



**HAL**  
open science

# Novel chromatin pathway controlling transposable elements and impact on genome stability

Emeline Roger

► **To cite this version:**

Emeline Roger. Novel chromatin pathway controlling transposable elements and impact on genome stability. Cellular Biology. Université Paris sciences et lettres, 2023. English. NNT : 2023UPSLS023 . tel-04315843

**HAL Id: tel-04315843**

**<https://theses.hal.science/tel-04315843>**

Submitted on 30 Nov 2023

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

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



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Institut Curie, Unité de Génétique et Biologie du Développement

**Novel chromatin pathway controlling transposable elements and impact on genome stability**

**Nouvelle voie de contrôle des transposons et implication pour la stabilité des génomes**

Soutenue par

**Emeline ROGER**

Le 7 juillet 2023

Ecole doctorale n° 515

**Complexité du Vivant**

Spécialité

**Biologie Cellulaire et Développement**

**Composition du jury :**

Dr Antonin, MORILLON Institut Curie	<i>Président</i>
Dr Julie, COCQUET Institut Cochin	<i>Rapportrice</i>
Dr Helen, ROWE Queen Mary University of London	<i>Rapportrice</i>
Dr Slimane, AÏT-SI-ALI Université Paris-Cité	<i>Examineur</i>
Pr Rob, KLOSE University of Oxford	<i>Examineur</i>
Dr Deborah, BOURC'HIS Institut Curie	<i>Directrice de thèse</i>
Dr Gaël, CRISTOFARI IRCAN	<i>Co-directeur de thèse</i>

# **ACKNOWLEDGMENTS**



This is the first part of this manuscript I wanted to write and of course, the last one I am actually writing. This is just to say that I have given it some thoughts, because acknowledging how anyone participated in this work really matters to me. To give it a more personal touch and give credit to something that accompanied me all along this adventure, I will associate key songs and artists to each paragraph.

### *Gretel – Snarky Puppy*

A masterpiece.

First, I want to thank the people whose work was necessary for this manuscript to be achieved, starting with the ones who will actually read it, namely, the members of the jury: Dr Helen Rowe and Dr Julie Cocquet (Referees); Pr Rob Klose, Dr Slimane Aït-Si-Ali and Dr Antonin Morillon (Examiners). I feel very grateful for the attention you gave to my work and I hope you enjoyed it as much as I enjoyed writing it. I deeply value all inputs you might give on it.

### *Get Lucky – Daft Punk*

Simple, neat, efficient and to the point, with party vibes.

Close to them lies someone who has already read and commented everything; who taught me tons of things from what details to focus on that actually matter, who broadened my views on the science we are doing and on science in general, including how to accurately employ “which” and “that” (although I might still be doing it wrong), my PhD supervisor, Deborah. Besides the knowledge gained during this PhD, I also feel very grateful for benefiting from a work environment and mentoring that allowed me to pursue my ideas without going too far deep in non-sense. I am still impressed by Deborah’s achievements and the trust she had in me when offering this project. I definitely wouldn’t have bet on me at that time. I am convinced that Deborah’s trust in each member of the lab is part of what makes it so safe and enjoyable, a good place to grow. The past 4 years in the lab were really a turning point for me, a period I truly enjoyed (despite a pandemic) and for all of that, I thank you Deborah.

### *“The Office” theme*

Just because I watched it all. Twice. And I loved it.

Lucky enough to have two PhD supervisors, I couldn’t have hoped for someone more easy-going than Gaël. Here, I want to thank you for making me part of the team as far

as I could be, for your advices along the way and for allowing me to meet the brilliant minds you gathered in your team. Who could have guessed the proportion of people coming from Verneuil-sur-Avre in this lab..!

### *Spin – Les Zôtres Fanfare*

A dynamic “You Spin Me Round” cover, definitely a apropos.

The work I present here include the time, ideas, and attempts of a great soon-to-be PhD student, Alexis. I really enjoyed being able to discuss all the technicalities of the project, trying to find solutions to the least expected problems. With these little lines, I want to thank you for all the energy and ideas you put in our work, no matter what. These are qualities I wish you to nurture, alongside your deep curiosity. I am going pass to you the same piece of wisdom I received to enjoy at best your time as a PhD student: “Read and have fun.”

### *La Belle-Hélène, Acte 3 No 21b, Tyrolienne – Jacques Offenbach*

For all the fun it brought me, and the subversion.

The “Read and have fun” motto was given to me by Tomek, previous Post-doc and (lucky) supervisor of my past-self, to whom I owe most of my technical skills in the lab, to say the least. In fact, this project not only relies on Tomek’s skills as a teacher but also as a scientist, as it is in direct continuation of his work. I must say, I have only fully understood the importance of that motto during the two last years of my PhD but I will definitely keep it in mind wherever my next adventure brings me. Thank you for all the discussions while culturing cells, for infusing ideas that took my years to fully develop, for pushing me just the right amount and for your trust.

### *Never gonna give you up – Scary Pockets*

Because that is what she will do for you.

Every project launched in this lab is based on the belief (and right belief!) that Aurélie, amazing bioinformatician, will be able to handle the generated data. One of the amazing parts of working with Aurélie (despite the fact that she can, indeed, handle anything) is that she truly understands your biological questions and will even ask some of her own, while following several projects at a time. Thinking/discussing with Aurélie was enlightening and enjoyable, including how far from science-based the

discussion might have been. Hence, for your presence in this lab, your work and easy-going personality, thank you, Aurélie.

### *YMCA – Village People*

In souvenir of all the songs that would not leave your minds.

Even though their work in this project is not fully disclosed in the following pages, I owe a big thank to Julian and Mélanie, for all the weighing they shared with me, for all the immunofluorescence attempts and microscopy hours, their understanding and involvement.

### *Spain – Chick Corea, Bobby McFerrin*

Simple to describe, but so essential and appreciated.

Fatima's work also deserves to be acknowledged, for the mice generation, her constant interest in my work and her kindness. Nicolas' understanding, efficiency, and flexibility in the handling of all the mice lines the Bourc'his lab possesses also deserves to be mentioned.

### *The Year After – French79*

Because it accompanied me all along.

Here, I also wish to thank the Margueron team and especially Samuel for all the protocols (and reagents...) sharing. I hope to be able to give it back to you in some ways later on.

### *Where I end and you begin – Radiohead*

One of my foundation.

While their work is not directly represented here, they definitely made it happen in many ways: this big thank is dedicated to the Bourc'his lab. Nearly all the lab members changed over my time there, each combination of people giving a different feeling to the lab, also always enjoyable. It started with the end of a post-doc era when the lab was furnished with knowledgeable (Max and Joan), in the sweetest way possible (Raquel) and unconventional people (Tomek). I must be honest and say that I was way too impressed by this crowd. Going through this ever evolving lab journey with me were amazing PhD student fellows: Mathilde, Mathieu, Lorraine. All of us different, brought closer along the way by sharing all the important steps of our PhD adventure, and very

basic day-to-day struggles. A collective thanks to the three of you for the inspiration you gave me, for the strength you have shown to pursue the path of your choice. More specifically, I want to thank Mathilde for being so on point and caring; Mathieu for taking the time whenever I had a question and bringing his energy and beliefs in the lab; Lorraine for all her many (and impressive) talents and sharing abilities, which created a space where you knew you were not alone. I wish all of you to always keep in mind how amazing you are, even in the hardest moments you might go through.

### *I want it that way – Backstreet boys*

Good reminder of this time period.

Among the people who came and left and I felt lucky to have met are Charlotte P, Charlotte D, Anissa and Germaine. Thanks for being part of this experience, and wishing you all the best in your adventures. Hopefully, I will soon enjoy Charlotte D.'s company and together, we will raise our glass and leg up.

### *La Gaillarde – Jocelyne Lafaille*

Do I really have to explain this one?

Nowadays, I share my lab days with a different crowd, a crowd I am really proud to be part of. Thanks to them, for the first time in my life, I experience the feeling to belong to a group, and to be a legitimate part of it. And that is really huge to me. Let's give each of them the appropriate thanks they need.

Elena, junior PhD student, who is simply the most joyful and sharing person I know; thank you for bringing this energy in the lab, for always wanting to "level up" and obviously put your leg up. I wish you to enjoy this PhD journey ahead of you like you enjoy a beautiful summer morning. Elena forms with Julian what has been collectively called the "beauf" corner, a corner that gives much laugh to the lab and needs to be acknowledged. Like Mélanie's little dance moves, they are part of makes this lab very comfy, a place to do your work while enjoying yourself. Like Lolita's appearances that we all appreciate, and thank for being so easy to discuss with. And what would be this team without the constant struggle of defining lunchtime according to Nico's estimation? Besides that, Nico's love for random mammals from Pallas' cat to raccoon gives a hint about how a sweet colleague he is. I also want to acknowledge how I enjoyed discussing our projects together or planning "the best" pranks. Again, what would it be without food-based discussion and Urvashi's love/hate relationship with



French cuisine? Apart from this, I thank you for being such an easy-going person and all the credit you gave me. Keep all your strength for the PhD ahead! The newest comers, Manuel, Marie, Aleksandra and Markus, fit well in this mix of working well in deep extravaganza and weirdness. For being my partners in the early mornings and late evenings in the lab, for trusting me with your questions, ideas and “artistic” abilities; yes, you guessed it, thank you guys. The last year of PhD has been really enjoyable thanks to this heart-warming environment I was happy to come to every day; which deserves a last big shout out to the Bourc’his lab.

### *Celebration – Kool & The Gang*

Because this unit deserves a celebration.

In these few lines, I also want to mention that this lab is implanted in a really great research unit, the BDD, filled with scientist willing to help and people willing to know each other and have a good time. Thanks to all the unit for allowing such wills to exist and be fulfilled.

### *Nightclub Amnesia – Ratatat*

Different life styles, same kindness.

Apart from my life lab, what made this time period of PhD a complete adventure is what I call my support system, filled with people I just want to thank for being themselves and sweet enough to share some of their amazingness with me. In this support system figures my flat mate, Vickie, who I definitely have to thank for her understanding during intense times of PhD writing, her delicious food and ways to making home feel like home.

### *Sonne – Rammstein*

For “die Sonne scheint” in Antoine’s eyes.

### *La purée – Salut c’est cool !*

For they made me discover this under-estimated group.

I notice now that I am surrounded by amazing cooks, like Antoine who I want to thank for sharing the experience of being new Parisian PhD students coming from nowhere, Normandy, around every possible meal. Antoine belongs to a group of people that is very dear to me, who fueled several, completely unexpected and yet memorable experiences: Clarence (again, amazing cook!), Louis (your transport expert in all

situations), David (this guy can do anything. For all your weirdest request, I would definitely advise you to go to him) and Bastien (full of surprises!). Thank you for being part of my life and to partner with me whatever plans we elaborate.

### *The Clock – Thom York*

Time passes but people stay.

In my support system, you will also find stand-alone people that I am very happy to count in, despite time and distance that separate us: Mailys, Man and Karl. I want to thank Mailys for sharing with me the journey of studying biology since the bachelor and for being a friend who always gives me perspective and easy discussions. Man, for taking the time, despite a very busy and intense schedule, to share discussions, games and whatever else during his Parisian visits. And finally, Karl, for just still being part of my life after all this time.

### *Breath – Pink Floyd*

*“It’s good to warm my bones besides the fire”*

### *Chic planète – L’affaire Louis’Trio*

Legendary.

My family is also part of this system for indeed having supported me since my early times of the journey that closes with the writing of this manuscript. I want here to thank my parents for always getting on board with my various life decisions no matter how they could understand them, for providing me the comfort of a home (not just a house, a home), and being always on the exponential part of their learning curves. They also provided me with siblings who made sure I would become as understanding and flexible as I could be. To Gabriel for his filmic and music culture and abilities to make anything funny. A Ethan, pour être cette personne que j’admire et dont je me sens si proche, merci à toi pour me laisser être moi-même. To Camille, for being the best sister and friend I could hope for. Je voudrais aussi remercier Alain et Anne-Marie, pour être là, toujours, en toutes circonstances.

### *Tout le monde veut devenir un cat – José Germain*

Whatever happens, you can dance it away.

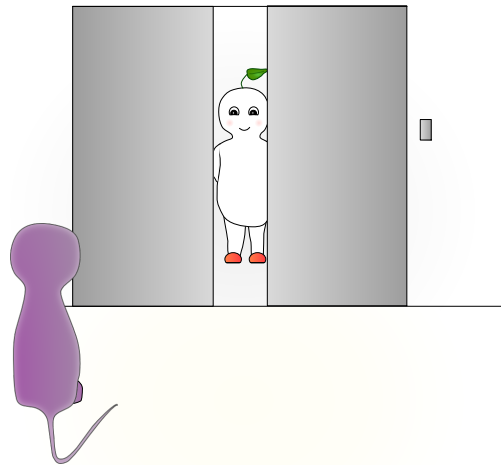
This journey would have been completely different if I hadn’t been pushed and accompanied by Maxence, the first one to believe, “Yes, of course, you can do a PhD”

or even “Yes, of course, you can study science”. For learning my PhD title, for all the way done together, for how much we gained and grew from knowing each other, and for the newest way to continue sharing our lives, a really deep thank you, Maxence.

