Table 2. Oligonucleotides required for IAPEz/MMERVK10C/MMERVK9C/MMETn shRNA plasmid cloning. 31
Table 3. Oligonucleotides required for Cthrc1 locus ERV deletion in the TRIM28-FKBP mESC line 33
Table 4. qRT-PCR primer sequences 36
Table 5. Sequences of RNA-FISH probes 38

List of abbreviations

3C	chromosome conformation capture
AML	acute myeloid leukemia
ATP	adenosine triphosphate
C_{sat}	saturation concentration
DMSO	dimethyl sulfoxide
dTAG	degradation tag
eRNA	enhancer RNA
ERV	endogenous retrovirus
FBS	fetal bovine serum
FISH	fluorescence in situ hybridization
FKBP	FK506-binding protein
FUS	fused in sarcoma
gDNA	genomic DNA
IAP	intracisternal A-type particle
IDR	intrinsically disordered region
iPSC	induced pluripotent stem cell
KRAB-ZFP	Krüppel-associated box zinc finger proteins
LIF	leukemia inhibitory factor
LLPS	liquid-liquid phase separation
lncRNA	long non-coding RNA
LTR	long-terminal repeat
MED1	Mediator of RNA polymerase II transcription subunit 1
MEF	mouse embryonic fibroblast
mESC	mouse embryonic stem cell
ncRNA	non-coding RNA
OSKM	Oct4, Sox2, Klf4 and c-Myc
P body	processing body

PALM	photoactivated localization microscopy
PCR	polymerase chain reaction
PML	promyelocytic leukemia protein
qRT-PCR	quantitative Real-Time PCR
RBD	RNA-binding domain
rDNA	ribosomal DNA
RNAPII	RNA polymerase II
RNP	ribonucleoprotein
rRNA	ribosomal RNA
SE	super-enhancer
TAD	topologically associating domain
tracrRNA	trans-activating crispr RNA
TT-SLAM-seq	transient transcriptome thiol(SH)-linked alkylation for the metabolic sequencing
uaRNA	upstream antisense RNA

Summary

Endogenous retroviruses (ERVs) are repetitive elements that constitute around 10 percent of mammalian genomes. ERVs need to be kept transcriptionally silenced during development and ERV upregulation is linked to loss of pluripotency and mouse embryonic lethality. Precise mechanistic understanding of this process remains incomplete. To investigate effects of ERV derepression, I utilized degradation tag (dTAG) system to acutely deplete endogenous levels of TRIM28, a heterochromatin protein that recruits silencing machinery to ERVs in mouse embryonic stem cells (mESCs). Nascent transcriptome sequencing indicated TRIM28 depletion induces acute transcriptomic changes: upregulation of ERVs and other targets of TRIM28-mediated repression and simultaneous downregulation of key pluripotency super enhancers (SEs) and associated genes. High-resolution microscopy data indicated ERV derepression reduces association of SEs with RNA polymerase II (RNAPII) and Mediator (MED1), the key components of transcriptional condensates. In turn, derepressed ERVs associated with RNAPII and MED1, indicating transcriptional condensate components are redistributed upon TRIM28 depletion. Presence of transcriptional machinery at derepressed ERV loci upregulated nearby genes, including *Cthrc1*. Ectopic overexpression of pluripotency factors enriched at SEs prevented the reduction in transcriptional condensate association with SEs. ERV knockdown rescued transcriptional condensate localization indicating RNAs produced at ERV play an important in transcriptional condensate redistribution. Evidence presented here shows derepressed ERVs have the capacity to 'hijack' transcriptional condensates from key pluripotency genes upon TRIM28 degradation. This may be the molecular mechanism contributing to embryonic lethality associated with ERV derepression in TRIM28 knockout mice.

Zusammenfassung

Endogene Retroviren (ERVs) sind repetitive Elemente, die etwa 10 Prozent des Säugetiergenoms ausmachen. ERVs müssen während der Entwicklung transkriptionell heruntergeregelt werden. In der Maus geht eine Hochregelung von ERVs mit dem Verlust der Pluripotenz und embryonaler Letalität einher. Das Verständnis des zugrunde liegenden Mechanismus ist jedoch unvollständig. Um die Auswirkungen der ERV-Hochregelung zu untersuchen, habe ich das Degradations-Tag (dTAG)-System verwendet, um die endogenen Level von TRIM28, einem Heterochromatin-Protein, das ERVs in embryonalen Stammzellen von Mäusen (mESCs) stummschaltet, akut zu verringern. Die Sequenzierung des entstehenden Transkriptoms deutete darauf hin, dass die TRIM28-Depletion akute transkriptomische Veränderungen induziert: Hochregelung von ERVs und anderen Zielen der TRIM28-vermittelten Repression und die gleichzeitige Herunterregelung wichtiger Pluripotenz-Super-Enhancer (SEs) und ihrer assoziierten Gene. Hochauflösende Mikroskopiedaten deuten darauf hin, dass die ERV-Hochregelung die Assoziation von SEs mit der RNA-Polymerase II (RNAPII) und Mediator (MED1), den Schlüsselkomponenten von Transkriptionskondensaten, verringert. Allerdings assoziierten ERVs mit RNAPII und MED, was darauf hinweist, dass sich Transkriptionskondensatkomponenten bei TRIM28-Depletion neu verteilen. Das Vorhandensein einer Transkriptionsmaschinerie an hochgeregelten ERV-Loci führte zu einer Hochregelung benachbarter Gene, einschließlich *Cthrc1*. Die ektopische Überexpression von Pluripotenzfaktoren die vermehrt an SEs binden, verhinderte die Verringerung der assoziation von Transkriptionskondensatkomponenten mit SEs. ERV-Knockdown rettete die Lokalisierung des Transkriptionskondensats, was darauf hindeutet, dass am ERV produzierte RNAs eine wichtige Rolle bei der Umverteilung des Transkriptionskondensats spielen. Die hier präsentierten Daten zeigen, dass hochgeregelte ERVs die Fähigkeit haben, Transkriptionskondensate von wichtigen Pluripotenzgenen beim Abbau „kapern“, wenn TRIM28 nicht vorhanden ist. Dies

könnte der molekulare Mechanismus sein, der zu embryonaler Letalität im Zusammenhang mit der ERV-Hochregelung bei TRIM28-Knockout-Mäusen beiträgt.

1 Introduction

1.1 Membraneless organelles

Functional compartmentalization of cellular processes is achieved through the formation of dedicated organelles, typically separated by a lipid bilayer membrane. Metabolic and gene regulatory pathways concentrate their catalytic and other functional constituents in mitochondria and nuclei, thereby decreasing distances between these constituents, reducing biological noise and maximizing efficiency (Sawyer et al., 2019). In addition to these and other canonical organelles, such as endoplasmic reticulum and the Golgi apparatus, cells comprise numerous other organelles that lack membranes completely. Some of these membranelles structures have been known to biologists for over a century, since they showed a granulated pattern in electron microscopy or enriched immunofluorescence signal of their protein components in fluorescence microscopy (Hirose et al., 2023). Nucleoli were first formally described in 1830s, and in 1903 neurobiologist Ramón y Cajal identified membraneless nuclear spots we now call Cajal bodies (Cajal, 1903; Pederson, 2011). To date, multiple other membraneless organelles have been described and investigated, including nuclear speckles, paraspeckles, processing bodies (P bodies), stress granules and promyelocytic leukemia protein (PML) nuclear bodies (Fig. 1.1) (Hirose et al., 2023).

Recent developments in the field suggest that for processes such as gene transcription in the nucleus it may be advantageous to crowd functionally related macromolecules without forming a distinct membrane. Spontaneous self-assembly of these membraneless organelles is suggested to occur by demixing from the solution, through a process called liquid-liquid phase separation (LLPS) (Courchaine et al., 2016; Shin & Brangwynne, 2017). Resulting structures are also known as droplets, coacervates, cellular/nuclear bodies, or biomolecular condensates (Hirose et al., 2023).

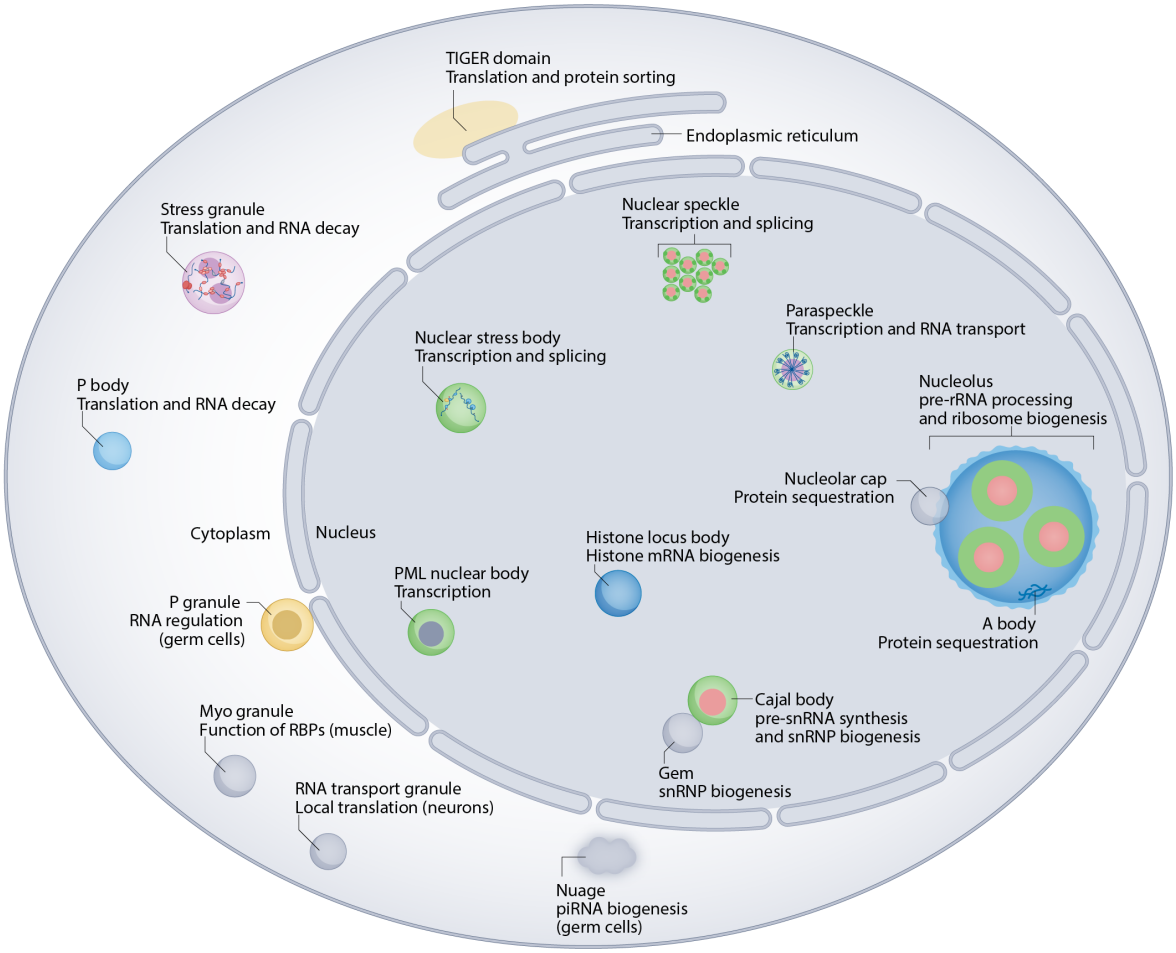


Figure 1.1. Membraneless organelles in eukaryotic cells. Membraneless organelles (MLOs) in the nucleus and cytoplasm, and the biological functions they participate in. Figure extracted from Hirose et al., 2023. Figure created using Adobe Illustrator.

1.2 Liquid-liquid phase separation

First evidence to demonstrate the formation of biomolecular condensates by aqueous phase separation came from Cliff Brangwynne, Tony Hyman and colleagues. Using *Caenorhabditis elegans* P granules as a model, they showed these protein-RNA assemblies exhibit liquid-like properties in cells, including fusion, wetting and dripping (Brangwynne et al., 2009). Same team later demonstrated this phenomenon is not unique to P granules – nucleoli in *Xenopus laevis* oocytes were similarly found to behave as viscous fluids through an adenosine triphosphate (ATP)-dependent process

(Brangwynne et al., 2011). These examples indicated cytoplasmic organelles are dynamic assemblies composed of protein and RNA that concentrate through transient molecular interactions, and liquid phase separation is not an isolated occurrence but rather a fundamental mechanism in cellular biology (Hyman & Brangwynne, 2011; Hyman & Simons, 2012).

Further work in characterizing ribonucleoproteins (RNPs) showed that RNAs found in cellular granules isolated from mouse brain extracts and human cell lysates phase separate together with low-complexity sequence domains of RNA binding proteins (for example, Fused in sarcoma (FUS)) (T. W. Han et al., 2012). Weak multivalent interactions and low-complexity, intrinsically disordered regions (IDRs) emerged as possible drivers of phase separation in cells (Hyman & Simons, 2012; Kato et al., 2012; P. Li et al., 2012). Interestingly, low complexity sequences and disorder are overrepresented in proteins shown to phase separate in vitro (Elbaum-Garfinkle et al., 2015; Kato et al., 2012; Kwon et al., 2013; Mitrea & Kriwacki, 2016; Nott et al., 2015). Low complexity domains are enriched for amino acids featuring polar residues such as glycine, glutamine, and serine, or aromatic residues, typically tyrosine (J. Wang et al., 2018). IDRs are protein regions that do not form stable secondary and tertiary structures and are conformationally heterogeneous and dynamic (Mitrea & Kriwacki, 2016). Proteins found in membraneless organelles can be either part-structured and part-disordered, or entirely disordered. The RNA-binding domains (RBDs) consist of folded RNA recognition motifs but can include intrinsically disordered regions, that are abundant in glycine and positively charged amino acids, for instance arginine (J. Wang et al., 2018) are comprised of one or more folded RNA recognition modules (RNA recognition motif, RRM), but they also carry regions of significant intrinsic disorder. The intrinsically disordered regions in RBDs are typically enriched in glycine and positively charged residues such as arginine. Abundance and patterning of selected amino acids seem to be relevant for the formation and physical properties of biomolecular condensates (Lyon et al., 2021; J. Wang et al., 2018).

Multivalent interactions drive liquid-like properties of both single- and multi-component biomolecular condensates (Elbaum-Garfinkle et al., 2015; Fromm et al., 2014; P. Li et al., 2012; Nott et al., 2015). Repetitive occurrence of either structured or low complexity disordered domains in protein sequences contributes to multivalency (Mitrea & Kriwacki, 2016). Both types of multivalent domains are sufficient on their own for a protein to phase separate *in vitro.* Formation of well-studied multicomponent condensates, for instance, nucleoli, is strongly dependent on the concentration of its components indicating the importance of establishing specific threshold concentrations as well as the order in which they are achieved (Fig. 1.2) (Weber & Brangwynne, 2015). Since phase separation thresholds vary from below micromolar to hundreds of micromolar, it is possible that components with lower threshold concentrations condense first, and increase the local concentrations of other components enabling them to reach their respective concentration thresholds and be incorporated inside the multicomponent condensate (Burke et al., 2015; P. Li et al., 2012; Mitrea & Kriwacki, 2016; Molliex et al., 2015).

Structured protein domains are associated with specific protein-DNA, protein-RNA and protein-protein interactions and could contribute to the recruitment of essential structural components of condensates (Allain et al., 2000; Johansson et al., 2004; Yang et al., 2003). On the other hand, low complexity domains most likely enable more dynamic interactions with a spectrum of possible partners. Structured domains and low complexity intrinsically disordered regions act synergistically to lower threshold concentrations, enable the formation of multicomponent biomolecular condensates and recruit additional components that perform specific functions within condensates (Banjade et al., 2015).

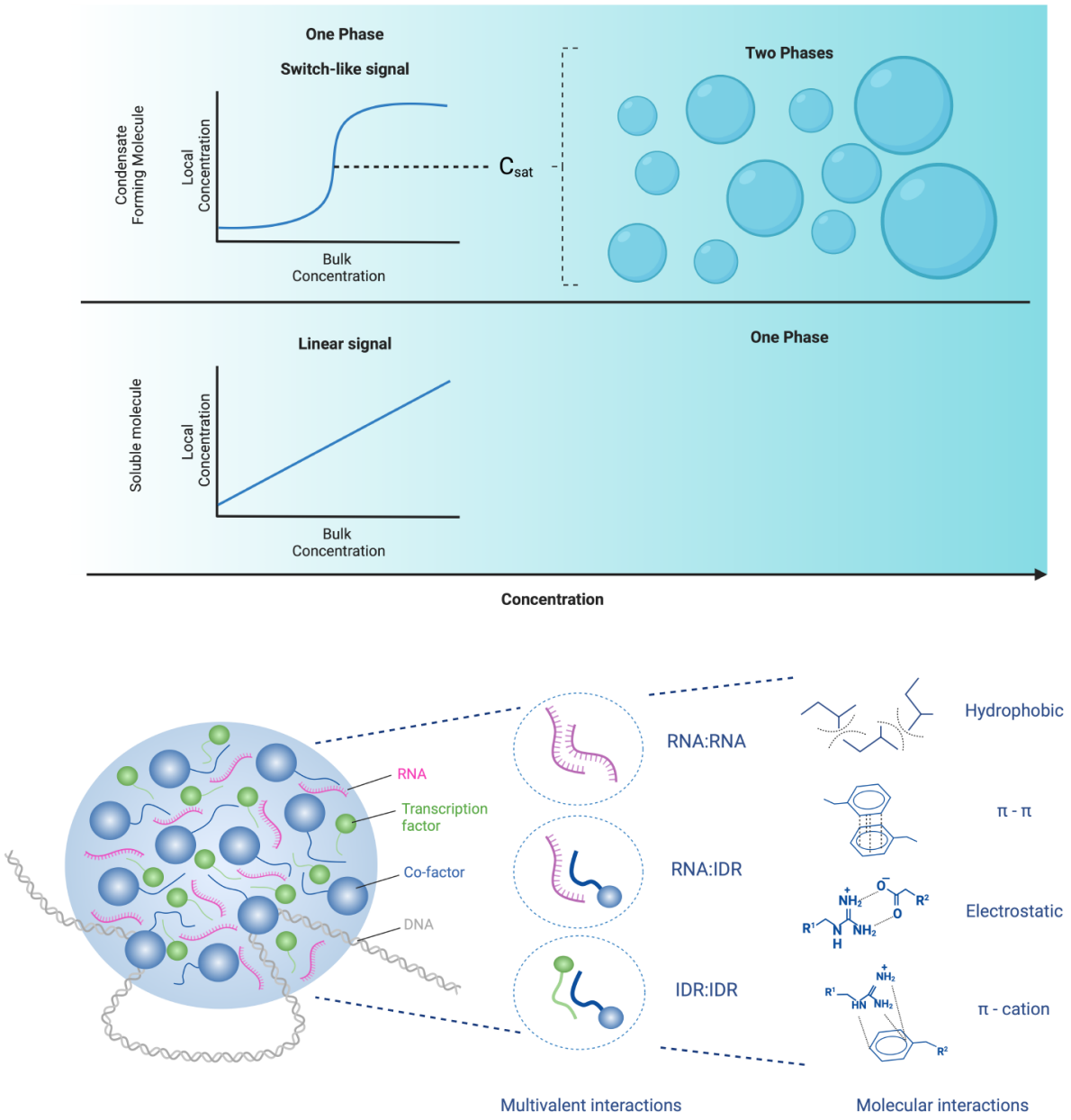


Figure 1.2. Principles of transcriptional condensate formation. (top) Concentration dependent phase transitions. Once the concentration of condensate forming molecules reaches saturation concentration (C_{sat}) it enables a switch-like transition into a two-phase system. Without phase separation, soluble molecule's local concentration scales linearly with the solvent concentration (bulk). **(bottom)** Representative model of transcriptional condensates, including components participating in multivalent interactions (DNA can be used instead of RNA) and the molecular interactions that drive condensation. Figure adapted from Boija et al., 2021 and Mitrea et al., 2022. Figure created using Adobe Illustrator and BioRender.com.

1.3 Nuclear organization

Hereditary information in eukaryotes is found in the nucleus, a ~10□m organelle that is the site of DNA transcription, replication, and RNA processing. Nucleus is a crowded environment since it contains almost all cellular DNA (2m in linear length), RNA and all the associated regulatory proteins. Nucleus needs to be organized in a way that facilitates precise gene expression programs during development.

From a coarse-grained perspective, the interior of the eukaryotic cell nucleus can be divided into two primary compartments (Rippe, 2022). First is chromatin, which is a sizable supramolecular complex made up of genomic DNA wrapped around histone proteins and associated with a variety of chromosomal proteins and chromatin-associated RNAs (Kornberg, 1974; Rippe, 2022). The soluble, liquid nucleoplasmic portion, also known as the nucleoplasm, is the other compartment, a fluid that envelops the chromatin compartment and is extremely viscous and abundant in dissolved proteins and RNAs (Rippe, 2022). Historically, investigations of the physical structure of nucleus were relying on evidence obtained with light microscopy which severely limited our understanding, considering that the size of the nucleosome is ~10 nm and the resolution limit of the light microscope is ~250 nm (Lakadamyali & Cosma, 2020).

Relatively recent emergence of techniques such as chromosome conformation capture (3C) and ChIP-seq launched a series of innovations in method-development space that now allow mapping of DNA contacts, chromatin-binding proteins and epigenetic modifications at unprecedented resolution (Furey, 2012; Grob & Cavalli, 2018; Lakadamyali & Cosma, 2020). Hi-C, a 3C method that uses DNA crosslinking and proximity-based ligation coupled with deep sequencing, probes genome-wide chromatin interactions and provides insights into the three-dimensional genome architecture that were previously not possible due to limitations in throughput and resolution of existing methods (Lieberman-Aiden et al., 2009). It provides a probabilistic measure for genome-wide contacts between any given loci on the chromosome and reveals

additional levels of genome organization. Two spatially segregated regions of active and silenced chromatin were found and named A and B compartment, respectively. Regions in the A compartment typically comprise transcribed genes and active histone modifications, but also some non-transcribed genes, whereas the B compartment contains inactive genes and repressive histone modifications (Jordan Rowley & Corces, 2018). Topologically Associating Domains (TADs) are sub-megabase regions with increased frequency of self-interactions as opposed to inter-TAD interactions (Dixon et al., 2012). Regulatory interactions between gene promoters and enhancers occur within such domains, and experimental evidence suggests this is relevant during development (Bolt et al., 2022; Glaser & Mundlos, 2021; Lupiáñez et al., 2015). Sub-TADs, and loops, which are self-interacting physical domains at more local distance scales, are additional levels of organizational units of the genome that have been discovered through Hi-C combined with deep sequencing (Lakadamyali & Cosma, 2020).

1.4 Heterochromatin

Heterochromatin was first identified in the 1920s by Emil Heitz (Heitz, 1928). By in situ staining chromosomes of *Pellia epiphylla*, a liverwort species, Heitz discovered that certain chromosomal sections remained condensed during interphase and named them heterochromatin (Passarge, 1979). These chromosomal areas were tightly packed for most of the cell cycle, only relaxing momentarily before mitosis, and then condensing once again. Heitz also recognized that there were parts of chromatin that would decondense during interphase and suggested the term euchromatin. Heterochromatin was later linked to gene-poor chromosome region and divided into facultative and constitutive heterochromatin.

Facultative heterochromatin silences developmental genes and can form at diverse chromosomal regions. Constitutive heterochromatin is more static and forms

consistently at gene-poor regions that contain repetitive elements, across all cell types. Most of constitutive heterochromatin forms around centromeres (pericentromeric) and at telomeres (Allshire & Madhani, 2018; Saksouk et al., 2015). Besides repeat DNA, proteins, RNAs, and various epigenetic modifications are enriched inside constitutive heterochromatin (Singh & Newman, 2020). Main epigenetic features of constitutive heterochromatin include histone H3 lysine 9 methylation (H3K9me3) and its reader protein heterochromatin protein 1 (HP1). Mammalian HP1α phase separates in vitro, indicating that constitutive heterochromatin may have features of biomolecular condensates (Larson et al., 2017; Strom et al., 2017).

1.5 Retrotransposons

Retrotransposons replicate via RNA Polymerase II (RNAPII) transcription followed by reverse transcription and reintegration into the genome, and account for more than 40% of human and mouse genomes (Thompson et al., 2016; Venter et al., 2001). They can be split into two groups based on whether they still contain long-terminal repeats (LTRs). Retrotransposons without LTRs comprise long and short interspersed nuclear elements (LINEs and SINEs) and SINE-Variable number tandem repeat-Alu (SVA) elements (Thompson et al., 2016). Those with LTRs are also called endogenous retroviruses (ERVs) or LTR retrotransposons and they comprise 10% of the mouse genome (Stocking & Kozak, 2008). ERVs are remnants of past retroviral infections that have permanently integrated into the host genomes. There are several important sub-classes of ERVs including MMERVK, MMETN and intracisternal A-type particles (IAPs) (Fig. 1.3). Unlike human, mouse ERVs are still capable of transposition and they contribute to about 12% of pathological mutations (Gagnier et al., 2019; Goodier & Kazazian, 2008). ERVs that gain retrotransposition activity but are not pathogenic have a strong gene regulatory potential (Mager & Stoye, 2015). They are capable of co-opting as ERV enhancers, lncRNAs or novel chimeric proteins (Enriquez-Gasca et al., 2022; Mager & Stoye, 2015). ERV inserted upstream of the agouti gene can cause

ectopic expression of agouti protein in mice, and LTRs derived from ERVs can drive transcriptional programs that move 2C-like cells away from pluripotency (Macfarlan et al., 2012; Morgan et al., 1999). In addition, perturbation of epigenetic repressors of retrotransposons can lead to de novo retrotransposition and have a tumor-suppressive role in acute myeloid leukemia (AML), and retrotransposon-derived RNA can adopt the role of viral replication intermediates and stimulate inflammation (Ahmad et al., 2018; Gu et al., 2021).

ERVs contain several essential components including: 5' and 3' LTRs (with identical sequences), pro-viral elements that encode proteins necessary for viral replication (*gag,* encoding a group-specific retroviral antigen, *pol*, encoding a reverse transcriptase, and *env*, encoding an envelope protein) (Fig. 1.3). Recombination between 5′ and 3′ LTRs removes the internal region, producing "solo" LTRs, which contain unique 3′ (U3) and 5′ (U5) regions and a regulatory region that includes the transcription start site (TSS) (Thompson et al., 2016). LTRs frequently contain combinations of several transcription factor binding sites (TFBSs), in addition to core polymerase II promoter elements (for instance, TATA box) and may also include a splice donor site within the U5 region (Thompson et al., 2016).

LTRs comprise regulatory elements necessary for proviral transcription including TFBSs. Cellular TFs are recruited to LTRs and boost transcription of proviral mRNA (Thompson et al., 2016). Genome-wide studies show about 20% of all TFBSs are found within TEs in both mouse and human cell lines (Sundaram et al., 2014). These TFs include key pluripotency factors, including octamer-binding transcription factor 4 (OCT4), SRY (sex determining region Y)-box 2 (SOX2), homeobox protein NANOG, and p53 (Kunarso et al., 2010; T. Wang et al., 2007).

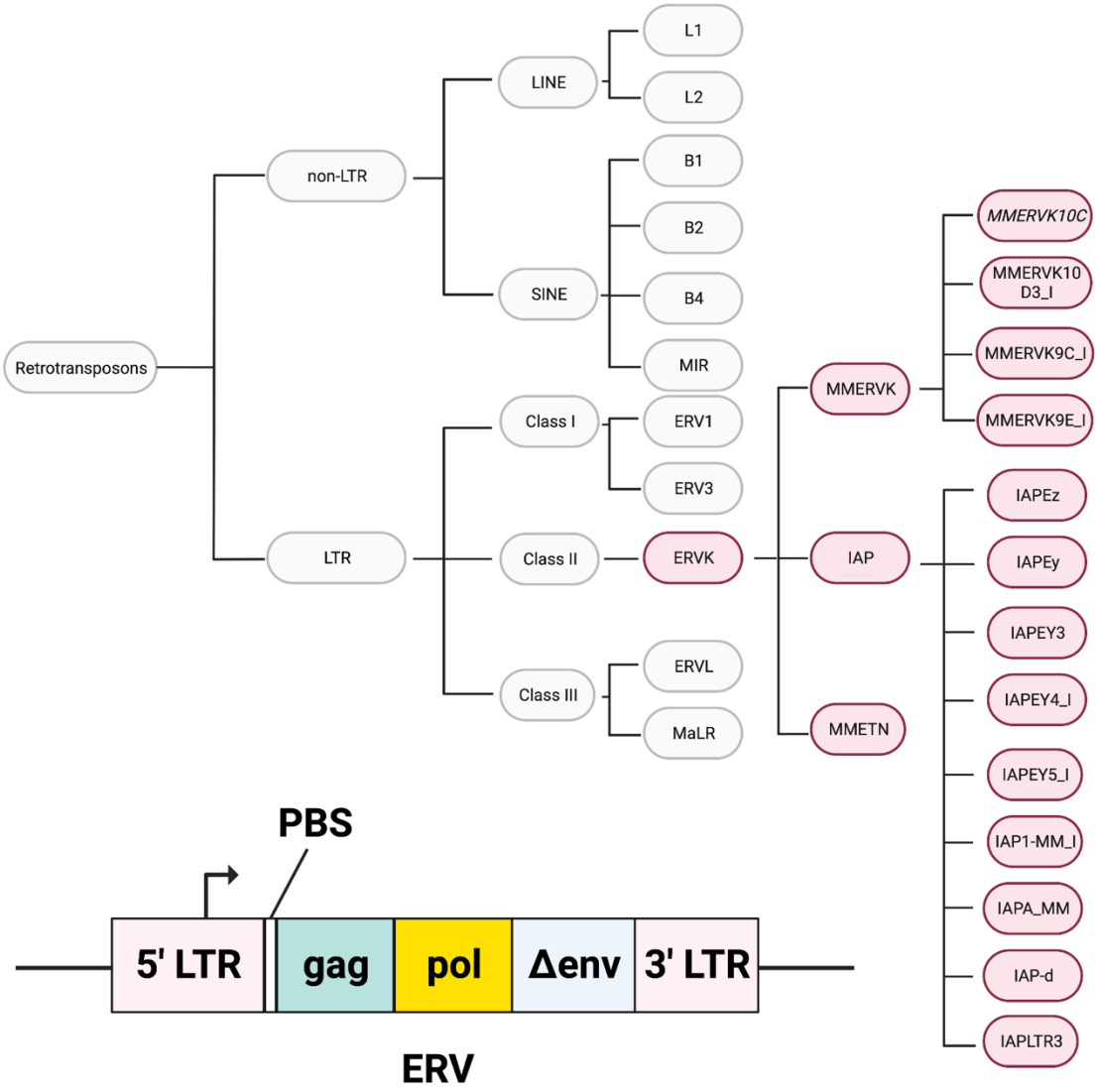


Figure 1.3. Classes of retrotransposons in the mouse. (top) ERV sub-classes in red are relevant for this study. **(bottom)** Elements found in typical ERVs. Full-length ERVs have 5′ and 3′ LTRs, and an "internal" region that includes a primer-binding site (PBS) that plays a role in the priming of reverse transcription and retroviral proteins *gag*, *pol*, and a truncated or mutated *env* gene (*Δenv*) (Thompson et al., 2016). Figure was adapted from Thompson et al., 2016. Figure created using Adobe Illustrator and BioRender.com.

In mouse embryonic stem cells (mESCs), ERVs are transcriptionally silenced by DNA sequence-specific Krüppel-associated box zinc finger proteins (KRAB-ZFPs) that recruit the co-repressor TRIM28 which in turn recruits the histone H3 lysine 9 (H3K9) methyltransferase SETDB1 and the heterochromatin protein HP1α (Enriquez-Gasca et al., 2022; Lachner et al., 2001; Matsui et al., 2010; Thompson et al., 2016; Wolf et

al., 2015). Indeed, chromatin immunoprecipitation sequencing (ChIP-seq) analysis in mESCs reveals that the IAP elements are marked by H3K9me3, TRIM28 and HP1α, but not typically occupied by active chromatin marks and pluripotency factors (Figure 1.4). Heterochromatin components, in turn, do not typically occupy enhancers enriched in pluripotency TFs (OCT4, SOX2, and NANOG) that control the cell-specific transcriptional program of mESCs (Asimi et al., 2022).