### *Springtij – Gallowstreet*

Sweet and strong, featuring the unexpected.

Finally, most of my PhD times benefited from the constant support, curiosity and care from Tom. I feel lucky, every day, to have met someone so genuine and free, to share feelings, life moments and sets of brand-new experiences with. Thank you for bringing me back to myself, and creating a safe space for my inner child to take all the place it deserves. As a very special thank, I give you this drawing of two inspirational characters. It’s not much, but I know you will enjoy it.



*To Colette, an endless source of inspiration*



# **ABSTRACTS**



## Abstract

Transposable elements (TEs) are mobile genetic elements present in every living species. To limit their harmful effects, species have developed over evolutionary time-scales several mechanisms to control TE expression and mobilization, which is of utmost importance to protect genomic integrity in germ cells and pluripotent cells. By suppressing their initial transcription, DNA methylation is a potent TE repressor. However, during the epigenetic reprogramming that accompanies pre-implantation development, embryonic cells lose most of their genomic methylation, except at a few regions including some TEs. These observations raise two overarching questions: How are TEs repressed in absence of DNA methylation at this particular time? How some TEs resist the global DNA methylation erasure?

My PhD work focused on the study of SPINDLIN1 (SPIN1) and SPINDOC—a transcriptional activator and its specific co-factor—as a new chromatin reading complex involved in TE control in embryonic cells. Using cultures of embryonic stem cells as a modeling system and CRISPR-Cas9-based genome editing tools, I generated knock-out and rescue cell lines for either or both of these factors and analyzed their transcriptome (including TE expression), chromatin and DNA methylation profile. I found both factors to be required for TE control, in part through promoting the function of the DNA methylation maintenance machinery genome-wide. I also found SPINDOC to be involved in growth and fertility in mice. Altogether, my work reveals a previously unknown link between SPIN1-SPINDOC and transposon biology, and a new role for this complex: the regulation of DNA methylation.

## Résumé

Les éléments transposables (ET) sont des éléments génétiques mobiles présents chez toutes les espèces vivantes. Pour limiter leurs effets néfastes, les organismes hôtes ont développé différents mécanismes de contrôle des ET au cours de l'évolution, ce qui est particulièrement essentiel pour assurer l'intégrité génomique des cellules germinales et pluripotentes. En supprimant leur transcription initiale, la méthylation de l'ADN est un répresseur efficace des ET. Cependant, pendant la phase de reprogrammation épigénétique qui accompagne le développement, la méthylation de l'ADN est largement perdue dans les cellules embryonnaires, à l'exception de quelques régions qui incluent certains ET. Ces observations soulèvent deux questions fondamentales : Comment les ET sont-ils réprimés en absence de méthylation pendant cette période ? Et comment certains ET résistent à la perte de méthylation globale ?

Pendant ma thèse, j'ai étudié le rôle de SPINDLIN1 (SPIN1) et SPINDOC, un activateur transcriptionnel et son co-facteur, comme nouveau complexe de lecture chromatinienne impliqué dans le contrôle des ET dans l'embryon. En utilisant des cultures de cellules souches embryonnaires comme système modèle et des outils d'édition du génome basés sur la technologie CRISPR-Cas9, j'ai généré des lignées mutantes pour l'une ou l'ensemble de ces deux protéines et analysé leur transcriptome (y compris aux ET), leur profil chromatinien et leur méthylation de l'ADN. J'ai pu ainsi démontrer que SPIN1 et SPINDOC sont nécessaires au contrôle des ET, au moins en partie en assurant le bon fonctionnement de la machinerie de maintenance de la méthylation de l'ADN sur l'ensemble du génome. J'ai aussi mis en évidence l'implication de SPINDOC dans la croissance et la fertilité chez la souris. Dans l'ensemble, mes travaux mettent en évidence un lien précédemment inconnu entre SPIN1-SPINDOC et la biologie des transposons et un nouveau rôle pour ce complexe : la régulation de la méthylation de l'ADN.





# RÉSUMÉ EN FRANÇAIS



Les éléments génétiques mobiles sont une source d'innovation et d'adaptation des espèces. Parmi eux figurent les éléments transposables (ET), constituants quasi-majoritaires du génome des mammifères. Le potentiel d'innovation des ET étant stochastique, leur présence et mouvement peuvent être aussi bien responsables de l'émergence de nouvelles structures et fonctions favorisant la survie de l'hôte, comme le placenta, que de conséquences pathologiques néfastes à l'hôte. En effet, sans contrainte exercée par l'hôte, l'expression et le mouvement des ET s'accompagnent d'une forte instabilité génomique, associés au développement de cancers, de maladies neurodégénératives ou auto-immunes. A l'échelle de l'espèce, il est de la plus haute importance d'assurer l'intégrité génomique des cellules germinales qui transmettent leur patrimoine génétique à la génération suivante. Aussi, les mouvements non maîtrisés des ET dans ces cellules peuvent engendrer des infertilités ou être à l'origine de mutations transmissibles à la descendance. Chez les mammifères, les cellules germinales émergent à partir de cellules embryonnaires pluripotentes présentes au début du développement, dans lesquelles le contrôle des ET a donc une importance majeure.

En effet, pour diminuer l'impact nocif des ET et profiter de leurs effets bénéfiques, de nombreux mécanismes sont en place pour limiter l'expression des ET à des intensités et temporalités déterminées par l'hôte. Ainsi, de nombreux facteurs produits par l'hôte sont impliqués dans le contrôle transcriptionnel des ET et en particulier, des ET jeunes ayant encore la capacité de transposer, comme les éléments LINE-1 ou IAP chez la souris. Des voies majeures de régulation concourent à créer un environnement hostile à l'expression des ET par la formation d'hétérochromatine. Deux marqueurs de ce contrôle sont la méthylation de la lysine 9 de l'histone 3 (H3K9me, et en particulier H3K9me3), et de l'ADN. La méthylation de l'ADN en particulier participe aussi au contrôle transcriptionnel des gènes de l'hôte et ainsi, à la consolidation épigénétique des destins cellulaires. Aussi, cette marque est naturellement reprogrammée au cours du développement à deux reprises, lors des transitions entre états somatiques et germinaux. Malgré cette perte développementale de la méthylation de l'ADN, le contrôle des ET est maintenu dans les cellules souches embryonnaires murines (ES). Les mécanismes de compensation mis en place pour maintenir le contrôle des ET restent cependant mal connus.

Afin d'identifier de nouveaux facteurs impliqués dans le contrôle des ET dans les cellules ES murines, un crible génétique par perte de fonction utilisant l'outil d'édition du génome CRISPR-Cas9 a été réalisé au sein de mon laboratoire d'accueil, en se focalisant sur les ET de type IAP. Cette expérience a permis de révéler la possible implication de deux protéines associées en complexe : SPINDLIN1 (SPIN1) et SPINDOC. D'après la littérature, SPIN1 est une protéine de lecture de la chromatine associée à une activation de la transcription, ce que SPINDOC semble contre-carrer. Les travaux présentés dans ce manuscrit sont les premiers à (1) étudier le rôle de SPIN1 et SPINDOC dans les cellules ES et à les associer au contrôle des ET. De plus, SPINDOC étant majoritairement exprimé dans le testicule où le contrôle des ET est essentiel, (2) j'ai aussi généré et caractérisé un modèle murin mutant pour *Spindoc*.

### **Implication de SPIN1-SPINDOC dans le contrôle des ET en cellules ES**

Pour comprendre le rôle joué par SPIN1 et SPINDOC, j'ai généré plusieurs lignées de cellules ES où *Spin1* (*Spin1-KO*) et *Spindoc* (*Spindoc-KO*) ont été délétés individuellement ou en combinaison. Une analyse transcriptomique de ces lignées a permis de confirmer les résultats du crible génétique et d'asseoir le rôle de SPIN1 et SPINDOC en tant que nouveaux régulateurs des ET. Une cartographie chromatinienne des cellules *Spin1-KO* et *Spindoc-KO* montre que la réactivation transcriptionnelle des ET ne s'associe pas d'une perte de la marque H3K9me3, représentative de l'hétérochromatine, ou d'un gain en H3K4me3, liée à l'activité de transcription. En revanche, la perte de SPIN1 ou de SPINDOC s'accompagne d'une perte substantielle de la méthylation de l'ADN sur l'ensemble du génome des cellules ES, mesurée par des techniques utilisant une modification chimique (bisulfite suivi de pyrosequençage) ou enzymatique (EM-seq) de la méthylation de l'ADN. J'ai pu attribuer ce phénotype moléculaire à la déstabilisation d'UHRF1, un co-facteur essentiel de l'ADN méthyltransférase de maintenance, DNMT1. Cependant, la comparaison de différents milieux de culture de cellules ES suggère que le défaut de méthylation de maintenance ne participe qu'en partie à la réactivation des ET en absence de *Spin1* ou de *Spindoc*. Bien que les mécanismes exacts liant SPIN1 et SPINDOC à la régulation des ET restent à détailler, mes résultats démontrent l'importance de ces deux nouveaux facteurs et leur implication multi-facette dans la biologie des ET dans les cellules ES.

### **Fonction biologique de SPINDOC chez la souris**

SPIN1 était connue comme étant une protéine vitale, sa délétion entraînant une létalité immédiatement après la naissance chez la souris. En revanche, j'ai observé que SPINDOC n'est pas une protéine vitale. La délétion de *Spindoc* n'entraîne pas de phénotype dommageable, les souris *Spindoc*<sup>-/-</sup> sont viables mais présentent un phénotype de fertilité sub-optimale, chez les mâles et femelles. L'absence de SPINDOC entraîne également un défaut de croissance chez les mâles juvéniles. Cependant, malgré sa forte expression dans le testicule et son rôle en cellules ES, SPINDOC ne semble pas impliqué dans la répression des ET dans cet organe.

L'ensemble de ces travaux crée de nouveaux liens entre SPIN1, SPINDOC et des fonctions essentielles à la survie de l'espèce : régulation des ET, et croissance et fertilité. Ils viennent ainsi compléter une précédente base de connaissance qui liait SPIN1 et SPINDOC à des instabilités génomiques et au développement de cancers, apportant des conclusions plus fines quant à leur fonction biologique et leur possible utilisation en tant que cibles thérapeutiques.



# TABLE OF CONTENTS

<b>INTRODUCTION</b>	<b>25</b>
<b>I. Preface</b>	<b>27</b>
<b>II. Transposable elements</b>	<b>29</b>
A. General Overview	29
1. Transposable elements: a feature of all genomes	29
2. Transposable elements: a diversity of origin and sequences	33
3. Transposable elements: a presence with side effects	36
a) TE implication in physiological processes	37
b) TE implication in pathological processes	41
B. Focus on two TE families with mobilization potential	42
1. Intracisternal-A-Particles	43
2. Long Interspersed Nuclear Element 1	46
<b>III. Control mechanisms</b>	<b>51</b>
A. Non-targeted control mechanisms	51
1. Innate immune response	51
2. Degradation through stress granules and processing bodies	53
3. Other degradation and editing systems	55
B. Selective targeting of TEs for repression	56
1. Targeting DNA to control transcription, the case of KZFPs	56
2. Targeting RNA to control TE expression	61
a) The HUSH complex	61
b) piRNAs	63
C. Chromatin determinants of TE targeted control	68
1. DNA methylation	68
a) DNA methylation maintenance happens through collaboration between DNMT1 and UHRF1	70
b) Defining DNA methylation patterning with de novo DNMTs	71
2. H3K9me3	74
a) SETDB1 and SETDB2	74
b) SUV39H1 and SUV39H2	76
c) G9a/EHMT2 and GLP/EHMT1	77
3. Relative dependency over H3K9me3 or DNA methylation for TE repression	78
<b>IV. New factors involved in TE regulation</b>	<b>81</b>
A. Identification of TE regulators through loss-of-function screens	81
B. SPIN1	84
1. SPIN1 reads several chromatin modifications	85
2. SPIN1 is involved in distinct biological processes	88
a) SPIN1 is an essential protein	88
b) SPIN1 is involved in viral defense	88
c) SPIN1 is linked to cancer development	89
C. SPINDOC	90
1. SPINDOC interacts with SPIN1	91
2. SPINDOC interacts with PARP1	93
3. SPINDOC is highly expressed in the testis	93
<b>RESULTS</b>	<b>95</b>
<b>I. Involvement of SPIN1 and SPINDOC in TE control in mESCs</b>	<b>97</b>
ABSTRACT	100
INTRODUCTION	102
RESULTS	104
DISCUSSION	117
METHODS	127
REFERENCES	135
<b>II. The biological relevance of SPINDOC in the mouse</b>	<b>139</b>
RESULTS	140



1. SPIN1 and SPINDOC expression pattern.....	140
2. Generation of a <i>Spindoc</i> -KO mouse line .....	141
3. SPINDOC is necessary for physiological male growth .....	143
4. SPINDOC loss leads to a minor sub-fertility .....	145
METHODS .....	147
<b>DISCUSSION.....</b>	<b>151</b>
<b>I. General summary .....</b>	<b>153</b>
<b>II. The unexpected triangle: SPIN1-SPINDOC, DNA methylation and TE control in ESCs .....</b>	<b>155</b>
A. SPIN1-SPINDOC and UHRF1 stability.....	155
1. Hypotheses regarding the role of SPIN1 and SPINDOC in UHRF1 stability .....	155
2. Loss of UHRF1 and TE control .....	156
3. Long-term culture and passive loss of DNA methylation .....	157
B. SPIN1-SPINDOC role in TE control is independent of DNA methylation.....	159
1. SPIN1 binding to chromatin in ESCs .....	159
2. Hypotheses concerning the role of SPIN1 and SPINDOC in TE control .....	161
<b>III. Studies of TE control and TE expression analyses .....</b>	<b>163</b>
A. TE annotation.....	163
B. Unique mapping and control groups .....	164
<b>IV. SPIN1 and SPINDOC biological relevance.....</b>	<b>165</b>
A. SPIN1 as a therapeutic target in cancer development.....	165
B. Going further with SPINDOC role <i>in vivo</i> .....	166
C. Going further with SPIN1 role <i>in vivo</i> .....	168
<b>BIBLIOGRAPHY.....</b>	<b>171</b>



# INTRODUCTION



## I. Preface

Biology is defined as the study of living organisms, relying on the proposed by the cellular theory. Formulated during the XIX<sup>th</sup> century, this theory defines a cell as the simplest living unit separated by a membrane from its environment and with abilities to duplicate autonomously when given access to right nutrients. This definition excludes therefore viruses from living organisms for their lack of autonomous duplication. However, despite being non-living species and often considered as parasites, they are studied through the lens of similar principles used to study living species: proliferation, management of resources and energy (even though they do not provide their own) and evolution. Due to the dependence of viruses toward living organisms, their evolution is intertwined with the biology of their hosts. These considerations regarding viruses extend to all mobile DNA sequences. Despite a parasitic nature, mobile DNA sequences have been essential drivers of evolution for living species, from antibiotic resistance in bacteria to placentation in mammals. Transposable elements, integrated in genomes, constitute a functional part of living organisms that should not be overlooked anymore. While mobile elements do not live, they are definitely part of biology.

As a first dogma, it was proposed that the DNA molecule contained all the essential genetic information, transcribed in the form of RNA used as a template for translation in proteins, the functional effectors in cells. This linear view, similar to a one-way elevator, completely missed the importance of transcriptional regulation by proteins and RNAs, which is absolutely necessary to define which genes are expressed. Indeed, in pluricellular organisms, all cells contain the same DNA template while having very different functions and abilities defined by their distinct sets of expressed genes. Being able to switch between several transcriptional programs also allows adaptability to the environment for unicellular organisms. What drive decisions regarding a specific transcriptional program is a central question in developmental biology. How these decisions are conserved or lost with divisions or sexual reproduction drives the epigenetic field.

As transposable elements are part of genomes, decisions regarding their expression are made. Evolution has conserved species that use transposable element transcription in a time and cell-type specific manner and benefit from the large quantity of regulatory sequences they provide. Control of transposable elements is not only necessary to tame their mobility but also to ensure correct transcriptional programs decided by host cells.

During my PhD work, I aimed at elucidating a new transcriptional control pathway targeted to transposable elements in pluripotent embryonic cells, and further described other biological roles for these understudied proteins. During this first Introductory chapter, I will focus on defining transposable elements, their origins, diversity and importance for host biology in mammals. A specific attention will be given to young transposable elements that are still able to transpose, via mechanisms that will be described. Then I will present the major pathways and actors known for their involvement in transposable element control in mammals, and their interplay. Finally, I will introduce the two proteins that my work allowed to link to transposable element regulation, SPIN1 and SPINDOC.

## II. Transposable elements

### A. General Overview

#### 1. Transposable elements: a feature of all genomes

Mobile DNA sequences are present in almost every species. Two distinct forms of mobility can be found: extra-cellular mobility that allows for transmission of genetic information between cells, organisms or species from the same generation, and intra-cellular mobility where the DNA molecule from the same cell is modified. Both forms can be found in prokaryotes that can exchange genetic information via plasmids (extra-cellular mobility) and also contain transposable elements (TEs). In both cases, the mobility is not due to homologous recombination occurring between almost identical sequences, but to active mechanisms at least partially encoded by the elements. As such, plasmids and TEs are recognized as autonomous. However, they highly depend on their host cells for their transmission, using the ATP or nucleotides they produce, for example. This ambivalence between autonomy and dependency makes them perfect candidates to be considered as parasites in host cells, while not being considered as living entities.

The success of TEs as parasites can be seen by their persistence in living organisms and especially in pluricellular organisms, as opposed to plasmids only present in unicellular organisms. Of note, the few species in which no TE could be found are obligatory unicellular intrinsic parasites (*Plasmodium falciparum*, *Toxoplasma gondii*, *Encephalitozoon intestinalis* and *Theileria parva*) whose fitness might be too endangered by TE presence (Kissinger & DeBarry, 2011). Looking at genomes only as the information necessary for one organism to function, one would expect to find small genomes in simple organisms (unicellular or with low number of function or specialized cells), and larger genomes in more complex species. Observations were not in line with this hypothesis and gave birth to the enunciation of the C-paradox: DNA content (or C-value) is not in line with a species complexity. Moreover, the C-value is neither in line with the number of genes encoded. Large portions of genomes being free of genes, it was suggested that these DNA locations could be used by TEs. In fact, smaller genomes tend to have a low TE contribution due to the need of coding every essential information in a small DNA content and bigger

genomes can use more DNA content for innovation. That said, this factor alone is not enough to understand the variability of genome sizes in the tree of life and to explain the C-paradox. As an example, angiosperms are known to have a high TE contribution. Among them, *Zea Mays*, a species from which the idea that such things as TEs could exist, has a 2.3Gb genome, with an estimate of 85% of TE contribution and around 32,000 coding genes (Schnable et al., 2009), while only 58% of TEs contribute to the even larger *Hordeum Vulgare* genome of 5.1Gb (Oliver et al., 2013).

The birth of research in transposon biology as we know it nowadays comes from the association between two observations: a protein encoded by cells can complement mutant viruses (Weiss, 1969), and mutations happening at high frequencies in maize crop can lead to variable phenotypes that are not inherited from parents (McClintock, 1950). In fact, the first observation was explained by the integration of viral DNA in host genomes, which had been suggested but not proven or seen before. The second observation led to the hypothesis that some kind of transcriptional controlling element could have the ability to “jump” inside the genome. The stochasticity of viral integration and excision is in accordance with the initial idea of jumping genes. Connecting these two fields of research led to the discovery of long integrated viral particle DNA and how they could impact host gene expression. Although the field of transposon research emerged from the virology field, not all sequences considered as TEs today are originating from viruses, as will be discussed in Part I.A.2 from introduction.

Autonomous TEs encode for the core proteins essential for their movement: to get excised or copied from bigger DNA molecules, to move to their next insertion location and to get integrated there. One type of enzymes that can accomplish all these functions are transposases. Due to the multiple forms of transposases and TEs and to their abundance in genomes, transposases were qualified as “the most prevalent gene in nature” (Aziz et al., 2010). Transposases have the ability to cut DNA, which is needed by TEs at least for their new integration. In theory, a first strategy to move a DNA sequences is to cut it out as is and to insert it somewhere else. Another possibility to move for a DNA sequence is to be copied, and therefore use either a DNA-directed DNA polymerase that all organisms possess for replicating their DNA, or a DNA-directed RNA polymerase that again is quite frequent due to the need for transcription. However, an RNA-directed DNA polymerase (or reverse transcriptase) is not useful to

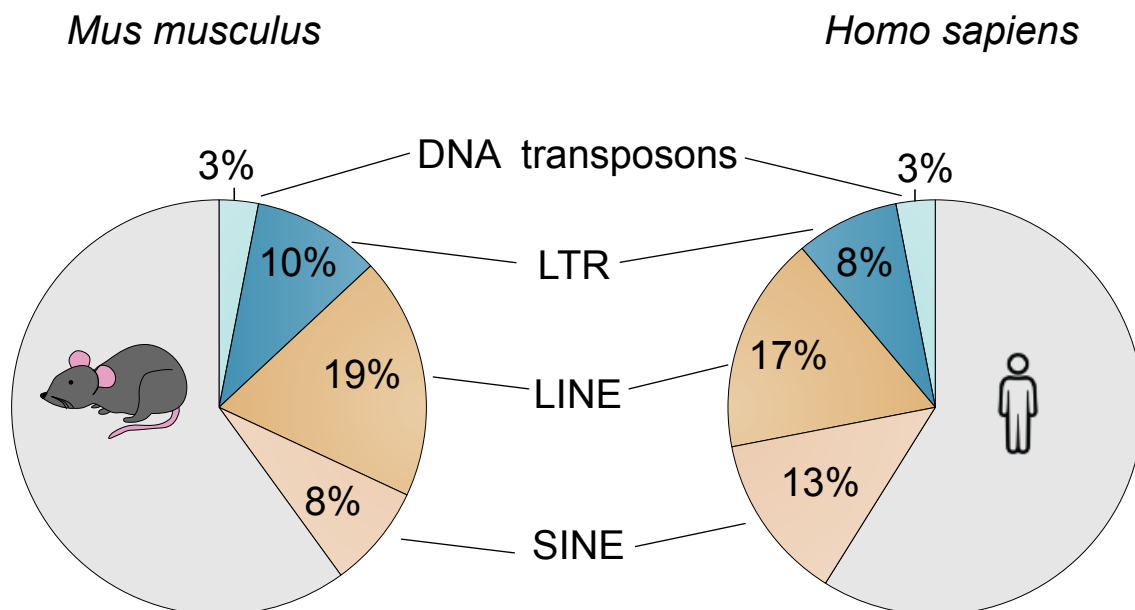


a large majority of organisms but is necessary for TE transposing with the latter strategy. Whether these kind of enzymes (transposase and reverse transcriptase) emerged from/for viruses or from/for TEs is a question in debate that will be difficult to answer for researchers who can only see the product of billion years of co-evolution. Some pieces of evidence, in the form of living species or viral elements, might have been lost along the way, leaving incomplete the picture we have today.

A separate and yet cumbersome issue considering the annotation and study of TEs in genomes is their multi-copy presence. At the moment, all available sequencing techniques are only reliable for a genome fragment of a certain size. Moreover, the quantity of DNA needed for these techniques supposes that the extracted DNA comes from a cell population. Short read sequencing technologies (Illumina) are able to accurately sequence fragments of hundreds base pairs (bp). However, full-length TEs can span several kb. Therefore, two identical copies of TEs might look just the same in regard to sequencing and could be considered as a duplicate of a TE sequence in a first locus while being present in a second one. Hence the number of TE copies present in genomes is in general under-estimated and global estimations of the number of TE copies in genomes have to be taken as approximations. However, emerging long read sequencing techniques (Oxford Nanopore) have a mean read length that can extend from 10 to 30 kilo bp (10-30kbp), allowing precise localization of TE copies and to correct previous estimations (M. Jain et al., 2018). Nevertheless, long read sequencing techniques are more prone to base calling error than short read technologies (<1% to <5% for SMRT-seq and nanopore respectively versus <0.1% for Illumina), a limitation under improvement (Amarasinghe et al., 2020). More and more accurate estimations of TE copy numbers are hoped to arise in the years to come. The ATLAS technique (amplification typing of L1 active subfamilies) based on targeted amplifications from sequences surrounding TEs has been used to be able to accurately locate them (Badge et al., 2003; Philippe et al., 2016), though these techniques can only allocate elements that did not accumulate mutations at the precise site of targeting, hence young TEs. As these copies share high homology, their copy number and location would have been the most under-estimated. Besides this targeted approach, new pipelines are proposed to accurately perform *de novo* assembly of genomes while taking repetitive elements into account and these constitute an emerging area of research. This approach is not restricted to TEs but to every kind of repeated

sequences, with the additional complexity for some of them that contain stretches of repeated nucleotides, which also represents a sequencing challenge.

The mutational drift endured by TE sequences throughout evolution also leads to an under-estimation of their copy number: with time the genetic drift is too wide to recognize them as TEs anymore. In addition to this technical or methodological challenges, as TE are mobile elements, they are highly polymorphic in populations. Some copies will be considered fixed because present in most of the population, while others will be only present and inherited in some individuals, and others could even be the product of a new transposition events in single individuals. Reference genomes were obtained from the sequencing of several individuals that might not be representative of the whole species population in terms of fixed or polymorphic copies, in addition to probably missing copies in these two categories. Therefore, proving new TE insertions requires a certain number of controls. Albeit estimating accurately the number of TE copies in genomes is a hard task, their diversity is quite well described.