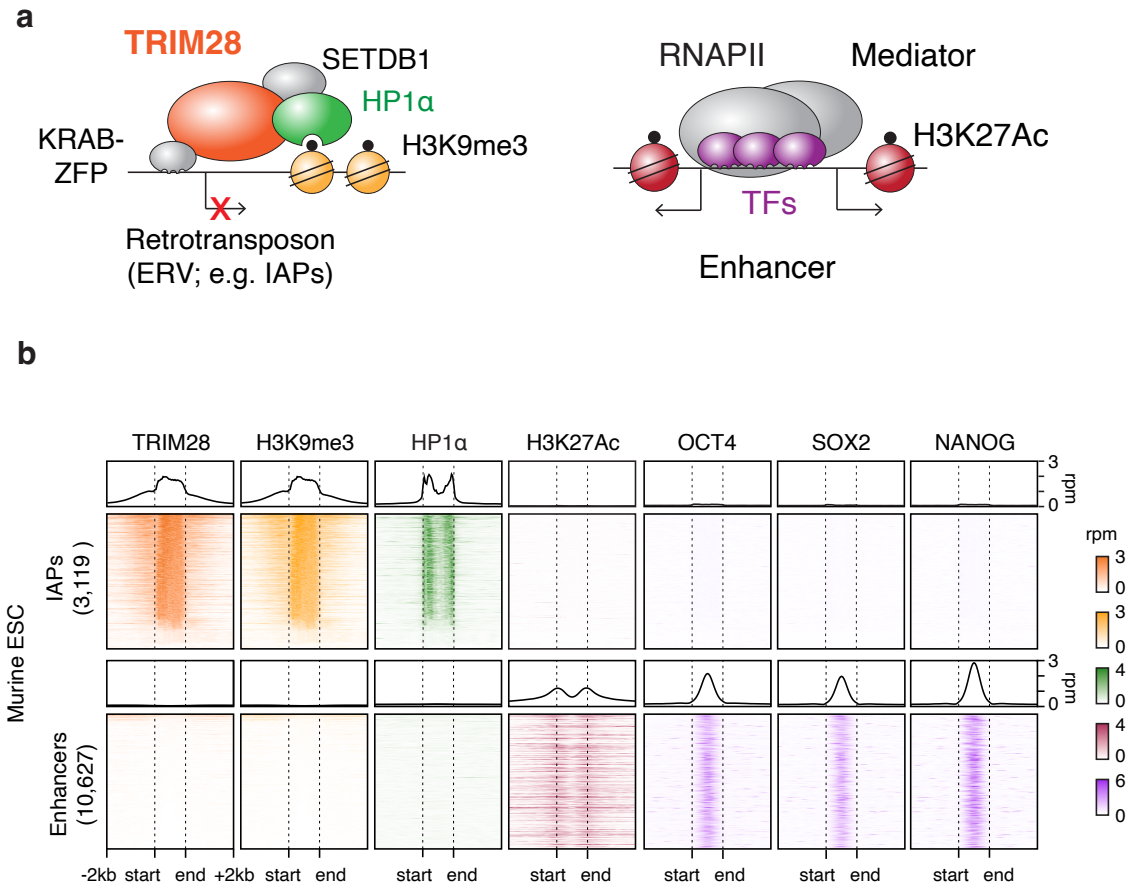


Figure 1.4. Epigenetic regulators of ERV repression and transcription. a. (left) TRIM28/HP1 pathway for retrotransposon repression implicates several epigenetic regulators and modifications including: TRIM28, H3K9me3, SETDB1, HP1α and KRAB-ZFPs. **(right)** Active transcription involves transcription factors (TFs) that occupy enhancer sites and recruit Mediator complex and RNAPII. **b.** Overview of epigenetic regulator occupancy at full-length IAPs and enhancers (heatmaps of ChIP-seq read densities within a 2kb window). The genomic elements were length-normalized. Rpm: reads per million. Figure adapted from Asimi et al., 2022.

ERV expression is regulated via the TRIM28/HP1α silencing pathway. ERVs follow specific expression patterns at precise developmental stages, and surges or almost complete shutdowns of transcription are essential for development (Gerdes et al., 2016). IAP transcripts are transferred from the oocyte to the zygote, but they bottom out during the two-cell stage, followed by a peak during the blastocyst stage (Fig. 1.5). ERV expression is subsequently kept downregulated, including by DNA methylation mechanisms. This dynamic expression curve needs to be closely controlled for proper development, for example, deletion of maternal TRIM28 alone leads to embryonic lethality (Messerschmidt et al., 2012).

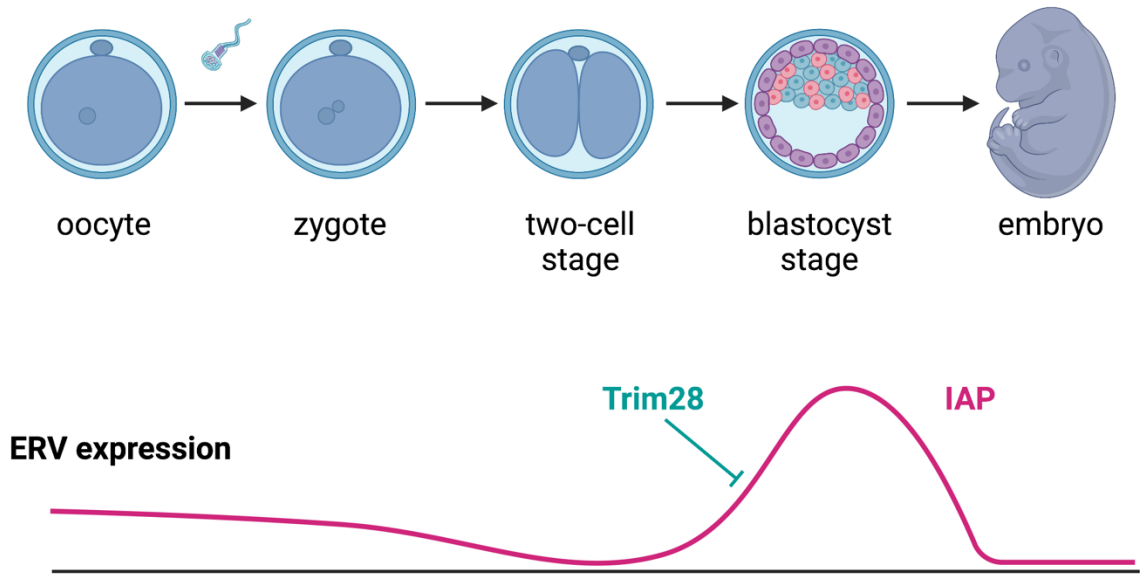
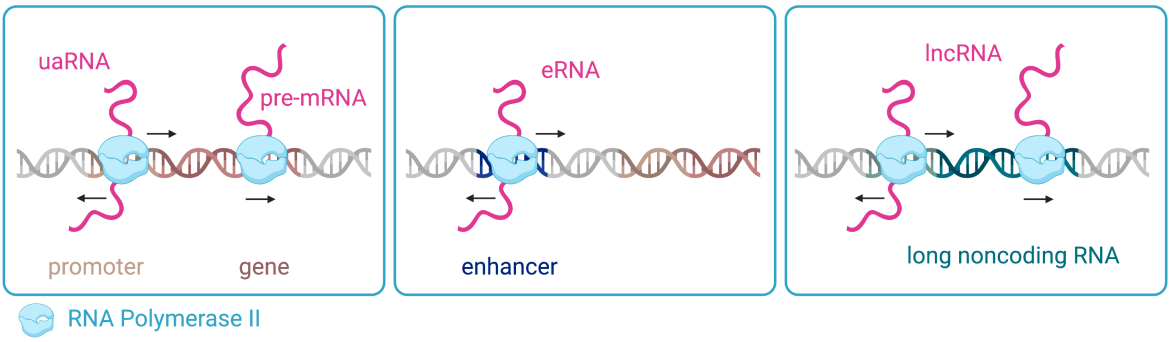


Figure 1.5. ERVs undergo waves of highly controlled epigenetic derepression during early development. IAP expression is carried from the oocyte into early embryos. IAP expression downregulates until two-cell stage and then peaks at the blastocyst stage before being silenced again into the late embryonic stages. Figure created using Adobe Illustrator and BioRender.com.

1.6 RNA in formation and regulation of transcriptional condensates

RNA may play an important role in nuclear organization (Nickerson et al., 1989; J. Rinn & Guttman, 2014). Thousands of non-coding RNAs (ncRNAs) are enriched in the nucleus where they participate in important regulatory processes (Frankish et al., 2019; J. L. Rinn & Chang, 2012). NcRNAs play a role in transcriptional regulation, cleavage and modification of pre-rRNAs, and splicing of pre-mRNAs (Black, 2003; Egloff et al., 2018; Kiss-László et al., 1996; Nilsen & Graveley, 2010; Plath et al., 2002; Watkins & Bohnsack, 2012). Moreover, ncRNAs seem to be preferentially enriched in specific regions of the nucleus and many localize to the sites of their production (Quinodoz et al., 2021; Quinodoz & Guttman, 2022). Most of the well-characterized biomolecular condensates to date contain RNA as one of the components (Sabari et al., 2020).

RNA may nucleate condensates by locally concentrating RNA-binding proteins, by promoting electrostatic interactions with proteins or by establishing secondary structures that facilitate RNA–protein or RNA–RNA interactions (Aumiller et al., 2016; Boeynaems et al., 2017; Burke et al., 2015; Drobot et al., 2018; Jain & Vale, 2017; Langdon et al., 2018). Active gene transcription produces nascent RNAs that are involved in condensate formation at nuclear compartments like nucleoli, Cajal bodies, and histone locus bodies (Berry et al., 2015; Sawyer et al., 2017). By proxy, it is conceivable that RNAs produced at transcribed enhancers, DNA damage sites and repetitive elements are involved in the nucleation of nuclear condensates (Arner et al., 2015; Frank & Rippe, 2020; Novo et al., 2022; Sabari et al., 2020).



	gene	enhancer	lncRNA gene
Number of genomic loci	19 954	≈40-60 000	≈30 000
Number of RNAs/cell	1-10 000	1-10	1-10
Half-life	min to hr	min	min to hr

Figure 1.6. Major types of RNA molecules produced by RNA polymerase II. (uaRNA - upstream antisense RNA, eRNA - enhancer RNA, lncRNA - long noncoding RNA). Numbers and half-life properties of given RNA species were extracted from the following sources: Frankish et al., 2019; Hon et al., 2017; W. Li et al., 2016; McStay & Grummt, 2008; Schwalb et al., 2016. Figure adapted from Sharp et al., 2022. Figure created using Adobe Illustrator and BioRender.com.

RNA properties such as length, structure, concentration and modifications could contribute to the formation and function of biomolecular condensates by regulating their biophysical features including size, shape, viscosity, surface tension and composition (Roden & Gladfelter, 2021). Indeed, recent evidence suggests nascent RNA is participating in the formation of transcriptional condensates via electrostatic interactions (Henninger et al., 2021). RNAs stemming from repetitive sequences are particularly promising due to their high copy number, clustered distribution and partial self-complementarity (Frank & Rippe, 2020). Repetitive RNA sequences could facilitate phase separation in several ways including by forming a scaffold for the recruitment of interacting proteins by increasing the number of binding sites, by using

sequence self-complementarity to increase interactions within protein-RNA complexes, or by interacting with IDRs (Frank & Rippe, 2020).

1.7 Thesis rationale and aims

Gene transcription occurs in the cell nucleus, the crowded biophysical environment composed of DNA, RNA and regulatory proteins. Precise spatial and temporal coordination of nuclear components is required to execute lineage-specific gene expression programs. Recent advances in the field suggest transcription may be mediated by the formation of transcriptional condensates, membranelles nuclear structures that concentrate functionally related macromolecules required for transcription (Hnisz et al., 2017). Epigenetic regulators are frequently investigated via knockout systems, but acute effects on transcription and genomic distribution of transcriptional machinery get overlooked. TRIM28 is a ubiquitously expressed adaptor protein that loads transcriptional silencing apparatus onto ERVs. Derepression of ERVs is correlated with mouse embryonic lethality in a process that is independent of transposition, indicating that ERV transcription and ERV RNA action may contribute to this phenotype.

By using a rapid degradation system (dTAG), this thesis aims to analyze acute effects of TRIM28 depletion on ERV expression, global nascent transcription as well as the genomic distribution of transcriptional condensates in mESCs. Combination of nascent transcriptome profiling and high-resolution imaging of SE/ERV RNA and RNAPII protein was used to determine whether ERV derepression affects pluripotency associated SE transcription and alters the association of transcriptional condensates with SEs and ERVs. Rescue experiments using pluripotency TF overexpression and shRNA knockdown of major ERV repeat classes are used to elucidate the role of TFs and ERV RNA in transcriptional condensate association with SEs and ERVs.

2 Methods

This section contains detailed descriptions of methods and analyses used to generate data presented in the Results section. Majority of protocols have been published in Asimi et al., 2022.

2.1 Cell culture

V6.5 mouse Embryonic Stem Cells (mESCs) and induced Pluripotent Stem Cells (iPSCs) were cultured on irradiated primary Mouse Embryonic Fibroblasts (MEFs) under standard Serum/LIF conditions (knockout DMEM containing 15% fetal bovine serum (FBS), supplemented with 1X penicillin/streptomycin, 1X GlutaMAX supplement, 1X non-essential amino acids, 0.05mM ß-mercaptoethanol (all from Gibco) and 1000 U/ml leukemia inhibitory factor (LIF)).

For ChIP-Seq, TT-SLAM-Seq and RNA-Seq experiments, mESCs were depleted from MEFs by incubating them on gelatin-coated cell culture plates for 45 min at 37°C allowing MEFs to attach to the plate while mESCs remain in suspension. MEF depletion was performed two times after which mESCs were seeded on gelatin-coated plates and maintained in Serum/LIF conditions with 2000 U/ml LIF.

For RNA FISH combined with immunofluorescence, MEF-depleted cells were grown on round 18mm glass coverslips (Roth LH23.1). Coverslips were coated with 5 µg/ml of poly-L-ornithine (Sigma-Aldrich, P4957) for 30 min at 37°C and with 5µg/ml of Laminin (Corning, 354232) overnight at 37°C.

To perturb RNAPII condensates, cells were treated 30 minutes with 1.5% 1.6-hexanediol (Sigma) in Serum/LIF conditions with 2000U/ml LIF.

2.2 Generation of the TRIM28-FKBP mESC line

For knock-in of the degradation-sensitive FKBPF36V tag at the N-terminus of TRIM28, a repair template containing homology arms spanning upstream and downstream of the target site was cloned into a pUC19 vector (NEB) (Fig. 3.1). The repair template included a mRuby2 fluorescent protein sequence, P2A linker and the FKBP tag sequence (Fig. 3.1) (Nabet et al., 2018). mRuby2-N1 plasmid (Addgene #54614) was used to amplify mRuby2 sequence and P2A-FKBP sequence was amplified from the PITCh dTAG donor vector (Addgene #91792). A guide RNA (Table 1) targeting the N-terminus of TRIM28 was cloned into the sgRNA-Cas9 vector pX458 (Addgene: #48138). The repair template and the sgRNA-Cas9 vector were transfected into V6.5 mESCs and iPSCs by nucleofection using Amaxa 4D Nucleofector X Unit (Lonza) according to the manufacturer's instructions. To screen for positive integrations, the transfected cells were sorted for mRuby2 fluorescent protein expression with flow cytometry. The sorted cells were seeded as single cells and expanded for a few days. Single colonies were picked and genotyped for the correct integration with PCR and Western blot.

Name	Oligo sequence (5'-3')
Guide RNA	CGTGTGAATGGCGGCCTCGG
5′ homology arm cloning forward	CCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCCCGAGTTTAACAATAAGAAAGTTTGGG
5′ homology arm cloning reverse	TCTTCGCCCTTAGACACCATGGCCGCCATTCACACG
3′ homology arm cloning forward	GGTCGGGTGGCGGCGGATCTGCCTCGGCGGCAGC
3′ homology arm cloning reverse	CGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCTCCCAGACACTGAGACC
mRuby2 forward	GCGGCGTGTGAATGGCGGCCATGGTGTCTAAGGGCGAAG
mRuby2 reverse	AAGTTAGTAGCTCCGCTTCCCTTGTACAGCTCGTCCATCCC
mCerulean forward	GCGGCGTGTGAATGGCGGCCATGGTGAGCAAGGGCGAGGAGCTGTTC
mCerulean reverse	GCAGTCGCTGCCGCCGAGGCAGATCCGCCGCCACCCG
P2A FKBP forward	GGATGGACGAGCTGTACAAGGGAAGCGGAGCTACTAACTTCAG
P2A FKBP reverse	GCAGTCGCTGCCGCCGAGGCAGATCCGCCGCCACC
Genotyping primer forward 1	CATCTCGACAGCGCCTGG
Genotyping primer forward 2	ACAAAGGCCACACCCGGG
Genotyping primer reverse	GTAGCAGATGCCCTACACAC
Guide RNA	CGTGTGAATGGCGGCCTCGG
5′ homology arm cloning forward	CCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCCCGAGTTTAACAATAAGAAAGTTTGGG
5′ homology arm cloning reverse	TCTTCGCCCTTAGACACCATGGCCGCCATTCACACG
3′ homology arm cloning forward	GGTCGGGTGGCGGCGGATCTGCCTCGGCGGCAGC
3′ homology arm cloning reverse	CGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCTCCCAGACACTGAGACC
mRuby2 forward	GCGGCGTGTGAATGGCGGCCATGGTGTCTAAGGGCGAAG
mRuby2 reverse	AAGTTAGTAGCTCCGCTTCCCTTGTACAGCTCGTCCATCCC

Table 1. Oligonucleotides required for TRIM28-FKBP mESC line generation.

2.3 Generation of the TRIM28-FKBP iPSC line

For mouse secondary induced pluripotent stem cells (iPSCs) generation, secondary MEFs that contain Doxycycline (Dox)-inducible Oct4, Sox2, Klf4 and c-Myc (OSKM) transgenes, and a NANOG::GFP reporter, were reprogrammed (Wernig et al., 2008). MEFs were seeded at a low density of about 10,000-25,000 cells per well of a 12-well

plate that was coated with 0.2% gelatin. Prior to transgene induction, cells were synchronized using a 12-16h incubation in 2.5% FBS containing medium. Medium was subsequently switched to 15% FBS/LIF with 2µg/ml Doxycycline. Media was changed every day; colonies emerged after about a week and clonal iPSC lines were generated within 2 weeks. TRIM28 FKBP knock-in alleles were integrated into the iPSC line following the protocol described above.

2.4 Inactivation of NANOG::GFP in the TRIM28-FKBP iPSC line

For immunofluorescence combined with RNA FISH experiments in the TRIM28-FKBP iPSC line, GFP at the Nanog::GFP locus was inactivated using guide RNA (GAAGTTCGAGGGCGACACCC) targeting against the GFP sequence. CRISPR-Cas9-guide RNA Ribonucleoprotein complex (RNPs) was assembled following the manufacturer's guidelines (Alt-R CRISPR; IDT). 2µl of 200µM crRNA and 2µl of 200µM tracrRNA were mixed, denatured at 95°C for 5 minutes and allowed to anneal while cooling down at room temperature for 20 minutes. 3µl of crRNA-tracrRNA was mixed with 1µl of 61µM Cas9 for 20 minutes prior to transfection. TRIM28-FKBP iPSCs were transfected with the assembled RNPs by nucleofection using Amaxa 4D Nucleofector X Unit (Lonza) according to the manufacturer's instructions. The cells were then sorted for non-GFP cells using flow cytometry. The sorted cells were expanded and cultured as a clonal line.

2.5 Generation of shRNA IAPEz/MMERVK10C/ MMERVK9C/MMETn knockdown mESC lines

Consensus sequences of inner parts of IAPEz, MMERVK10C, MMERVK9C, and MMETn were used to design shRNAs (Table 2). shRNA sequences were cloned into an inducible plasmid (Tet-pLKO-puro) following the standard cloning protocol recommended for this vector. Lentiviruses with the shRNA constructs were generated by transfecting HEK293 cells with packaging and envelope plasmids. TRIM28-FKBP mESCs were transduced with the virus with 8µg /mL polybrene for 48 hours followed by treatment with 2µg/ml puromycin to select for the transduced cells. Single colonies were picked, characterized, and expanded as a bulk line for the quadruple knockdown line for IAPEz, MMERVK10C, MMERVK9C, MMETn. shRNA expression was induced by treatment with 2µg/ml Doxycycline. For the experiment with pre-induction of shRNAs the cells were treated for 24 hours with Doxycycline followed by DMSO, dTAG-13, or Dox+dTAG-13 for another 24 hours. Tet-pLKO-puro was a gift from Dmitri Wiederschain (Novartis, Cambridge US) (Addgene plasmid # 21915). psPAX2 was a gift from Didier Trono (EPFL Lausanne) (Addgene plasmid # 12260). pCMV-VSV-G was a gift from Bob Weinberg (WIBR, Cambridge US) (Addgene plasmid # 8454). S2 work was performed following all relevant guidelines and regulations, approved by the Max Planck Institute for Molecular Genetics and the local authorities LAGeSo, Berlin (license number: 222/15-17a).

Name	Oligo sequence (5'-3')
shRNA_MMETn_1_Fwd	CCGGTACAGATGGTTAGTGTTAAATCTCGAGATTTAACACTAACCATCTGTATTTTTG
shRNA_MMETn_1_Rev	AATTCAAAAATACAGATGGTTAGTGTTAAATCTCGAGATTTAACACTAACCATCTGTA
shRNA_MMETn_2_Fwd	CCGGGACATCCAAACTTCGTAATATCTCGAGATATTACGAAGTTTGGATGTCTTTTTG
shRNA_MMETn_2_Rev	AATTCAAAAAGACATCCAAACTTCGTAATATCTCGAGATATTACGAAGTTTGGATGTC
shRNA_MMERVK9c_1_Fwd	CCGGCCCAGAGGCAGAGAGAAATTTCTCGAGAAATTTCTCTCTGCCTCTGGGTTTTTG
shRNA_MMERVK9c_1_Rev	AATTCAAAAACCCAGAGGCAGAGAGAAATTTCTCGAGAAATTTCTCTCTGCCTCTGGG
shRNA_MMERVK9c_2_Fwd	CCGGTTCTCTGCAATCGTGATTATTCTCGAGAATAATCACGATTGCAGAGAATTTTTG
shRNA_MMERVK9c_2_Rev	AATTCAAAAATTCTCTGCAATCGTGATTATTCTCGAGAATAATCACGATTGCAGAGAA
shRNA_MMERVK10c_1_Fwd	CCGGCCGGTAAGTGTAAAGAAATTACTCGAGTAATTTCTTTACACTTACCGGTTTTTG
shRNA_MMERVK10c_1_Rev	AATTCAAAAACCGGTAAGTGTAAAGAAATTACTCGAGTAATTTCTTTACACTTACCGG
shRNA_MMERVK10c_2_Fwd	CCGGCAAGGCCGTTATGGAAATTAACTCGAGTTAATTTCCATAACGGCCTTGTTTTTG
shRNA_MMERVK10c_2_Rev	AATTCAAAAACAAGGCCGTTATGGAAATTAACTCGAGTTAATTTCCATAACGGCCTTG
shRNA_IAPez_1_Fwd oligo	CCGGACTGCGGGTCGCGGTAATAAACTCGAGTTTATTACCGCGACCCGCAGTTTTTTG
shRNA_IAPez_1_Rev oligo	AATTCAAAAAACTGCGGGTCGCGGTAATAAACTCGAGTTTATTACCGCGACCCGCAGT
shRNA_IAPez_2_Fwd oligo	CCGGGTAGGAGCAAAGAGGAAATATCTCGAGATATTTCCTCTTTGCTCCTACTTTTTG
shRNA_IAPez_2_Rev oligo	AATTCAAAAAGTAGGAGCAAAGAGGAAATATCTCGAGATATTTCCTCTTTGCTCCTAC
shRNA_MMETn_1_Fwd	CCGGTACAGATGGTTAGTGTTAAATCTCGAGATTTAACACTAACCATCTGTATTTTTG
shRNA_MMETn_1_Rev	AATTCAAAAATACAGATGGTTAGTGTTAAATCTCGAGATTTAACACTAACCATCTGTA
shRNA_MMETn_2_Fwd	CCGGGACATCCAAACTTCGTAATATCTCGAGATATTACGAAGTTTGGATGTCTTTTTG
shRNA_MMETn_2_Rev	AATTCAAAAAGACATCCAAACTTCGTAATATCTCGAGATATTACGAAGTTTGGATGTC
shRNA_MMERVK9c_1_Fwd	CCGGCCCAGAGGCAGAGAGAAATTTCTCGAGAAATTTCTCTCTGCCTCTGGGTTTTTG

Table 2. Oligonucleotides required for IAPEz/MMERVK10C/ MMERVK9C/MMETn shRNA plasmid cloning.

2.6 Integration of PiggyBac transposon encoding Dox-inducible ERVs

The PB-tetO-lox-GFPpA-lox-IAPEz, and -MMERVK10C constructs (Dox-inducible GFP) were created by digesting an "all-in-one" PiggyBac, TREG/Tet-3G plasmid (Addgene plasmid # 97421; a gift from Steven Carr & Samuel [Broad Institute, Cambridge US]) with restriction enzymes NcoI and KpnI and cloning the PB_ATG_GFP oligonucleotide to insert a start-codon "ATG" in front of the existing GFP via NEBuilder HiFi Assembly according to manufacturer's instructions. To integrate the loxP sites, the plasmid was digested with NheI and BamHI and integrated the sequences loxP_upstream and loxP_downstream were integrated. To integrate various IAP and MMERVK10C sequences, sequences from gDNA were amplified with primers binding in conserved regions of the ERVs and overhangs to the PiggyBac plasmid. After size selection (~900bp) fragments were cloned into the PiggyBac construct by cutting with BamHI and cloning via NEBuilder HiFi Assembly according to the manufacturer's instructions. The constructs were randomly integrated in the mESC V6.5 Trim28-FKBP subclone by co-transfecting 4x105 cells with 8.5µg PB-tetO-lox-GFPpA-lox-IAPEz, or -MMERVK10C and 1.5µg Super PiggyBac transposase expression vector (SBI, PB210PA-1) using FuGENE HD Transfection Reagent (Promega). After 4 days of Puromycin selection (2µg/ml) single colonies were picked, expanded and tested for doxycycline-inducibility of the GFP construct monitored by GFP detection with FACS. Clones expressing high levels of GFP were selected and expanded for subsequent experiments. A plasmid encoding Cre recombinase was transfected to catalyze recombination between the two homologous loxP sites, removing the GFP sequence to generate isogenic cell lines, with the same copy number and insertion sites of either GFP or IAPEz/MMERVK10C. The sequence of GFP until the polyA is ~900bp long and comparable with the length of the IAPEz and MMERVK10C sequences.

2.7 Deletion of ERVs in the TRIM28-FKBP mESC line

To generate a TRIM28-FKBP mESC line with three ERV deletions at the Cthrc1 locus, deletions of MMETn, MMERVK and IAP sequences were performed sequentially with CRISPR/Cas9. Guide RNAs (Table 3) flanking the individual ERV elements were cloned into the sgRNA-Cas9 vector pX458 and the two vectors for each cut were delivered to cells with Lipofectamine 2000 (Thermo Fisher, 11668027) according to manufacturer's instructions. Transfected cells were sorted based on GFP expression two days later with flow cytometry. Single colonies were picked, genotyped and clones with homozygous deletion were selected for the next ERV deletion. Deletions were also confirmed by sequencing the PCR products.

Name	Oligo sequence (5'-3')
MMETn deletion gRNA1 Fw	CACCGAAACATCAGTGAAACAGAGT
MMETn deletion gRNA1 Rev	AAACACTCTGTTTCACTGATGTTTC
MMETn deletion gRNA2 Fw	CACCGTTGCTGTGGCAAACATGTGA
MMETn deletion gRNA2 Rev	AAACTCACATGTTTGCCACAGCAAC
MMETn deletion genotyping Fw	CAGTTTGGTGGAGTATGGATTTT
MMETn deletion genotyping Rev	ACACAGGTCAAATGCACAGA
MMETn non-deletion genotyping Fw	CACCACCACCTAGCCCATAA
MMERVK9C deletion gRNA1 Fw	CACCGATATGTCTGTAATCACACAC
MMERVK9C deletion gRNA1 Rev	AAACGTGTGTGATTACAGACATATC
MMERVK9C deletion gRNA2 Fw	CACCGACTGTTTTGTCATTCTGTGG
MMERVK9C deletion gRNA2 Rev	AAACCCACAGAATGACAAAACAGTC
MMERVK9C deletion genotyping Fw	TTTCCCAGACTGAGACGGAG
MMERVK9C deletion genotyping Rev	AAGTGACACAGGGCTTATGG
MMERVK9C non-deletion genotyping Rev	ATCTTTCTCTCACACCGCCA
IAP deletion gRNA1 Fw	CACCGTGCTGTCTAATAGAGACCGG
IAP deletion gRNA1 Rev	AAACCCGGTCTCTATTAGACAGCAC
IAP deletion gRNA2 Fw	CACCGGTGGCCGACTCAGTAGTTGG
IAP deletion gRNA2 Rev	AAACCCAACTACTGAGTCGGCCACC
IAP deletion genotyping Fw	CTTCAATGAGATCTGATGCCCA
IAP deletion genotyping Rev	AAGTCCACTGAGAACCCACC
IAP non-deletion genotyping Fw	GCCGGTTTAGATGGTCCTATT

Table 3. Oligonucleotides required for Cthrc1 locus ERV deletion in the TRIM28-FKBP mESC line

2.8 TRIM28 degradation

Across all experiments involving TRIM28 degradation, prior to treatment, cells were seeded onto 0.2% gelatin-coated plates after two rounds of MEF depletion. For TRIM28 degradation, 500nM of dTAG-13 compound (Nabet et al., 2018) was mixed with mESC media (Supplemented with 2000 U/mL LIF) and incubated for the indicated time duration. Media was changed daily using fresh dTAG-13.

2.9 Western blot

Cell lysis of cultured cells was performed in RIPA buffer for 30 minutes at 4^{o}C, followed by centrifugation for 20 minutes at maximum speed. The supernatant was then transferred to a new tube and quantified by BCA assay (Thermo Scientific). Extracted protein (10μg) was run on a 4-12% NuPAGE SDS gel and transferred onto a PVDF membrane using iBlot2 Dry Gel Transfer Device (Invitrogen) according to manufacturer's instructions. The blots were blocked with 5% skim milk in TBST and incubated with primary antibodies. Primary antibodies used in this study include TRIM28 (ab22553; 1:500), ACTB (ab8226; 1:1000), OCT4 (ab19857; 1:500), OCT4 (sc-5279; 1:500), HSP90 (BD610419; 1:4000), SOX2 (ab79351; 1:500). HRP-conjugated secondary antibodies (115-035-174, 211-032-171, Jackson Immuno) were used against the host species at 1:3000 – 1:5000 dilution and visualized with HRP substrate SuperSignal West Dura (Thermo Scientific).

2.10 Proteomics sample preparation and liquid chromatography with tandem mass spectrometry (LC-MS/MS)

Proteomics sample preparation was performed according to an available protocol with minor modifications (Kulak et al., 2014). Three biological replicates of dTAG-13 - treated samples with 4 million cells per replicate were lysed under denaturing

conditions and sequentially digested with LysC and Trypsin (Roche, Basel, Switzerland). Peptide desalting was performed according to the manufacturer's instructions (Pierce C18 Tips, Thermo Scientific, Waltham, MA). Desalted peptides were further separated into four fractions by strong cation exchange chromatography (SCX, 3M Purification, Meriden, CT). LC-MS/MS was carried out by nanoflow reverse phase liquid chromatography (Dionex Ultimate 3000, Thermo Scientific) coupled online to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Scientific), as reported previously 64. Briefly, the LC separation was performed using a PicoFrit analytical column (75 µm ID × 50 cm long, 15 µm Tip ID; New Objectives, Woburn, MA) in-house packed with 3-µm C18 resin (Reprosil-AQ Pur, Dr. Maisch, Ammerbuch, Germany). Raw MS data were processed with MaxQuant software (v1.6.10.43) and searched against the mouse proteome database UniProtKB with 55,471 entries, released in May 2020. The MaxQuant processed output files can be retrieved (Supplemental Table 3 in Asimi et al., 2022), including peptide and protein identification, accession numbers, % sequence coverage of the protein, q-values, and label-free quantification (LFQ) intensities. The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PDX021895 (Martens et al., 2005). The correlation analysis of biological replicates and the calculation of significantly different proteins were done with Perseus (v1.6.14.0). Only groups with valid values in at least one group were used, missing values were replaced by values from the normal distribution.

2.11 RNA isolation and quantitative Real-Time PCR (qRT-PCR)

RNeasy kit (Qiagen) was used to isolate RNA from cultured cells, according to manufacturer's instructions. RNA (1µg) was used for cDNA synthesis using RevertAid First Strand cDNA synthesis kit (Thermo Scientific) with random hexamer primers

according to manufacturer's instructions. qRT-PCR was performed with primers (Table 4) and 2X PowerUP SYBR green master mix (Applied Biosystems).

Name	Oligo sequence (5'-3')
pri-miR290-295 transcript	TTTCAAGGGAGGAACGAGCC
	TAGACTCCCTAGACCTGCCC
miR290-295 super-enhancer	GAGTCGAAGGCAGGAGGATG
	CTCCGAGTAGAGGCTGTCC
Klf4 transcript	CGGGAAGGGAGAAGACACT
	GAGTTCCTCACGCCAACG
Klf4 super-enhancer	TCGTTATGGGAGAAGGAGCC
	CAGATGCCTTCCACTTGCTG
IAPez	TCTTTTAGAAGTGGCTGGAGTGA
	TGCTGAAAAATATTGGCAACTC
Cthrc1 nascent transcript	ACACTAGCGCTTACAGATGGT
	GTGGCCCAGTTAGTTTCACG
Cthrc1 transcript	GACCTCTTCCCATCGAAGCC
	CACAGAGTCCTTCCACAGAGG
MMERVK10c	GCCACCAGAGACATGGTTTT
	CGGGCTTCTTTTCTTGTGAG
Slc25a32	TATCAGGTTGTGAGAGCCCG
	ACCGATGCCTTCTTTCCTCC
Dcaf13	AGCTGGACATACAAAGAGTCCC
	GCTTTGCAAACACTCGTTCCA
Fzd6	CCCCCACCGAAAGCTCG
	CTGGGGCAACTGCTCGG
Rims2	GGAGAGGAAAATCATCCTGGCT
	AAAGGGAAACCACTGTGTCG

Table 4. qRT-PCR primer sequences

2.12 RNA-FISH combined with Immunofluorescence (IF)

RNA-FISH combined with Immunofluorescence was performed as previously described (Sabari et al., 2018). For immunofluorescence, dTAG-13 or DMSO treated cells were

fixed in 4% PFA for 10 minutes at RT and stored in PBS at 4°C. All buffers and antibodies were diluted in RNase-free PBS (Thermo Fisher, AM9624). Cells were permeabilized with 0.5% Triton X-100 (Thermo Fisher, 85111) for 10 min at RT, followed by three consecutive 5 min PBS washes. Cells were then incubated in the primary antibody (RNAPII (abcam, ab817) at 1:500, NFY-A (Santa Cruz, sc-17753 X) at 1:250, NRF1 (abcam, ab55744) at 1:500, MED1 (abcam, ab64965) at 1:500 and MED23 (Bethyl Labs, A300-425A) in PBS overnight. After two 5 min PBS washes, cells were incubated in the secondary antibody (Invitrogen, goat anti-mouse Alexa 488 (A-11001) or goat anti-rabbit Alexa 488 (A-11008)) at 1:500 in PBS for 60 min at RT. Cells were washed twice in PBS for 5 min and re-fixed with 4% PFA in PBS for 10 min at RT. Following two 5 min PBS washes, cells were washed once with 20% Stellaris RNA FISH Wash Buffer A (Biosearch Technologies, Inc., SMF-WA1-60), 10% Deionized Formamide (EMD Millipore, S4117) in RNase-free water (Invitrogen, 10977035) for 5 min at RT. Cells were hybridized with 90% Stellaris RNA FISH Hybridization Buffer (Biosearch Technologies, SMF-HB1-10), 10% Deionized Formamide and 12.5 or 25 µM Stellaris RNA FISH probes. Probes were hybridized in a humidified chamber O/N at 37°C. Cells were washed with Wash Buffer A for 30 min at 37°C and stained with 0.24µg/mL DAPI in Wash Buffer A for 3 min at RT. Cell were washed with Stellaris RNA FISH Wash Buffer B (Biosearch Technologies, SMF-WB1-20) for 5 min at RT, mounted onto glass microscopy slides with Vectashield mounting medium (Vector Laboratories, H-1900) and sealed using transparent nail polish. Images were acquired with LSM880 Airyscan microscope equipped with a Plan-Apochromat-63x/1.40 oil DIC objective or Z1 Observer (Zeiss) microscope with 100X magnification, with Zen2.3 version 2.3.69.1016 (blue edition) or Zen (version black). Images were processed with ZEN 3.1 (Zeiss) and ImageJ software version 2.1.0/1.53i. ImageJ co-localization plugins were used for co-localization analysis of ERV IAP RNA FISH with RNAPII and MED1 IF (Bolte & Cordelières, 2006; Gilles et al., 2017). For nearest RNAPII cluster distance analysis in the *miR290-295* RNA FISH dataset, z-projections consisting of ±4.5 slices around the FISH spot were obtained in both

channels, thresholded to allow detection of individual RNAPII clusters. Center of mass distances to the nearest cluster were calculated using FIJI (DiAna) (Gilles et al., 2017). RNA FISH probes were designed and generated by Biosearch Technologies Stellaris RNA FISH to target introns of *miR290-295* primary transcript and *Cthrc1*, and IAPEz transcripts. Sequences of RNA-FISH probes are available in Table 5.