**Figure 1 | TE content varies between species.**  
Genome proportion occupied by TEs according to their classes in mouse and human genomes

## 2. Transposable elements: a diversity of origin and sequences

As mentioned above, TEs use several ways to transpose: through excision (equivalent to a “cut-and-paste” mechanism) or through a DNA or RNA copy of an inserted element (equivalent to a “copy-and-paste” mechanism). Probably due to the inevitable increase of copy number linked to copy-and-paste mechanisms, elements using this kind of transposition tend to be more numerous and successful in invading genomes. Importantly, copy-and-paste can be used through a DNA or an RNA intermediate that depends on the type of proteins encoded by the element. Hence a first criterion used to classify TEs is their use or not of an RNA intermediate. The class I of TEs contains both TEs transposing through cut-and-paste or copy-and-paste as a DNA form, thus often referred to as DNA transposons. Class II groups all TEs using a copy-and-paste through an RNA form, and are termed RNA transposons or Retrotransposons for their obligatory use of an RNA-mediated DNA polymerase or reverse transcriptase. The genomic proportion of these two classes varies in great measure between distant related species. While most of the TE content is due to Class II elements in mammals, the majority of TEs in *C. elegans* belongs to Class I. Interestingly, most eukaryotic DNA transposons have close relatives in prokaryotes among the Insertion Sequences (IS) group (Siguier et al., 2014). However, they only represent 3 to 5% of mammalian genomes in size, against approximately 40% for retrotransposons (Figure 1). Moreover, they mostly are considered as transcriptionally inactive, with the exception of piggyBat elements in little brown bats (Mitra et al., 2013). Therefore, I will focus the rest of the manuscript on mammalian RNA transposons.

Retrotransposons are divided in two distinct classes depending on their origins or what they encode for, which is reflected by the presence or not of a repeated regulatory sequence at both ends of the element: the long terminal repeat (LTR) (Figure 2). LTR-elements probably emerged from past viral infections of the germ line DNA and transmission to every cell in the next generation. Strictly, LTR-elements are termed endogenous retroviruses (ERVs) when it can be proven that there is at least a copy possessing all the hallmark of a retrovirus, with a necessity to have a partly extra-cellular lifestyle. However, practically, most of LTR-elements are considered as ERVs, even when this final and absolute evidence is missing. In fact, the extra-cellular lifestyle relies on one of the proteins encoded by LTR-transposons, the envelope Env. Because

this protein is not needed for an exclusively intra-cellular cycle of transposition, this coding part tends to derive quite fast after the invasion of a new family of ERVs. Accumulation of mutations impede the task of finding the proper evidence of a past viral existence of a specific LTR-transposons family. As previously said, the LTR sequence is originally present at both sides of each element, which often leads to homologous recombination between the two sequences and complete excision of the coding sequence of the element, leaving one solo-LTR. Despite having lost their coding abilities, solo-LTR elements are still regulatory sequences with abilities to attract and activate the transcriptional machinery. Moreover, they can still transpose, although in a non-autonomous manner, though proteins produced by intact elements. The exact mechanism by which solo-LTRs hijack from other elements the factors needed for their mobilization is not completely understood.

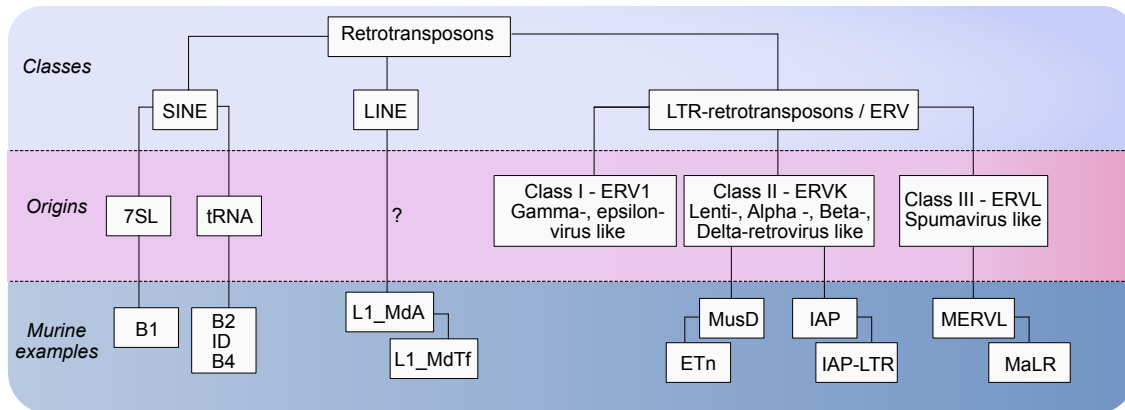
Non-LTR retrotransposons also regroup autonomous and non-autonomous elements: Long and short interspersed nuclear elements (LINEs and SINEs) ([Figure 2](#)). The origin of LINEs is quite discussed and they might even predate viruses with evidence of LINE presence in the earliest multi-cellular organism ([Malik et al., 1999](#)). In this scenario, LINEs could even be the ancestors of viruses. It is undeniable that every families of LINEs present in a genome come from the same ancestor element from that genome: the distinct families of LINEs emerged solely from mutational evolution, in contrast to LTR-families that originate from distinct types of viruses inserted at different times ([Boissinot & Furano, 2001](#)). Therefore, dating LINE elements and tracing back their evolution—although a hard task for aforementioned reasons—is considered easier than for LTR-elements. While several ERV and LINE-1 families are still competent for transposition in mice, only LINE elements are still capable of transposition in the human genome, and even more specifically from the human-specific L1 family (L1-HS). Specifics of LINE-1 sequences will be discussed in part I.B.2.

Due to their diversity of origins, ERV sequences might be completely different between one another and so their classification includes one additional level, by resemblance to modern exogenous retroviruses. A tRNA binding sequence that is present between the LTR and the coding sequence (a region known as primer binding site, PBS) also serves as a way to differentiate retroviruses lineages. The PBS is one

of the most conserved sequence in transposon families, as this tRNA binding is used as reverse-transcription starting point. The most ancient class of ERVs nowadays present in mice and humans are class III ERVs, for their resemblance to spumaviruses. They are also referred to as ERVL, as they use a Leucine tRNA to bind their PBS. Their presence in both species implicates that their invasion occurred before the primate-rodent split and indeed, ERVL have been found in all placental mammals. However, ERVL copies in both species diverged in their own way and are thus somewhat different. Notably, some ERVLs in the mouse genome (MERVL) are still mobile, while this is not the case in humans. Beside these autonomous elements, the ERVL family also includes Mammalian apparent LTR-retrotransposons (MaLR), which are the most common ERVs in the mouse genome (almost 5%) despite being non-autonomous. Class I ERVs are in close proximity to gamma- and epsilonviruses and contain few members in mammals (Figure 2). Finally, the most important ERV families in terms of transcription and transposition are contained in Class II or ERVK (K standing for the Lysine tRNA binding motif at their PBS). These elements are in close proximity to lenti-, alpha-, beta- and deltaretroviruses and are relatively young compared to other ERVs (Figure 2). Two families of ERVs present in this class are specific to the *Mus* genus: MusD/ETn and intracisternal-A-particles (IAP). Both MusD and IAP families contain full-length autonomous elements, while ETn derived from MusD and have lost their ability to encode proteins. IAPs are responsible for more than half of the deleterious insertions found in laboratory mouse strains and represent therefore the most active TE family in the mouse genome (Gagnier et al., 2019). Part I.B.1 is dedicated to their biology.

The origin of SINEs is quite different and heterogenous. The two main families in mice and humans, B1 and Alu respectively, emerged from the 7SL gene (Figure 2), at the origin of an RNA that forms a cytoplasmic ribonucleoprotein called *signal recognition particle* involved in translation of secreted proteins. Three other lineages of SINEs (B2, ID and B4) are still active in the mouse genome, originating from transfer RNAs (tRNAs) (Ade et al., 2013). Consequently, SINEs are transcribed by RNA polymerase III (RNAPol III) whereas all other TEs are produced by the action of RNA pol II. A last category of SINEs, present specifically in hominids and still mobilizable, emerged from the fusions between a SINE element and an Alu-element, with the fusion consisting of a Variable number of tandem repeat (VNTR), hence their name as SVA.

SINEs, like solo-LTRs, are non-autonomous elements and rely on proteins provided by LINEs for their transposition. Notably, it is estimated that 5% of newborn children have a Alu insertion (Hancks & Kazazian, 2012). Not only can TEs mobilize but they are transcriptionally competent which leads to a wide variety of consequences.



**Figure 2 | Diversity and origins of retrotransposons.**

Diagram presenting all the families and classes of elements mentioned in this manuscript according to their origins.

### 3. Transposable elements: a presence with side effects

Undoubtedly, TE presence impacts host genomes in various ways and at various levels. As mentioned previously, their presence might have played a role in genome size. The body of literature on the influence of TEs on their host is tremendous, which reflects its importance. For this part, I will present a direct or “theoretical impact” due to the presence, the expression or the transposition of TEs (Figure 3), providing examples taken from the mouse or other mammals. However, this will merely scratch the surface of their physiological impact on living species (or even on extinct species for which we lack informative data). While polymorphic copies tend to be linked to pathological outcomes, fixed copies are more linked to general processes at the level of the population. I will focus hereafter on the influence of fixed copies in physiological processes, which are very often related to immunity and development in their broad accession. I will also mention some pathological outcome due to polymorphic copies.

## a) TE implication in physiological processes

One specificity of TEs is their presence in multiple copies across the genome. These copies might be slightly different, because under different selective pressure, they acquire mutations over time at a specific rate. However, they will conserve for a long time a high homology allowing for intra- or inter-chromosomal homologous recombination, which can lead to deletion of large genomic regions (Muotri, 2023) (Figure 3). But the influence of TE sequences on nuclear architecture goes beyond homologous recombination, through impacting 3D chromatin folding. Chromatin is arranged in loops of sequences under the same transcriptional status called topologically associated domains (TADs) (Nora et al., 2012). Formation of these loops is maintained by cohesins bound to CTCF, a DNA binding protein. CTCF can bind to specific sequences that are notably present in B2 SINE elements in mice, giving them the role of TAD boundaries (Sundaram et al., 2014). Removing one particular B2 SINE could lead to TAD disruption and transcription modifications. Conversely, adding a new CTCF binding site defines new TAD boundaries and can also change the transcriptional status of host gene expression. If TEs did not exist, we could envision that these TAD domains would still exist by using other systems. However, this is a case where TE sequences have been repurposed by the host at its own advantage and now depends on it.

As mentioned before, a large number of TE copies are non-autonomous and lost or never had the ability to transpose by themselves. However, they are still able, for a large number, to modify transcription of host genes either by providing a promoter signal for RNA polymerase II or, in an opposite manner, by inducing transcriptional silencing by being themselves a target for repression (see II. Targeting TE for repression) (Figure 3). Promoting transcription of neighboring genes can have different manifestations: *i*) promoter-like effects, leading to the production of chimeric transcripts containing both TE and gene sequences; *ii*) enhancer-like long distance effects, *iii*) activation of transcription in a global manner by promoting chromatin opening. Depending on the role of the gene and its standard expression pattern in the considered cell or organism, this might have a positive, negative or no physiological outcome. Moreover, the observed physiological outcome might come from diverse effects of TE transcription: cis-acting or trans-acting effects of TE transcripts

themselves, modification of nearby gene expression pattern, production of chimeric TE-gene transcripts and proteins, or production of TE proteins. Deciphering the relative part of all these effects requires a large panel of experiments.

Since the early ages of TE biology, these elements have been shown to be expressed in the early mammalian embryo, during pre-implantation development (Kuff & Lueders, 1988). With the use and development of new technologies, notably NGS and single-cell sequencing techniques, the pattern of TE expression has been refined, as well as the presence of chimeric transcripts and roles they might carry during this period (Evsikov et al., 2004; Ge, 2017). For decades, MERVL elements have been reported to be highly expressed in the mouse 2-cell stage embryo, in correlation with Zygotic Genome Activation (ZGA), which corresponds to the first autonomous transcription of the embryonic genome after relying on maternally inherited transcripts and proteins (Q. Wang et al., 2001). The paucity of material available at early stages of development combined with the multiplicity of MERVL sequences across the genome made it technically challenging to prove the necessity of MERVL expression for ZGA. However, different studies overcame these challenges and showed that full-length MERVL transcription (not translation) is indeed essential for pre-implantation development but not for ZGA (Huang et al., 2017; Kigami et al., 2003; Sakashita et al., 2023). The two first studies used anti-sense oligos (ASOs) or siRNAs, leading to degradation of nuclear or cytoplasmic MERVL RNA, respectively. Using siRNAs, one can assess the need for MERVL to be transcribed, while using ASO allows evaluating the possible trans-acting role of the targeted RNA. In the last publication, MERVL were transcriptionally silenced using a CRISPRi system, allowing here to conclude on the necessity of MERVL transcription or acting as a transcriptional activator for neighboring sequences. ASO or siRNA against MERVL provoked a development arrest or delay at the 4-cell stage, coincidentally with ZGA and it was therefore thought that MERVL RNAs were needed for starting transcription and/or degrading maternally inherited transcripts. Blocking transcription of MERVL resulted in a similar developmental delay with only a proportion of embryos reaching the blastocyst stage, but RNA-seq of these embryos showed the presence of a large number of transcripts not inherited from the oocyte. Developmental arrest or delay was there attributed to a continuous expression of totipotent genes rather than incapacity to progress through ZGA.



Controlling TEs allows to limit mobilization and to use them in physiological processes essential to the host. The time-restricted expression of MERVL is essential to early development: lack of expression or prolonged expression have both deleterious effects (Sakashita et al., 2023; F. Yang et al., 2020). While TE expression control is needed, as gene expression, it has to be in a time and cell-type dependent manner, and dynamically regulated by the host. Allowing some level of TE expression favors innovation, through the emergence of new phenotypes and structures. These innovations might be deleterious, beneficial or even have neutral effects. A very visual example is the case of the *agouti* gene in mice. Inside can be inserted an IAP element whose control depends upon DNA methylation (see Part II). Variable DNA methylation levels at this specific IAP determines the expression level of the *agouti* gene involved in fur color (Michaud et al., 1994), bringing some inter-individual variety. This illustrates a case of position-effect variegation whereby the presence of a TE influences the local chromatin context, including genes.

To promote their expression, similarly to genes, TEs make use of transcription factors (TFs), a class of DNA-binding proteins that locally enhance or restrict RNA pol II binding. Transcription factors usually recognize a canonical DNA sequence called motif and are more or less permissive to sequence variation from the canonical motif. Several studies agreed on the global finding that some TF binding motifs are found embedded in TE sequences at a higher rate than random (Bourque et al., 2008; Sundaram et al., 2014). Consequently, TE expression can follow the trend of expression of these TFs. A curious observation is that although a TE family may be present in two species and probably inherited from a common ancestor together with a common TF, the association between the two will not necessarily occur. In other words, each species developed their own diversity of TF-bound-TEs since speciation, either evolving TE sequences to use already available TFs, or by evolving TFs to use binding sequences already present at a high number at TEs.

Once expressed, full-length TE transcripts can be translated. They produce proteins or polypeptides with specific functions. Because these coding sequences are not under selective pressure for the survivability of the host, they can “freely” derive and acquire mutations over time, which can lead to the creation of proteins with new functions for the host (Figure 3). This represents one way for TEs to be evolution

drivers. Notably, upon their new insertion in genomes, ERVs still possess and express their envelope gene whose product promotes membrane fusion. Envelope-encoding sequences have led to the emergence of SYNCYTINs, a protein family involved in cell-cell fusions at the core of placenta development and physiology (Dupressoir et al., 2011). SYNCYTIN-driven cell fusions are responsible for the formation of the syncytiotrophoblast, a placental tissue structure involved in materno-fetal exchanges of nutrients and gases. This is referred to as a co-option or domestication process: a TE sequence has derived and encodes now for a protein that only serves the host. Importantly, the role of TEs in the development of placenta in mammals is not limited to this process of protein domestication. Their capacity to promote or enhance transcription of genes in their vicinity has also been co-opted in this tissue, as response elements to specific hormones involved in the expression of genes involved in placentation (Chuong et al., 2013). Expression of these genes is notably essential to modulate immunity regarding the fetus, which despite being non-self has to be protected from the mother's immune system.

If TEs can be expressed, they might even be capable of mobilizing and duplicating. While these new insertions can in principle happen anywhere in the genome, they tend to be not completely random or at least, the observed and conserved insertions present some insertional bias. Exon insertion is rarely seen compared to intron insertions: either they do not occur (or at a low rate) or are too deleterious to be conserved and propagated. On the contrary, TE sequences present in introns are quite frequent and can even be used as new exons (or alternative exons), and give rise to new proteins and functions in a process entitled exonization (Figure 3). Such a process can lead to transcript diversity from a single gene, as observed for the prolactin-encoding *plr3c1* gene that contains a composite TE element in its first intron. The TE-free *plr3c1* transcript isoform is expressed in specific cells of the uterus, while the TE-included *plr3c1* isoform is expressed in Leydig cells of the testis and plays a role in their growth (Bu et al., 2017).

New insertions occurring in somatic cells are not transmitted to the next generation and so will not be domesticated, co-opted or exonized. However, LINE-1 mobilization has been shown to occur in physiological conditions, and with a high rate, in human and mouse neuron progenitor cells (NPCs) (Coufal et al., 2009; Muotri et al.,

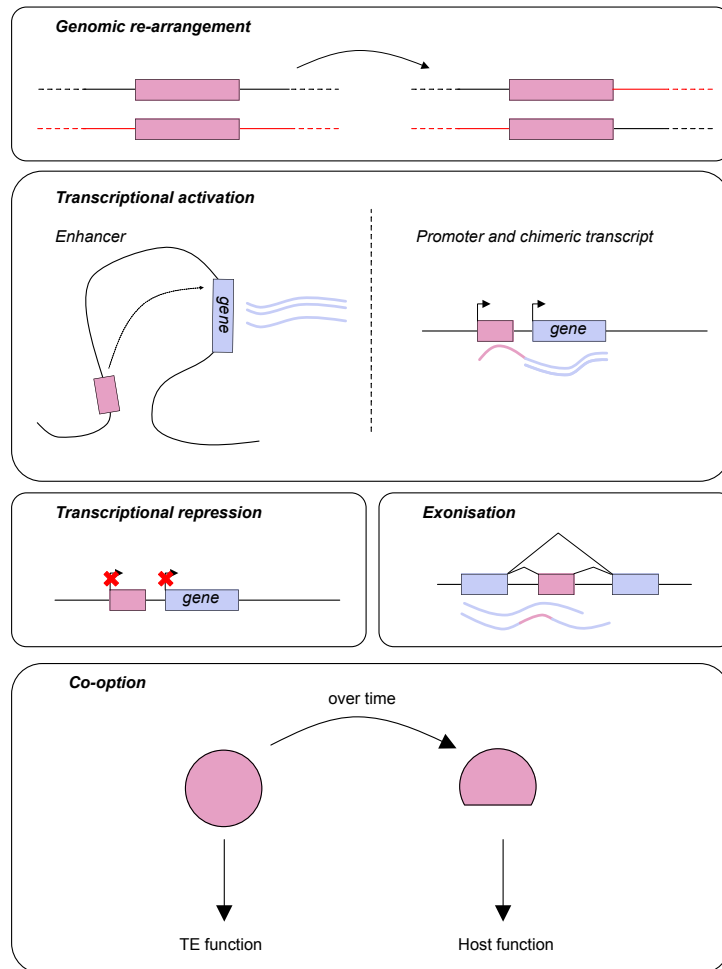
2005). It is thought to participate in neurons population diversity and to be especially important in primates. This is the tip of the iceberg of the impact of TE biology on brain and neuron development, a link that has been reviewed extensively, both in physiological and pathological aspects (Popova et al., 2022).

In the cases reported above, TE expression or presence not only has an impact on the host but it has become essential for its physiology. This raises fundamental questions: whether the use of TEs by the host is *i)* an active process as opposed to a consequence of evolution, or *ii)* a consequence of TE presence and activity promoted by TEs themselves or by a by-product of their selfish behavior or *iii)* promoted by the host to use available DNA sequences to its advantages or to ensure proper control of TEs. These conceptual notions are difficult (if not impossible) to test experimentally because what we observe nowadays is the product of million years of co-evolution between TEs and their host genomes, and new processes of TE co-option cannot be observed at our time scale. All things considered, it seems reasonable to say that TEs cannot be only seen as intra-genomic parasites anymore and that their contribution to host biology should not be overlooked.

#### b) TE implication in pathological processes

Pathological outcomes mainly arise when host repressive mechanisms against TEs are inoperative. As an example, failure to repress TEs in male germ cells typically prompt meiotic arrest, smaller testis and sterility (Ding et al., 2017.; Aravin et al., 2007; Shoji et al., 2009; Barau et al., 2016; Watanabe et al., 2018; Zoch et al., 2020). Global chromatin remodeling in cancer cells often leads to loss of control over TEs and a burst of TE expression compared to normal cells. In humans, high LINE-1 expression is a hallmark and could even serve as a biomarker for early detection of cancer (Sato et al., 2023; Taylor et al., 2023). Few examples of TE mobilization or expression being one of the initial events eliciting cancer development exist (Babaian et al., 2016; Scott et al., 2016; Ukai et al., 2003). Their number is probably underestimated for technical reasons mentioned earlier regarding de novo assembly and accuracy of TE localization as well as for the hindrance of proving the early event leading to cancer development. Permissive TE activity in neural cells also explains why several neurodegenerative disorders like Alzheimer's or Huntington's diseases are linked to TE expression or

mobilization (Popova et al., 2022). Controlling where and when TEs are expressed is then an essential process to ensure correct physiological functions using them (presence or transcription) and avoid the development of pathologies due to their selfish behavior.



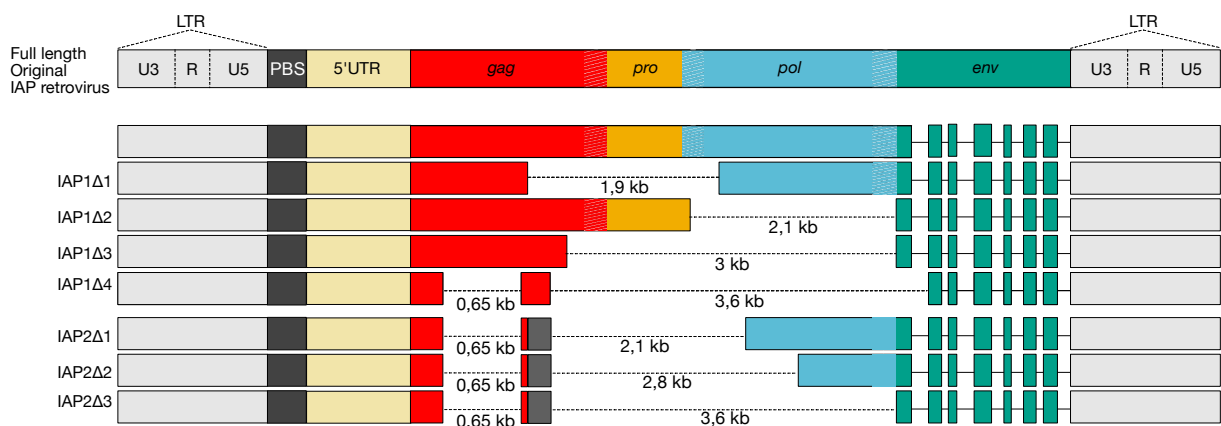
**Figure 3 | Instabilities due to TE presence.** Schemes showing different effects of TE presence on host gene regulation beside their mobilization

## B. Focus on two TE families with mobilization potential

LINE-1 and IAP elements are members of distinct TE classes with mobilization potential in the mouse genome. To get a better understanding of what distinguishes these two classes of elements and why some controlling pathways are efficient on one but not the other, I propose hereby a more detailed description of these families.

## 1. Intracisternal-A-Particles

IAP is one of the few ERV family that can still mobilize in mammals and is only present in the mouse genome and closely related rodent species (Magiorkinis et al., 2012). Their name is inherited from a very specific observed behavior: their localization inside the cisternae of the endoplasmic reticulum that no other ERV is known to localize to. The retrovirus they originated from encodes for four proteins: Group specific Anti-Gene (GAG), Proteinase (PRO), Polymerase (POL) and Envelope (ENV) (Figure 4). Although not having their own transcriptional start site or transcription termination because they are produced from the same transcript, each of these open reading frames are considered as genes encoding for their own protein (Figure 4). Localization of IAP to the endoplasmic reticulum is due to a short sequence at the N-terminal part of the *gag* sequence, which proved also to be important for control of these elements (Fehrmann et al., 2003; Enriquez-Gasca et al., 2022).

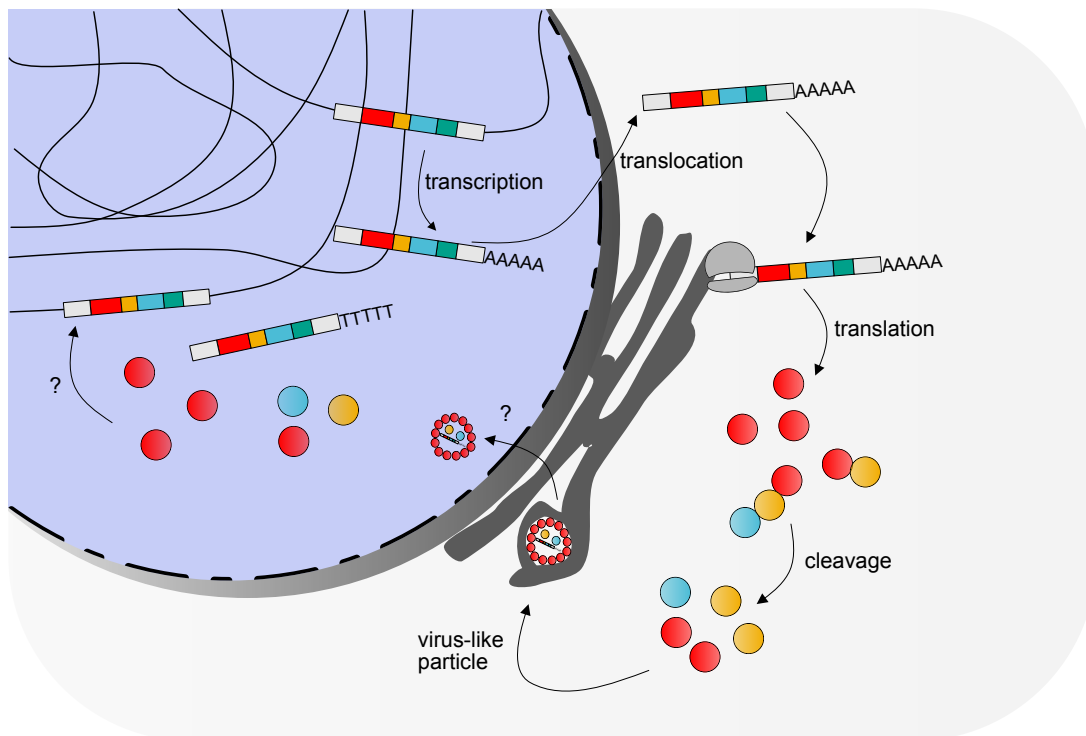


**Figure 4 | Functional organization of IAP elements.**

Schemes representing the functional organization of the different types of IAPs described in (Kuff & Lueders, 1988).