	Mir290-295	*IAPez*	*Cthrc1*	*Fgf4*
1	gctagcctgccttttaaaaa	taacagtctgctttacggga	cgtagcggttcaagactctg	tcggcgcgcccagaactcac
2	gagcgaggaaggctgagttc	tccttagaaagttcaaggcc	atgtggtgacctctcatgat	actactgcccgggaccctat
3	aatgtcttctttggagacca	agattgccctgaatttcatc	acttacacggatttgggcat	gggagctcgactctactcag
4	actctttttccacacacatt	cttggaaaggcctgtatact	tggagccagaaattactggg	ctgactctcgcgtgggccgc
5	ttcctcccttgaaattatgt	atctagtgacgagtatctcc	ctcgggctttctttgaaat	cgccgggtgcgctctcagtc
6	tactcactttccccacatag	ctgtttcctcagaggagaat	ctcacaaaggacacgttgca	gagagaccaggcttccgaag
7	taactcctagctttggtttc	ttttggctttttttcttgacc	aacaaatgctttgtgggggt	gggtgaaggcgatggaattc
8	aatgtactgcatagactccc	tatatttgcccatactaggg	cccagggagatgacacaaag	cccgggctcccgtgggtgcg
9	cttaaaattcactccaacct	gtagtttgtctgatcagcag	tgtggtcctctaaggagtaa	cagcctcagtctgggcgctt
10	ccaggaggaaagaacgtgga	cttgtaagccttttgttcttt	ttcagcttcaaagcctctga	gtctaccggtggccctcaag
11	gcggtccagacgttaaaaca	cttgccacacttagagcaaa	agatcaggcaaaatgggtcc	aagctttaagagctcatctc
12	gctggtaaatgtgccagata	ggaagaattctgccctttat	ctgactgctcatctgtttta	tctcgatgctggggtgaaaa
13	cagttaacccggaacacgtg	taataggaagtcggaggcgg	agtgttcctgactggttaag	acagaccctccgactctaag
14	tttcttcgaatccgtactca	ctatcctatctcctttacta	gaaagcccttgggatatttg	tcagaaaagcctaggtgacc
15	tcgctatactcagtctcatt	gtattatcagggaggagcag	cagactttcccacagattag	agccagacccgctggcactc
16	tacaacgaccacctcagtta	tatctcaggtcctgcaaatt	ctccactaggtttattgact	ctagtcttgcacactgtatc
17	taacagctccaagcagcgac	actcttttccgttgatctta	gtgggacagacaaactctct	cctctgtgaccaacacacaa
18	gcgtcagatgcaaagctatg	tatacttttatctgctccgg	caaatctatcgatggctggc	ctgaaaacgtgagacctggc
19	taaactccaagcctaaaccc	atgagatgactctgtggtgg	cagaagtttatgatgtgcca	tctagaaacccagcctggct
20	aactgaaccgccctctttag	gcccaaatgctgcataatat	agtgcctgacaaccatatag	ttggatggggtagttccctt
21	acgactgccttacatccatc	gggcgttttcattggacaaa	aactagctgtcaaggagctt	aagtacctcactcaggagga
22	caatctacaatgcacctgga	taattgttcctctaccagtt	ccagtgcttttgaatcagtt	cttaatatctcctgctggag
23	ttagttcttagccgttttga	gagtattccaaggtgaggta	tgcgtgaaacagcagggatt	ataagtccacgctaggcata
24	agaaatgcaaccccagtgaa	cccttggacacaaaggtata	gtataatccttagaggggga	tgtcccagatttcgaagagg
25	gactcaaacccacatgtgac	tggtaagaccttccattgat	cacagtatgcttttccaagt	taggtaggcggagtgtcaaa
26	aacgcggaaagcctttagta	ctttatggcacaggaggatg	cctaatgactgtgtggactt	caatcgggacccaagagatc
27	tccaacttccaagacctgag	actttttctgtggctatctg	gcgtgaaagacacgctttga	aaagagggtactgggtgagc
28	aggtaagcgattccaggttg	gatctcttcttatctctacc	gctagttccaataggtattt	
29	agcacacatacctgtttcaa	ccttaactcagcggaaggaa	cctctcaggaaagatgctac	
30	tagccagtggcaacgaattc	ggggaggagatatgaggatc	ttgcttatcaattcctgcac	

Table 5. Sequences of RNA-FISH probes.

	Mir290-295	*IAPez*	*Cthrc1*	*Fgf4*
31	taatatggcggccacgtgag	attctgtaaggctttttcca	gcattcttgccatgaagtta	
32	gcaactacagtagtcaagca	ttttagctggggatacgttt	cctcagcaggtgtgtaataa	
33	ccaactacagtagtcaagca	taagggcaagctgtgcaatt	ccagttagtttcacgtgatc	
34	ttaaagtcagctacagccag	tgtttggataccactttacc	cccatttttatgcagaggaa	
35	aagcttgtttgtgctaggag	taacaatgtcacgggcttct	ctaagttgtatgtgcacctc	
36	ttatgggtattatctacccg	agaactcacagcagctttga	agtagctgtcttgtatcaca	
37	ctgggctattgtaaagccaa	aataccgcgtgggttaattc	ggacttggttcttggaactt	
38	agattatgcttagggcacac	tttctccggttaacggagaa	cctgacattctatgttagct	
39	gctaggcaggattacattca	ctggctctgaacaatgacgt	ggcgaactcattttaagggt	
40	ttgaaggcaagtaagtaccc	aaggctgtctgaatcaaggc	gtccgtgttctaatggatta	
41	ccacagatgacacccaaatg	tgttaatggtgtcagtcgtc	gtcaaatgatcacagctgcg	
42	cacctcagcttttacttttg	atcttgacattagctgaggc	ggggacagtaccattactta	
43	ctgtcaaatctgggtcactt	tgacaaattagctgccctag	ccagtcacgtggtaattttt	
44	gccaaaaggataaatgcagc	aattcagccgtccaattttg	ggcgaccaaggacaaaagga	
45	ttcgctagatccaaacatgc	ctgaagtctcaattcccgaa	ccctaatttcagcagtttta	
46	gttgattgaagttccgatgc	agatccaattgggtaatcct	cagaatcaccatcgtcacta	
47	gatgagcaagcaaggagtct	aatgagactagcctcgtgag	ctgacatgctattttgcgtg	
48	aaagcagccgacctgtgaat	tgtaggtaaaggagccgtac	tccctggatcattcatcaaa	

Table 5. (continued) Sequences of RNA-FISH probes.

2.13 TrueSeq Stranded mRNA-seq

mESCs were cultured with either DMSO or 500 nM dTAG-13 for 2, 6, 24 and 96 hours. RNA was isolated using RNeasy kit (Qiagen) and 1μg of RNA was used for preparing the libraries. TrueSeq Stranded mRNA capture kit was used to prepare the libraries (KAPA biosystems) according to manufacturer's instructions. Unique Dual-Indexed (UDI; KAPA biosystems) adapters were ligated and the library was amplified for 8 cycles. The libraries were then sequenced as Paired-end 100 (PE100) on a Novaseq6000 with 50 million fragments per library.

2.14. TT-SLAM-Seq

TT-SLAM-Seq was performed as described previously (Reichholf et al., 2019). Briefly, cells were treated with DMSO or 500 nM dTAG-13 for 2, 6 or 24 hours and subjected to 15 minutes of 4-Thiouridine (4sU) labeling using 500 µM 4sU. Total RNA was extracted with Trizol (Ambion) and 24:1 chloroform:isoamylalcohol (Sigma), while using 0.1 mM DTT in isopropanol precipitation and ethanol washes. For each sample, 50 µg of total RNA was fragmented with Magnesium RNA Fragmentation Module (NEB), and fragmentation buffer was removed from samples with ethanol precipitation in presence of 0.1 mM DTT. RNA was then resuspended in 350 µl RNAse free water, diluted in biotinylation buffer (200 mM HEPES pH 7.5, and 10 mM EDTA) and topped up with 5 µg MTS-Biotin (previously diluted to 50 µg/ml in dimethylformamide) to reach a final volume of 500 µl. Biotinylation reaction was incubated for 30 minutes at room temperature while keeping samples in rotation and protected from light. Unbound biotin was removed with Acid-Phenol:Chloroform extraction (124:24:1, Ambion) and isopropanol precipitation. Biotinylated RNA was resuspended in 100 µl RNAse-free water, denatured in 65 □C for 10 minutes and then cooled on ice for 5 minutes. The biotinylated RNA was captured with 100 µl µMACS streptavidin beads (Miltenyi) by incubating for 15 minutes in rotation while keeping samples protected from light. µMACS columns were equilibrated on magnetic stand with nucleic acid equilibration buffer and two times with biotinylation buffer (20 mM HEPES, 1 mM EDTA, pH 8). Beads were transferred to columns, washed three times with wash buffer (100 mM Tris-HCl pH7.5, 10 mM EDTA, 1 M NaCl and 0.1 % Tween 20) and labeled RNA was eluted two times with total 200 µl of 100 mM DTT. RNA was cleaned up with RNEasy Minelute columns (Qiagen) and eluted to RNAse-free water with 1 mM DTT. 4sU residues of RNA were alkylated with iodoacetamide treatment (10 mM iodoacetamide in 50 mM NaPO4, pH 8, and 50 % DMSO) by incubating samples in 50 □C for 15 minutes, followed by quenching with 20 mM DTT. RNA samples were purified with ethanol precipitation and treated with Turbo DNAse (Invitrogen). Sequencing libraries were prepared with NEBNext Ultra II Directional RNA Library

Prep Kit and NEBNext Multiplex Oligos (NEB), according to manufacturer's instructions, except for using 8 minutes incubation time in fragmentation step.

2.15. H3K27Ac and H3K9me3 ChIP-Seq

For ChIP-Seq experiments, DMSO and dTAG-13 treated cells were detached with TrypLE Express (Gibco), washed once with PBS, fixed in rotation with 1 % formaldehyde for 10 minutes in room temperature followed by 5 minutes of quenching with 125 mM glycine. For H3K27Ac and H3K9me3 ChIPs, three million mESCs were used per replicate sample, and 750,000 S2 cells were added for exogenous genome spike-in normalization (Orlando et al., 2014). Cells were lysed in LB1 (50 mM HEPES-KOH, 140 nM NaCl, 1 mM EDTA, 10 % glycerol, 0.5 % Igepal CA-630 and 0.25 % Triton X-100, 5 mM Na-butyrate and 1x protease inhibitor cocktail) and collected by centrifugation. Lysis was continued in LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 5 mM Na-butarate and 1x protease inhibitor cocktail) followed by centrifugation. Nuclei were lysed in LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-deoxycholate, 0.5 % N-Lauroylsarcosine, 5 mM Na-butyrate and 1x protease inhibitor cocktail) and chromatin was fragmented with Bioruptor NextGen for 35 cycles (high setting). Lysates were clarified and 10 % of the sample was set aside as input. The remaining sample was split in two to capture protein-DNA complexes with 1µg H3K27Ac (ab4729; Abcam) and 1µg H3K9me3 (ab8898; Abcam) antibodies by incubating them in rotation overnight in 4 □C. This was followed by 24 hours of incubation with Protein A Dynabeads (Invitrogen) that had been washed three times with 0.25 % BSA in PBS. Beads from immunoprecipitation were washed 7 times with RIPA buffer (50 mM HEPES-KOH pH7.5, 1 mM EDTA, 1 % Igepal CA-630, 0.7 % Na-deoxycholate, 500 mM LiCl, 5 mM Na-butyrate and 1 x protease inhibitor cocktail), once with TE buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA and 50 mM NaCl) and eluted from beads with Elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 1% SDS). Samples were

decrosslinked for 16 hours at 65 □C in presence of 550 mM NaCl and proteinase K (Ambion), treated with RNAse A (Thermo Scientific) and DNA was extracted with Phenol:Chloroform:Isoamylalcohol followed by chloroform extraction and ethanol precipitation. Sequencing libraries were prepared from 10 ng of DNA with KAPA HyperPrep Kit (Roche) and paired-end sequenced with NovaSeq 6000 (Illumina) to produce ~100 million fragments for each library.

2.16. RNAPII and MED23 ChIP-Seq

Cells were treated with DMSO or 500 nM dTAG-13 for 24 hours. RNAPII and MED23 ChIP-Seq samples were prepared as described for H3K27Ac and H3K9me3 ChIP-Seq, with an exception for MED23 samples: protein-protein crosslinking was performed by incubating cells in rotation with 2 mM disuccimidyl glutarate (Thermo Scientific; Pierce) in PBS, followed by two washes with PBS and formaldehyde fixation as above. 30 million and 10 million cells were used for RNAPII and MED23 IPs, respectively. RNAPII samples included an exogenous genome spike-in of 7.5 million S2 cells. Cell lysis was performed as above and samples were sonicated in sonication buffer (50 mM HEPES-KOH pH 7.5, 140 mM EDTA, 1 mM EDTA, 1 mM EGTA, 1 % Triton X-100, 0.1 % Na-deoxycholate, 0.1 % SDS and 1x protease inhibitors (Roche)) for 35 cycles in case of RNAPII samples and 50 cycles for MED23 samples by using Bioruptor NextGen (high setting). When preparing beads for IPs, 50 µl of Protein A Dynabeads (Invitrogen) was washed three times in blocking buffer (0.5 % BSA in PBS) and then bound to 5 µg of RNAPII antibody (8WG16; Biolegend) and 2 µg of Spike-in antibody (61686, Active Motif) or 5 µg of MED23 antibody (A300-425A; Bethyl Laboratories) during an overnight incubation in blocking buffer. Antibody-conjugated beads were washed three times with blocking buffer, resuspended to 100 µl of blocking buffer prior to adding them to sheared chromatin in a total volume of 2.5 ml sonication buffer for RNAPII and 1 ml for Med23. Immunoprecipitation was performed in rotation overnight in 4 □C. Beads were washed twice with sonication buffer, once with sonication

buffer containing 500 nM NaCl, once with LiCl wash buffer (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5 % Igepal CA-630, 0.5 % Na-deoxycholate, 1x protease inhibitors) and once with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1x protease inhibitors). Samples were eluted from beads with 200 µl elution buffer as above and crosslinks were reversed by incubating samples in 65 □C for 16 hours. Samples were topped up with 200 µl TE buffer and 8 µl of RNAse A, incubated in 37 □C for 2 hours followed by addition of 7 µl of 300 mM CaCl2, 2 µl Proteinase K (Ambion) and 30 minutes incubation in 37 □C. DNA purified from samples with phenol:chloroform:isoamyl alcohol extraction followed by ethanol precipitation. Libraries were prepared and sequenced as above for total ~50 million fragments for each library.

2.17. Average image and radial distribution analysis

Image analysis pipeline used for the colocalization analysis of RNA FISH combined with IF was described previously (Sabari et al., 2018). MATLAB scripts were used to identify RNA FISH foci in z stacks through intensity thresholding (same threshold was used for image sets shown on the same figure panels) and create RNA FISH signal centroids (x, y, z) that were stitched together and positioned in a box of size l = 1.5 µm. For identified FISH foci, signal from corresponding location in the IF channel was collected in the l x l square centered at the RNA FISH focus at every corresponding z-slice. The IF signal centered at FISH foci for each FISH and IF pair were then combined to calculate an average intensity projection, providing averaged data for IF signal intensity within a l x l square centered at FISH foci. The same process was carried out for the FISH signal intensity centered on its own coordinates, providing averaged data for FISH signal intensity within a l x l square centered at FISH foci. As a control, this same process was carried out for IF signal centered at random nuclear positions which were generated using custom Python scripts. These average intensity projections were then used to generate 2D contour maps of the signal intensity or

radial distribution plots. Contour plots are generated using in-built functions in MATLAB. The intensity radial function ((r)) is computed from the average data. For the contour plots of the IF channel, an intensity colormap consisting of 14 bins with gradients of black, violet and green was generated. For the FISH channel, black to magenta was used. The generated colormap was employed to 14 evenly spaced intensity bins for all IF plots. The averaged IF centered at FISH or at randomly selected nuclear locations were plotted using the same color scale. For the radial distribution plots, the Spearman correlation coefficients r were computed and reported between the FISH and IF (centered at FISH) signal. A two-tailed student's t-test, comparing the Spearman correlation calculated for all pairs, was used to generate P-values.

2.18. Bioinformatics

All analyses were carried out using R 3.6.3 unless stated otherwise.

2.19. RNA-Seq processing

Adapter and quality trimming of raw reads was performed using cutadapt (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC -A AGATCGGAAGAGC), followed by poly-A trimming with cutadapt (parameters: --overlap 20 --minimum-length 25 --adapter "A[100]" --adapter "T[100]") (M. Martin, 2011). Reads were aligned to the mouse reference (mm10) using STAR (version 2.7.5a; parameters: --runMode alignReads --chimSegmentMin 20 --outSAMstrandField intronMotif --quantMode GeneCounts) and transcripts were assembled using StringTie (version 2.0.6; parameters: -e) with GENCODE annotation (VM19) (Dobin et al., 2013; Pertea et al., 2015). For the repeat expression quantification, reads were re-aligned with additional parameters '--outFilterMultimapNmax 50'.

2.20. TT-SLAM-Seq processing

Raw TT-SLAM-Seq reads were trimmed by quality, Illumina adapter content and poly-A content similar to RNAseq processing and aligned using STAR with parameters '--outFilterMultimapNmax 50 --outReadsUnmapped Fastx' to the SILVA database (download: March 6, 2020) in order to remove rRNA content (Quast et al., 2013). Unaligned reads were afterwards reverse-complemented using the seqtk 'seq' command (https://github.com/lh3/seqtk, version 1.3-r106; parameters: -r). Reverse-complemented reads were processed using SLAM-DUNK with the 'all' command (version 0.4.1; parameters: -rl 100 -5 0) with the GENCODE gene annotation (VM19) as '-b' option (Neumann et al., 2019). Reads with a 'T>C' conversion representing nascent transcription were filtered from the BAM files using alleyoop (provided together with SLAM-DUNK) with the 'read-separator' command. Counts per gene were quantified based on the 'T>C'-converted reads using htseq-count (version 0.11.4; parameters: --stranded=yes, --nonunique=all) (Anders et al., 2015). FPKM values were calculated based on the resulting counts. Technical replicates were merged using samtools 'merge' to obtain genome-wide coverage tracks (Ramírez et al., 2014). Single and merged replicate coverage tracks were generated using deepTools bamCoverage (version: 3.4.3; parameters: --normalizeUsing CPM) separately for the forward and reverse strand based on the 'T>C'-converted reads (Ramírez et al., 2014).

2.21. Public ChIP-Seq data

Fastq files of public ChIP-Seq data were downloaded from GEO using following IDs: H3K27Ac - GSM1526287, H3K9me3 - GSM1429903, HP1α - GSM1375159, TRIM28 - GSM1555120, OCT4 - GSM1082340, SOX2 - GSM1082341 and NANOG - GSM1082342 (Bulut-Karslioglu et al., 2014; Elsässer et al., 2015; Ji et al., 2015; Whyte et al., 2013).

2.22. ChIP-Seq processing

Adapter and quality trimming of raw reads from treatment and input samples was performed using cutadapt (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC -A AGATCGGAAGAGC). BWA with the 'mem' command (version 0.7.17, default parameters) was used to align reads individually to the mouse genome (mm10) and to the fly genome (D. Melanogaster, dm6) (H. Li & Durbin, 2009). A sorted BAM file was obtained and indexed using samtools with the 'sort' and 'index' commands (version 1.10). Duplicate reads were identified and removed using GATK (version 4.1.4.1) with the 'MarkDuplicates' command and default parameters (McKenna et al., 2010). Technical replicates of treatment and input samples were merged respectively using samtools 'merge'. Peaks were called with reads aligning to the mouse genome only using MACS2 'callpeak' (version 2.1.2; parameters --bdg --SPMR) using the input samples as control samples (Zhang et al., 2008). For H3K9me3 only, the '--broad' option was used. Genome-wide coverage tracks for single and merged replicates normalized by library size were computed using deepTools bamCoverage (parameters: --normalizeUsing CPM --extendReads) and in addition normalized by the spike-in factor obtained from the reads aligning to the Drosophila genome as described (Orlando et al., 2014).

Publicly available paired-end ChIP-Seq data (HP1α) were processed as described above except spike-in and input normalization.

Publicly available single-end ChIP-Seq data (H3K9me3, TRIM28, H3K27ac, NANOG, OCT4, SOX2) were trimmed using cutadapt (version 2.4; parameters: --nextseq-trim 20 --overlap 5 --minimum-length 25 --adapter AGATCGGAAGAGC), and aligned and de-duplicated analogous to the paired-end data. In order to estimate the fragment size, only reads with a mapping quality of at least 15 were considered and used as input for spp (version 1.2.2) (Kharchenko et al., 2008). All reads were then used to create

coverage tracks using bamCoverage and to call peaks with MACS2 'callpeak' (version 2.1.2; parameters --bdg --SPMR --nomodel) with '--extsize' set to the estimated fragment length divided by two and input samples used as control for H3K27ac, NANOG, OCT4 and SOX2. For H3K9me3 only, the '--broad' and '--nolambda' options were used.

The co-ordinates displayed for ChIP-Seq tracks in figures all correspond to mm10 genome assembly co-ordinates.

2.23. Enhancer and super-enhancer annotation

The annotation of super-enhancers, enhancers and enhancer constituents was adopted from the original publication (Whyte et al., 2013). Coordinates were lifted from mm9 to mm10 using UCSC liftOver. These coordinates were used throughout this study for all enhancer-associated analyses (Supplementary Table 2 in Asimi et al., 2022).

2.24. Detection of eRNA expression and preparation of TT-SLAM-Seq histograms

TT-SLAM-Seq signal was quantified at typical enhancer and SE constituents. Enhancers within 3kb of TSS were excluded and intragenic enhancers were only considered on opposing strand relative to genes (GENCODE VM19, excluding "bidirectional_promoter_lncRNA") to measure eRNA without interference from gene transcription. Gene coordinates were extended for 10kb from transcription termination site to prevent overlap of transcriptional readthrough signal. The detection of eRNA was performed at 2 kb regions centered by the constituent coordinates. TT-SLAM-Seq counts were quantified from .bam files with htseq-count (parameters: --stranded = yes, --nonunique = all). Differential expression analysis was performed on enhancer elements with DEseq2 (parameters: test="LRT", reduced=~1) and excluding regions

with less than 10 counts across samples (Love et al., 2014). To visualize fold changes, regions with low eRNA expression (FPKM < 1) were excluded (Fig. 3.7a). In Fig. 3.7a, enhancer constituents with significant transcription (FPKM > 1) are considered (n = 117 for super-enhancers, n = 153 for typical enhancers).

Quantifications for eRNA histograms were done with HOMER software (version 4.10) and only intergenic enhancers were considered (Heinz et al., 2010). Reads from .bam files were prepared for quantifications with 'makeTagDirectory', three replicates were merged, and histograms for enhancer and SE constituents were prepared with 'annotatePeaks.pl' (parameters: -size 4000, -hist 50) and metagene plots for SEs with 'makeMetaGeneProfile.pl' (parameters: -min 500, -size 4000). Histograms were smoothened by taking a rolling mean with window size 5 for enhancer constituents and 3 for SEs before plotting. To visualize TT-SLAM-Seq signal across genes, protein-coding genes were considered and metagene plot was prepared using 'makeMetaGeneProfile.pl' after removing outliers (FPKM <0.5 or >100).

To compare relative levels of eRNAs and transcripts derived from repeat elements (Fig. 3.10), reads were prepared for quantifications using 'makeTagDirectory' with -keepOne option, allowing multimapping reads to be considered only once. Reads at intergenic retrotransposons belonging to LTR class (Repbase) were quantified using 'analyzeRepeats.pl' and LTRs with less than 5 reads across 0h and 24h samples were discarded (Jurka et al., 2005). To analyze putative upregulated LTRs ("UP LTRs"), differential expression was estimated using DEseq2 (parameters: test="LRT", reduced=~1). LTRs with fold change above 2 and adjusted p-value < 0.05 were considered. To visualize TT-SLAM-Seq signal, coverage tracks from both strands were prepared from HOMER TagDirectories with 'makeUCSCfile', converted to bigwig format using 'bedGraphToBigWig' and metaprofile plots were prepared using EnrichedHeatmap as described for ChIP-Seq (Gu et al., 2018).

To quantify eRNA expression from total RNA-Seq data (Fig. 4c), aligned reads were prepared for quantification with HOMER 'makeTagDirectory' and quantified with 'analyzeRepeats.pl' using the enhancer constituents defined above. Mean FPKM values from three replicates were calculated for each enhancer and regions with low read counts (FPKM < 0.05) were excluded from analysis.

2.25. Differential gene expression analysis

Only protein coding genes were included in differential gene expression analysis. Differential expression for RNAseq and TT-SLAM-Seq samples was measured using DESeq2 (parameters: test="LRT", reduced=~1) based on the raw counts per gene considering all time points per experiment type in one design. Only genes with at least 10 reads across all samples of the same experiment type were considered for the analysis. Genes with an absolute log2 fold change greater than 1 in comparison to the DMSO control and an adjusted p-value of less than 0.05 were termed differentially expressed per time point. Lowly expressed genes across all time points (average TPM across all RNAseq samples of less than 0.5 or average FPKM across all TT-SLAM-Seq samples of less than 0.25) were excluded from the analysis.

2.26. Retrotransposon element definition

The genome-wide retrotransposon annotation of LTR, LINE and SINE elements was downloaded from Repbase (Jurka et al., 2005). Based on the Repbase classification system, we used the element annotation as LTR, LINE or SINE as the retrotransposon classes. Retrotransposon families considered in this study were L1 and L2 elements (LINE), ERV1, ERV3, ERVK, ERVL and MALR (LTR), as well as Alu, B2, B4 and MIR elements (SINE). Repeat subfamilies used in this thesis were subdivided into IAP, MMERVK and MMETn (ERVK) elements. IAPs and MMERVKs consist of multiple different subfamilies as annotated by Repbase which were included under these broader subfamilies. The classification is consistent with retrotransposon classification

described in previous studies (Crichton et al., 2014; Stocking & Kozak, 2008; Thompson et al., 2016).

Full length retrotransposons were defined based on the Repbase repeat annotation. For full length ERVK elements, we required the element to consist of an inner part with two flanking LTRs. First, elements annotated as inner parts (containing the keyword 'int') were merged if they belonged to the same subfamily and were located within maximal 200 base pairs of each other. Second, only the merged inner parts with an annotated ERVK LTR within a distance of at most 50 base pairs on each side were selected as full-length element candidates. For IAPs specifically only LTRs that belonged to an IAP subfamily were considered. No size restrictions were applied on the inner parts or LTRs, which could lead to potential false positive candidates that are too truncated to be able to be transcribed, but on the other hand, provides an unbiased definition of full-length repeat elements. The subfamily per element was defined based on the inner part.

Inner parts flanked by only one LTR were termed half-length elements. LTRs without an inner part were termed solo LTRs.

To provide an overview of potential full-length L1 elements, exclusively annotated elements with a size of more than 6 kb were shown.

The genomic co-ordinates of retrotransposons can be retrieved from Supplementary Table 2a-e in Asimi et al., 2022.

2.27. ChIP-Seq enrichment analysis

Bigwig tracks of ChIP-Seq signal were imported into R with the package rtracklayer (Lawrence et al., 2009). Enriched heatmaps and metaprofile plots of ChIP-Seq signal

were generated using the R package EnrichedHeatmap (Gu et al., 2018). For this purpose, the signal was normalized to genomic features using the function 'normalizeToMatrix' (parameters: extend = c(2000, 2000), mean_mode = "w0", w = 50, target_ratio = 0.25). The resulting data matrix was visualized using the function 'EnrichedHeatmap'.

2.28. Motif enrichment

Enrichment of motifs was calculated using ame (version 5.3.0, default parameters) (McLeay & Bailey, 2010). The sequences of 5' full length IAP LTRs and full length IAP inner parts were analyzed separately using the sequences of super-enhancer constituents as control (Fig. 3.22). In Fig. 3.22, the top TFs whose motifs show enrichment in the IAP LTRs or inner parts are shown. Also displayed is the expression level of the TFs calculated from the RNA-Seq data. For further functional tests, NFY was selected as its motif is highly enriched in IAPs, and it is expressed above 50 TPM in mESCs.

2.29. Retrotransposon expression quantification

Global repeat expression quantification from RNAseq, TT-SLAM-Seq and scRNA-seq (Fig. 3.7b, 3.16a-b) was carried out as described previously (Grosswendt et al., 2020). Briefly, in order to estimate the expression for each retrotransposon subfamily without bias due to gene expression, only reads not overlapping any gene were considered for the analysis. Reads overlapping splice sites as well as reads with a high poly-A content were removed. The remaining reads were counted per subfamily only if they aligned uniquely or multiple times to elements of the same subfamily. Here, any annotated element of a specific subfamily from Repbase was considered independent of our full length ERVK annotations. Reads aligning to multiple elements were only counted once. For scRNA-seq samples, reads were counted per subfamily, sample and cell state.

The number of reads per subfamily were normalized by library size for RNAseq and TT-SLAM-Seq samples, and normalized by reads aligning to genes and repeats for scRNA-seq samples. Fold changes were calculated with respect to the DMSO or wild-type samples.

2.30. Statistics and reproducibility

Raw data and custom code are available at: https://doi.org/10.5281/zenodo.6521914 (Asimi et al., 2022).

For all RNA-FISH combined with IF experiments, the target combination of gene transcript and transcriptional co-factor was probed on one coverslip of mESCs and at least two viewpoints were acquired. Number of detected foci included in the radial plot analysis is indicated under N_{foci} in Fig. 3.11, Fig. 3.20, Fig. 3.25, Fig. 3.27. and Fig. 3.30. In Fig. 3.15a-b, n indicates the number of analyzed nuclei collected from at least three viewpoints; the total number of detected IAPez foci is indicated in Fig. 3.16 (1774 for RNAPII and 2735 for MED1). Colocalizing foci (distance $<$ 200 nm) from Fig. 3.15a-b are indicated in Fig. 3.16 (344/1774 for RNAPII and 381/2735 for MED1).

IAP RNA FISH – RNAPII IF experiments were repeated three times. Images and analysis of one representative experiment are visualized in Fig. 3.15a. Additional data from a replicate experiment is shown in Fig. 3.14. *IAP* RNA FISH – MED1 IF images were obtained from one biological replicate staining experiment (Fig. 3.15b).

For 1-6 HD-treatment RNAPII IF experiment (Fig. 3.21), and *Cthrc1* RNA FISH–NRF1 IF (Fig. 3.25) images are from one biological replicate staining experiment.

3 Results

3.1 Rapid and selective degradation of TRIM28 in mESCs

Previous studies of TRIM28 included Cre-loxP knockouts in mESCs and embryos, and brought important insights into TRIM28 function, but lacked kinetic resolution to resolve immediate effects on the transcriptional program of mESCs (Cammas et al., 2000; Rowe et al., 2010). To investigate direct effects of ERV derepression, we used the dTAG system to engineer an mESC line that endogenously expressed degradation-sensitive TRIM28-FKBP alleles (Nabet et al., 2018). The dTAG system uses CRISPR-mediated locus-specific knock-in of $FKBP12^{F36V}$ in frame with the protein of interest, which can then be targeted by the ubiquitin/proteasome system via a bifunctional degrader (dTAG-13) (Fig. 3.1a-b).