These four genes are the ones classically found in infectious retrovirus sequences and they play here the same functions. The IAP family has been shown to come from a pre-existing retrovirus, still present and inserted in the mouse genome (Ribet et al., 2008). Nowadays, IAP elements are known to have an obligate intracellular life cycle and therefore do not need their *env* gene. This coding sequence is still present in some copies of the family (IAPE) although mutated or truncated, but most copies have either lost their *env* or have it so divergent that it cannot be recognized anymore as *env* (Figure 4). One particularity of *gag*, *pro* and *pol*, in ERVs but also in some retroviruses, is that their coding sequences overlap, which leads to

the necessity to cleave the produced polyproteins (Figure 4). In fact, this system produces GAG as an alone protein with the first ORF, and two polyproteins GAG-PRO and GAG-PRO-POL (Figure 5). These polyproteins have to be cleaved to be fully functional, which is the role of PRO. Moreover, a frameshift is needed during translation at the junction between GAG and PRO to produce fully functional PRO and POL peptides (Fehrmann et al., 1997). GAG forms the inner shell of the virus-like particle, the outer shell being formed by the endoplasmic reticulum itself (Figure 5), while POL encodes the reverse-transcriptase and endonuclease activities.



**Figure 5 | Mobilization cycle of IAP elements.**

Scheme of IAP mobilization process. Colors used to depict functional elements are the same than in Figure 4.

The first step for IAP elements to initiate their retro-transposition cycle is to be transcribed by hijacking the host machinery. This is achieved through their LTR that can be divided in three sub-parts: U3, R and U5 (Figure 4). U3 contains the transcription termination sequence, R corresponds to a duplicated region on both LTRs, while U5 contains the transcriptional start site and a TATA box to help recruiting the RNA polymerase II. The LTR is repeated in the same orientation at both ends of full-length IAP elements, and therefore, only transcription starting from one LTR in the right orientation can produce a full-length transcript (Figure 4). However, transcription can start from the other LTR and even be pervasive enough to produce the anti-sense element. In the same way, pervasive transcription (or weak termination signal) can lead

to transcription of neighboring sequences, including host genes (Figure 3). In fact, transcription and transposition have to be considered as separate processes. Transcription can occur even if the coding sequence is incomplete or even completely missing in the case of solo-LTRs. They represent approximately half of the IAP copies (approx. 3000 solo LTRs and 2800 full-length or partially full-length elements) found in mouse reference genome.

One good reminder of this separation between transcription and transposition abilities is perfectly illustrated in the IAP super-family. In fact, several members possess a variety of deletions in their internal sequence, and these are known as the IAP $\Delta$  elements (Figure 4). A first category of elements (IAP1 $\Delta$ ) has only deletions, while the second one (IAP2 $\Delta$ ) acquired in addition to their deletions a new insertion compared to the ancestor IAP retrovirus (Figure 4). IAP1 $\Delta$  harbors three types of possible deletions depicted in Figure 4. A really interesting finding is that IAP1 $\Delta$ 1 is the type of IAP copies that propagates the most while being unable to transpose on its own due to an internal deletion removing part of *gag* and *pol* as well as the entire *pro* sequence (Figure 4) (Gagnier et al., 2019). This is possible only if two processes simultaneously happen: trans-complementation or the use of proteins produced by other fully competent IAP elements, and the expression of the non-autonomous element (Saito et al., 2008). Although trans-complementation has been proven, the exact reason why IAP1 $\Delta$ 1 elements are more successful than their autonomous counterparts or other IAP1 $\Delta$  elements is still not completely understood. The 120 kDa GAG-POL fusion protein that they produce seems to be important for their more efficient ability to use functional IAP proteins than other elements (Saito et al., 2008).

In the case of IAPs and other ERVs, two nomenclatures co-exist in the literature: *i)* one based on their DNA content inside the family mostly studied in the 80's without resort to sequencing and *ii)* one based on NGS data acquired during this century. Nomenclature chosen for all transposable elements and present in new databases does not consider observations made previously and are based on divergence between sequences. Therefore, IAP1 $\Delta$  elements are not annotated as such in databases and this annotation will not be present in our data analyses. Separate annotations were made for LTR and coding part of full-length elements because LTRs can be found on their own. Annotations diverge sometimes between the different

databases available: Dfam (publicly available) and RepBase (accessible with fees). Accordingly, IAP\_LTR1 and IAP\_LTR1a are the youngest and most potent LTRs found in the mouse genome, including solo-LTRs. The most abundant type of full-length element according to recent annotations belong to the IAPez subfamily (also annotated as IAPLTR1a in Repbase), whose distribution into ~6500 copies is quite homogeneous among autosomal and sex chromosomes. By contrast, the majority of other IAP subfamilies are highly enriched on the Y chromosome, such as IAPey.

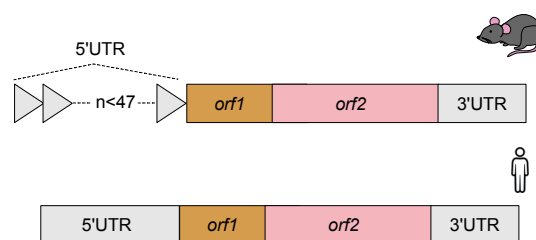
IAPs are expressed in oocytes, zygotes and 2-cell stage embryos and their expression declines afterward (Fadloun et al., 2013; Piko et al., 1984; Rothstein et al., 1992). However, reasons and needs for this expression have not been properly investigated, at least in comparison to the attention given to the developmental expression pattern of other ERV families like MERVL. Moreover, a study based on RT-qPCR propose an increased number of IAP RNAs during early development as opposed to previously mentioned publications (Inoue et al., 2012). In mouse embryonic stem cells (mESCs) derived from the inner cell mass of the blastocyst, although IAP expression can be detected, they are not highly expressed compared to oocytes or zygotes. Some IAPs also seem to be expressed in several adult tissues, like brain, heart, liver and kidney, although with a higher expression in brain compared to the other tissues (Gaubatz et al., 1991). While research on how IAP elements are repressed, especially in early development is quite prolific, study of their actual role *in vivo* is an unlooked area. Moreover, available expression data come from different kinds of experiments, and some can only reveal specific sub-families and completely miss the expression of solo-LTRs due to use of primers or probes binding to internal sequences (RNA-FISH or RT-qPCR). Overall, a unified study of the expression of IAPs during mouse development is missing.

## 2. Long Interspersed Nuclear Element 1

Full-length LINE-1 elements (approx. 6kb in mice) encode at least two proteins from two open reading frames (ORFs), called ORF1p and ORF2p (Figure 6). Some copies can also encode for ORF0, which has only been found in primates. ORF1p and ORF2p are essential to LINE-1 transposition cycle, while the action of ORF0 is quite unclear although it seems to stimulate retro-transposition (Denli et al., 2015). This two ORFs



are bordered by 5'UTR and 3'UTR sequences, which serve the same function in all species but can be different in sequences (Figure 6). In mouse, the L1 5'UTR is formed by variable repetitions of monomers, whose range spans 2 to 50 copies (Figure 6), although a mean of 2-3 monomers is observed for active elements (Goodier et al., 2001). Active Human Specific LINE-1 (L1Hs) (and any human L1s in general) do not possess this monomeric structure on their 5'UTR (Figure 6). Mouse 5'UTR monomers have been shown to possess sequences bound by transcription factors (such as YY1) with an active role in LINE-1 transcription (Cusack et al., 2020; Deberardinis & Kazazian, 1999). The nomenclature used today for mouse LINE-1 is based on early studies identifying divergences between elements in these monomers. L1-MdA (L1-A, A-type promoter), L1\_MdF (F-type, promoter) and L1\_MdTf (L1-T, F-type promoter), therefore possess different monomer sequences (Fanning, 1983; Wincker et al., 1987). A phylogeny of L1 was also based on the divergence between these tandem repeats, putting L1\_MdA as descendant of L1\_MdF (Adey et al., 1994). However, LINE-1 tend to lose their 5'UTR sequences containing these monomers during their reverse-transcription starting at their 3'UTR. Therefore, these sequences are not highly conserved and even recombined between different LINE-1 elements along evolution and might not be the sequence of choice to decipher mouse LINE-1 evolution. Another phylogenetic tree of mouse LINE-1 was proposed based on ORF2p sequence which is highly conserved due to its essential role in transposition (Sookdeo et al., 2013). Here, it is proposed that L1\_MdF gave raise to L1\_MdA while continuing to evolve on its own in two separated lineages. L1\_MdA elements then are proposed to be source of the active LINE-1 lineage in mouse and L1\_MdTf diverged from it.



**Figure 6 | Functional organization of LINE-1 elements.**

Scheme representing the main composition of LINE-1 elements in mouse and human genomes

ORF1p is a 40kDa protein that organizes as trimers (Figure 7). While its exact role in the retro-transposition process is still unclear, it is highly expressed and binds to RNA or DNA and might act as a nucleic chaperone. Despite difficulties to catch the

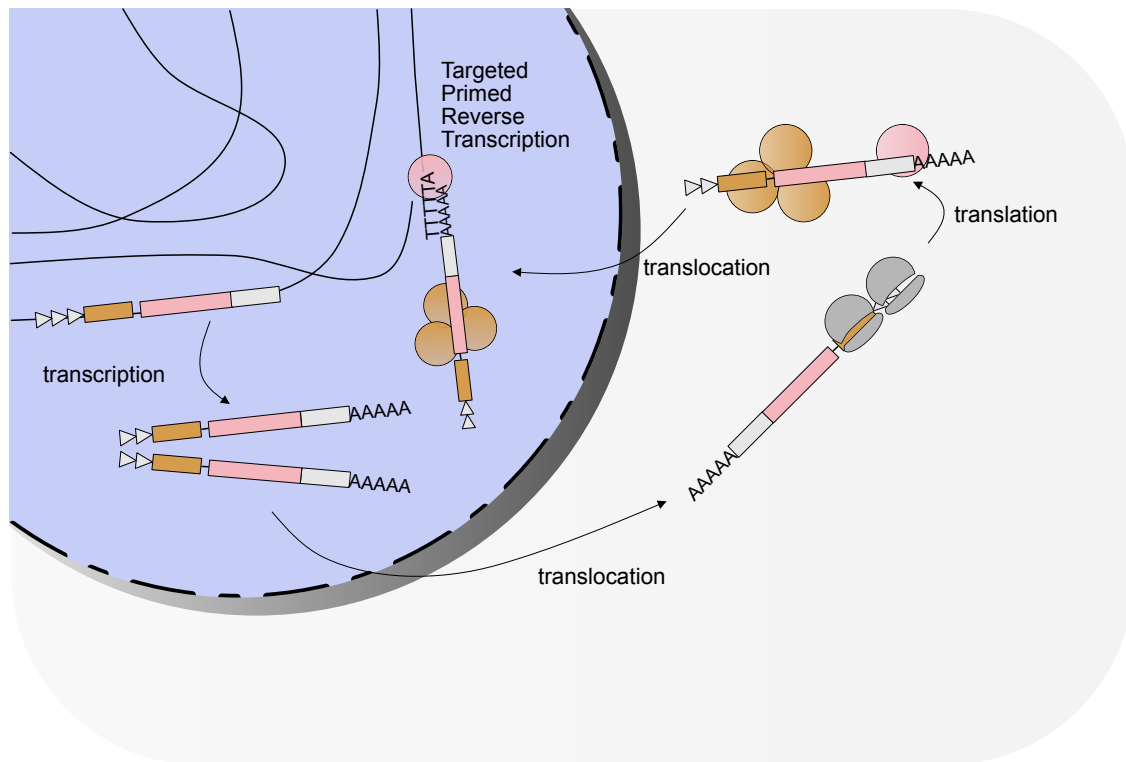
149 kDa expressed ORF2p in living cells, this protein bears two essential functions for retro-transposition: endonuclease and reverse-transcriptase. After transcription, the LINE-1 RNA has to be translocated in the cytoplasm to be translated. LINE-1 presents as a bicistronic RNA, which is commonly translated by bacterial ribosomes but is not the way eukaryotic RNAs are formed. Therefore, the question of how ORF2p is translated (the second ORF from 5' to 3', [Figure 6](#)) is still in question. Accordingly, detection of ORF2p in living organisms has been a challenge due to its low translation rate. ORF1p and ORF2p have been shown to strongly bind to LINE-1 RNA forming a LINE-1 ribonuclear particle (L1-RNP) ([Martin & Bushman, 2001](#)) with a preference for their own RNA, although binding to other RNAs, such as SINEs, has been shown ([Callahan et al., 2012](#)). This partially explains how non-autonomous elements can hijack autonomous element machinery. L1-RNPs are found to localize at cytoplasmic stress granules ([Figure 8](#)). This specific localization might be actively regulated by the host, by the LINEs or it might reflect a secondary effect of ribonuclear particle formation ([Arora et al., 2022](#); [Goodier et al., 2007](#)). However, if this localization is needed to perform retrotransposition or plainly due to host control is not completely known.

To retrotranspose, this RNA-Protein complex has to re-enter the nucleus ([Figure 7](#)). How this is exactly achieved is still in debate however, a recent paper suggests that LINE-1-RNPs might enter the nucleus by localizing to chromatin during mitosis when the nuclear membrane is gone, and be entrapped in the nucleus once the nuclear membrane builds up again ([Mita et al., 2018](#)). This would be the only way of LINE-1 to enter the nucleus, non-dividing cells should not suffer from its mobilization. However, mobilization of LINE-1 elements has been observed in post-mitotic neurons, meaning that other processes might be used by LINE-1-RNPs to enter the nucleus ([Muotri et al., 2005](#)).

Once in the nucleus, what is needed for a LINE-1 element to retro-transpose is a free single stranded 3'-OH, which can be obtained directly from its own ORF2p-dependent endonuclease activity (canonical pathway) or from already present DNA single strand breaks (non-canonical pathway, endonuclease independent or non-classical L1 insertion), a process used by host machinery in order to maintain telomeres length ([Figure 7](#)) ([Kowald, 1997](#)). Depending on the amount of available DNA breaks at a particular moment, finding the breaks might not be as efficient as

creating new ones. Notably, ORF2p has a preferred site for its endonucleolytic activity (TTTTA), which allows an easy start of reverse-transcription by hybridization with LINE-1 poly-A tail. Then, ORF2p will take LINE-1 RNA as a template to produce and incorporate its DNA sequences at the site of the DNA break. This process is also known as Target-Primed Reverse Transcription (TPRT) (Figure 7). When already present DNA breaks are used, stretches of Ts might not be available to start this process. It is then worth noting that although reverse-transcription is more efficient when the DNA template starts with 4Ts, ORF2p can achieve it inefficiently with other initiation sequences. Although this process can occur at any time once LINE-1-RNP enters the nucleus, it seems to be more frequent at the S phase of the cell cycle (Mita et al., 2018), maybe in relationship with DNA accessibility. ORF2p has been shown to bind PCNA, which is present at replication fork, an easy place to find single-stranded or accessible DNA (Mita et al., 2018).

LINE-1s are the only autonomous elements still able to transpose in humans (around a hundred copy of active L1HS), according to the current detection tools (Brouha et al., 2003). New insertions have been estimated to occur in 1/60 births (Feusier et al., 2019). Mobilization of LINE-1 also still occurs in the mouse genome with a ratio going as high as 1 new insertion in 8 births, probably linked to higher number of active LINE-1 copies in this species (Richardson et al., 2017). New somatic insertions of L1HS are biased toward AT-rich and early-replicating regions (Sultana et al., 2019). The observed AT rich bias is expected considering their target sequence for insertion. However, it seems that these new insertions are not biased toward a specific chromatin pattern (Flasch et al., 2019). Despite the need for LINE-1s to be transcribed for new insertions to arise, open chromatin or genic regions are not favored for insertions. Selection toward these regions may happen overtime.



**Figure 7 | Mobilization of LINE-1 elements.**

Scheme of LINE-1 mobilization process. Colors used are the same than in [Figure 6](#).

LINE-1 elements are also dynamically expressed in early development. L1 retrotranscription and translation has been shown to happen in mouse zygotes and 2-cell embryos ([Richardson et al., 2017](#); [Vitullo et al., 2012](#)). LINE-1 transcript levels were shown to increase from zygote to morula stages ([Inoue et al., 2012](#)), as opposed to results obtained using RNA Fluorescent hybridization (FISH) instead of RT-qPCR ([Fadloun et al., 2013](#)) Discrepancies between these two studies might come from the distinction between active transcription ([Fadloun et al., 2013](#)) and conserved transcript load ([Inoue et al., 2012](#)). LINE-1 expression timing in pre-implantation development proved to be essential with a prolonged expression beyond up to 4 and 8-cell stage leading to developmental arrest and chromatin silencing in zygotes also having a negative impact on development progression ([Jachowicz et al., 2017](#)). This phenotype is very similar to the one observed concerning MERVL expression, also concerning the proper progression of ZGA of LINE-1 silenced zygotes ([Sakashita et al., 2023](#)).

## III. Control mechanisms

Control of TE transcription and mobilization is especially important in pluripotent and germ cells due to the transmission of their genome to the next generation. However, once acquired in host genomes, TEs also have the ability to mobilize in somatic cells. Although a lot of efforts have been made to understand how TE regulation is specifically achieved in germ cells and their precursors, other pathways are involved in the regulation of TEs in various cell types, their expression and roles being broader and often also involved in limiting viral infections. I will briefly present some of these general restraining pathways in a first part, before focusing on targeted control of active TEs in ESCs and male germ cells.

### A. Non-targeted control mechanisms

Under this banner of non-targeted control mechanisms, I want to mention several pathways that are involved in the regulation of TE transposition without the need to adapt to their specific sequences. In all pathways described herein, discrimination between self and non-self is achieved by other means. Investigations often starts with the idea of confirming if factors known to regulate viruses are involved in TE repression, due to their common features. However, after confirmation obtained, the exact mechanism is often not detailed and TE features triggering this particular control not investigated. Before detailing these active methods of TE repression, let's remember that TE sequences have a high mutation rate that often dampen their transposition abilities or in the case of LINE-1s, new insertions are often 5' truncated. These newly inserted copies might therefore never be expressed, causing no harm without the need for any specific active mechanism of repression.

#### 1. Innate immune response

The first line of defense against ERVs is to avoid viral infection or viral integration, and indeed, the immune system is a barrier against TE transposition. Like viral infection, the TE transposition cycle or invasion triggers one specific innate immune system: the type I interferon response (IFN) (Figure 8). In response to detection of double-strand RNA (dsRNA), double-strand DNA (dsDNA) or RNA-DNA hybrids in the cytoplasm, a

specific transcriptional program, partially due to the expression of the transcription factor IRF3, is launched (Schneider et al., 2014). The aim of IFN response is to restrict viral replication within infected cells, alert neighboring cells of the presence of infections and expand lymphocytes to provide long-term and specific protection against viruses. IFN responses can be triggered by various ways and in the case of TEs, this will be mainly achieved by detection of ribonuclear species that they produce in the cytoplasm: RNA would be detected by OAS/RNaseL system, while DNA would be recognized by cGAS/STING path.

Both pathways have been implicated in the control of LINE-1 and IAP elements. In the first case, 2',5'-Oligoadenylate synthase (OAS) senses dsRNA, which promotes its activity towards the production of 2'-5' oligoadenylates. These molecules activate RNaseL that cleaves in turn single-strand RNAs (ssRNAs). Both host and viral ssRNAs are cleaved leading to translational arrest and in the long-term, to cell apoptosis. While this pathway has been shown to reduce LINE-1 and IAP retrotransposition (Figure 8) (A. Zhang et al., 2014), it is unclear how, as these elements are not known to produce dsRNA species during their transposition cycle. A proposed hypothesis for LINE-1 is the binding of small antisense RNAs, produced from the bi-directional transcriptional potential of the 5'UTR (N. Yang & Kazazian, 2006). Another TE transposition byproduct could be responsible for the activation of this system, or it could be triggered by secondary structures of TE-derived mRNAs. In contrast, dsDNA presence in the cytoplasm is a trace of retrotransposition activity independently of their insertion in the genome by ERVs. dsDNAs are sensed by the cyclic GMP-AMP synthase (cGAS) generating a metabolite activating stimulator of interferon genes (STING), the cGAS/STING pathway. cGAS/STING was also shown to sense DNA/RNA hybrids which are also formed during ERV transposition (Mankan et al., 2014). Moreover, response to viral-derived DNA in the nucleus was reported (Lahaye et al., 2018). In a nutshell, IFN response is activated by ribonuclear species produced either by viruses, retroviruses or LINE elements,

Although this system is efficient against bursts of unwanted TE expression, it leads to cell death and therefore has to be activated only as a last resort. Moreover, constant activation of IFN response is associated with auto-immunity in the long-term, which is highly detrimental to the host, as observed in the Aicardi-Goutières Syndrome.

Activity of Three Prime Repair Exonuclease 1 (TREX1) has been shown to counteract this auto-immune phenotype by targeting reverse-transcribed DNA from LINEs and SINEs for their degradation (Stetson et al., 2008). This prevents its accumulation in the cytosol and activation of immune system, by degrading the problem inside without calling for help from outside.

## 2. Degradation through stress granules and processing bodies

Degradation of viral or TE RNA can be achieved by other means, notably by targeting RNA to stress granules or processing bodies (P-bodies) present in the cytoplasm. Both structures are membrane-less condensates that coalesce various factors. While P-bodies are mainly involved in RNA decay and are constitutively present, stress granules concentrate translation-associated factors and RNAs, and specifically form under stress conditions (Riggs et al., 2020). Calcium-binding and coiled-coil domain-containing protein 2 (CACO2 alias NDP52) and p62 target P-bodies and stress granules for autophagy, respectively, leading to the degradation of all proteins and RNAs contained in this condensate type (Figure 8). Not only TE repression depends on the formation of these structures, but also on all the proteins needed for the autophagy pathway, like ATPG6 (Guo et al., 2014). IAP RNAs are often found localizing in P-bodies, while L1 proteins were found to localize to stress granules (Figure 8) (Doucet et al., 2010; C. Lu et al., 2011). Blockade of LINE-1 and IAP retro-transposition depends thus on the different proteins involved in the formation of these granules. Sterile alpha motif and histidine-aspartate domain-containing protein 1 (SAMHD1), a deoxynucleoside triphosphate triphosphohydrolase usually involved in repressing viral infection and insertion through degradation of dNTPs below the requirement for DNA replication, was also shown to play a role in the formation of stress granules and furthermore, in LINE-1 repression (Figure 8) (Hu et al., 2015).

How ORF1 and ORF2 are targeted to stress granules while involved in a complex with L1-RNA is not resolved. However, some recent advances have been made regarding how IAPs may be addressed to P-bodies. A possibility is that it involves an RNA modification, and more specifically, the methylation of the 6<sup>th</sup> carbon of adenosine (m<sup>6</sup>A). This RNA modification is deposited co-transcriptionally by writer-complexes, including the RNA methyltransferase METTL3 that targets the internal part

of mRNA species, including IAP elements (Figure 8) (Chelmicki et al., 2021; Geula et al., 2015). An increasing list of readers can bind to m<sup>6</sup>A-modified RNAs, with implications in all RNA processing steps: splicing, nuclear export, stability, translation (Shi, Wei, et al., 2019). The binding specificity of each reader is not completely resolved but may be linked to the position of the modified adenosine along the mRNA. While several publications agree on a repressive role of m<sup>6</sup>A on IAP elements, both negative and positive effects have been reported on LINE-1 elements. On one hand, removal of the m<sup>6</sup>A mark by the m<sup>6</sup>A eraser called fat mass and obesity-related protein (FTO) favors L1 RNA abundance in mESCs (Wei et al., 2022), while addition of the m<sup>6</sup>A mark was reported to enhance L1 transcription and retrotransposition in human cell lines by another study (F. Xiong et al., 2021). Moreover, claims have been made toward a role for METTL3 in enhancing H3K9me3 enrichment at IAP elements, which would make this pathway involved in transcriptional control of TEs too (Liu et al., 2021; Xu et al., 2021).

LINE-1 RNAs might also be targeted for degradation through the Non-sense Mediated Decay (NMD). The RNA helicase MOV10 was shown to both interact with LINE-1 RNA and to colocalize with ORF1 and ORF2, which cover LINE-1 RNA in the cytoplasm (Goodier et al., 2012; Gregersen et al., 2014). MOV10 further interacts with regulator of non-sense transcript 1 (UPF1) that also co-localizes with LINE-1 complexes. Interaction between MOV10 and UPF1 suggests that LINE-1 RNAs would be targeted for degradation, which may prevent their re-entry into nucleus and might even destabilize ORF1 and ORF2. Although MOV10 binds the coding sequence of LINE-1 RNA, its binding to other RNAs is biased to long 3'UTR regions (Gregersen et al., 2014). Of note, MOV10 also interacts with YTH domain-containing family 2 and 3 (YTHDF2 and YTHDF3), two m<sup>6</sup>A readers, and with YLP motif-containing 1 (YLPM1, also known as ZAP). Although the underlying mechanism is still unknown, ZAP participates in the post-transcriptional repression of LINE-1 elements and locates in stress granules, which promotes MOV10 to stress granules too (Figure 8) (Goodier et al., 2015). MOV10 might then be implicated in addressing LINE-1 to stress granules, in complement to its role in NMD (Arora et al., 2022).