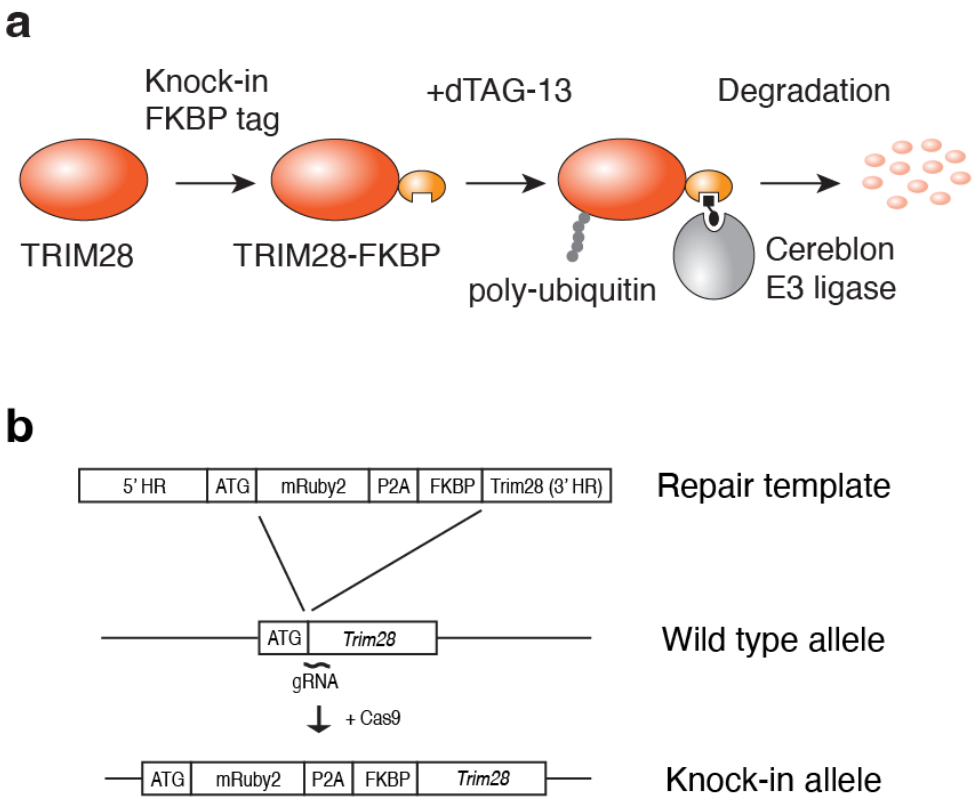


Figure 3.1. TRIM28-FKBP knock-in strategy a. Scheme of TRIM28 dTAG system in mESCs. **b.** Scheme of FKBP knock-in strategy at the Trim28 locus.

Depletion of the targeted protein is acute and reversible, thus providing an effective strategy to study the immediate consequence of protein loss (Nabet et al., 2018). Cells expressing endogenously tagged TRIM28 showed reversible and inducible proteolysis with near-complete degradation after 6h exposure to the dTAG-13 ligand (Fig. 3.2).

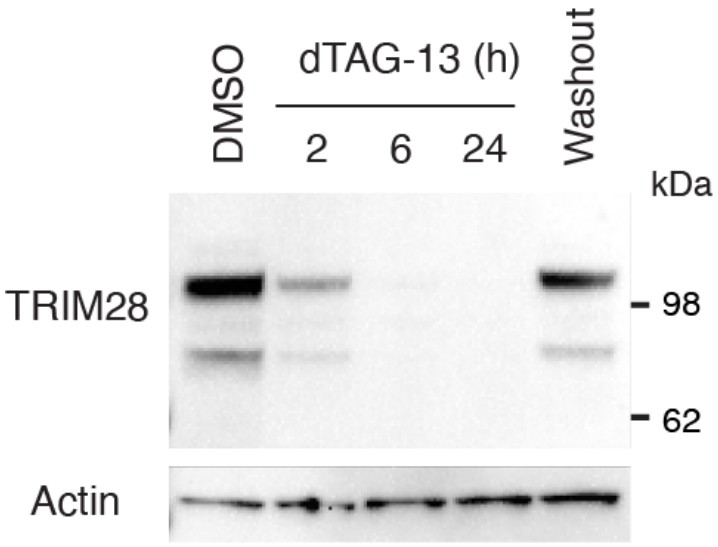


Figure 3.2. Acute and reversible degradation of TRIM28 in mESCs. Western blot validation of the FKBP degron tag and its capacity to acutely and reversibly degrade TRIM28. Protein levels in control mESCs are compared to 2h, 6h, 24h of 500nM dTAG-13 treatment and reversibility is demonstrated using washout.

TRIM28 degradation did not significantly affect protein levels of pluripotency markers including OCT4 and SOX2. Time series analysis of protein levels, consisting of 2h, 6h, 8h, 12h, and 24h of dTAG-13 treatment validated rapid TRIM28 depletion and did not significantly affect levels of OCT4 and SOX2, suggesting that up to 24h of TRIM28 depletion does not affect pluripotency (Fig. 3.3a). In addition, quantitative mass-spectrometry validated that TRIM28 depletion was highly selective up to 24h of dTAG-13 treatment (Fig. 3.3b). These validations of the TRIM28 degradation demonstrated a fast and selective system to investigate effects of acute TRIM28 depletion in mESCs.

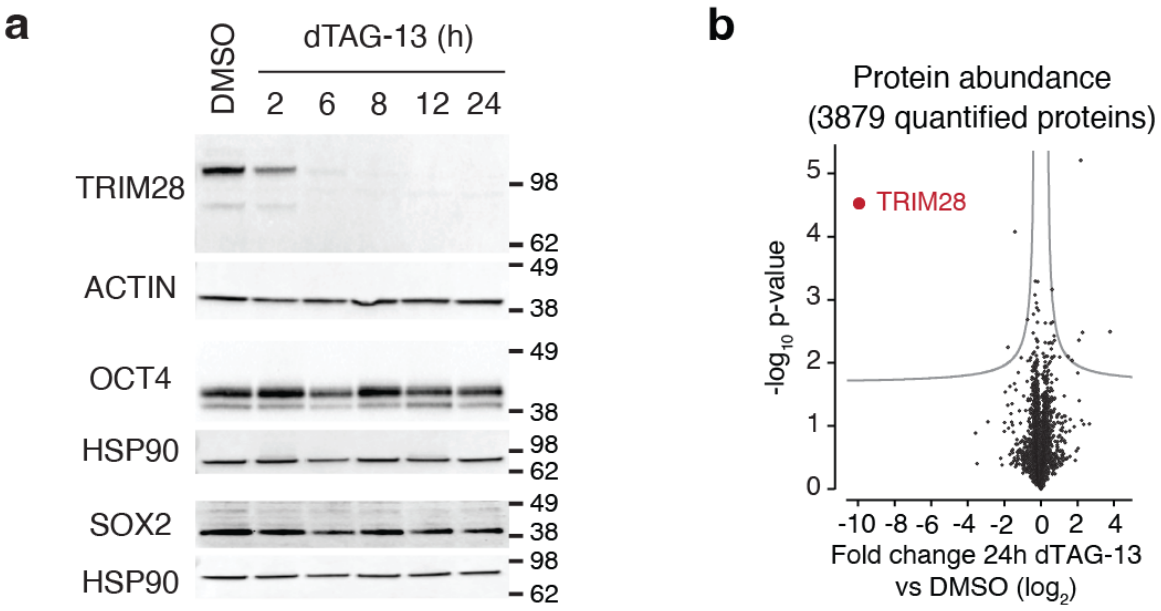


Figure 3.3. Selective and acute TRIM28 degradation does not significantly impact levels of pluripotency factors in mESCs. a. OCT4 and SOX2 protein levels in TRIM28 degraded cells with matching controls (ACTIN and HSP90) upon 2h, 6h, 8h, 12h, and 24h of dTAG-13 treatment. **b.** Mass spectrometry analysis of protein levels after 24h of dTAG-13 treatment. Every dot corresponds to a quantified protein. The degradation appears highly selective for TRIM28. Mass spectrometry experiment was performed with Abhishek Sampath Kumar.

3.2 Reduced SE transcription in TRIM28-degraded mESCs

To inspect changes in transcriptional activity following TRIM28-degradation, we used TT-SLAM-Seq, a method that combines transient transcriptome sequencing (TT-seq) with metabolic sequencing of small RNA using thiol-linked alkylation (SLAM-seq) (Figure 3.4a) (Herzog et al., 2017; Reichholf et al., 2019; Schwalb et al., 2016). Using this approach that utilizes 4sU nucleotide conversion, unlabeled transcripts that often contaminate RNA libraries generated by TT-seq can be bioinformatically filtered allowing a more precise measurement of nascent transcripts (Reichholf et al., 2019). TT-SLAM-seq detects strand-specific nascent transcription at genes, enhancers, and super-enhancers (Figure 3.4b).

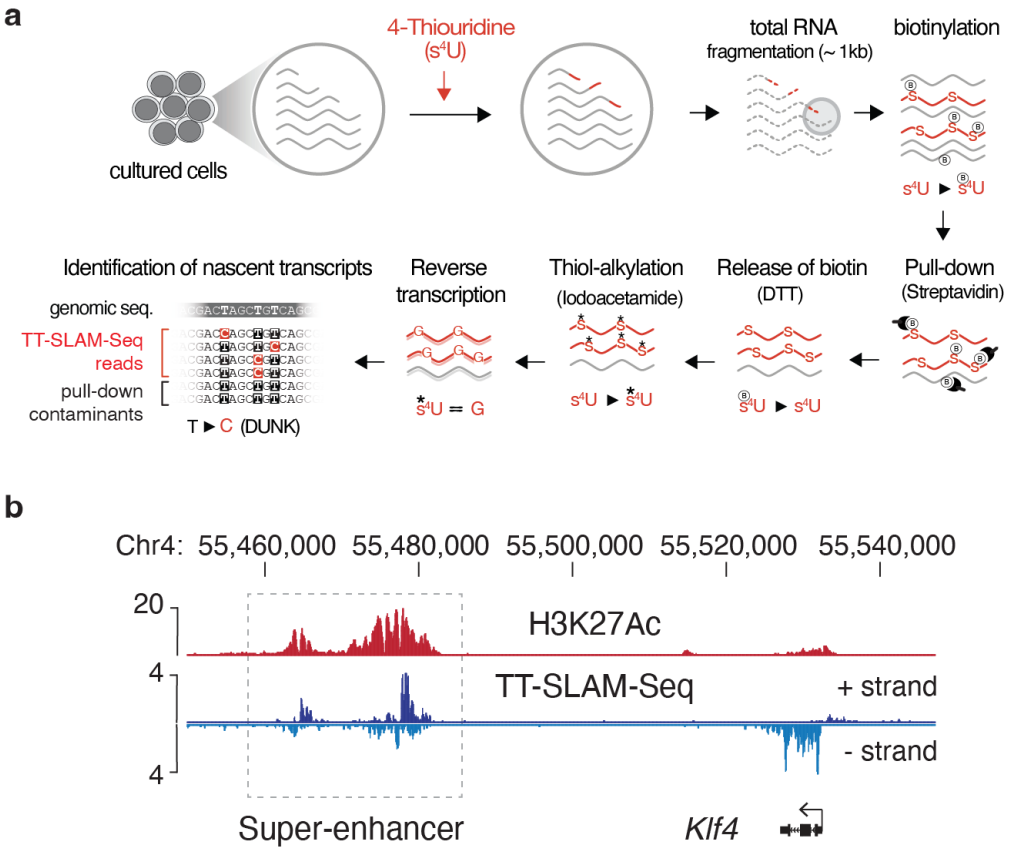


Figure 3.4. TT-SLAM-Seq detects nascent transcription. a. Schematic overview of the TT-SLAM-Seq experimental and analytical procedure. **b.** TT-SLAM-Seq and H3K27Ac ChIP-Seq browser tracks at the Klf4 super-enhancer locus. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates. TT-SLAM-seq experiment was performed with Henri Niskanen.

Closer inspection of the Klf4 SE locus revealed significant reduction in nascent transcription of the SE and reduction in Klf4 gene transcription upon 2-6h and 24h of TRIM28 degradation, respectively (Fig. 3.5). H3K27ac, OCT4, SOX2 and NANOG occupancy tracks from publicly available datasets were used to demonstrate enrichment typical for SEs in mESCs. TT-SLAM-seq data indicated transcription at selected SEs and SE controlled genes is reduced rapidly upon TRIM28 degradation, and progressively reaches minimal levels at 24h.

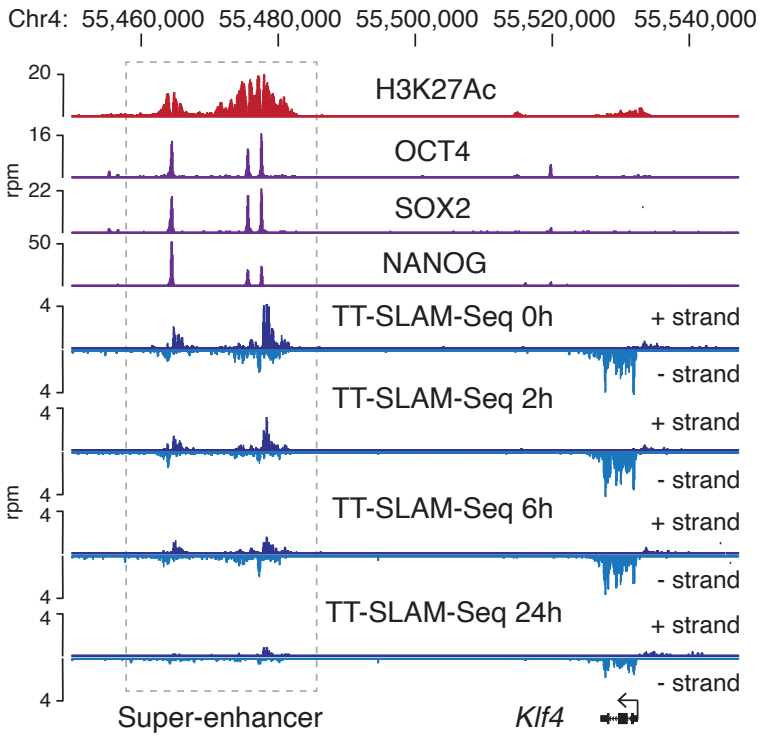


Figure 3.5. TRIM28 degradation leads to reduction of SE transcription. a. Genome browser tracks of ChIP-Seq data (H3K27Ac, OCT4, SOX2, NANOG) in control mESCs and TT-SLAM-Seq data upon 0h, 2h, 6h and 24h dTAG-13 treatment, at the Klf4 locus. TT-SLAM-seq experiment was performed with Henri Niskanen.

Genome-wide nascent transcriptome analysis showed there were 260 genes that were significantly induced, and 328 genes that were significantly reduced after 24h of TRIM28 degradation (>2-fold, FDR<0.05) (Fig. 3.6). Gene transcription induction effects were expected since TRIM28 is a co-repressor protein with a role in H3K9me3 mediated gene silencing, but there were equally as many genes that were significantly reduced. Many of the downregulated genes, for instance, Mycn, Fgf4, and Sox2, included genes relevant for pluripotency indicating that TRIM28 depletion influenced pluripotency genes at the transcriptional level.

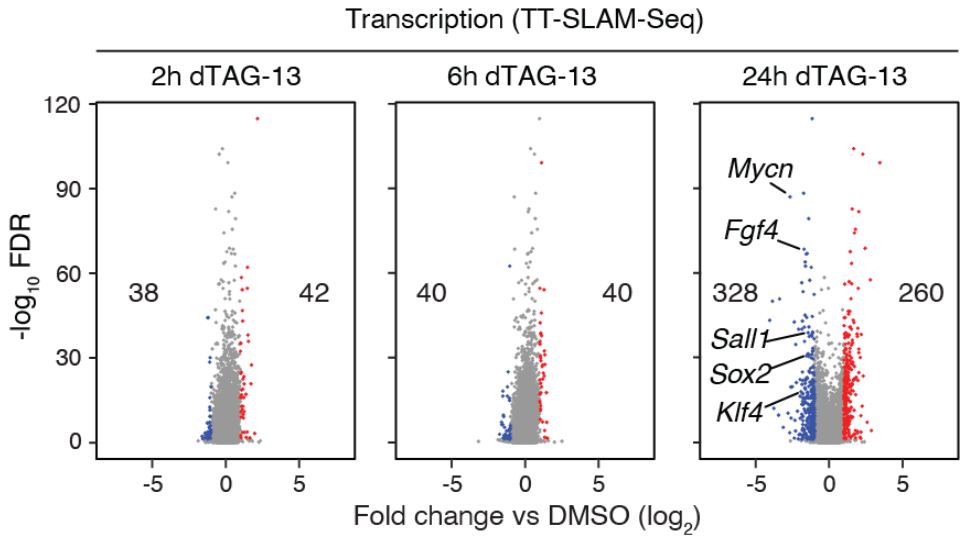


Figure 3.6. TRIM28 degradation leads to genome-wide changes in nascent transcription. Fold change of gene transcription (TT-SLAM-Seq data) upon dTAG-13 treatment. The number of significantly de-regulated genes (DESeq2), and example pluripotency genes are highlighted. TT-SLAM-seq experiment was performed with Henri Niskanen.

In addition, transcription was globally reduced at SEs (Fig. 3.7a). At the same timepoint (24h), TT-SLAM-seq and RNA-seq showed IAP transcription induction by several fold (Fig. 3.7b). This evidence indicated that TRIM28 degradation leads to ERV derepression and reduction of SE transcription, followed by downregulation of SE-controlled genes.

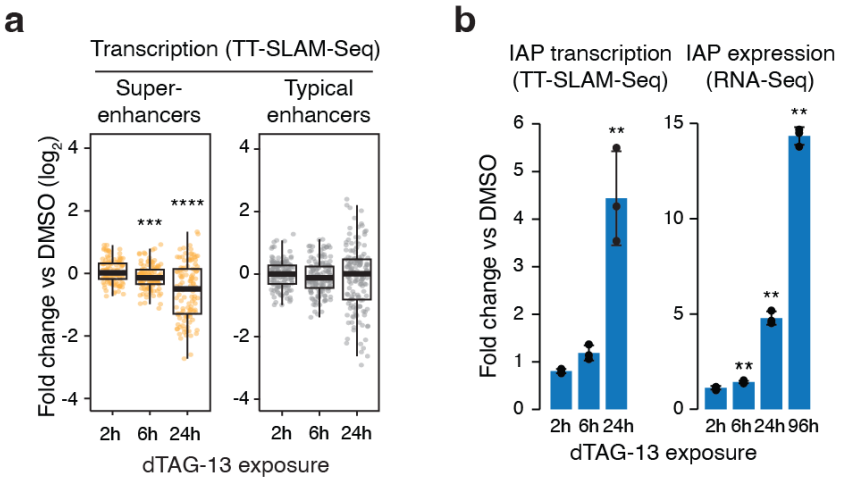


Figure 3.7. TRIM28 degradation leads to reduction of nascent SE transcription and IAP derepression. a. Change (Log2 fold) in TT-SLAM-Seq read density at super-enhancers and typical enhancers upon dTAG-13 treatment (normalized to DMSO-treated control mESCs.) P-values are from two-sided Wilcoxon-Mann-Whitney tests. ****: P=5*10-8, ***: P=5*10-4. Following elements are shown in boxplots: middle line - median; box limits - upper and lower quartile; whiskers - 1.5x interquartile range. TT-SLAM-seq analysis was performed with Henri Niskanen. **b.** Fold change in read density of TT-SLAM-Seq and RNA-Seq data after the indicated duration of dTAG-13 treatment, normalized to the level in DMSO control. Data are presented as mean values +/- SD from three biological replicates. P-values are from unpaired two-sided t-tests. **P<0.01 (Asimi et al., 2022). TT-SLAM-seq analysis was performed with Henri Niskanen.

Additional SE loci (including miR290-295) showed striking reductions in SE and SE-driven gene transcription (Fig. 3.8). Similar to the Klf4 locus, miR290-295 SE showed progressive reduction in nascent transcription upon TRIM28 depletion.

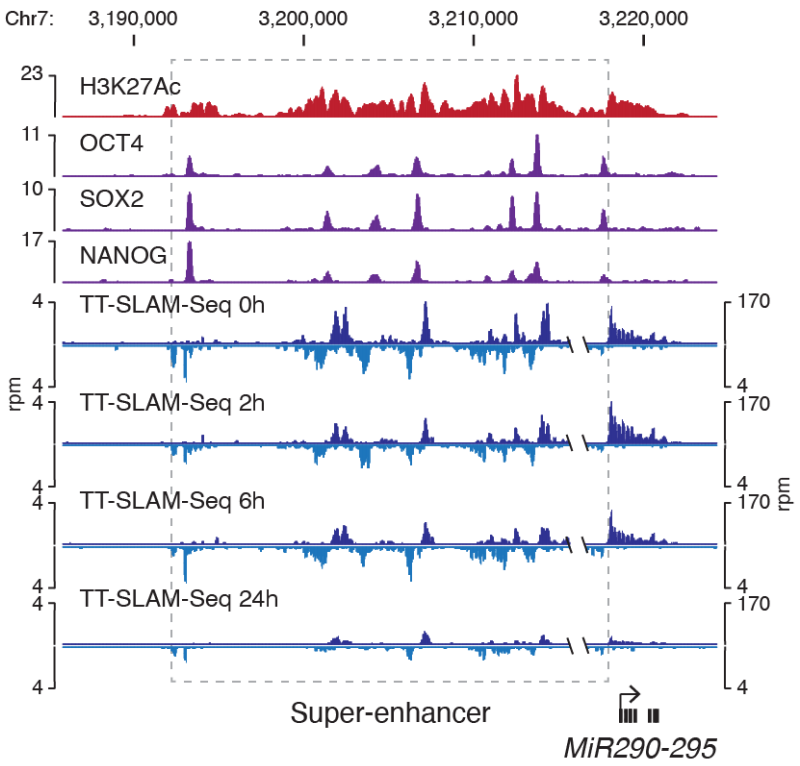


Figure 3.8. TT-SLAM-seq validates the effects of TRIM28 degradation on nascent transcription at the miR290-95 SE. a. Acute reduction of transcription at the miR290-295 super-enhancer locus upon TRIM28-degradation. Displayed are genome browser tracks of ChIP-seq data (H3K27Ac, OCT4, SOX2, NANOG) in control mESCs, and TT-SLAM-seq data upon 0 h, 2 h, 6 h and 24 h dTAG-13 treatment at the miR290-295 locus. Rpm: reads per million. Co-ordinates are mm10 genome assembly co-ordinates.

Supplementary qRT-PCR analysis confirmed reduction of SE transcription, indicating it is likely more dynamic and precedes reduction in transcription of SE controlled genes (Fig. 3.9). At both miR290-295 and Kl4 SE locus, transcription reduction of SE levels was stronger than the reduction of the gene itself.

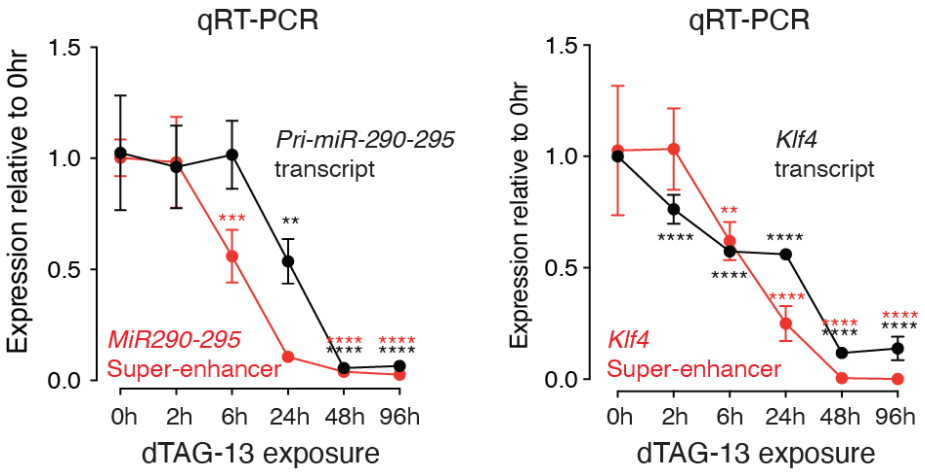


Figure 3.9. qRT-PCR validation of the TT-SLAM-seq data at the miR290-295 and Klf4 loci. Displayed are transcript levels after the indicated duration of dTAG-13 treatment. Values are displayed as mean ± SD from three independent experiments and are normalized to the level at 0 h. P values are from two-tailed t-tests. ****: $P < 10{-}4$, ***: $P < 10{-}3$, **: $P < 10{-}2$, *: $P < 0.05$.

Global analysis of intergenic SEs revealed a reduction of nascent transcription upon 24h of TRIM28 degradation – this reduction was more moderate at typical enhancers. Differentially expressed LTRs on the other hand, displayed noticeable induction of nascent transcription (Fig. 3.10). These analyses further validated the effect of TRIM28 degradation on nascent transcription of SEs and ERVs.

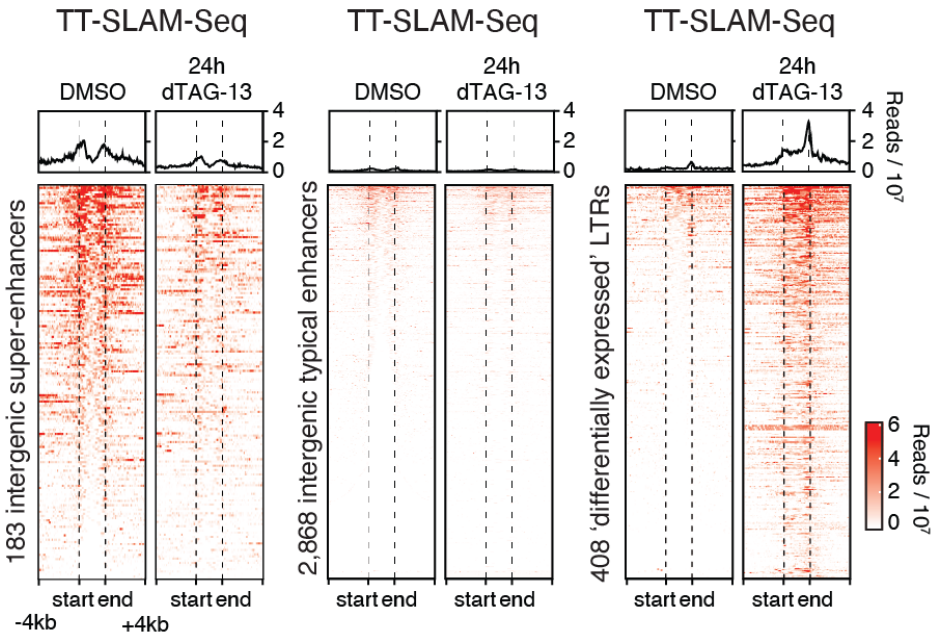


Figure 3.10. Global analysis of TRIM28 depletion effects on nascent transcription at intergenic SEs, intergenic typical enhancers and de-repressed LTRs. Displayed are TT-SLAM-seq read densities from both strands within 4 kb around the indicated sites. The genomic features (the middle part of the plot) were length normalized. Meta-analyses of the mean signals are displayed above the heatmaps. TT-SLAM-seq analysis was performed with Henri Niskanen.

3.3 Reduced SE-condensate association in TRIM28-degraded mESCs

SEs in ESCs are occupied by TFs and coactivators including Mediator complex. Considering that actively transcribed loci are associated with transcriptional condensates formed by these factors and RNAPII, we were interested in resolving what occurs at SE loci upon TRIM28 degradation and the subsequent reduction in nascent transcription. We used a high-resolution microscopy approach that combines RNA Fluorescence in situ hybridization (FISH), that labels the nascent RNA produced at the specific SE-controlled locus, with immunofluorescence (IF), to estimate presence of transcriptional condensates at the labeled locus. RNA FISH probes were purposedly designed against introns of SE-driven genes so that the fluorescence signal is enriched

for nascent RNAs (active transcription sites) and not for mature mRNA. We imaged the miR290-295 and Fgf4 loci and observed clear RNA FISH signal that tended to colocalize with RNAPII puncta, frequently with significant overlap (Fig. 3.11a-b). Upon 24h of TRIM28 degradation, RNA FISH signal was still detectable but colocalized with RNAPII less frequently, with most RNA FISH foci not directly overlapping high intensity RNAPII puncta (Fig. 3.11a-b, lower left panels). Metaplots that represent computed average colocalizations between RNA FISH and IF signal in 3D and across multiple nuclei, showed that there is a reduced frequency of colocalization in 24h dTAG-treated cells (Fig. 3.11a-b, lower right panels).

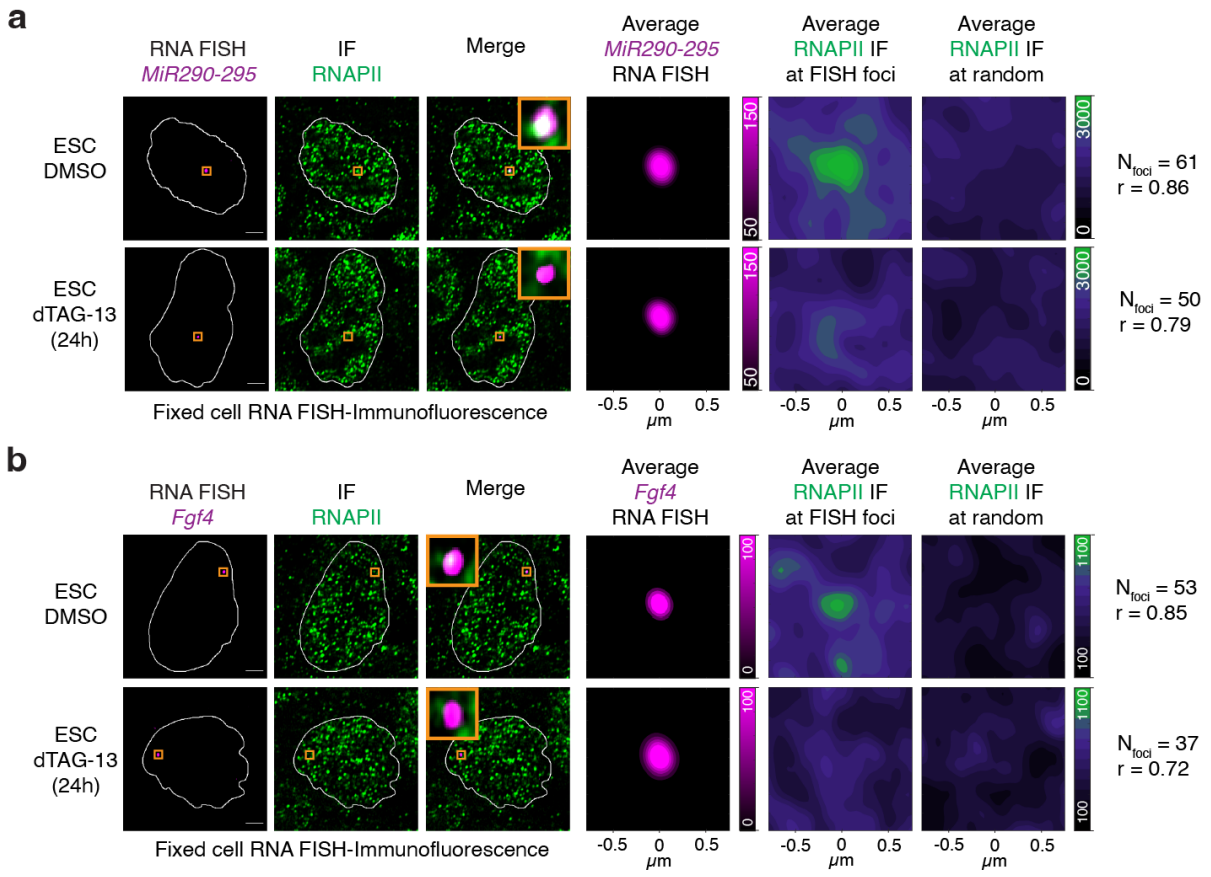


Figure 3.11. Reduced SE-condensate association in TRIM28-degraded mESCs. a. and b. Representative images of individual z-slices (same z) of RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery was determined using DAPI staining and labeled using the white contour. Also shown are averaged signals of either RNA FISH or RNAPII IF centered on the FISH foci or randomly selected nuclear positions. r denotes a Spearman's correlation coefficient. Scale bars: 2.5μm.

Further quantification of mean RNAPII IF fluorescence intensities at miR290-295 and Fgf4 RNA FISH signal at all foci detected in Fig. 3.11a-b showed statistically significant reduction in 24h dTAG-treated cells (Fig. 3.12a-b, left panels). Mean RNAPII IF intensities at random nuclear position were not significantly different between the treated and control cells (Fig. 3.12a middle panel and Fig. 3.12b right panel), indicating that the levels of RNAPII are not affected by TRIM28 degradation. Additionally, we generated 3D projections and measured distances of the miR290-295 RNA FISH foci from the nearest RNAPII puncta and observed statistically significant increase in distance in 24h dTAG-treated cells (Fig. 3.12a right panel). These data demonstrated reduced association of SE condensates with RNAPII in TRIM28-degraded mESCs.

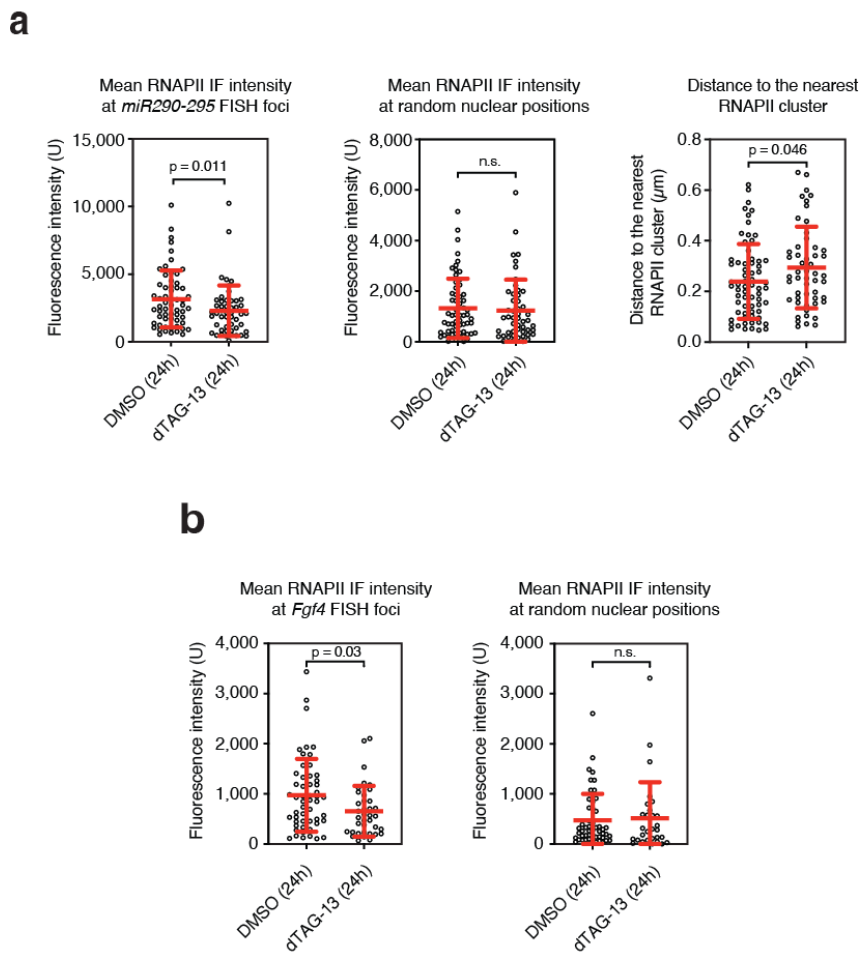


Figure 3.12. Additional quantifications of RNAPII colocalization at the miR290-295 and Fgf4 locus. a-b. Further analyses of cells used in Fig. 3.11a. (top) and Fig. 3.11b. (bottom) RNAPII IF intensity at the miR290-295 FISH foci (nDMSO = 61, ndTAG-13 = 50). (middle) RNAPII mean fluorescence intensity at random nuclear positions (nDMSO = 61, ndTAG-13 = 50). (right) Distance between the FISH focus and the nearest RNAPII puncta (nDMSO = 67, ndTAG-13 = 53). Data presented as mean values ± SD from one staining experiment. P values are from two-sided Mann-Whitney tests. NS: not significant.

To gain additional insights into colocalization between RNAPII and SEs, and mitigate limitations of fixed-cell confocal microscopy (which include potential fixation artefacts, antibody specificity issues and the spatial resolution limit of ~250nm), we combined the TRIM28 dTAG system with a mESC cell line compatible with live-cell super-resolution photoactivated localization microscopy (PALM) (Cho et al., 2018). To

visualize nascent transcription, this cell line contains endogenously integrated MS2 stem loops at the Sox2 gene; and to track transcriptional machinery it expresses RNAPII that is endogenously labeled with Dendra2, a green-to-red photo-convertible fluorophore (Cho et al., 2018). It additionally stably expresses MS2 capsid protein (MCP) fused to a SNAP-Tag, which is used to visualize MS2 stem loops. Resulting setup allows simultaneous TRIM28 degradation (Fig. 3.13a-b), nascent transcription tracking of the SE-controlled Sox2 gene, and RNAPII tracking in super resolution in live cells.

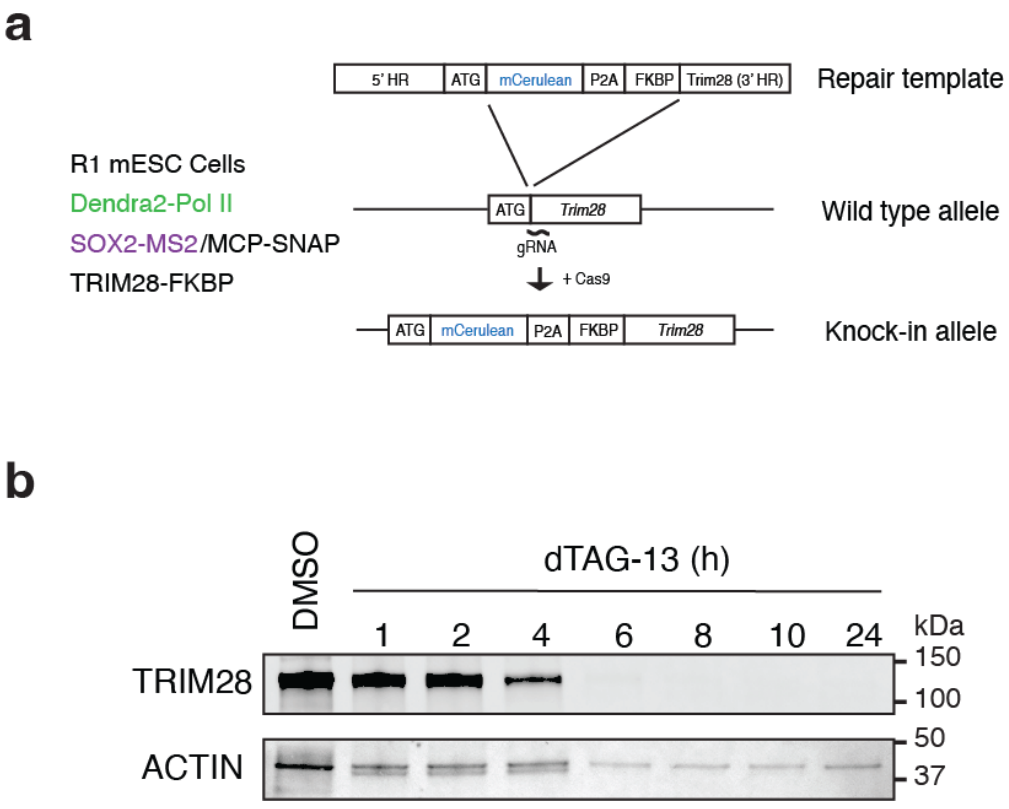


Figure 3.13. PALM microscopy cell line setup. a. Scheme of FKBP knock-in strategy in the R1 mESCs used for PALM experiments. MS2 stem loops are endogenously integrated at the Sox2 gene, RNAPII is endogenously labeled with Dendra2, and TRIM28-FKBP alleles allow TRIM28 degradation **b.** TRIM28 Western blot in the R1 mESCs. Western blot was done once. Live-cell PALM microscopy quantifications are reported in Asimi et al., 2022.

RNAPII clusters were recorded for 2 min, and size and distance of the RNAPII cluster closest to the Sox2 locus were measured. In 24h TRIM28-degraded cells, RNAPII clusters were located further away from the Sox2 locus and smaller in size (than those found in control cells (Asimi et al., 2022). Global number and size of RNAPII clusters did not change. Together with RNA FISH-IF data, this confirmed TRIM28 degradation reduced association of SEs with RNAPII condensates in both live and fixed mESCs.

3.4 Derepressed ERVs form nuclear foci that overlap with RNAPII condensates

Combining TRIM28-dTAG system with TT-SLAM-seq and high-resolution RNA FISH-IF proved to be a powerful tool for the dissection of dynamic regulation of nascent transcription and condensate localization at SEs. Since transcriptional condensates associate with SEs less frequently upon TRIM28 degradation, and RNAPII levels seem to be unaffected, this raises a question about the redistribution of transcriptional condensates to other genomic loci. TRIM28 silences ERVs so their derepression is expected upon TRIM28 degradation. To establish if transcriptional condensates are associating with derepressed ERVs, I tested the colocalization of IAP ERVs with RNAPII via RNA FISH-IF.

I targeted a consensus *IAPez* sequence using RNA FISH probes. In DMSO treated cells, RNA FISH signal was weakly present and difficult to discern from the background (Fig. 3.14, top row). Following prolonged dTAG-13 treatment, discrete nuclear *IAPez* foci started appearing. After 24h of TRIM28 degradation, I detected tens of nuclear *IAPez* of variable intensity that frequently colocalized with RNAPII (Fig. 3.14, middle row). Additional dTAG-13 treatment (48h total) yielded higher intensity foci that were easy to identify and distinguish from the background (Fig. 3.14, bottom row).

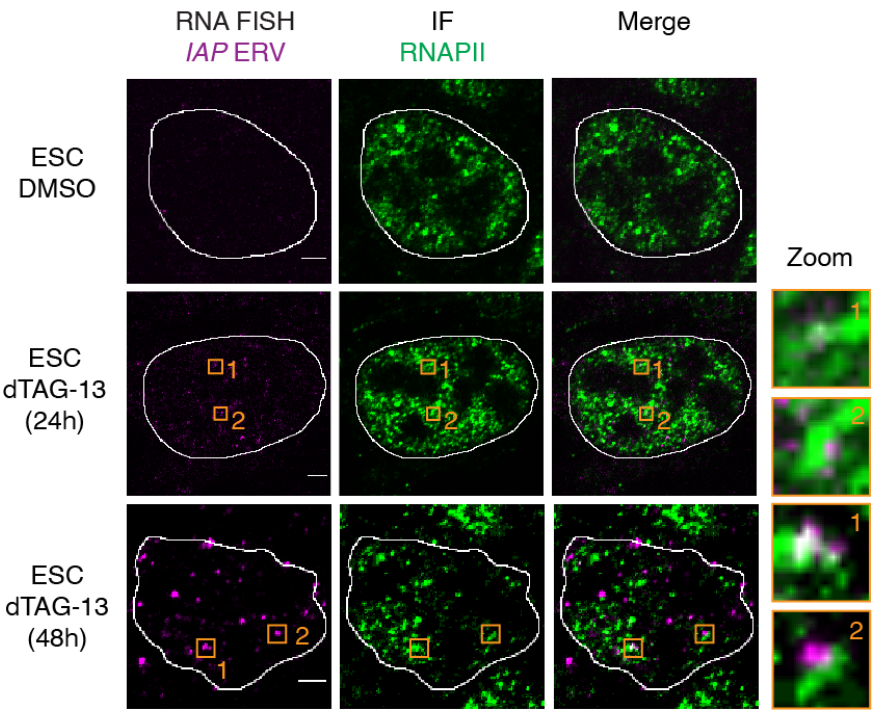


Figure 3.14. TRIM28 degradation induces derepression of IAPs. Prolonged dTAG-13 treatment progressively increases the intensity of *IAPez* RNA FISH foci. Representative images of individual z-slices (same z) of *IAPez* RNA-FISH and RNAPII IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. The zoom column displays the region of the images highlighted in a yellow box zoomed in for greater detail. Images were acquired and processed using identical settings across all conditions.

To allow more feasible colocalization analysis, I focused on *IAPez* foci obtained after 48h of dTAG-13 treatment. Upon 48h of TRIM28 degradation, numerous IAP foci that occasionally colocalized with RNAPII and MED1 were observed (Figure 3.15a-b). To obtain a measure of colocalization frequency I calculated Manders' overlap coefficient (MOC) values (Fig. 3.15a (RNAPII): 0.193; n = 24 cells; Fig. 3.15b (MED1): 0.135; n = 24 cells). MOC measures the proportion of pixels in one fluorescent channel (IAP) that is also occupied by the pixels in the other fluorescent channel (RNAPII or MED1).

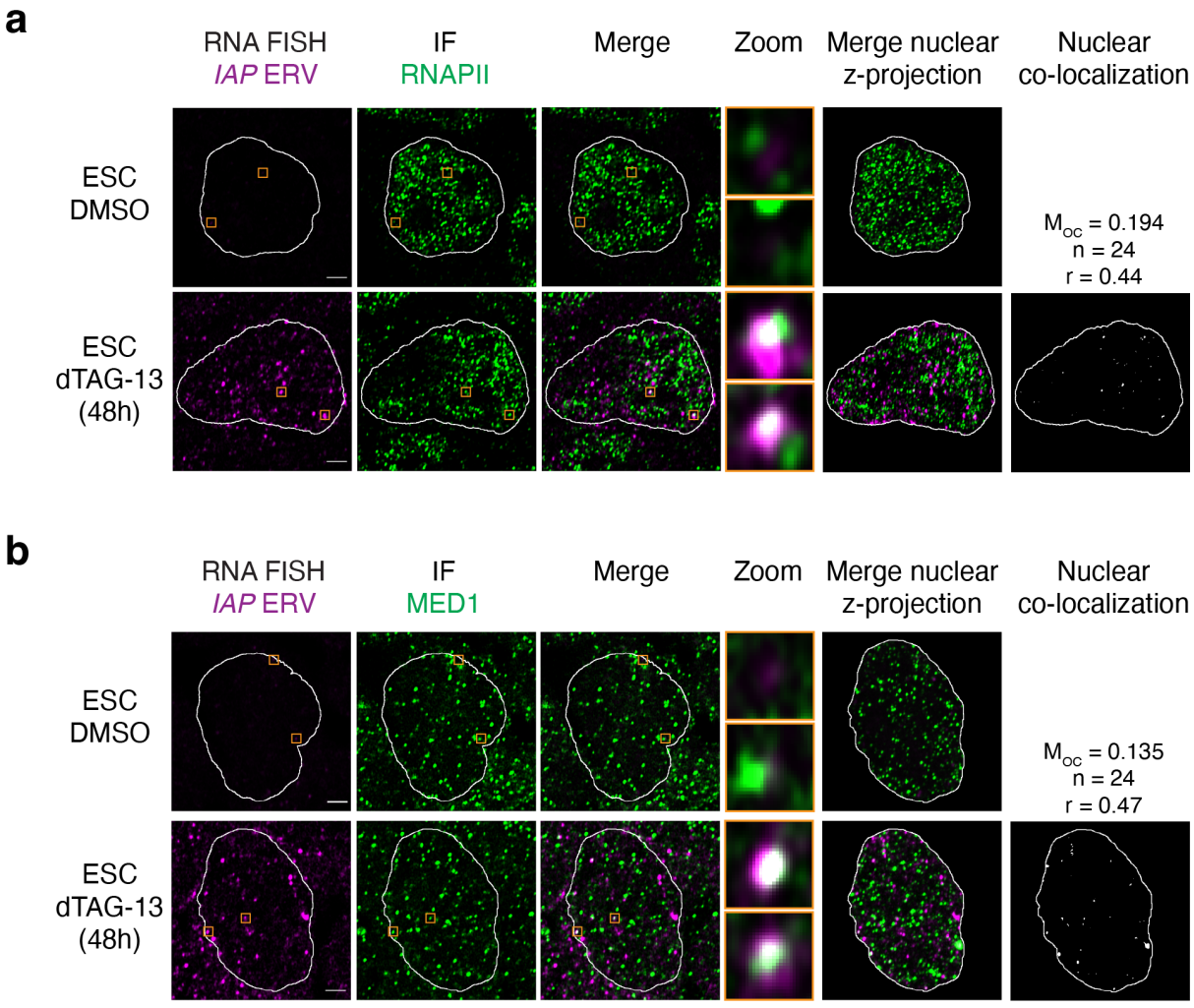


Figure 3.15. De-repressed IAPs form nuclear foci that associate with RNAPII condensates. a-b. Representative images of individual z-slices (same z) of RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. The zoom column displays the region of the images highlighted in a yellow box zoomed in for greater detail. Merge of the nuclear z-projections is displayed and overlapping pixels between the RNA-FISH and IF channels are highlighted in white. Displayed Manders' overlap coefficient (MOC) and Pearson's correlation coefficient (r) values are an average obtained from 24 analyzed nuclei. Scale bars: 2.5μm.

To additionally test colocalization extents, I obtained maximum intensity projections of RNA FISH-IF images, extrapolated geometrical centers of mass of RNA FISH and IF puncta and calculated distances between all detected *IAP* RNA foci and their nearest RNAPII/MED1 partners. This analysis revealed distances of ~20% of RNAPII

and ~15% of MED1 puncta were below 200nm, and therefore located in the regulatorily relevant vicinity of *IAP* foci (Figure 3.16) (Cho et al., 2018). These data suggested IAP RNA stemming from derepressed IAP ERVs could be colocalizing with key transcriptional condensate components and potentially contribute to their redistribution of away from SEs in the mESC nuclei.

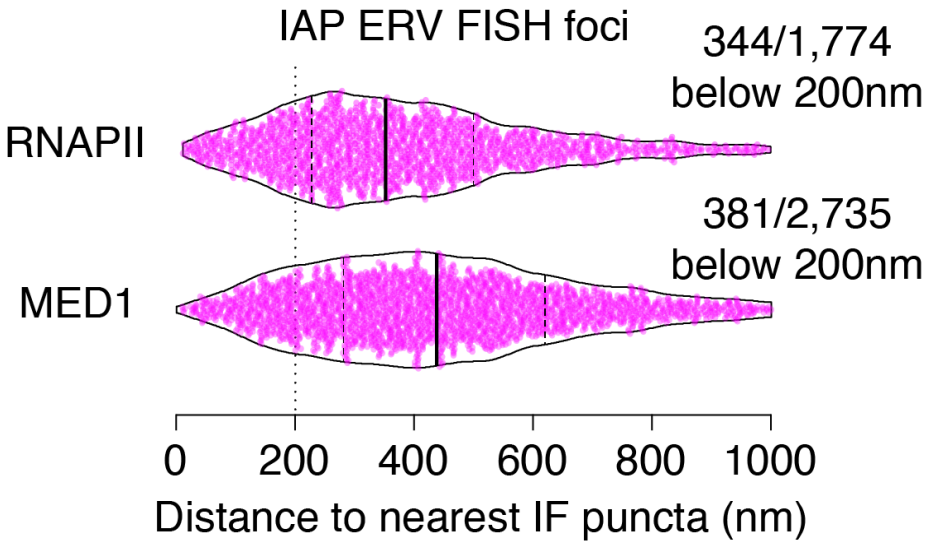


Figure 3.16. Further quantification of RNAPII and MED1 colocalization with de-repressed IAPs. Distance of IAP RNA FISH foci to the nearest RNAPII or MED1 IF puncta observed in images analyzed in Fig. 3.15. Each dot represents one IAP RNA FISH focus.

TT-SLAM-seq and RNA-seq data revealed TRIM28-degradation dependent de-repression of ERV, whereas other retrotransposon classes were unaffected after 24h of TRIM28 degradation (Fig. 3.17a). Progressive de-repression of all major sub-classes of IAPs was detected in both TT-SLAM-seq and RNA-seq data (Fig. 3.17b).

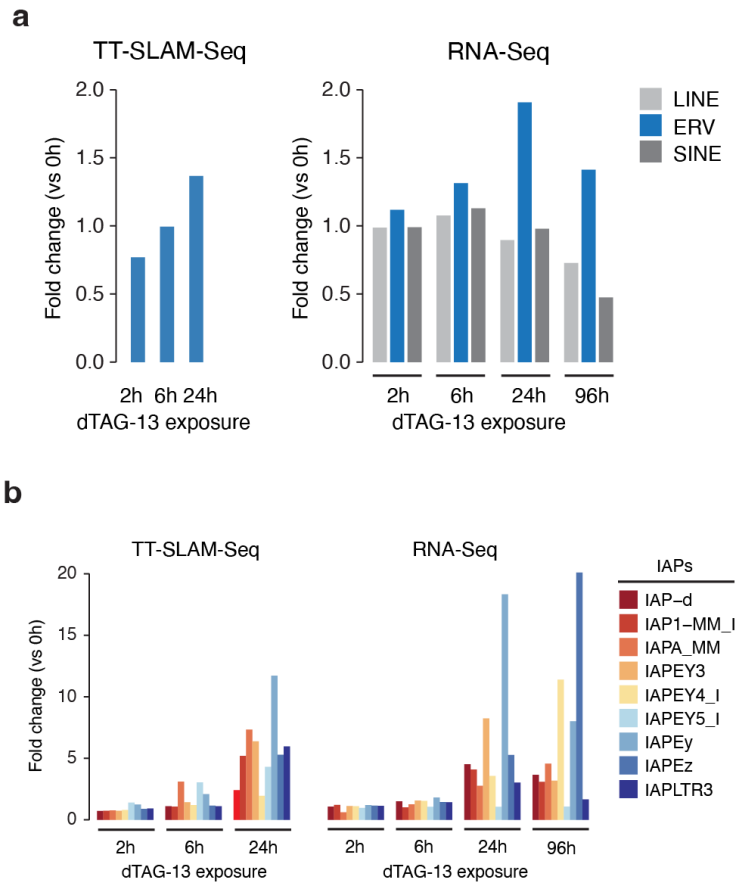


Figure 3.17. Retrotransposon de-repression in TRIM28-degraded ESCs. a. Transcript levels of ERVs, LINEs and SINEs detected with TT-SLAM-Seq and RNA-Seq upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h. Data are from three biological replicates. **b.** Transcript levels of IAP subfamilies detected with TT-SLAM-Seq and RNA-Seq upon dTAG-13 treatment in mESCs. Values are normalized to the levels detected at 0h. Data are from three biological replicates. Repeat RNA-Seq analysis was performed with Sara Hetzel.

Loss of average H3K9me3 occupancy at IAP genomic sites was detected upon 24h of TRIM28 degradation (Fig. 3.18a). Simultaneously the ERV classes (IAP, MMERVK, MMETN) showed moderate increases in RNAPII, MED23 and the active transcription mark H3K27ac (Fig. 3.18b), indicating increased presence of transcriptional machinery components at ERV elements. Together with microscopy colocalization data, this showed derepressed ERVs associate with transcriptional condensates upon TRIM28 degradation.

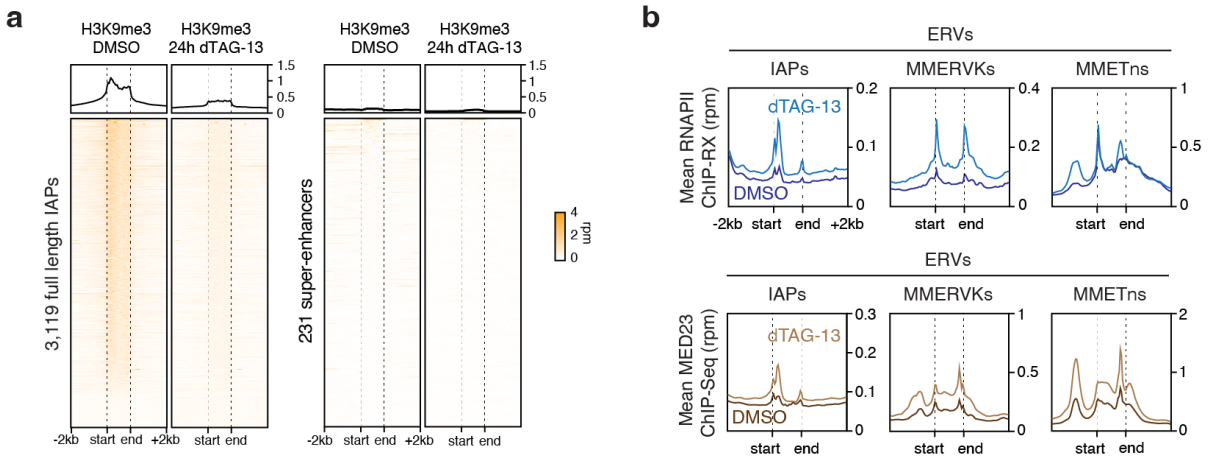


Figure 3.18. TRIM28-degradation leads to reduced H3K9me3 and increased MED23/RNAPII occupancy at ERVs a. Reduced H3K9me3 occupancy at IAPs in dTAG-13 treated (24h) mESCs. Displayed are heatmap representations of H3K9me3 ChIP-seq read densities within a 2kb window around the indicated genomic features. The genomic features (the middle part of the plot) were length normalized. Meta-analyses of the mean binding profile of the indicated factors are displayed above the heatmaps. Rpm: reads per million **b.** Meta representations of RNAPII ChIP-RX (top) and MED23 ChIP-Seq (bottom) read densities at IAP, MMERVK and MMETn ERVs in control (DMSO) and dTAG-13 (24h) treated mESCs. The mean read densities are displayed +/-2kb around the indicated elements. The genomic elements were length normalized.

Due to their number and evolutionary origin, retrotransposons are inevitably found in the vicinity of protein-coding genes (Thompson et al., 2016). Considering that ERVs get derepressed upon TRIM28 degradation and simultaneously contribute to redistribution of transcriptional condensates which can associate with multiple linearly distant DNA elements, I wanted to understand if ERV-associated transcriptional condensates could incorporate neighboring genes. Upon the examination of highly upregulated genes in the 24h TRIM28 degradation TT-SLAM-seq data, *Cthrc1*, a gene located within 100kb of three ERV elements emerged as an interesting candidate (Fig. 3.19). Repeat elements at the *Cthrc1* locus (MMETn, MMERVK and IAP) showed a reduction in H3K9me3 occupancy, and gains in either RNAPII, MED23 or nascent transcription signal, while *Cthrc1* itself was upregulated by several fold.

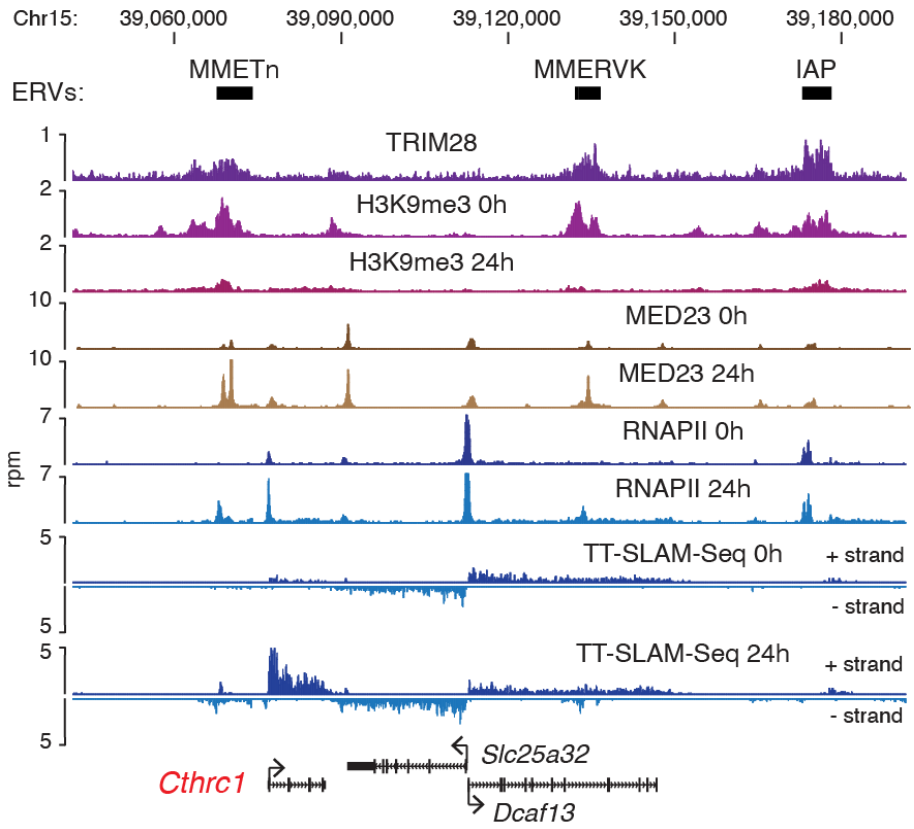


Figure 3.19. ERV-associated transcriptional condensates incorporate neighboring genes. Genome browser tracks at the *Cthrc1* locus. Note the independent transcription initiation events at *Cthrc1* and MMETn, ruling out that the MMETn acts as an alternative *Cthrc1* promoter. Rpm: reads per million.

To investigate if *Cthrc1* locus associated with transcriptional condensates, I tested the colocalization of *Cthrc1* RNA FISH signal (designed against introns of Cthrc1) and RNAPII IF. RNA FISH signal was not readily detectable in control cells, possibly due to low basal transcription levels of *Cthrc1* in these cells, but upon 24h TRIM28 degradation discrete *Cthrc1* RNA FISH foci were detectable (Fig. 3.20, left panels), which on average showed frequent association with RNAPII puncta (Fig. 3.20, right panels).

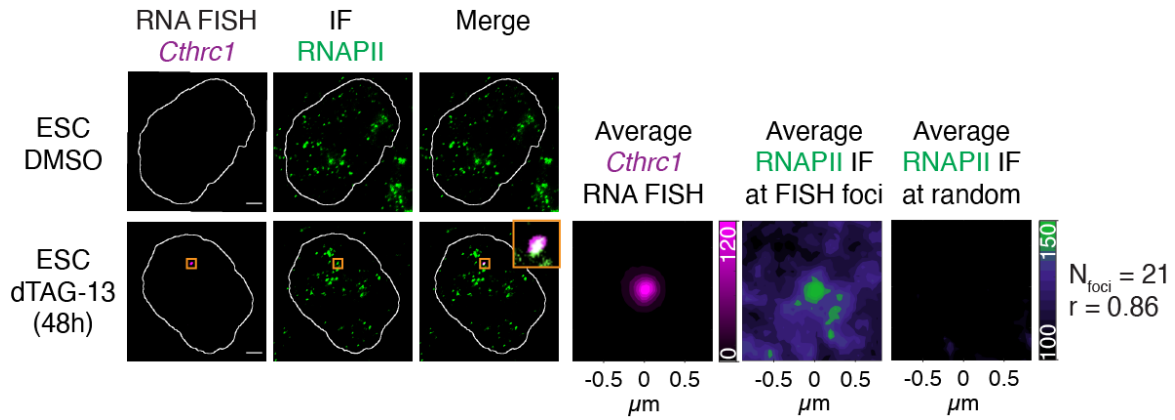


Figure 3.20. ERV-neighboring genes associate with transcriptional condensates. Representative separate images of individual z-slices (same z) of RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. Also shown are averaged signals of either RNA FISH or IF centered on the *Cthrc1* FISH foci or randomly selected nuclear positions. r denotes a Spearman's correlation coefficient. Scale bar: 2.5μm

To gain additional insights into the condensate character of this process, control or dTAG-treated cells were transiently (30min) treated with 1.5% 1,6 hexanediol (1-6 HD)—a short chain aliphatic alcohol that was reported to dissolve various biomolecular condensates including RNAPII condensates (Fig. 3.21a) (Cho et al., 2018). The 1-6 HD treatment affected the RNAPII IF pattern, by dissolving the discrete puncta found in the untreated condition (Fig. 3.21b) and reduced the levels of *Cthrc1* nascent RNA (twofold, $P < 0.05$, t-test) in cells that received 1-6 HD and dTAG compared to cells that only received dTAG (Fig. 3.21c). This suggested presence of RNAPII condensates could be contributing to the upregulation of *Cthrc1*.

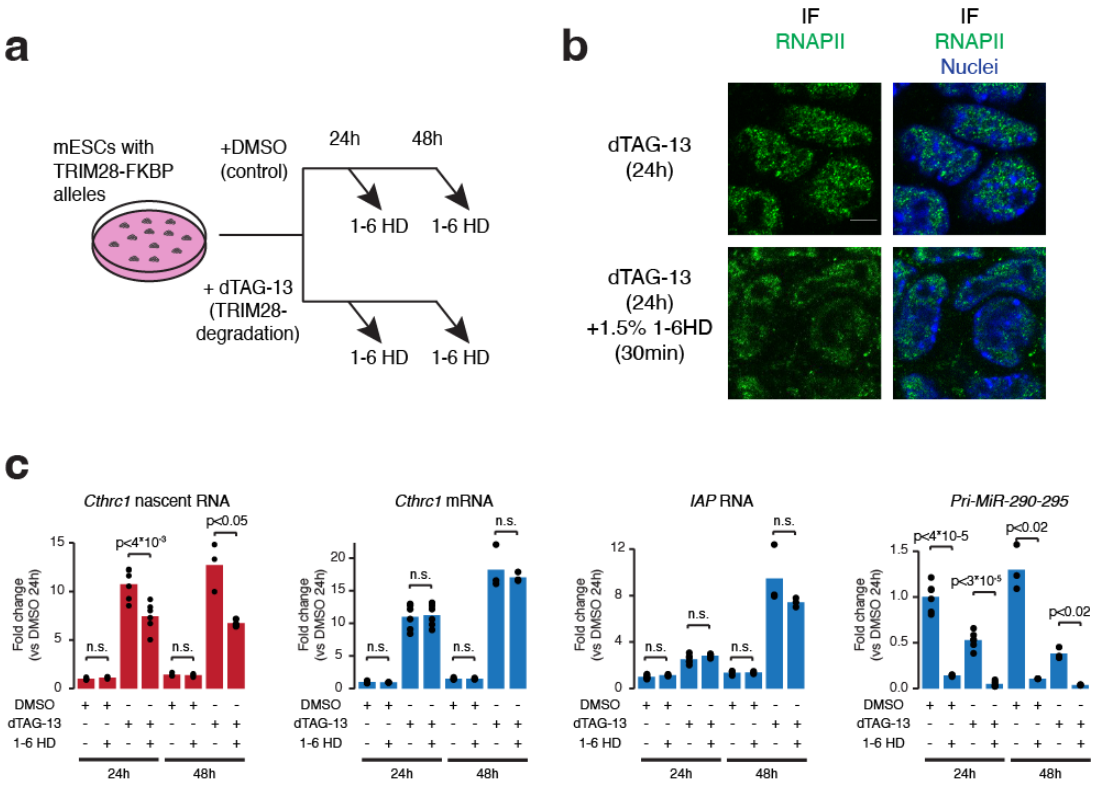


Figure 3.21. RNAPII condensates contribute to Cthrc1 upregulation. a. Scheme of the 1-6 hexanediol (1-6 HD) treatment experiments. **b.** Representative images of RNAPII immunofluorescence in control and 1-6 HD-treated cells. Addition of 1-6 HD partially dissolved the punctate localization of RNAPII. Scale bars: 5 μm. **c.** Transcription of the nascent Cthcr1 RNA is reduced by 30 min 1% 1-6 hexanediol-treatment in TRIM28-degraded cells. The bar plots show qRT-PCR data as fold change normalized to the DMSO control across 6 and 3 biological replicates for 24 h and 48 h timepoints, respectively. Note that the IAP RNA does not contain introns; thus, the IAP RNA qRT-PCR detects the steady state pool of IAP RNAs. Each dot represents a data point, and bar indicates the mean. P values are from two-tailed t tests. NS: not significant. Hexanediol experiments were performed with Henri Niskanen.

To further dissect the *Cthrc1* locus and establish if ERVs are indeed important for the association with transcriptional condensates, CRISPR–Cas9 was used to delete the three ERVs at the *Cthrc1* locus and generate an ERV triple knockout (TKO) cell line (Fig. 3.22a). Deletions were validated by genotyping (Fig. 3.22b).

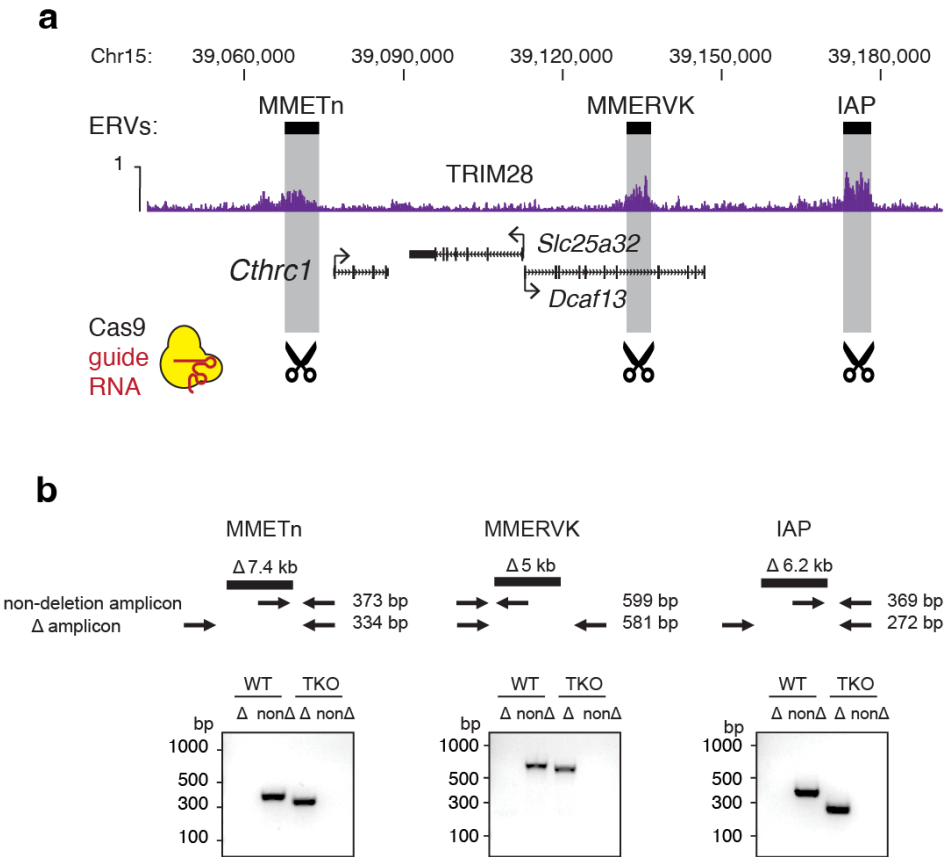


Figure 3.22. CRISPR/Cas9-mediated deletion of ERVs at the *Cthrc1* locus. Highlighted ERVs, shown within the extended *Cthrc1* locus, were deleted in the ERV TKO mESC line. TRIM28 ChIP-Seq tracks are included for the entire locus. **b.** Genotyping PCR in the ERV TKO cell line. Genotyping was performed once.

ERV deletions caused significant reduction in *Cthrc1* induction upon TRIM28 degradation compared to wildtype TRIM28-FKBP cells (Fig. 3.23). Under identical conditions, TRIM28 degradation in ERV TKO cells induced typical upregulation of *IAPEz* and reduction of *pri-MiR290-295* transcription, similar to that seen in TRIM28-FKBP cells. Genes found in the vicinity of deleted ERVs, *Dcaf13 and Slc25a3*, were unaffected. Reduction in *Cthrc1* induction upon TRIM28 depletion in ERV TKO cells indicated ERVs may be essential for the association of transcriptional machinery with the locus and the incorporation of ERV proximal genes in transcriptional condensates.

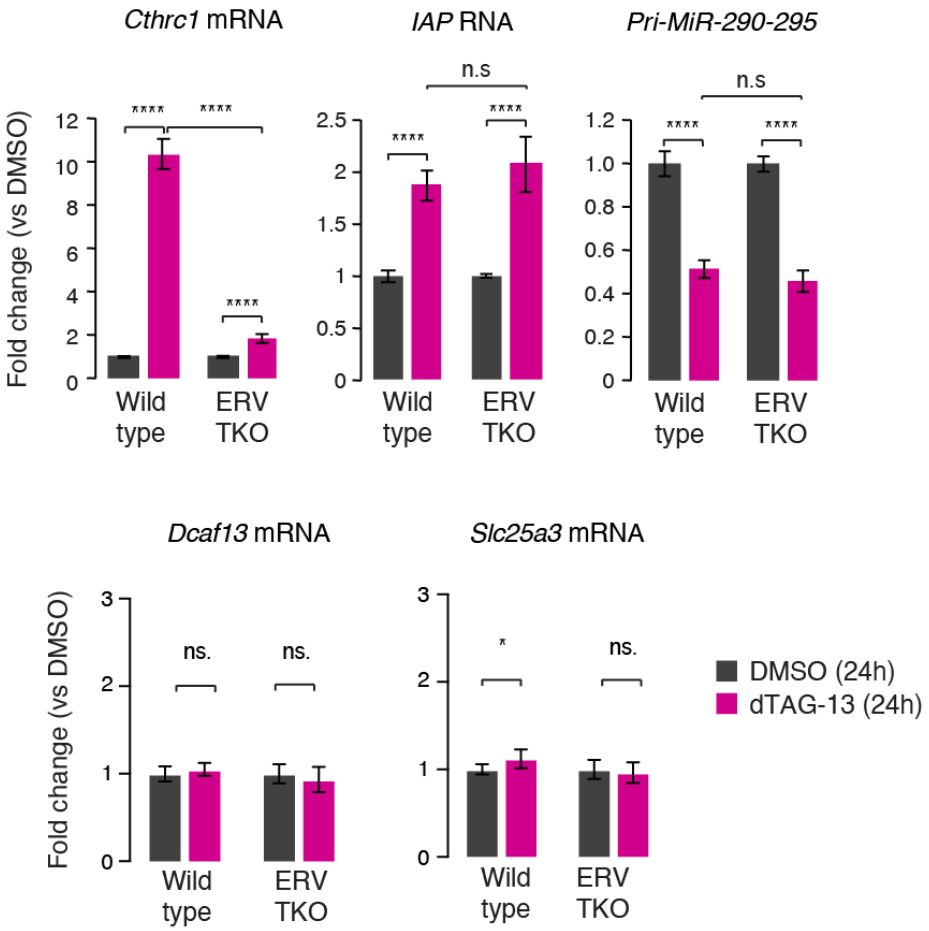


Figure 3.23. Gene expression at the Cthrc1 locus in ERV deletion cell line. RT-qPCR analysis of gene expression of *Cthrc1* and proximal genes upon 24h dTAG-13 treatment from six biological replicates. *IAPEz* and *pri-MiR290-295* are shown as controls. Values were normalized to the DMSO control. Bar charts depict the mean and error bars standard deviation. P-values are from two-tailed t-tests. This experiment was performed with Henri Niskanen.

Transcription factors phase separate with coactivators to form transcriptional condensates (Boija et al., 2018). Since derepressed IAPs are implicated in transcriptional condensate redistribution and associate with RNAPII and Mediator, we investigated whether additional TFs could be a part of the IAP RNA-coactivator multicomponent condensates at the *Cthrc1* locus. As discussed above, LTRs of IAPs harbor binding sites for many important TFs. TFs with enriched motifs were examined and two (NFY and NRF1) with highest expression levels in mESCs were chosen for closer investigation (Fig. 3.24).

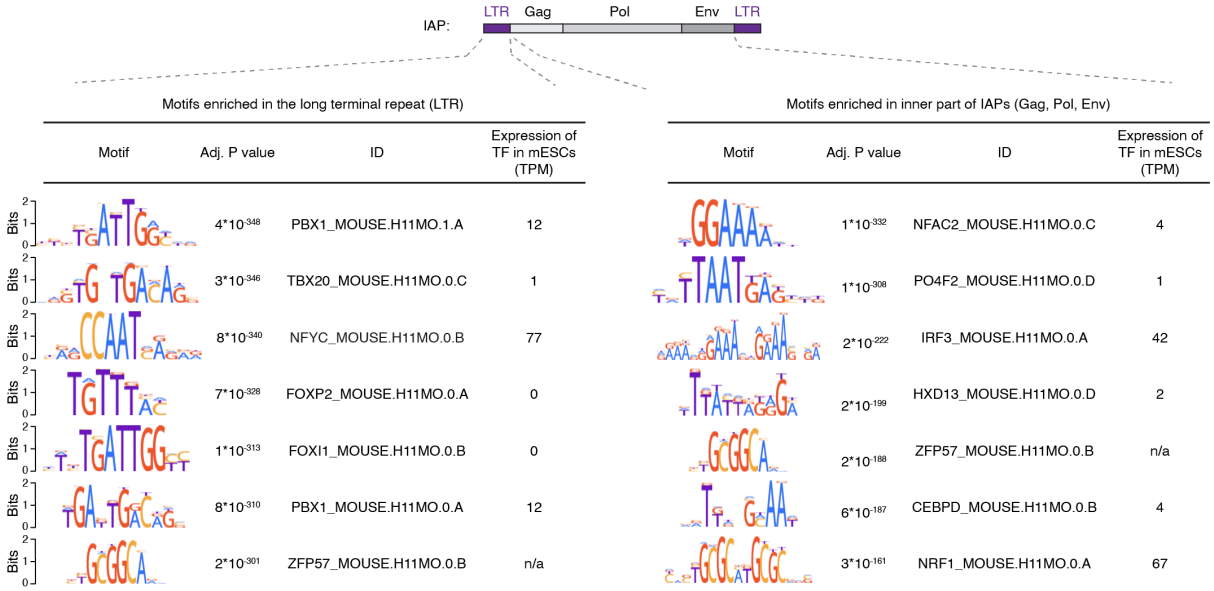


Figure 3.24. IAP sequences are enriched for TF binding motifs. NFY and NRF1 TF binding motifs have highest expression levels in mESCs. **Top:** schematic of an IAP element. **Bottom:** motif images, adjusted P values and motif IDs, and the expression level of the TF in mESC RNA-seq data. Displayed are the top-scoring motifs based on adjusted P-value. Motifs were filtered for redundancy. This analysis was performed with Sara Hetzel.

RNA FISH-IF revealed NFY forms puncta that colocalized with Cthrc1 RNA foci upon TRIM28 degradation (Fig. 3.25a). Although NRF1 forms discrete clusters in IF, those clusters did not colocalize with *Cthrc1* RNA FISH signal (Fig. 3.25b). These data indicated that specific TFs, with binding sites in derepressed ERV sequences, participate in transcriptional condensate formation at ERVs and neighboring genes.

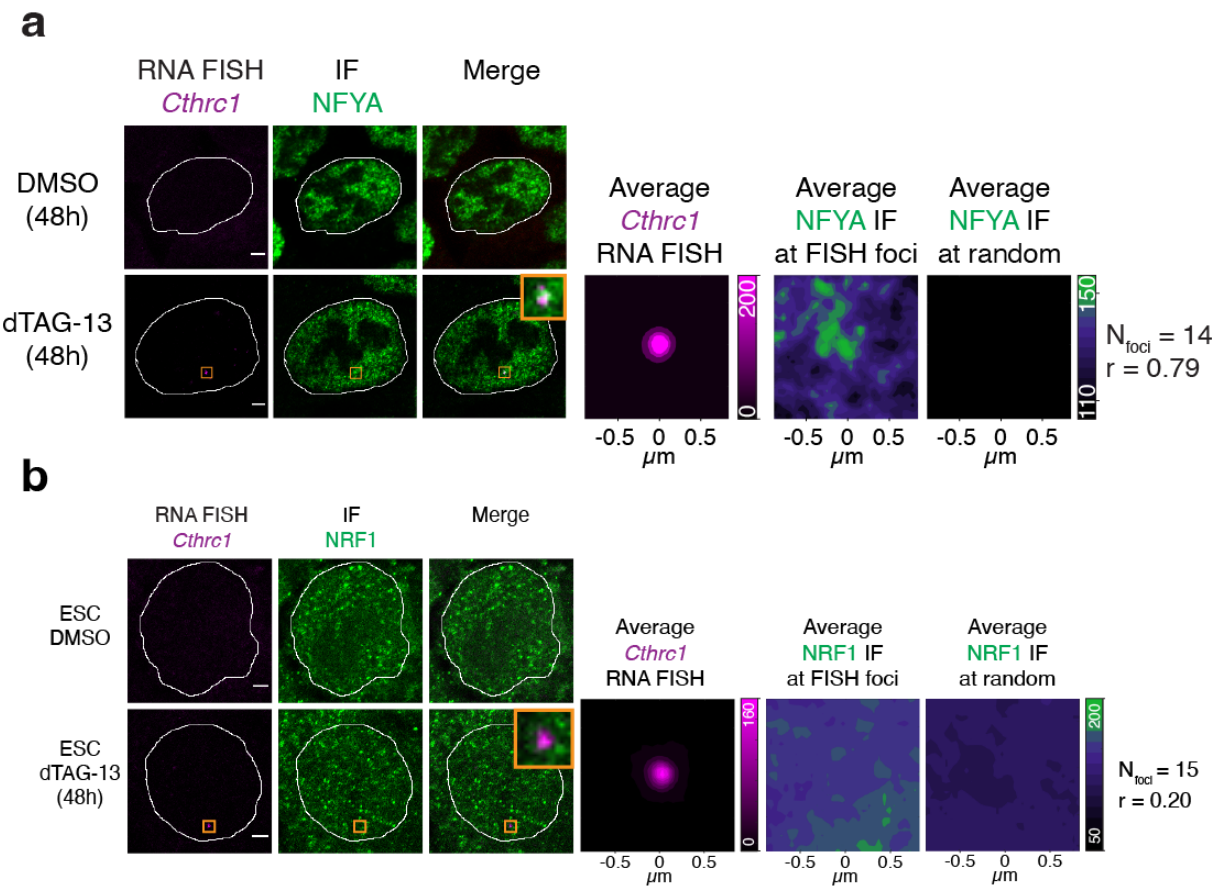


Figure 3.25. Characterization of NFY and NRF1 co-localization with *Cthrc1* RNA. **a.** NFYA puncta co-localize with *Cthrc1* nascent RNA FISH signal. Representative separate images of individual z-slices (same z) of RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. Also shown are averaged signals of either RNA FISH or IF centered on the *Cthrc1* FISH foci or randomly selected nuclear positions. r denotes a Spearman's correlation coefficient. Scale bars: 2.5μm. **b.** NRF1 puncta do not co-localize with the nascent RNA *Cthrc1* in TRIM28-degraded mESCs. Displayed are separate images of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. Also shown are averaged signals of either RNA FISH or NRF1 IF centered on the *Cthrc1* FISH foci or randomly selected nuclear positions. Scale bars: 2.5 μm.

3.5 SE-enriched TFs rescue transcriptional condensate localization

Since TFs are an essential component of SE-associated transcriptional condensates, it is conceivable that the overexpression of SE TFs may rescue the localization of transcriptional condensates in TRIM28-degraded mESCs. To test this hypothesis, the dTAG-sensitive TRIM28 FKBP alleles were endogenously integrated into an induced pluripotent stem cell (iPSC) line with integrated transgenes encoding *Oct4, Sox2, Klf4 and Myc* (OSKM) under a doxycycline-inducible promoter (Fig. 3.26a). OCT4, SOX2 and KLF4 are pluripotency TFs that are highly enriched at SEs in ESCs, and OCT4 IDR has previously been investigated in the phase separation context (Boija et al., 2018; Whyte et al., 2013). Treatment of TRIM28 FKBP iPSCs with dTAG-13 degraded TRIM28 and reduced OCT4 and SOX2 levels (upon 48h dTAG treatment); however simultaneous doxycycline induction rescued OCT4 and SOX2 levels (Fig. 3.26b).

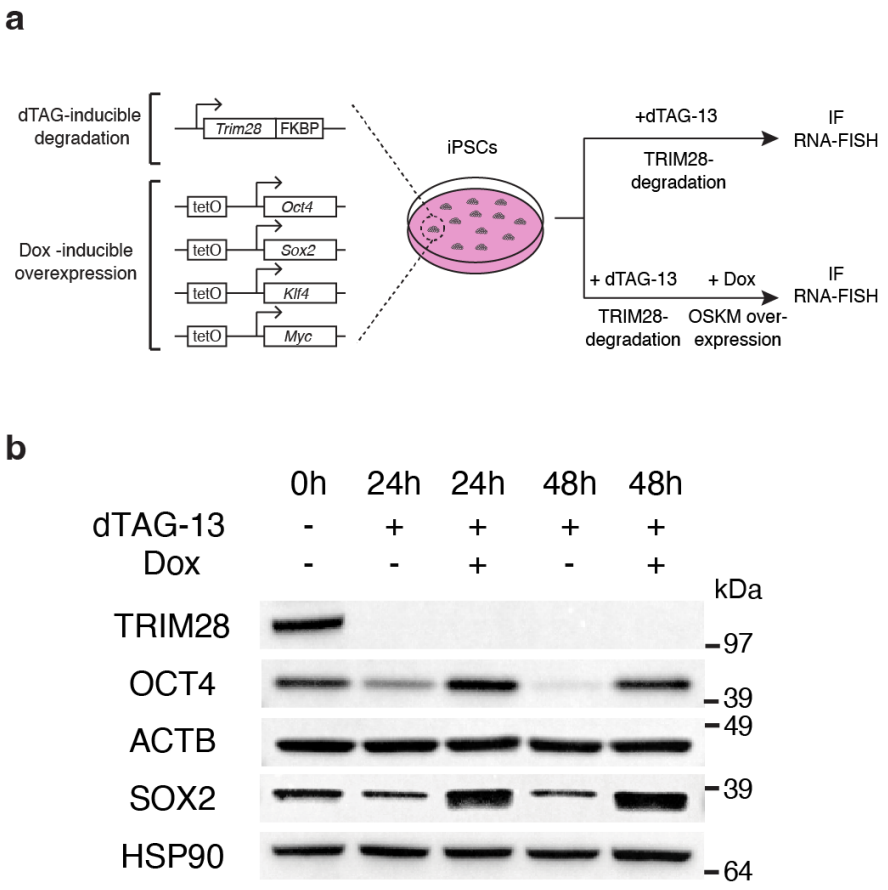


Figure 3.26. Pluripotency TF overexpression system in TRIM28-FKBP iPSCs. a. Genotype scheme of the iPSC line and scheme of the experimental setup. The iPSC cell line includes dTAG-inducible Trim28-FKBP alleles, and Doxycycline-inducible *Oct4*, *Sox2*, *Klf4*, *c-Myc* (OSKM) transgenes. **b.** Western blot validation of TRIM28 degradation and OSKM ectopic overexpression in iPSCs. Doxycycline induction boosts OCT4 and SOX2 levels and dTAG-13 treatment depletes TRIM28. The iPSC cell line was generated and validated by Christina Riemenschneider.

The TRIM28-FKBP iPSC cell line showed frequent co-localization between *miR290-295* SE RNA FISH foci and RNAPII puncta (Fig. 3.27, top row). This colocalization was reduced upon TRIM28-degradation, indicating the collapse of transcriptional condensates at SE loci (Fig. 3.27, middle row). Overexpression of OSKM rescued colocalization of *miR290-295* SE RNA FISH foci with RNAPII puncta in TRIM28-depleted iPSCs (Fig. 3.27, bottom row). These data demonstrated that overexpression of SE-associated TFs can rescue transcriptional condensate colocalization at SE loci in TRIM28 degraded cells.

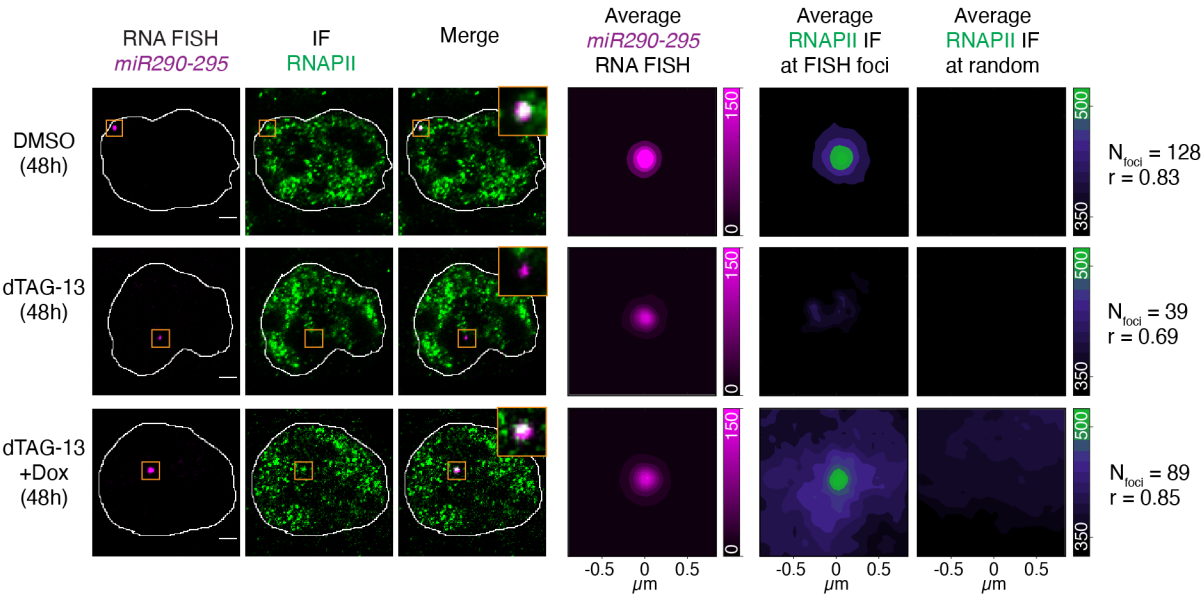


Figure 3.27. Pluripotency TFs rescue transcriptional condensate localization in TRIM28-degraded iPSCs. Co-localization between the nascent RNA of *miR290-295* and RNAPII puncta in TRIM28-degraded iPSCs that ectopically express OSKM factors. In DMSO treated cells, RNAPII puncta frequently co-localize with *miR290-295* foci, this co-localization is reduced following TRIM28 depletion. Doxycycline induction of OSKM factors rescues the colocalization between the RNAPII puncta and *miR290-295* foci in TRIM28-degraded iPSCs. Displayed are separate images of individual z-slices (same z) of the RNA-FISH and IF signal, and an image of the merged channels. The nuclear periphery determined by DAPI staining is highlighted as a white contour. Also shown are averaged signals of either RNA FISH or RNAPII IF centered on the *miR290-295* FISH foci or randomly selected nuclear positions. r denotes a Spearman's correlation coefficient. Scale bars: 2.5μm.

3.6 IAP RNA contributes to transcriptional condensate localization in TRIM28-degraded mESCs

To assess whether ERV RNA contributes to the redistribution of transcriptional condensates in TRIM28-degraded mESCs, shRNAs system to knock down four subclasses of ERV retrotransposons (IAPs, MMERVK10Cs, MMERVK9Cs, MMETns)

was used (Fig. 3.28a). Expression of shRNAs against ERVs for 24h prior to 24h TRIM28 degradation completely rescued the derepression of *ERVs* and the induction of *Cthrc1*, and significantly reduced the downregulation of the *miR290-295* SE locus (Fig. 3.28b).

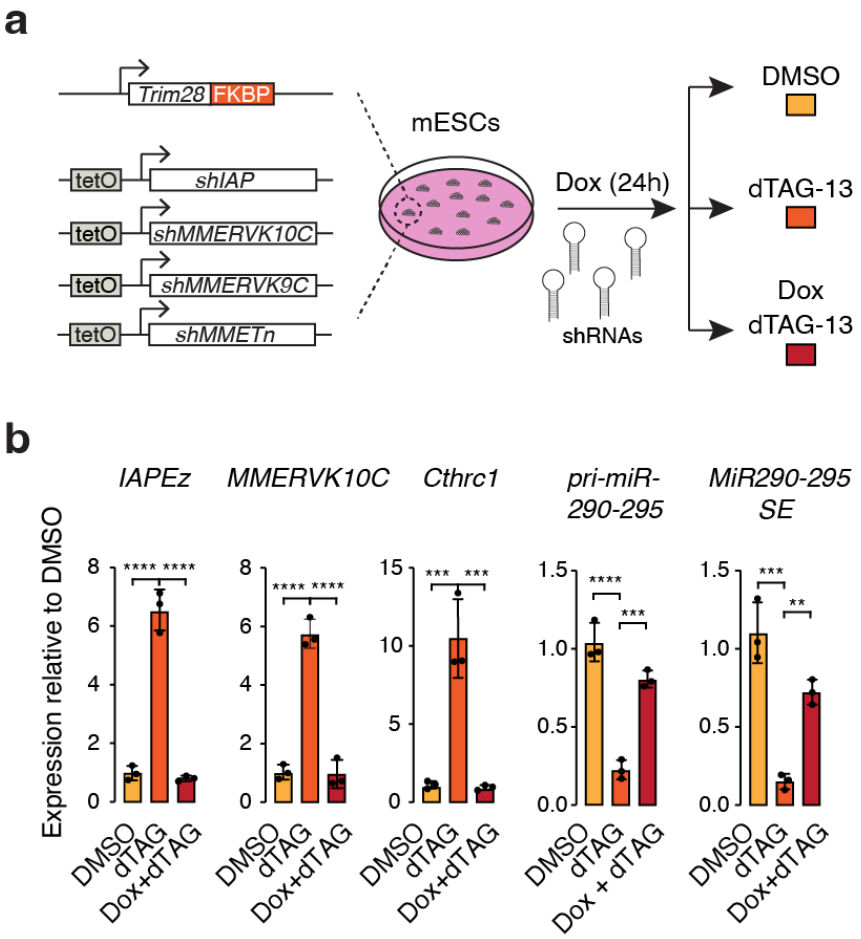


Figure 3.28. ERV knockdown rescues transcriptional effects of TRIM28 depletion. **a.** Scheme of the ERV shRNA knockdown experiments. Cells were doxycycline-induced for 24h prior to DMSO, dTAG-13, or doxycycline plus dTAG-13 treatment. **b.** qRT-PCR data as fold change normalized to the DMSO treatment control. ERV knockdown prevents TRIM28-depletion-induced *IAPez* and MMERVK10c derepression, and rescues downregulation of *MiR290-295* and *Cthrc1*. Data are presented as mean values +/- SD from three biological replicates. P-values are from two-tailed t-tests. ****: P<10-4, ***: P<10-3, **: P<10-2. The ERV shRNA cell line was generated and validated by Abhishek Sampath Kumar.

In addition, ERV knockdown rescue of *IAPez* derepression was validated via RNA FISH imaging. Compared to high intensity RNA FISH foci that appear upon TRIM28 degradation, doxycycline-induced ERV knockdown reduced the number and intensity of *IAPez* foci in TRIM28-degraded cells to levels comparable to DMSO control (Fig. 3.29). This indicated *IAPez* RNA FISH agrees with qRT-PCR expression data.

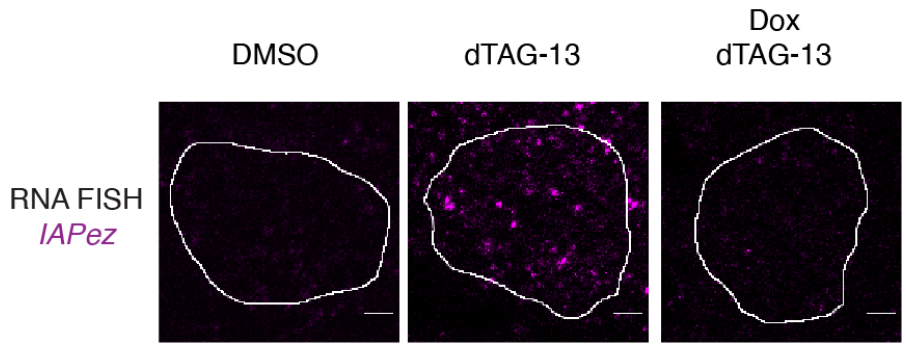


Figure 3.29. ERV knockdown counteracts ERV derepression in TRIM28-depleted cells. Representative images of *IAPEz* RNA FISH in the ERV knockdown experiments (48h of either DMSO, dTAG-13, or doxycycline plus dTAG-13, treatment). TRIM28 depletion induces derepression of *IAPEz*, as evidenced by the increase in RNA FISH foci intensity. ERV knockdown reduces the intensity of *IAPEz* foci to levels comparable to DMSO control cells. Images in all experiments were acquired at and processed under identical settings. The nuclear periphery determined by DAPI staining is highlighted as a white contour.

Further microscopy data indicated ERV knockdown rescues RNAPII localization at the *miR290-295* RNA FISH foci upon TRIM28 degradation in mESCs (Fig. 3.30). RNAPII IF signal showed similar focal nuclear pattern across all treatment conditions. DMSO-treated cells showed frequent co-localization between RNAPII and *miR290-295* RNA FISH, as indicated by representative images and average co-localization analysis (Fig. 3.30, top row). This association was reduced in dTAG-13 treated cells (Fig. 3.30, middle row), but simultaneous doxycycline induction of shRNA rescued the co-localization to levels comparable to DMSO control cells (Fig. 3.30, bottom row). This

indicated presence of ERV RNA plays an important role in altered localization of transcriptional condensates upon TRIM28 degradation.

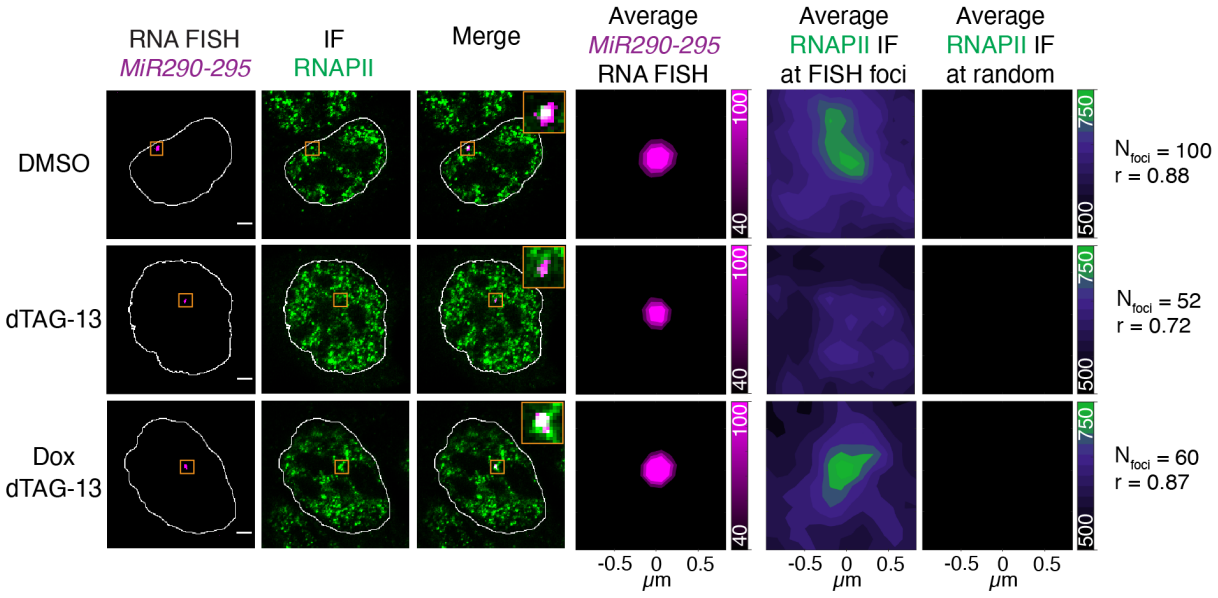


Figure 3.30. ERV knockdown rescues transcriptional condensate localization in TRIM28-degraded mESCs. Representative images of individual z-slices (same z) of RNA-FISH and IF signal, and an image of the merged channels. Cells were treated according to the experimental setup described in Fig. 3.27a. The nuclear periphery determined by DAPI staining is highlighted as a white contour. Also shown are averaged signals of either RNA FISH or RNAPII IF centered on the *miR290-295* FISH foci or randomly selected nuclear positions. r denotes a Spearman's correlation coefficient. Scale bars: 2.5µm.

4 Discussion

4.1 Significance of the transcriptional hijacking model

Retrotransposons occupy significant portions of mammalian genomes and coexist with genes required for development. Although some retrotransposon classes retained the capacity for retrotransposition, once the repression machinery is depleted, it is the upregulated transcription of ERVs that compromises pluripotency and causes lethality in mouse embryos (Matsui et al., 2010; Rowe et al., 2010). Precise mechanistic understanding of this process has so far been missing. Evidence presented in this thesis indicates ERV derepression hijacks transcriptional condensates from key pluripotency genes upon TRIM28 degradation in mESCs (Fig. 4.1).

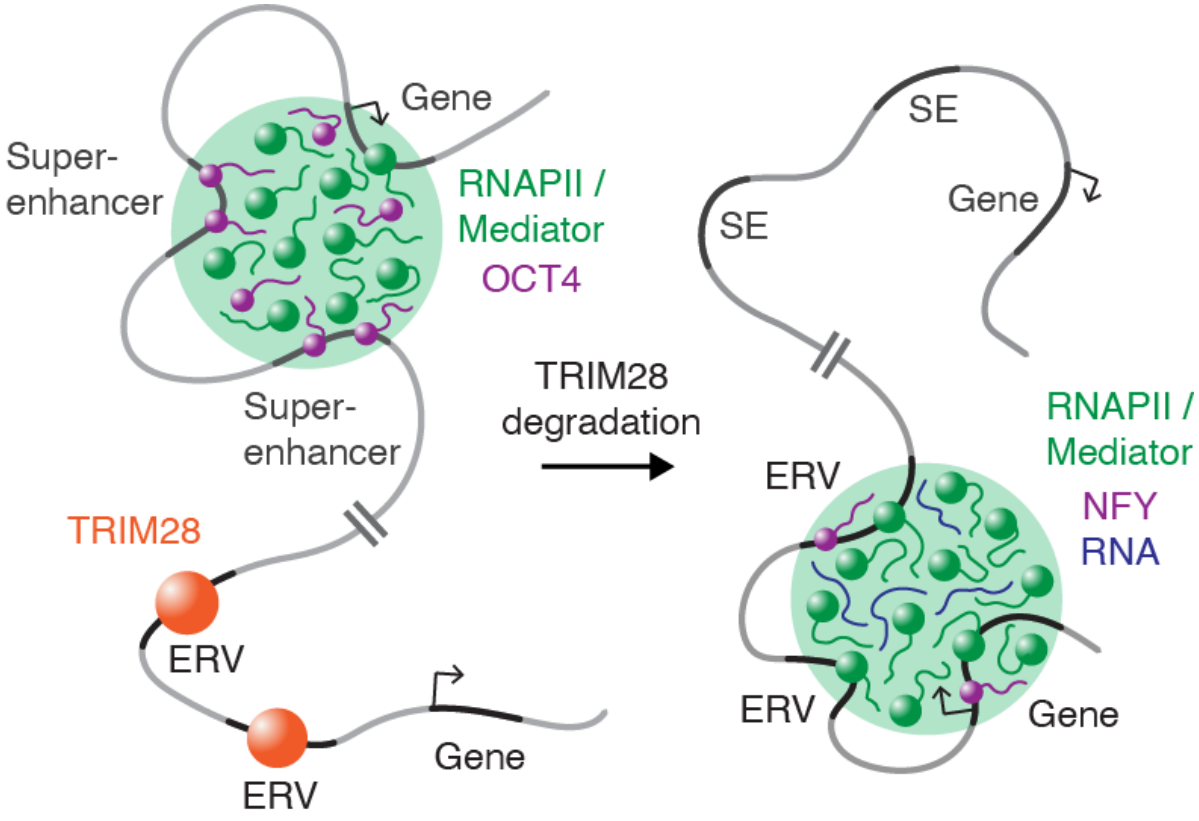


Figure 4.1. Transcriptional hijacking model. In wildtype mESCs, transcriptional condensates are formed at super-enhancers where they promote transcription of pluripotency genes, while ERVs are silenced by TRIM28-mediated heterochromatin. TRIM28 degradation disrupts genomic distribution of transcriptional condensates, leading to their formation at ERV loci.

Using a range of genomic tools and methods including the degron system, TT-SLAM-seq and multiplexed RNA FISH-IF imaging, we demonstrated the power of using an integrated genomics approach to mechanistically dissect the function of a key epigenetic regulator. TRIM28 degradation induced acute transcriptomic changes with a pronounced downregulating effect on key pluripotency SEs and associated genes. It simultaneously upregulated key targets of TRIM28/HP1α repressive pathway, including ERV retrotransposon classes (ERVK, MMERVK, MMETN). High-resolution microscopy targeting nascent RNAs of key SE genes and transcriptional cofactors (RNAPII and MED1) indicated transcriptional condensates associate less frequently with SEs upon TRIM28 degradation in mESCs. Consequently, components of transcriptional machinery are redistributed to the derepressed, transcriptionally potent ERV elements. ERV RNA knockdown experiments indicated RNAs stemming from repeat elements play an important role in the hijacking mechanism. Knockdown of ERV RNAs prior to and concurrent with TRIM28 degradation rescued transcriptional condensate localization. While these results bring important new insights into our understanding of ERV regulation and present a possible molecular mechanism that explains the loss of pluripotency in TRIM28 degraded mESCs, it is imperative to further explore the antagonistic relationship between ERVs and SEs, the specificity of interactions between transcriptional condensate components at repeat loci, as well as the role of hijacking in mouse embryonic development and various disease phenotypes.

4.2 Future perspective: RNA in transcriptional condensates

RNA has been an important research focus of both condensate and transcription biology, but RNA role at the intersection of these two fields is still relatively underexplored. Various RNA species are produced during RNAPII transcription, and they should, by proxy, influence biomolecular condensates formed in the vicinity of transcription sites. Considering condensates are formed by low-affinity interactions,

the high negative charge originating from RNA phosphate backbones should contribute to overall electrostatic balance of transcriptional condensates. Net negative charge of an RNA molecule increases with length, suggesting length could be an important driver of condensate formation and maintenance (Boeynaems et al., 2019). Recent study suggested RNAs produced during the course of RNAPII transcription may regulate transcriptional condensates via a non-equilibrium feedback mechanism (Henninger et al., 2021).

In this model, transcriptional initiation is associated with low concentrations of shorter RNAs produced at gene enhancers and promoters, and the electrostatic contribution of RNAs in such configuration facilitates transcriptional condensate formation. Once the transcription starts bursting during elongation, longer mRNAs are generated at increased rates thus increasing the RNA concentrations and the net negative electrostatic charge, and transcriptional condensates are dissolved. Although this model presents an elegant solution to the dynamic nature of transcription, since it allows rapid formation and dissolution of condensates by manipulating one internal component (RNA), its application to the hijacking model is currently somewhat limited.

RNAs produced at ERVs most likely reside at the nuclear sites of their production for extended periods of time, but the imaging setup currently does not allow us to discriminate repeat RNAs according to their number, length, or half-life. Since retrotransposons do not possess introns, it is difficult to target their nascent RNA transcripts exclusively in microscopy, the way this can be done for SE-controlled genes. Instead, we rely on targeting full length sequences, subsequently isolate the signal that localizes in the nucleus and assess the relative association with transcriptional condensates at the global level. The time dimension additionally complicates the picture since it is not possible to determine what stage of transcription and RNA lifetime most of the signal is derived from. It is therefore possible that the distinct

nuclear foci displayed in IAP RNA FISH imaging upon TRIM28 degradation are either predominantly nascent RNAs, more mature mRNAs or a combination of these and other intermediate RNA species.

TRIM28 degradation progressively leads to more intensified RNA FISH signal, which definitively indicates higher concentration and number of IAP RNA molecules in the local environment, and this is supported by transcriptomics data. Further analysis of RNA-seq and TT-SLAM-seq datasets could potentially be used to extrapolate half-life values of various RNAs, but this is additionally complicated by difficulties in read mapping to unique repetitive sequences (Teissandier et al., 2019). In addition, RNA condensation studies in vitro should be taken with a degree of caution. Recent study demonstrated RNA condensation is unspecific in the presence of macromolecular crowding agents like PEG, but sequence- and structure-specific in the absence of PEG (Poudyal et al., 2021). It is therefore currently challenging to completely reconcile the RNA feedback model with the transcriptional hijacking model, simply due to limitations in our understanding of biochemical properties of ERV RNA including length, concentration, half-life, and number.

Some of these constraints could be overcome if there was a practical way to apply existing live-cell microscopy techniques with improved sensitivity (including the MS2-MCP system) to specific derepressed ERV elements (Vera et al., 2019). This would allow single molecule tracking of RNAs produced at a specific ERV locus which could then be combined with RNAPII or MED1 endogenous fluorescence tagging to observe real-time interactions and validate presence of repeat RNA in redistributed transcriptional condensates upon TRIM28 degradation. RNA length is another feature that could be investigated in more detail. Additional in vitro reconstitution experiments, not included in this thesis, indicated addition of purified IAP RNA facilitates condensation of RNAPII CTD and MED1 IDR in a dose-dependent manner (Asimi et al., 2022). In agreement with the RNA feedback model, partitioning of

RNAPII CTD peaked at a specific concentration and increasing IAP RNA beyond that level is not additionally beneficial for condensation. The length of in vitro transcribed IAP RNA sequence was standardized at 800 bp, therefore the in vitro experiments reflect electrostatic charge generated by that specific molecular length. Input from an improved classification of RNAs generated at derepressed ERVs could improve the in vitro experimental setup, since different lengths of IAP RNA could be tested.

Identical RNA (IAP) species can facilitate in vitro condensation of both activator (MED1) and repressor (HP1α) cofactors at different optimal concentrations, forming a sort of a bimodal concentration gradient (Asimi et al., 2022). This possibility is extremely interesting both from the repeat element and developmental gene activation perspective. For repeat elements, basal transcription levels of several repeat classes seem to be relevant for heterochromatin maintenance, but the same elements are actively transcribed once derepressed. On the other hand, from the developmental gene standpoint, many developmentally relevant decisions are brought about by dynamic transcriptional activation and silencing of developmental genes many of which are transcription factors. The opportunity to control these transitions in part through modulating RNA concentration levels that then facilitate formation of either transcriptional or heterochromatin condensates may be an important regulatory mechanism and deserves further investigation. More advanced live-cell microscopy setups could potentially be used to monitor if ERV-proximal nuclear environment (for instance, *Cthrc1* locus) transition from higher concentration of heterochromatic factors to active transcription co-factors, as IAP RNA concentration is increased.

Biomolecular condensates are regulated by covalent modifications of their components, including protein phosphorylation, acetylation, methylation and sumoylation, and DNA and RNA methylation (Banani et al., 2016; Conti & Oppikofer, 2022; Gibson et al., 2019; Guo et al., 2019; C. H. Li et al., 2020; X. Su et al., 2016). Considering ERV RNA plays an essential role in the proposed transcriptional hijacking model,

determining the contribution of RNA modifications to the biochemical properties of ERV RNAs could be relevant for our understanding of their condensate behavior. N6-methyadenosine (m6A) is the most abundant RNA modification and it impacts mRNA and lncRNA localization, translation, half-life and splicing (N. Liu et al., 2017). It may influence RNA-mediated phase separation by affecting RNA-RNA and RNA-protein interactions (Y. Cheng et al., 2021). Number and spacing of m6A sites in mRNAs can alter the phase separation of m6A readers, including the cytosolic m^6A-binding proteins, YTHDF1–3, which form condensates in vitro and in cells (Ries et al., 2019; J. Wang et al., 2020). Polymethylated mRNAs seem to facilitate YTHDF1-3 condensation at lower concentrations when compared to mono- or un-methylated mRNAs. It would be significant to check if the manipulation of m6A site number and distribution would have a strong impact on IAP RNA in the in vitro reconstitution experiments.

Mounting evidence suggests a strong link between m6A methylation and ERV regulation (Chelmicki et al., 2021). METTL3, a methytransferase that catalyzes the methylation of the N6-adenosine of RNA, localizes specifically to IAPs in mESCs, where it is supposed to form a complex with YTHDC1 and IAP RNA (Xu et al., 2021). YTHDC1 is an m6A reader that binds IAP, ERVK and LINE1 transcripts in mESCs (J. Liu et al., 2021; Xiao et al., 2016). YTHDC1 knockout causes reprogramming of mESCs to a 2C-like state and induces upregulated transcription of repeat elements, indicating that m^6A RNA modification of lowly transcribed repeats plays an important role in heterochromatin maintenance at those genomic regions (J. Liu et al., 2021). METTL3 has itself been implicated in phase separation processes (D. Han et al., 2022). It physically interacts with TRIM28 and SETDB1, and its knockout results in reduced presence of heterochromatin marks and increased transcription of IAPs in mESCs, indicating it is important for the repression of IAPs (Xu et al., 2021). IAP transcripts produced at METTL3-occupied IAPs are nuclear, chromatin-bound and m6A methylated. It would be important to examine the role played by the m6A methylation

and its readers and writers in the transcriptional condensate hijacking model, that is whether these components are potentially phase separating together with heterochromatin machinery at IAPs and reduced IAP RNA methylation leads to IAP RNA association with transcriptional condensates. Interestingly, small molecule inhibitors of METTL3 are currently in clinical trials for acute myeloid leukemia and solid cancers, and combining METTL3 inhibition with TRIM28 degradation could potentially offer an elegant way to test some of these interactions in mESCs (Yankova et al., 2021).

4.3 Future perspective: Competition for transcriptional machinery in vivo

One open question that arises from the hijacking model is whether, independent of TRIM28 degradation or the loss of pluripotency phenotype, ERV elements have the potential to outcompete SE controlled genes in the competition for transcriptional machinery. Additional data not included in this thesis showed that forced induction of randomly integrated ~900 bp fragments of IAPEz ERVs can significantly undermine SE and SE associated gene transcription when compared to forced induction of a GFP construct of similar length (Asimi et al., 2022). PiggyBac transposon system was used to generate two mESC lines that contain either the Dox-inducible GFP sequence or the Dox-inducible IAPEz fragments, both with the equal number of copy numbers and insertion sites. Dox treatment induced expression of GFP and IAPEz in their respective cell lines, but while the effect of GFP induction was minor, IAPEz induction significantly reduced SE and SE associated gene transcription at the miR290-295, Klf4 and Fgf4 loci, and did not significantly affect typical enhancer and associated genes. In addition, cellular fractionation assays demonstrated that the nuclear enrichment of IAPEz RNA was about double that of GFP RNA indicating that IAPEz are more likely to be retained in the nucleus. This experiment was significant because it implied forced induction of specific RNAs (in this case ERV) may have a more pronounced

effect on transcriptional hijacking, most probably due to the inherent features of such RNA species including nuclear enrichment and the capacity to nucleate transcriptional condensates.

Data presented here demonstrates ERV de-repression affects pluripotency in mESCs by compromising the transcription of key pluripotency genes. This mechanism should translate to mouse embryonic development and provide some of the missing links in our understanding of embryonic lethality of TRIM28 KO. Additional results, beyond the scope of this thesis, demonstrated that pluripotent lineages are depleted in TRIM28 KO mouse embryos (Asimi et al., 2022). Single-cell RNA-seq analysis of E6.5 TRIM28 KO embryos showed transcriptional upregulation of major ERV classes and a marked shortage of epiblast cells which typically originate from the inner cell mass, structure that contains pluripotent stem cells during embryonic development (G. R. Martin, 1981). Inner cell mass of E3.5 TRIM28 KO embryos was almost completely depleted of pluripotency factors NANOG, OCT4, SOX2 and KLF4, and replaced by cells expressing GATA6, the endoderm marker, tested by immunofluorescence. ERV expression levels are tightly regulated during embryonic development, IAP expression is minimal during the two-cell stage, peaks during the blastocyst stage and downregulates again in the later stages. The ability to follow this expression pattern seems to be essential for normal development, and absence of TRIM28 mediated silencing hinders this. Transcriptional hijacking provides a molecular mechanism that could explain the correlation between the loss of pluripotency and the upregulated ERV transcription in TRIM28 KO embryos.

While single-cell RNA-seq and immunofluorescence imaging data in mouse embryos form solid initial evidence that our in vitro and mESC data on condensate hijacking may translate to live embryos, there are additional ways to directly test this model. Targeted protein degradation using the dTAG system has recently been successfully adapted to mouse embryos to investigate the role of RNAPII pausing in embryonic

development (Abuhashem et al., 2022). A similar strategy would allow acute depletion of TRIM28 and help dissect the effects of dynamic ERV upregulation during the various embryonic stages. Additionally, more systematic profiling of ERV RNAs using RNA FISH and nascent transcriptomics combined with TRIM28 degradation at each embryonic stage would be a valuable resource since it may provide information on ERV transcript levels and localization and give a more complete picture of the dependency between TRIM28 repression and ERV upregulation phenotype.

One of the direct outcomes of the transcriptional hijacking model is the capacity of newly formed ERV-associated RNAPII condensates to incorporate spatially adjacent genes. This may be particularly relevant in processes such as embryonic development or cancer, where enrichment of specific subsets of genes in the pool of ERV-incorporated genes could drive cell decisions towards specific embryonic lineages or pathologies. Indeed, single-cell RNA-seq data in TRIM28 KO embryos showed epiblast cells were replaced by parietal endoderm-like cells and endoderm markers were accordingly upregulated in TRIM28 degraded mESCs (Asimi et al., 2022). This thesis focused on one prominent example of ERV-proximal genes, *Cthrc1*, to demonstrate a proof of principle. Data showed that *Cthrc1* gets significantly upregulated and associates with RNAPII and NFYA upon TRIM28 degradation. In addition, CRISPR KO of all three ERVs in *Cthrc1* gene's vicinity showed that its upregulation is conditional upon the presence ERV RNA. Deeper analysis of TT-SLAM-seq data and subsequent CRISPR characterization of candidate loci may show that this is a general mechanism that contributes to TRIM28 depletion phenotypes. Moreover, the nearby gene incorporation by rearranged transcriptional condensates may apply to knockouts of other regulators of retrotransposon repression, or knockouts of any other general epigenetic factors that cause the redistribution of transcriptional machinery.

Imaging data presented here shows upregulated nascent *Cthrc1* transcripts frequently associate with RNAPII and NFYA protein upon TRIM28 degradation in mESCs, but

they do not directly test colocalization of ERV RNAs and *Cthrc1*-containing transcriptional condensates. Dual color RNA FISH targeting *IAPEz* transcripts and *Cthrc1* nascent transcripts multiplexed with RNAPII immunofluorescence would be one way to test the frequency of such interactions, but this proved to be technically difficult using the current setup. An additional option would be to use the readily available triple ERV KO cell line and test the association of the *Cthrc1* locus with RNAPII via DNA FISH in wildtype and triple ERV KO condition with or without TRIM28 degradation. Such experimental setup would be independent of direct *Cthrc1* transcriptional output and indicate the probability of *Cthrc1* DNA locus' increased interaction with RNAPII when ERVs are present. These experiments would provide definitive evidence that ERV RNAs co-reside in identical nuclear compartments with RNAPII and other components of transcriptional condensates and corroborate the nearby gene incorporation hypothesis of the hijacking model.

4.4 Future perspective: TRIM28 condensates

Repetitive DNA and RNA sequences are known drivers of nuclear organization (Frank & Rippe, 2020). Well studied examples include the nucleolus which is shaped around clusters of repetitive ribosomal DNA (rDNA) and the histone locus bodies which form around clustered histone genes (Quinodoz et al., 2018). Within the biomolecular condensate context, DNA can serve as multivalent binding sites for proteins to nucleate condensates at both repressed and active chromatin loci. Considering majority of ERVs are targeted by the repressive complex that includes TRIM28, this process may be facilitated by long range interactions between ERVs and the concomitant formation of heterochromatin condensates at these loci (Singh & Newman, 2020). In situ Hi-C experiments indicated TRIM28 degradation does not cause major disruptions of genome organization, but the most-induced ERV classes do transition from the inactive to the active chromatin compartment and ERVs tend to contact SEs and transcribed genes more frequently in mESCs upon 24 h TRIM28 (Asimi et al., 2022). Whether

prolonged TRIM28 degradation in mESCs may have other major effects on nuclear organization is still an open question.

Unpublished data from this PhD work suggest TRIM28 protein may itself undergo phase separation in vitro and form condensates that repel transcription factor and coactivator condensates in cells. This process may likely be mediated by disordered sequences located on the N-terminal and part of the RBCC domain of TRIM28. Interestingly, recent crystal structure and size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS) investigations of TRIM28 identified the disordered B-Box 1 portion of the RBCC domain as the main driver of TRIM28's concentration-dependent higher-order oligomerization (Stoll et al., 2019; Sun et al., 2019). In this model TRIM28 has the capacity to self-assemble through weak homotypic interactions between the B-boxes of the RBCC domain. High local concentrations of TRIM28 induce self-assembly of dimers into tetramers, octamers, and potentially higher-order polymers (Stoll et al., 2019). This intrinsic capacity to form higher-order complexes is nevertheless not required for transcriptional silencing of retrotransposons, since oligomerization-deficient mutants are still efficient at repressing SVA-D and LINE-1 retrotransposons in a luciferase reporter assay (Stoll et al., 2019). On the other hand, mutations in the KRAB-binding region of the RBCC domain and the HP1 binding motif significantly reduce the capacity of TRIM28 to transcriptionally silence retrotransposons. Multivalent interactions of TRIM28-HP1 complex with H3K9me2 and 3-modified nucleosomal arrays resulting in phase-separated droplets have also been reported as a mechanism behind heterochromatin formation (L. Wang et al., 2019). Our preliminary data indicates formation of TRIM28 condensates in overexpression imaging experiments redistributes endogenous HP1α signal from chromocenters to all regions occupied by TRIM28. This is significant because it indicates that although condensation or higher order oligomerization may not be essential for retrotransposon silencing, high local concentrations or reported overexpression of TRIM28, in for instance, tumor cell lines may, among other effects, lead to the dilution of HP1α away from its native targets, potentially resulting in target

derepression (C. Su et al., 2018; Wei et al., 2016). Additional analyses, combining inputs from condensate biology and structural biochemistry, should dissect the specific effects of TRIM28 phase separation on nuclear organization and transcriptional repression of retrotransposons.

The multi-domain structure of TRIM28 may translate to multifunctionality and explain TRIM28's implication in various other processes beyond retrotransposon silencing (Czerwińska et al., 2017; Iyengar & Farnham, 2011). These pleiotropic effects include TRIM28-mediated SUMOylation of physiologically relevant proteins like NPM1 or TRIM28 implication in DNA damage response (C.-T. Cheng et al., 2014; Neo et al., 2015). TRIM28 degron system has proved valuable in the investigation of acute transcriptional consequences of TRIM28 depletion in mESCs, but it could additionally be utilized to study TRIM28 role in differentiated cell types. Moreover, it provides a new modality to attempt rescue experiments, for example by simultaneously degrading TRIM28 whole protein and overexpressing individual TRIM28 domains or combinations of domains. The proposed condensation driving RBCC domain together with HP1 binding motif would be one such construct that could be used test the capacity of TRIM28 localization rescue using parts of its sequence. Iterative domain overexpression may also help separate the various anticipated phenotypes of TRIM28 dysregulation. Similarly, the RBCC domain could be used to design synthetic constructs that include well-studied TF IDRs or DBDs and inspect if this potentially enhances condensation capacity or transactivation activity of these sequences.

4.5 Implications for disease and open questions

TRIM28 overexpression was reported in breast, gastric, pancreatic and ovarian cancers (Addison et al., 2015; Cui et al., 2014; Y.-Y. Wang et al., 2013; Yu et al., 2014). TRIM28 haploinsufficiency, on the other hand, predisposes patients to Wilms' tumor and dysregulates a gene network capable of driving obesity in mice (Dalgaard et al., 2016; Diets et al., 2019). Recent in vivo CRISPR-Cas9 screens in mouse cancer models

identified TRIM28 as one of the top chromatin regulating suppressors of tumor-intrinsic immunity (Griffin et al., 2021). Targeting TRIM28 directly may additionally offer a new modality for breast cancer treatment (Wei et al., 2016). Transcriptional hijacking and TRIM28 phase separation offer new conceptual inputs that could be used to review these disease phenotypes and plan new analyses that may elucidate the molecular interplay between retrotransposons and transcriptional regulation in disease.

Beyond TRIM28 associated pathologies, upregulated expression of human ERVs has been detected in ALS patient neurons and associated with the development of schizophrenia (Karlsson et al., 2001; W. Li et al., 2015). Retrotransposons, particularly the LINE-1 elements, are upregulated in cancer and their activity can lead to somatic acquisition of insertions in cancer genomes (Burns, 2017; Miki et al., 1992). LTRs are immobile but their derepression can lead to activity as alternative promoters for expression of oncogenes like CSF1R in human lymphoma or IRF5 in Hodgkin lymphoma (Babaian et al., 2016; Lamprecht et al., 2010). Recent data from the TRACERx consortium showed mouse and human B cells inside lung cancer environment unexpectedly generate antibodies against ERV envelope proteins, and ERV expression correlates with positive response to immune checkpoint blockade therapy in human lung adenocarcinoma (Ng et al., 2023). These studies collectively indicate retrotransposon regulation is becoming a promising pre-clinical research focus in oncology.

Condensates have been linked to disease pathologies such as cancer, neurodegeneration and rare genetic disease (Boija et al., 2021; Fasciani et al., 2020; C. H. Li et al., 2020; Patel et al., 2015). Targeting condensate properties may provide new therapeutic modalities. Broadly speaking, condensate level interventions could be grouped into three main categories: drugs that concentrate in specific condensates, drugs that selectively change condensate properties, and drugs that alter condensate component's post-translational modifications (Boija et al., 2021). Most existing pharmaceutical targets are structured proteins including G-protein coupled receptors, enzymes,

kinases, ion channels and nuclear receptors, but over half of the genome, including proteins with IDRs, do not fall into these categories and were long considered undruggable (Dang et al., 2017; Wheeler, 2020). Possible therapies that target physicochemical properties of biomolecular condensates formed by multivalent interactions between IDRs are attractive because they do not involve conventional drug binding sites.

Building on the hijacking model, it may be possible to use RNA, either produced directly at or delivered to the condensate site, to nucleate, alter or dissolve transcriptional condensates and achieve more optimal outcomes, for instance, activate immune response in cancer. ERV knockdown data showed depleting repeat RNAs using shRNAs rescues transcription of pluripotency genes and delays the acute effects of TRIM28 degradation. Several drug development programs are currently ongoing in the direction of modulating non-coding RNA levels to alter transcriptional programs associated with disease (including using ASOs to either boost local RNA levels, by targeting regulatory non-coding RNAs, or reduce levels of lncRNAs associated with cardiac disease) (Alexanian & Ounzain, 2020; Aydemir et al., 2022). Condensate processes associated with non-coding RNAs are emerging as active therapeutic targets with high translational potential.

Biomolecular condensate research has entered its second decade, unquestionably revolutionizing our understanding of cellular biology. Nevertheless, several credible research labs are pointing out the lack of causal evidence that relates biochemical behavior observed in vitro with in vivo phenotypes (McSwiggen et al., 2019; Musacchio, 2022; B. Wang et al., 2021). Even though local enrichment of a particular transcription factor inside discrete nuclear foci is suggestive of phase separation, future studies must go beyond descriptive research and phenomenological evidence, and focus on quantitative perturbation experiments of both the individual condensate components and the physicochemical features of condensates (Frank & Rippe, 2020; McSwiggen et

al., 2019). Many fundamental questions regarding the function and importance of biomolecular condensates remain elusive, including the association with pathophysiology (Conti & Oppikofer, 2022). Hnisz lab made pioneering efforts to mechanistically link condensate dysfunction to disease, including rare genetic disorders and castration resistant prostate cancer (Basu et al., 2020, 2022; Mensah et al., 2023). This guides future research to leverage basic condensate biology into viable therapeutic strategies. Small molecules and RNA based therapeutics that modulate condensate properties are primary intervention candidates and optimizing delivery and tissue selectivity will be key hurdles en route to the clinic (Conti & Oppikofer, 2022). Common chemotherapy agents like cisplatin were already found to concentrate within condensates in cells (Klein et al., 2020). Implications for drug discovery are multifaceted since existing pharmaceuticals may modulate condensate properties without having previously been investigated in this context. Condensate assays are therefore becoming a viable screening strategy in drug discovery pipelines (Patel et al., 2022).

Mounting evidence suggests ERVs and other retrotransposons have unanticipated roles in gene regulation and disease. In addition, retrotransposons were an important instrument for evolutionary innovation and some of the evolutionary impetuses may have been driven by a mechanism independent of transposition but resembling transcriptional hijacking. Development of more tailored techniques that allow investigation of specific elements via microscopy and improved bioinformatic analyses should allow us to mechanistically resolve retrotransposon contribution to transcriptional regulation, evolution and disease (Lerat, 2022). Integrating the biomolecular condensate paradigm into existing retrotransposon research provides a novel framework to reevaluate ERV related phenotypes in developmental biology. Transcriptional condensate hijacking is an insightful example of such consolidative approaches. Follow up testing of the hijacking model including additional condensate perturbation and rescue experiments in vivo should demonstrate its mechanistic role

in embryonic development and stimulate exploratory studies as potential therapeutic modality in cancer.

5 Bibliography

Abuhashem, A., Lee, A. S., Joyner, A. L., & Hadjantonakis, A.-K. (2022). Rapid and efficient degradation of endogenous proteins in vivo identifies stage-specific roles of RNA Pol II pausing in mammalian development. *Developmental Cell*, *57*(8), 1068-1080.e6. https://doi.org/10.1016/j.devcel.2022.03.013

Addison, J. B., Koontz, C., Fugett, J. H., Creighton, C. J., Chen, D., Farrugia, M. K., Padon, R. R., Voronkova, M. A., McLaughlin, S. L., & Livengood, R. H. (2015). KAP1 Promotes Proliferation and Metastatic Progression of Breast Cancer CellsKAP1 Stabilizes KRAB-ZNFs and Promotes Mammary Tumorigenesis. *Cancer Research*, *75*(2), 344–355.

Ahmad, S., Mu, X., Yang, F., Greenwald, E., Park, J. W., Jacob, E., Zhang, C.-Z., & Hur, S. (2018). Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation. *Cell*, *172*(4), 797-810.e13. https://doi.org/10.1016/j.cell.2017.12.016

Alexanian, M., & Ounzain, S. (2020). Long Noncoding RNAs in Cardiac Development. *Cold Spring Harbor Perspectives in Biology*, *12*(11), a037374. https://doi.org/10.1101/cshperspect.a037374

Allain, F. H., Bouvet, P., Dieckmann, T., & Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *The EMBO Journal*, *19*(24), 6870–6881. https://doi.org/10.1093/emboj/19.24.6870

Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature Reviews. Molecular Cell Biology*, *19*(4), 229–244. https://doi.org/10.1038/nrm.2017.119

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., Freeman, T. C., Alhendi, A. M. N., Arner, P., Axton, R., Baillie, J. K., Beckhouse, A., Bodega, B., Briggs, J., Brombacher, F., … Hayashizaki, Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (New York, N.Y.)*, *347*(6225), 1010–1014. https://doi.org/10.1126/science.1259418

Asimi, V., Sampath Kumar, A., Niskanen, H., Riemenschneider, C., Hetzel, S., Naderi, J., Fasching, N., Popitsch, N., Du, M., Kretzmer, H., Smith, Z. D., Weigert, R., Walther, M., Mamde, S., Meierhofer, D., Wittler, L., Buschow, R., Timmermann, B., Cisse, I. I., … Hnisz, D. (2022). Hijacking of transcriptional condensates by endogenous retroviruses. *Nature Genetics*, *54*(8), Article 8. https://doi.org/10.1038/s41588-022-01132-w

Aumiller, W. M., Pir Cakmak, F., Davis, B. W., & Keating, C. D. (2016). RNA-Based Coacervates as a Model for Membraneless Organelles: Formation, Properties, and Interfacial Liposome Assembly. *Langmuir: The ACS Journal of Surfaces and Colloids*, *32*(39), 10042–10053. https://doi.org/10.1021/acs.langmuir.6b02499

Aydemir, Y. Y., Abdullatif, A. A., Ramaswami, G., Golczer, G., Akerberg, B., Weinstein, J., Nishi, Y., Pai, R., Cohick, E., Matthews, B., Liu, Y., Giagtzoglou, N., Bumcrot, D., & Sehgal, A. (2022). Progranulin upregulation by antisense oligonucleotides targeting regRNAs for treatment of FTD-GRN. *Alzheimer's & Dementia*, *18*(S10), e067563. https://doi.org/10.1002/alz.067563

Babaian, A., Romanish, M. T., Gagnier, L., Kuo, L. Y., Karimi, M. M., Steidl, C., & Mager, D. L. (2016). Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene*, *35*(19), 2542–2546. https://doi.org/10.1038/onc.2015.308

Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., & Rosen, M. K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell*, *166*(3), 651–663. https://doi.org/10.1016/j.cell.2016.06.010

Banjade, S., Wu, Q., Mittal, A., Peeples, W. B., Pappu, R. V., & Rosen, M. K. (2015). Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6426-6435. https://doi.org/10.1073/pnas.1508778112

Basu, S., Mackowiak, S. D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., & Ewers, H. (2020). Unblending of transcriptional condensates in human repeat expansion disease. *Cell*, *181*(5), 1062–1079.

Basu, S., Martínez-Cristóbal, P., Pesarrodona, M., Frigolé-Vivas, M., Lewis, M., Szulc, E., Bañuelos, C. A., Sánchez-Zarzalejo, C., Bielskutė, S., Zhu, J., Pombo-García, K., Garcia-Cabau, C., Batlle, C., Mateos, B., Biesaga, M., Escobedo, A., Bardia, L., Verdaguer, X., Ruffoni, A., … Salvatella, X. (2022). *Rational optimization of a transcription factor activation domain inhibitor* (p. 2022.08.18.504385). bioRxiv. https://doi.org/10.1101/2022.08.18.504385

Berry, J., Weber, S. C., Vaidya, N., Haataja, M., & Brangwynne, C. P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(38), E5237–E5245. https://doi.org/10.1073/pnas.1509317112

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V. H., Janke, A. M., Baatsen, P., Vercruysse, T., Kolaitis, R.-M., Daelemans, D., Taylor, J. P., Kedersha, N., Anderson, P., Impens, F., … Van Den Bosch, L. (2017). Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Molecular Cell*, *65*(6), 1044-1055.e5. https://doi.org/10.1016/j.molcel.2017.02.013

Boeynaems, S., Holehouse, A. S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P. S., Pappu, R. V., & Gitler, A. D. (2019). Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proceedings of the National Academy of Sciences*, *116*(16), 7889–7898. https://doi.org/10.1073/pnas.1821038116

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J., & Young, R. A. (2018). Transcription factors activate genes through the phase separation capacity of their activation domains. *Cell*, *175*(7), 1842-1855.e16. https://doi.org/10.1016/j.cell.2018.10.042

Boija, A., Klein, I. A., & Young, R. A. (2021). Biomolecular Condensates and Cancer. *Cancer Cell*, *39*(2), 174–192. https://doi.org/10.1016/j.ccell.2020.12.003

Bolt, C. C., Lopez-Delisle, L., Hintermann, A., Mascrez, B., Rauseo, A., Andrey, G., & Duboule, D. (2022). Context-dependent enhancer function revealed by targeted inter-

TAD relocation. *Nature Communications*, *13*(1), 3488. https://doi.org/10.1038/s41467-022-31241-3

Bolte, S., & Cordelières, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*, *224*(3), 213–232. https://doi.org/10.1111/j.1365-2818.2006.01706.x

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., & Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.)*, *324*(5935), 1729–1732. https://doi.org/10.1126/science.1172046

Brangwynne, C. P., Mitchison, T. J., & Hyman, A. A. (2011). Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4334–4339. https://doi.org/10.1073/pnas.1017150108

Bulut-Karslioglu, A., De La Rosa-Velázquez, I. A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G. E., Engist, B., Gerle, B., O'Sullivan, R. J., Martens, J. H. A., Walter, J., Manke, T., Lachner, M., & Jenuwein, T. (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular Cell*, *55*(2), 277–290. https://doi.org/10.1016/j.molcel.2014.05.029

Burke, K. A., Janke, A. M., Rhine, C. L., & Fawzi, N. L. (2015). Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular Cell*, *60*(2), 231–241. https://doi.org/10.1016/j.molcel.2015.09.006

Burns, K. H. (2017). Transposable elements in cancer. *Nature Reviews Cancer*, *17*(7), Article 7. https://doi.org/10.1038/nrc.2017.35

Cajal, S. R. (1903). Un sencillo metodo de coloracion seletiva del reticulo protoplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. *Trab. Lab. Invest. Biol.(Madrid)*, *2*, 129–221.

Cammas, F., Mark, M., Dollé, P., Dierich, A., Chambon, P., & Losson, R. (2000). Mice lacking the transcriptional corepressor TIF1β are defective in early postimplantation development. *Development*, *127*(13), 2955–2963. https://doi.org/10.1242/dev.127.13.2955

Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Rucli, S., Dossin, F., Fouassier, C., Lameiras, S., & Bourc'his, D. (2021). M6A RNA methylation regulates the fate of endogenous retroviruses. *Nature*, *591*(7849), 312–316. https://doi.org/10.1038/s41586-020-03135-1

Cheng, C.-T., Kuo, C.-Y., & Ann, D. K. (2014). KAPtain in charge of multiple missions: Emerging roles of KAP1. *World Journal of Biological Chemistry*, *5*(3), 308–320. https://doi.org/10.4331/wjbc.v5.i3.308

Cheng, Y., Xie, W., Pickering, B. F., Chu, K. L., Savino, A. M., Yang, X., Luo, H., Nguyen, D. T., Mo, S., Barin, E., Velleca, A., Rohwetter, T. M., Patel, D. J., Jaffrey, S. R., & Kharas, M. G. (2021). N6-Methyladenosine on mRNA facilitates a phase-separated nuclear body that suppresses myeloid leukemic differentiation. *Cancer Cell*, *39*(7), 958-972.e8. https://doi.org/10.1016/j.ccell.2021.04.017

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., & Cisse, I. I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (New York, N.Y.)*, *361*(6400), 412–415. https://doi.org/10.1126/science.aar4199

Conti, B. A., & Oppikofer, M. (2022). Biomolecular condensates: New opportunities for drug discovery and RNA therapeutics. *Trends in Pharmacological Sciences*, *43*(10), 820–837. https://doi.org/10.1016/j.tips.2022.07.001

Courchaine, E. M., Lu, A., & Neugebauer, K. M. (2016). Droplet organelles? *The EMBO Journal*, *35*(15), 1603–1612. https://doi.org/10.15252/embj.201593517

Crichton, J. H., Dunican, D. S., MacLennan, M., Meehan, R. R., & Adams, I. R. (2014). Defending the genome from the enemy within: Mechanisms of retrotransposon suppression in the mouse germline. *Cellular and Molecular Life Sciences*, *71*(9), 1581–1605. https://doi.org/10.1007/s00018-013-1468-0

Cui, Y., Yang, S., Fu, X., Feng, J., Xu, S., & Ying, G. (2014). High levels of KAP1 expression are associated with aggressive clinical features in ovarian cancer. *International Journal of Molecular Sciences*, *16*(1), 363–377.

Czerwińska, P., Mazurek, S., & Wiznerowicz, M. (2017). The complexity of TRIM28 contribution to cancer. *Journal of Biomedical Science*, *24*(1), 63. https://doi.org/10.1186/s12929-017-0374-4

Dalgaard, K., Landgraf, K., Heyne, S., Lempradl, A., Longinotto, J., Gossens, K., Ruf, M., Orthofer, M., Strogantsev, R., Selvaraj, M., Lu, T. T.-H., Casas, E., Teperino, R., Surani, M. A., Zvetkova, I., Rimmington, D., Tung, Y. C. L., Lam, B., Larder, R., … Pospisilik, J. A. (2016). Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity. *Cell*, *164*(3), 353–364. https://doi.org/10.1016/j.cell.2015.12.025

Dang, C. V., Reddy, E. P., Shokat, K. M., & Soucek, L. (2017). Drugging the “undruggable” cancer targets. *Nature Reviews Cancer*, *17*(8), Article 8. https://doi.org/10.1038/nrc.2017.36

Diets, I. J., Hoyer, J., Ekici, A. B., Popp, B., Hoogerbrugge, N., van Reijmersdal, S. V., Bhaskaran, R., Hadjihannas, M., Vasileiou, G., Thiel, C. T., Seven, D., Uebe, S.,

Ilencikova, D., Waanders, E., Mavinkurve-Groothuis, A. M. C., Roeleveld, N., de Krijger, R. R., Wegert, J., Graf, N., … Metzler, M. (2019). TRIM28 haploinsufficiency predisposes to Wilms tumor. *International Journal of Cancer*, *145*(4), 941–951. https://doi.org/10.1002/ijc.32167

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Drobot, B., Iglesias-Artola, J. M., Le Vay, K., Mayr, V., Kar, M., Kreysing, M., Mutschler, H., & Tang, T.-Y. D. (2018). Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nature Communications*, *9*(1), 3643. https://doi.org/10.1038/s41467-018-06072-w

Egloff, S., Studniarek, C., & Kiss, T. (2018). 7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features. *Transcription*, *9*(2), 95–101. https://doi.org/10.1080/21541264.2017.1344346

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C.-H., Eckmann, C. R., Myong, S., & Brangwynne, C. P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(23), 7189–7194. https://doi.org/10.1073/pnas.1504822112

Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D., & Banaszynski, L. A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature*, *522*(7555), 240–244. https://doi.org/10.1038/nature14345

Enriquez-Gasca, R., Gould, P. A., Tunbak, H., Conde, L., Herrero, J., Chittka, A., Gifford, R., & Rowe, H. M. (2022). *Co-option of endogenous retroviruses through genetic escape from TRIM28 repression* (p. 2022.06.22.497016). bioRxiv. https://doi.org/10.1101/2022.06.22.497016

Fasciani, A., D'Annunzio, S., Poli, V., Fagnocchi, L., Beyes, S., Michelatti, D., Corazza, F., Antonelli, L., Gregoretti, F., Oliva, G., Belli, R., Peroni, D., Domenici, E., Zambrano, S., Intartaglia, D., Settembre, C., Conte, I., Testi, C., Vergyris, P., … Zippo, A. (2020). MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome. *Nature Genetics*, *52*(12), Article 12. https://doi.org/10.1038/s41588-020-00724-8

Frank, L., & Rippe, K. (2020). Repetitive RNAs as Regulators of Chromatin-Associated Subcompartment Formation by Phase Separation. *Journal of Molecular Biology*, *432*(15), 4270–4286. https://doi.org/10.1016/j.jmb.2020.04.015

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*, *47*(D1), D766–D773. https://doi.org/10.1093/nar/gky955

Fromm, S. A., Kamenz, J., Nöldeke, E. R., Neu, A., Zocher, G., & Sprangers, R. (2014). In vitro reconstitution of a cellular phase-transition process that involves the mRNA

decapping machinery. *Angewandte Chemie (International Ed. in English)*, *53*(28), 7354–7359. https://doi.org/10.1002/anie.201402885

Furey, T. S. (2012). ChIP-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews. Genetics*, *13*(12), 840–852. https://doi.org/10.1038/nrg3306

Gagnier, L., Belancio, V. P., & Mager, D. L. (2019). Mouse germ line mutations due to retrotransposon insertions. *Mobile DNA*, *10*(1), 15. https://doi.org/10.1186/s13100-019-0157-4

Gerdes, P., Richardson, S. R., Mager, D. L., & Faulkner, G. J. (2016). Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service. *Genome Biology*, *17*(1), 100. https://doi.org/10.1186/s13059-016-0965-5

Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S., & Rosen, M. K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*, *179*(2), 470-484.e21. https://doi.org/10.1016/j.cell.2019.08.037

Gilles, J.-F., Dos Santos, M., Boudier, T., Bolte, S., & Heck, N. (2017). DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods*, *115*, 55–64. https://doi.org/10.1016/j.ymeth.2016.11.016

Glaser, J., & Mundlos, S. (2021). *3D or not 3D: shaping the genome during development. Cold Spring Harb Perspect Biol a040188*.

Goodier, J. L., & Kazazian, H. H. (2008). Retrotransposons revisited: The restraint and rehabilitation of parasites. *Cell*, *135*(1), 23–35. https://doi.org/10.1016/j.cell.2008.09.022

Griffin, G. K., Wu, J., Iracheta-Vellve, A., Patti, J. C., Hsu, J., Davis, T., Dele-Oni, D., Du, P. P., Halawi, A. G., Ishizuka, J. J., Kim, S. Y., Klaeger, S., Knudsen, N. H., Miller, B.

C., Nguyen, T. H., Olander, K. E., Papanastasiou, M., Rachimi, S., Robitschek, E. J., … Bernstein, B. E. (2021). Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. *Nature*, *595*(7866), Article 7866. https://doi.org/10.1038/s41586-021-03520-4

Grob, S., & Cavalli, G. (2018). Technical Review: A Hitchhiker's Guide to Chromosome Conformation Capture. *Methods in Molecular Biology (Clifton, N.J.)*, *1675*, 233–246. https://doi.org/10.1007/978-1-4939-7318-7_14

Grosswendt, S., Kretzmer, H., Smith, Z. D., Kumar, A. S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S., & Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. *Nature*, *584*(7819), Article 7819. https://doi.org/10.1038/s41586-020-2552-x

Gu, Z., Eils, R., Schlesner, M., & Ishaque, N. (2018). EnrichedHeatmap: An R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics*, *19*(1), 234. https://doi.org/10.1186/s12864-018-4625-x

Gu, Z., Liu, Y., Zhang, Y., Cao, H., Lyu, J., Wang, X., Wylie, A., Newkirk, S. J., Jones, A. E., Lee, M., Botten, G. A., Deng, M., Dickerson, K. E., Zhang, C. C., An, W., Abrams, J. M., & Xu, J. (2021). Silencing of LINE-1 retrotransposons is a selective dependency of myeloid leukemia. *Nature Genetics*, *53*(5), Article 5. https://doi.org/10.1038/s41588-021-00829-8

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., Spille, J.-H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., Abraham, B. J., Boija, A., Decker, T.-M., Rimel, J. K., Fant, C. B., Lee, T. I., Cisse, I. I., Sharp, P. A., Taatjes, D. J., & Young, R. A. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, *572*(7770), Article 7770. https://doi.org/10.1038/s41586-019-1464-0

Han, D., Longhini, A. P., Zhang, X., Hoang, V., Wilson, M. Z., & Kosik, K. S. (2022). Dynamic assembly of the mRNA m6A methyltransferase complex is regulated by METTL3 phase separation. *PLOS Biology*, *20*(2), e3001535. https://doi.org/10.1371/journal.pbio.3001535

Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., & McKnight, S. L. (2012). Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell*, *149*(4), 768–779. https://doi.org/10.1016/j.cell.2012.04.016

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Heitz, E. (1928). *Das heterochromatin der moose*. Bornträger.

Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M. M., Andrews, J. O., Zamudio, A. V., Lazaris, C., Hannett, N. M., Lee, T. I., Sharp, P. A., Cissé, I. I., Chakraborty, A. K., & Young, R. A. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell*, *184*(1), 207-225.e24. https://doi.org/10.1016/j.cell.2020.11.030

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*, *14*(12), Article 12. https://doi.org/10.1038/nmeth.4435

Hirose, T., Ninomiya, K., Nakagawa, S., & Yamazaki, T. (2023). A guide to membraneless organelles and their various roles in gene regulation. *Nature Reviews Molecular Cell Biology*, *24*(4), 288–304. https://doi.org/10.1038/s41580-022-00558-8

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, *169*(1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007

Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A. M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y. A., … Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, *543*(7644), 199–204. https://doi.org/10.1038/nature21374

Hyman, A. A., & Brangwynne, C. P. (2011). Beyond stereospecificity: Liquids and mesoscale organization of cytoplasm. *Developmental Cell*, *21*(1), 14–16. https://doi.org/10.1016/j.devcel.2011.06.013

Hyman, A. A., & Simons, K. (2012). Cell biology. Beyond oil and water—Phase transitions in cells. *Science (New York, N.Y.)*, *337*(6098), 1047–1049. https://doi.org/10.1126/science.1223728

Iyengar, S., & Farnham, P. J. (2011). KAP1 Protein: An Enigmatic Master Regulator of the Genome*. *Journal of Biological Chemistry*, *286*(30), 26267–26276. https://doi.org/10.1074/jbc.R111.252569

Jain, A., & Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature*, *546*(7657), 243–247. https://doi.org/10.1038/nature22386

Ji, X., Dadon, D. B., Abraham, B. J., Lee, T. I., Jaenisch, R., Bradner, J. E., & Young, R. A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-

marked genomic regions. *Proceedings of the National Academy of Sciences*, *112*(12), 3841–3846. https://doi.org/10.1073/pnas.1502971112

Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., & Feigon, J. (2004). Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *Journal of Molecular Biology*, *337*(4), 799–816. https://doi.org/10.1016/j.jmb.2004.01.056

Jordan Rowley, M., & Corces, V. G. (2018). Organizational Principles of 3D Genome Architecture. *Nature Reviews. Genetics*, *19*(12), 789–800. https://doi.org/10.1038/s41576-018-0060-8

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, *110*(1–4), 462–467. https://doi.org/10.1159/000084979

Karlsson, H., Bachmann, S., Schröder, J., McArthur, J., Torrey, E. F., & Yolken, R. H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proceedings of the National Academy of Sciences*, *98*(8), 4634–4639. https://doi.org/10.1073/pnas.061021998

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., & McKnight, S. L. (2012). Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell*, *149*(4), 753–767. https://doi.org/10.1016/j.cell.2012.04.017

Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*, *26*(12), 1351–1359. https://doi.org/10.1038/nbt.1508

Kiss-László, Z., Henry, Y., Bachellerie, J.-P., Caizergues-Ferrer, M., & Kiss, T. (1996). Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell*, *85*(7), 1077–1088. https://doi.org/10.1016/S0092-8674(00)81308-2

Klein, I. A., Boija, A., Afeyan, L. K., Hawken, S. W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J. E., Shrinivas, K., Sabari, B. R., Sagi, I., Clark, V. E., Platt, J. M., Kar, M., McCall, P. M., Zamudio, A. V., Manteiga, J. C., Coffey, E. L., Li, C. H., … Young, R. A. (2020). Partitioning of cancer therapeutics in nuclear condensates. *Science*, *368*(6497), 1386–1392. https://doi.org/10.1126/science.aaz4427

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA: Chromatin structure is based on a repeating unit of eight histone molecules and about 200 DNA base pairs. *Science*, *184*(4139), 868–871.

Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., & Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods*, *11*(3), Article 3. https://doi.org/10.1038/nmeth.2834

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., & Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, *42*(7), Article 7. https://doi.org/10.1038/ng.600

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J. L., & McKnight, S. L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*, *155*(5), 1049–1060. https://doi.org/10.1016/j.cell.2013.10.033

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, *410*(6824), Article 6824. https://doi.org/10.1038/35065132

Lakadamyali, M., & Cosma, M. P. (2020). Visualizing the genome in high resolution challenges our textbook understanding. *Nature Methods*, *17*(4), 371–379. https://doi.org/10.1038/s41592-020-0758-3

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D., Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I., … Mathas, S. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nature Medicine*, *16*(5), 571–579, 1p following 579. https://doi.org/10.1038/nm.2129

Langdon, E. M., Qiu, Y., Niaki, A. G., McLaughlin, G. A., Weidmann, C., Gerbich, T. M., Smith, J. A., Crutchley, J. M., Termini, C. M., Weeks, K. M., Myong, S., & Gladfelter, A. S. (2018). MRNA structure determines specificity of a polyQ-driven phase separation. *Science (New York, N.Y.)*, *360*(6391), 922–927. https://doi.org/10.1126/science.aar7432

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, *547*(7662), Article 7662. https://doi.org/10.1038/nature22822

Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics (Oxford, England)*, *25*(14), 1841–1842. https://doi.org/10.1093/bioinformatics/btp328

Lerat, E. (2022). Recent Bioinformatic Progress to Identify Epigenetic Changes Associated to Transposable Elements. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/articles/10.3389/fgene.2022.891194

Li, C. H., Coffey, E. L., Dall'Agnese, A., Hannett, N. M., Tang, X., Henninger, J. E., Platt, J. M., Oksuz, O., Zamudio, A. V., Afeyan, L. K., Schuijers, J., Liu, X. S., Markoulaki, S., Lungjangwa, T., LeRoy, G., Svoboda, D. S., Wogram, E., Lee, T. I., Jaenisch, R., & Young, R. A. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature*, *586*(7829), Article 7829. https://doi.org/10.1038/s41586-020-2574-4

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T., & Rosen, M. K. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature*, *483*(7389), 336–340. https://doi.org/10.1038/nature10879

Li, W., Lee, M.-H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., Campanac, E., Hoffman, D. A., von Geldern, G., Johnson, K., Maric, D., Morris, H. D., Lentz, M., Pak, K., Mammen, A., Ostrow, L., Rothstein, J., & Nath, A. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Science Translational Medicine*, *7*(307), 307ra153. https://doi.org/10.1126/scitranslmed.aac8201

Li, W., Notani, D., & Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nature Reviews. Genetics*, *17*(4), 207–223. https://doi.org/10.1038/nrg.2016.4

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Liu, J., Gao, M., He, J., Wu, K., Lin, S., Jin, L., Chen, Y., Liu, H., Shi, J., Wang, X., Chang, L., Lin, Y., Zhao, Y.-L., Zhang, X., Zhang, M., Luo, G.-Z., Wu, G., Pei, D., Wang, J., … Chen, J. (2021). The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature*, *591*(7849), Article 7849. https://doi.org/10.1038/s41586-021-03313-9

Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., & Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Research*, *45*(10), 6051–6063. https://doi.org/10.1093/nar/gkx141

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Lyon, A. S., Peeples, W. B., & Rosen, M. K. (2021). A framework for understanding functions of biomolecular condensates on molecular to cellular scales. *Nature Reviews. Molecular Cell Biology*, *22*(3), 215–235. https://doi.org/10.1038/s41580-020-00303-z

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., & Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, *487*(7405), 57–63. https://doi.org/10.1038/nature11244

Mager, D. L., & Stoye, J. P. (2015). Mammalian Endogenous Retroviruses. *Microbiology Spectrum*, *3*(1), MDNA3-0009–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0009-2014

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., & Apweiler, R. (2005). PRIDE: The proteomics identifications database. *Proteomics*, *5*(13), 3537–3545. https://doi.org/10.1002/pmic.200401303

Martin, G. R. (1981). Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(12), 7634–7638.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., & Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, *464*(7290), Article 7290. https://doi.org/10.1038/nature08858

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The

Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, *11*(1), 165. https://doi.org/10.1186/1471-2105-11-165

McStay, B., & Grummt, I. (2008). The epigenetics of rRNA genes: From molecular to chromosome biology. *Annual Review of Cell and Developmental Biology*, *24*, 131–157. https://doi.org/10.1146/annurev.cellbio.24.110707.175259

McSwiggen, D. T., Mir, M., Darzacq, X., & Tjian, R. (2019). Evaluating phase separation in live cells: Diagnosis, caveats, and functional consequences. *Genes & Development*, *33*(23–24), 1619–1634. https://doi.org/10.1101/gad.331520.119

Mensah, M. A., Niskanen, H., Magalhaes, A. P., Basu, S., Kircher, M., Sczakiel, H. L., Reiter, A. M. V., Elsner, J., Meinecke, P., Biskup, S., Chung, B. H. Y., Dombrowsky, G., Eckmann-Scholz, C., Hitz, M. P., Hoischen, A., Holterhus, P.-M., Hülsemann, W., Kahrizi, K., Kalscheuer, V. M., … Hnisz, D. (2023). Aberrant phase separation and nucleolar dysfunction in rare genetic diseases. *Nature*, *614*(7948), Article 7948. https://doi.org/10.1038/s41586-022-05682-1

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., & Knowles, B. B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. *Science*, *335*(6075), 1499–1502. https://doi.org/10.1126/science.1216154

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B., & Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Research*, *52*(3), 643–645.

Mitrea, D. M., & Kriwacki, R. W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, *14*(1), 1. https://doi.org/10.1186/s12964-015-0125-7

Mitrea, D. M., Mittasch, M., Gomes, B. F., Klein, I. A., & Murcko, M. A. (2022). Modulating biomolecular condensates: A novel approach to drug discovery. *Nature Reviews Drug Discovery*, *21*(11), 841–862. https://doi.org/10.1038/s41573-022-00505-4

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., & Taylor, J. P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*, *163*(1), 123–133. https://doi.org/10.1016/j.cell.2015.09.015

Morgan, H. D., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*, *23*(3), 314–318. https://doi.org/10.1038/15490

Musacchio, A. (2022). On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal*, *41*(5), e109952. https://doi.org/10.15252/embj.2021109952

Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A. L., Erb, M. A., Lawlor, M. A., Souza, A., Scott, T. G., Vittori, S., Perry, J. A., Qi, J., Winter, G. E., Wong, K.-K., Gray, N. S., & Bradner, J. E. (2018). The dTAG system for immediate and target-specific protein degradation. *Nature Chemical Biology*, *14*(5), 431–441. https://doi.org/10.1038/s41589-018-0021-8

Neo, S. H., Itahana, Y., Alagu, J., Kitagawa, M., Guo, A. K., Lee, S. H., Tang, K., & Itahana, K. (2015). TRIM28 Is an E3 Ligase for ARF-Mediated NPM1/B23 SUMOylation

That Represses Centrosome Amplification. *Molecular and Cellular Biology*, *35*(16), 2851–2863. https://doi.org/10.1128/MCB.01064-14

Neumann, T., Herzog, V. A., Muhar, M., von Haeseler, A., Zuber, J., Ameres, S. L., & Rescheneder, P. (2019). Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics*, *20*(1), 258. https://doi.org/10.1186/s12859-019-2849-7

Ng, K. W., Boumelha, J., Enfield, K. S. S., Almagro, J., Cha, H., Pich, O., Karasaki, T., Moore, D. A., Salgado, R., Sivakumar, M., Young, G., Molina-Arcas, M., de Carné Trécesson, S., Anastasiou, P., Fendler, A., Au, L., Shepherd, S. T. C., Martínez-Ruiz, C., Puttick, C., … Kassiotis, G. (2023). Antibodies against endogenous retroviruses promote lung cancer immunotherapy. *Nature*, *616*(7957), Article 7957. https://doi.org/10.1038/s41586-023-05771-9

Nickerson, J. A., Krochmalnic, G., Wan, K. M., & Penman, S. (1989). Chromatin architecture and nuclear RNA. *Proceedings of the National Academy of Sciences*, *86*(1), 177–181. https://doi.org/10.1073/pnas.86.1.177

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457–463. https://doi.org/10.1038/nature08909

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T., Forman-Kay, J. D., & Baldwin, A. J. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular Cell*, *57*(5), 936–947. https://doi.org/10.1016/j.molcel.2015.01.013

Novo, C. L., Wong, E. V., Hockings, C., Poudel, C., Sheekey, E., Wiese, M., Okkenhaug, H., Boulton, S. J., Basu, S., Walker, S., Kaminski Schierle, G. S., Narlikar, G. J., & Rugg-Gunn, P. J. (2022). Satellite repeat transcripts modulate heterochromatin

condensates and safeguard chromosome stability in mouse embryonic stem cells. *Nature Communications*, *13*(1), 3525. https://doi.org/10.1038/s41467-022-31198-3

Orlando, D. A., Chen, M. W., Brown, V. E., Solanki, S., Choi, Y. J., Olson, E. R., Fritz, C. C., Bradner, J. E., & Guenther, M. G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Reports*, *9*(3), 1163–1170. https://doi.org/10.1016/j.celrep.2014.10.018

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: Longitudinal chromosome differentiation was recognized fifty years ago. *American Journal of Human Genetics*, *31*(2), 106.

Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., & Alberti, S. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, *162*(5), 1066–1077. https://doi.org/10.1016/j.cell.2015.07.047

Patel, A., Mitrea, D., Namasivayam, V., Murcko, M. A., Wagner, M., & Klein, I. A. (2022). Principles and functions of condensate modifying drugs. *Frontiers in Molecular Biosciences*, *9*. https://www.frontiersin.org/articles/10.3389/fmolb.2022.1007744

Pederson, T. (2011). The Nucleolus. *Cold Spring Harbor Perspectives in Biology*, *3*(3), a000638. https://doi.org/10.1101/cshperspect.a000638

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), Article 3. https://doi.org/10.1038/nbt.3122

Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A., & Panning, B. (2002). Xist RNA and the Mechanism of X Chromosome Inactivation. *Annual Review of Genetics*, *36*(1), 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433

Poudyal, R. R., Sieg, J. P., Portz, B., Keating, C. D., & Bevilacqua, P. C. (2021). RNA sequence and structure control assembly and function of RNA condensates. *RNA*, *27*(12), 1589–1601. https://doi.org/10.1261/rna.078875.121

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Quinodoz, S. A., & Guttman, M. (2022). Essential Roles for RNA in Shaping Nuclear Organization. *Cold Spring Harbor Perspectives in Biology*, *14*(5), a039719. https://doi.org/10.1101/cshperspect.a039719

Quinodoz, S. A., Jachowicz, J. W., Bhat, P., Ollikainen, N., Banerjee, A. K., Goronzy, I. N., Blanco, M. R., Chovanec, P., Chow, A., Markaki, Y., Thai, J., Plath, K., & Guttman, M. (2021). RNA promotes the formation of spatial compartments in the nucleus. *Cell*, *184*(23), 5775-5790.e30. https://doi.org/10.1016/j.cell.2021.10.014

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*, *174*(3), 744-757.e24. https://doi.org/10.1016/j.cell.2018.05.024

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(Web Server issue), W187-191. https://doi.org/10.1093/nar/gku365

Reichholf, B., Herzog, V. A., Fasching, N., Manzenreither, R. A., Sowemimo, I., & Ameres, S. L. (2019). Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis. *Molecular Cell*, *75*(4), 756-768.e7. https://doi.org/10.1016/j.molcel.2019.06.018

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., Patil, D. P., Kwak, H., Lee, J. H., & Jaffrey, S. R. (2019). M6A enhances the phase separation potential of mRNA. *Nature*, *571*(7765), 424–428. https://doi.org/10.1038/s41586-019-1374-1

Rinn, J., & Guttman, M. (2014). RNA and dynamic nuclear organization. *Science*, *345*(6202), 1240–1241. https://doi.org/10.1126/science.1252966

Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, *81*(1), 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Rippe, K. (2022). Liquid–Liquid Phase Separation in Chromatin. *Cold Spring Harbor Perspectives in Biology*, *14*(2), a040683. https://doi.org/10.1101/cshperspect.a040683

Roden, C., & Gladfelter, A. S. (2021). RNA contributions to the form and function of biomolecular condensates. *Nature Reviews Molecular Cell Biology*, *22*(3), Article 3. https://doi.org/10.1038/s41580-020-0264-6

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, *463*(7278), Article 7278. https://doi.org/10.1038/nature08674

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., … Young, R.

A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)*, *361*(6400), eaar3958. https://doi.org/10.1126/science.aar3958

Sabari, B. R., Dall'Agnese, A., & Young, R. A. (2020). Biomolecular Condensates in the Nucleus. *Trends in Biochemical Sciences*, *45*(11), 961–977. https://doi.org/10.1016/j.tibs.2020.06.007

Saksouk, N., Simboeck, E., & Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin*, *8*(1), 3. https://doi.org/10.1186/1756-8935-8-3

Sawyer, I. A., Hager, G. L., & Dundr, M. (2017). Specific genomic cues regulate Cajal body assembly. *RNA Biology*, *14*(6), 791–803. https://doi.org/10.1080/15476286.2016.1243648

Sawyer, I. A., Sturgill, D., & Dundr, M. (2019). Membraneless nuclear organelles and the search for phases within phases. *WIREs RNA*, *10*(2), e1514. https://doi.org/10.1002/wrna.1514

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., & Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, *352*(6290), 1225–1228. https://doi.org/10.1126/science.aad9841

Sharp, P. A., Chakraborty, A. K., Henninger, J. E., & Young, R. A. (2022). RNA in formation and regulation of transcriptional condensates. *RNA*, *28*(1), 52–57. https://doi.org/10.1261/rna.078997.121

Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science (New York, N.Y.)*, *357*(6357), eaaf4382. https://doi.org/10.1126/science.aaf4382

Singh, P. B., & Newman, A. G. (2020). On the relations of phase separation and Hi-C maps to epigenetics. *Royal Society Open Science*, *7*(3), 191976. https://doi.org/10.1098/rsos.191976

Stocking, C., & Kozak, C. A. (2008). Murine endogenous retroviruses. *Cellular and Molecular Life Sciences : CMLS*, *65*(21), 3383–3398. https://doi.org/10.1007/s00018-008-8497-0

Stoll, G. A., Oda, S., Chong, Z.-S., Yu, M., McLaughlin, S. H., & Modis, Y. (2019). Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(30), 15042–15051. https://doi.org/10.1073/pnas.1901318116

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, *547*(7662), Article 7662. https://doi.org/10.1038/nature22989

Su, C., Li, H., & Gao, W. (2018). TRIM28 is overexpressed in glioma and associated with tumor progression. *OncoTargets and Therapy*, *11*, 6447–6458. https://doi.org/10.2147/OTT.S168630

Su, X., Ditlev, J. A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D. S., Taunton, J., Rosen, M. K., & Vale, R. D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*, *352*(6285), 595–599. https://doi.org/10.1126/science.aad9964

Sun, Y., Keown, J. R., Black, M. M., Raclot, C., Demarais, N., Trono, D., Turelli, P., & Goldstone, D. C. (2019). A Dissection of Oligomerization by the TRIM28 Tripartite Motif and the Interaction with Members of the Krab-ZFP Family. *Journal of Molecular Biology*, *431*(14), 2511–2527. https://doi.org/10.1016/j.jmb.2019.05.002

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P., & Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, *24*(12), 1963–1976. https://doi.org/10.1101/gr.168872.113

Teissandier, A., Servant, N., Barillot, E., & Bourc'his, D. (2019). Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mobile DNA*, *10*(1), Article 1. https://doi.org/10.1186/s13100-019-0192-1

Thompson, P. J., Macfarlan, T. S., & Lorincz, M. C. (2016). Long terminal repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire. *Molecular Cell*, *62*(5), 766–776. https://doi.org/10.1016/j.molcel.2016.03.029

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Vera, M., Tutucci, E., & Singer, R. H. (2019). Imaging Single mRNA Molecules in Mammalian Cells Using an Optimized MS2-MCP System. *Methods in Molecular Biology (Clifton, N.J.)*, *2038*, 3–20. https://doi.org/10.1007/978-1-4939-9674-2_1

Wang, B., Zhang, L., Dai, T., Qin, Z., Lu, H., Zhang, L., & Zhou, F. (2021). Liquid–liquid phase separation in human health and diseases. *Signal Transduction and Targeted Therapy*, *6*(1), 290. https://doi.org/10.1038/s41392-021-00678-1

Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., & Hyman, A. A. (2018). A molecular grammar governing the driving forces for phase

separation of prion-like RNA binding proteins. *Cell*, *174*(3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006

Wang, J., Wang, L., Diao, J., Shi, Y. G., Shi, Y., Ma, H., & Shen, H. (2020). Binding to m6A RNA promotes YTHDF2-mediated phase separation. *Protein & Cell*, *11*(4), 304–307. https://doi.org/10.1007/s13238-019-00660-2

Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., Zhang, G., Wei, Y., Qu, H., Li, Y., Allis, C. D., Li, G., Li, H., & Li, P. (2019). Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Molecular Cell*, *76*(4), 646-659.e6. https://doi.org/10.1016/j.molcel.2019.08.019

Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S. M., Brachmann, R. K., & Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences*, *104*(47), 18613–18618. https://doi.org/10.1073/pnas.0703637104

Wang, Y.-Y., Li, L., Zhao, Z.-S., & Wang, H.-J. (2013). Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. *World Journal of Surgical Oncology*, *11*(1), 1–8.

Watkins, N. J., & Bohnsack, M. T. (2012). The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. *WIREs RNA*, *3*(3), 397–414. https://doi.org/10.1002/wrna.117

Weber, S. C., & Brangwynne, C. P. (2015). Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Current Biology: CB*, *25*(5), 641–646. https://doi.org/10.1016/j.cub.2015.01.012

Wei, C., Cheng, J., Zhou, B., Zhu, L., Khan, M. A., He, T., Zhou, S., He, J., Lu, X., Chen, H., Zhang, D., Zhao, Y., & Fu, J. (2016). Tripartite motif containing 28 (TRIM28)

promotes breast cancer metastasis by stabilizing TWIST1 protein. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep29822

Wernig, M., Lengner, C. J., Hanna, J., Lodato, M. A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., & Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnology*, *26*(8), 916–924. https://doi.org/10.1038/nbt1483

Wheeler, R. J. (2020). Therapeutics—How to treat phase separation-associated diseases. *Emerging Topics in Life Sciences*, *4*(3), 331–342. https://doi.org/10.1042/ETLS20190176

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Wolf, G., Yang, P., Füchtbauer, A. C., Füchtbauer, E.-M., Silva, A. M., Park, C., Wu, W., Nielsen, A. L., Pedersen, F. S., & Macfarlan, T. S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes & Development*, *29*(5), 538–554.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., Wang, X., Ma, H.-L., Huang, C.-M., Yang, Y., Huang, N., Jiang, G.-B., Wang, H.-L., Zhou, Q., Wang, X.-J., … Yang, Y.-G. (2016). Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell*, *61*(4), 507–519. https://doi.org/10.1016/j.molcel.2016.01.012

Xu, W., Li, J., He, C., Wen, J., Ma, H., Rong, B., Diao, J., Wang, L., Wang, J., Wu, F., Tan, L., Shi, Y. G., Shi, Y., & Shen, H. (2021). METTL3 regulates heterochromatin in

mouse embryonic stem cells. *Nature*, *591*(7849), Article 7849. https://doi.org/10.1038/s41586-021-03210-1

Yang, W., Xu, Y., Wu, J., Zeng, W., & Shi, Y. (2003). Solution structure and DNA binding property of the fifth HMG box domain in comparison with the first HMG box domain in human upstream binding factor. *Biochemistry*, *42*(7), 1930–1938. https://doi.org/10.1021/bi026372x

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., Pilka, E. S., Aspris, D., Leggate, D., Hendrick, A. G., Webster, N. A., Andrews, B., Fosbeary, R., Guest, P., Irigoyen, N., Eleftheriou, M., Gozdecka, M., Dias, J. M. L., Bannister, A. J., … Kouzarides, T. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, *593*(7860), 597–601. https://doi.org/10.1038/s41586-021-03536-w

Yu, C., Zhan, L., Jiang, J., Pan, Y., Zhang, H., Li, X., Pen, F., Wang, M., Qin, R., & Sun, C. (2014). KAP-1 is overexpressed and correlates with increased metastatic ability and tumorigenicity in pancreatic cancer. *Medical Oncology*, *31*, 1–9.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137