### 3. Other degradation and editing systems

Degradation of ORF1 and ORF2 might be mediated through the action of Testis Expressed 19.1 (TEX19.1) too. This protein is expressed in the testis, as well as in ESCs and placenta. Deletion of TEX19.1 in testis leads to a meiosis defect but without reactivation of LINE-1 or IAP elements, while an IAP-related family of ERVK, MMERVK10C, is greatly over-expressed in these conditions (Öllinger et al., 2008). In contrast, deletion of TEX19.1 in ESCs lead to increased mobilization of LINE-1 elements (Maclennan et al., 2017). Interaction between TEX19.1 and ORF1 in ESCs is further thought to target ORF1 to degradation through the proteasome, via an interaction with the E3 ubiquitin ligase UBR2 (Figure 8) (Maclennan et al., 2017). As protein degradation prevents mobilization without impacting transcription, observations in ESCs are not necessarily discordant with the ones in *Tex19.1*-KO spermatocytes and spermatogonia. Increased L1 mobilization—in absence then of LINE-1 RNA overexpression—could create DNA breaks and perturb meiosis.

Several other processes are involved in restraining TE mobilization. For instance, the family of Apolipoprotein B mRNA editing catalytic polypeptide-like 3 (APOBEC3) proteins is involved in repressing LINE-1, IAP and other ERVs (Figure 8). APOBEC3G does so by enhancing Adenosine (A) deamination, causing a high rate of A to Guanosine (G) mutations in mRNAs and disrupting the coding sequence of the element (Arias et al., 2012). APOBEC3G also targets retrovirus RNAs in their capsids. Notably, the human APOBEC3G is not able to repress mouse IAP sequences, probably as a result of fast evolution and high adaptivity of APOBEC proteins to their host endo- and exogenous retroviruses. APOBEC3A is also implicated in LINE-1 and IAP repression, although through a deaminase-independent mechanism (H. Chen et al., 2006). Several proteins involved in DNA damage response as well as DNA repair were also shown to restrain new insertions of TEs, especially LINE-1 (Zamudio and Bourc'his, 2010). Finally, the deletion of RNA interference core factors, like DROSHA or DICER, also leads to enhanced mobilization of LINE-1 elements, although it is not clear whether the effects are direct or not (Goodier, 2016).

Mechanisms used by host cells to regulate their own RNA and protein metabolism or to fight viral infections contribute to the overall low mobilization rate of TEs. These defense systems allow cells to cope with production of TE transcripts, which might be needed, notably during development, to avoid unwanted transposition events. However, to allow time-, cell type- and family- dependent control of TEs, specific TE-targeted mechanisms are needed.

## B. Selective targeting of TEs for repression

TEs present a wide variety of DNA sequences, length, origins, ages and encoded proteins. A first challenge of an efficient controlling system is to use a characteristic that is shared by the highest number of TEs. All autonomous TEs start their transposition cycle by transcription, a step that can already influence the host genome. Acting at the front line, several mechanisms are used to prevent TE transcription. Three main complexes or pathways have been described in the literature, each of them targeting a different feature of TEs. One system is based on a vast family of proteins under positive selection for fast evolution with the ability to recognize TE sequences, the Kruppel Associated Box Domain Zinc Finger Proteins (KRAB-ZFP, KZFP). A second complex of proteins target RNA sequences lacking introns, which seems to be the most common feature of retrotransposons, the Human Silencing Hub (HUSH). Finally, a third system, specifically acting in germ cells, is based on the memory of past exogenous RNA invasion from which retrotransposons come from, the Piwi-interacting RNA (piRNA) pathway.

### 1. Targeting DNA to control transcription, the case of KZFPs

KZFPs constitute the largest family of transcription factors in mammals and even in higher vertebrates (Ecco et al., 2017). They contain C2H2-type zinc fingers in variable number (2 to 40 in humans), each of them recognizing three consecutive nucleotides on one DNA strand, and one nucleotide on the reverse strand. This particular form of zinc finger is also observed in PRDM9, a protein essential for meiotic recombination, which also harbors a domain catalyzing H3K4 trimethylation and is hypothesized to be the original protein KZFP (Birtle & Ponting, 2006). Mutating amino acids involved in nucleotide recognition can change KZFP sequence specificity. Therefore, to be able to

recognize 100 different sequences, one would need to encode hundred different KZFPs. Around 350 and 700 KZFPs are encoded in humans and mice, respectively, while their whole genome encodes for around 25,000 proteins (Kauzlaric et al., 2017; Urrutia, 2003). A large majority of KZFPs target TE sequences (Figure 8) and especially retrotransposons, but approximately one third of them target non-repetitive host genes (Imbeault et al., 2017). Of note, the presence of DNA methylation can affect the binding of some KZFPs, like ZFP57, which adds one more layer of regulation to the system (Shi, Strogantsev, et al., 2019).

KZFPs also encode for other domains to complete their targeting function and in particular a KRAB domain, recognized by Krab-Associated Protein 1 (KAP1) also known as Tripartite Motif Containing Protein 28 (TRIM28) (Friedman et al., 1996). KZFPs can also harbor *i)* a SCAN domain that interact with other SCAN domains, allowing interactions between different KZFPs, and *ii)* a DUF3669 domain whose functions or binding abilities are unknown. Most of SCAN-endowed KZFPs target genes, not retrotransposons. KAP1 binding to KZFPs is key to allow this targeting mode to become a transcriptionally repressive one (Figure 8). In fact, KAP1 recruits in turn an H3K9 trimethyltransferase, SETDB1, which transforms KZFP-bound sequences into a heterochromatic environment, refractory to the transcription machinery (Figure 8) (Rowe et al., 2010). To consolidate this chromatin compaction, other factors are also recruited, such as Heterochromatin Protein 1 (HP1), tightens the DNA around nucleosomes and excludes the RNA polymerase (Figure 8). The  $\alpha$ -thalassemia/mental retardation syndrome X-linked protein (ATRX) is also recruited to these regions to enhance the formation of compact chromatin at and in proximity to ERVs (Figure 8), and in particular at IAP elements (Sadic et al., 2015). ATRX interacts notably with death-domain associated protein (DAXX) itself involved in H3.3 loading (Lewis et al., 2010; Tang et al., 2004). H3.3 is a dynamic H3 variant that can be deposited outside of replication and is associated to both transcriptional activation and repression. Renewal of H3 by its variant H3.3 participates in IAP repression (Figure 8) (Elsässer et al., 2015). Loss of H3.3 at KAP1-bound regions is minimal in absence of ATRX but more important in absence of DAXX, while the presence of H3.3 also enhances recruitment of these two proteins. H3.3 variants together with ATRX-DAXX participate in the maintenance of H3K9me3 at ERVs (Figure 8) (Elsässer et al., 2015). Importantly, these partners also play a role in the maintenance of heterochromatin at

satellite repeats. Microorchidia family protein 3 (MORC3) was also recently involved in heterochromatin formation at KAP1 bound regions (Figure 8). Sumoylation-dependent interaction of MORC3 with DAXX proved to be necessary for H3.3 incorporation (Groh et al., 2021). This was surprising considering the binding of MORC3 to H3K4me3-enriched regions, indicating a rather open chromatin context (Desai et al., 2021). ATRX-DAXX and H3.3 are enriched at IAPs and other ERVs but not at LINE-1 elements, despite LINE-1 being also enriched in H3K9me3, DNA methylation and KAP1.

The nucleosome remodeling and deacetylation (NuRD) complex can also be recruited to KAP1-bound regions, as well as DNA methyltransferases (Figure 8) (DNMTs) (Quenneville et al., 2012). Notably, KAP1 also targets Human Immunodeficiency Virus (HIV) and allows the deacetylation of its Integrase DNA sequence by HDAC1, limiting its integration into the genome (Allouch et al., 2011). KAP1 has a variety of interactors, including SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1 (SMARCAD1). Deletion of SMARCAD1 increases IAP expression, concomitantly with decreased KAP1 binding and H3K9me3 (Sachs et al., 2019). SMARCAD1 is a chromatin remodeler whose action could help KAP1 to access chromatin (Figure 8) or other factors to sumoylate KAP1, a modification that is indeed necessary for KAP1 to recruit all its partners on chromatin (Ivanov et al., 2007). KAP1 binding is not limited to TE sequences and therefore, SMARCAD1 might be implicated in transcriptional repression of other sequences. However, it is noteworthy that the absence of SMARCAD1 leads to only mild IAP over-expression (2-fold) compared to loss of KAP1 (more than 100-fold), and does not impact the expression of other classes of KAP1-bound TEs (like LINE-1) (Rowe et al., 2010; Sachs et al., 2019). Mono-ADP ribosylation of KAP1 by SIRTUIN 6 (SIRT6) also improves the recruitment of HP1 to repeats in somatic cells and therefore favors their repression (Figure 8) (Van Meter et al., 2014).

The characterization of the targets and pattern of expression of all the KZFPs in different species is a massive task that several research laboratories tried to tackle. More precisely, on retrotransposons, KZFPs often target regulatory sequences necessary for their expression, which are usually lowly divergent sequences, in

opposition to other DNA sequences that TEs could lose to escape restriction. The PBS sequence, which I already mentioned, is for instance a good target in the sense that it is essential for reverse transcription-dependent expansion. Among the studied KZFPs, ZFP932 and Gm15446 target IAP elements and their relatives (RLTR44, MMERVK10D3, MMERV10C) at a sequence located upstream of their 3'LTR and important for their reverse transcription (Ecco et al., 2016). Hence, one particular KZFP can bind sequences present on different TE families. From the TE side, one element can also be bound by different KZFPs, although often at well distinct regions. This is particularly the case for old TE sequences that have been present for generations in genomes, while newly inserted TE families tend to be recognized by one specific KZFP, or even none. For example, L1 Human Specific (L1HS) elements that emerged recently are not targeted by any KZFP, in contrast to older primate L1 elements such as L1PA4 (Imbeault et al., 2017). Old TEs, that might have even lost their transcriptional ability, are still under the control of KZFPs through KAP1 recruitment, which seems like a waste of resources for cells. This can be seen as a picture taken along the long stride of evolution: with more time, these useless targeting will disappear. Alternatively, it might be a clue that these targeted TEs, have now been domesticated and serve as a host gene regulatory network.

As exemplified here, the evolution of TEs and KZFPs is tightly linked, even though the cause and consequences are not easy to disentangle. At first sight, it seems easy to imagine that KZFPs are under positive selection to create new targeting amino-acid sequences and adapt to new invasions. In fact, the positive correlation between the number of KZFP genes and LTR-retrotransposon supports the idea of co-evolution between TE invasions and KZFP diversification (Kauzlaric et al., 2017). Moreover, KZFP genes organize mainly in clusters, and contrast with adjacent genes that have a high degree of identity, with only point mutations or one substitution between them. These observations fit the hypothesis along which KZFPs evolve fast to adapt to new LTR-retrotransposon invasions. However, there is evidence that some TEs, although bearing sequences recognized by KZFPs already present in the host, were successful in invading and spreading in genomes, which does not fit with the one-way co-evolution or arm-race hypothesis (Imbeault et al., 2017). Another striking evidence that arm race may not be the only driver of KZFP evolution is the fact that several KZFPs have roles outside of TE replication, purely essential to the host. As an example, ZFP57 binds in

majority to IAP elements but has no effect on their regulation, while being essential for genes subject to genomic imprinting (Quenneville et al., 2011; Shi, Strogantsev, et al., 2019).

Another important question on this matter is whether the H3K9me3 heterochromatic pattern is transmitted to the next cell generation or kept during differentiation. The role of this pathway is mainly studied in ESCs, where silencing of TEs by this system is particularly important due to the absence of piRNAs and/or to low DNA methylation levels. However, KZFPs have various expression patterns and some of them happen to be expressed in somatic cells. Although previous studies have shown that ZFP809 deletion in somatic cells has no effect on TE expression if H3K9me3 is already established in ESCs (G. Wolf et al., 2015), this conclusion seems to be dependent on each KZFP. In fact, ZFP932 deletion in differentiated cells, in culture or *in vivo*, leads to increased expression of some TEs, associated with a loss of H3K9me3 but maintenance of DNA methylation (Ecco et al., 2016). Beyond KZFP targeting, loss of SETDB1, responsible for H3K9me3 deposition at KAP1-targeted sequences, does not impact TE repression in mouse embryonic fibroblasts but greatly in ESCs (Matsui et al., 2010).

The fact that a transcription factor can bind to multiple copies of TEs supports the hypothesis along which TEs could act as regulatory factors of host gene networks. A third of KZFP binding gene promoters are rather ancient and evolutionary conserved regions (Imbeault et al., 2017). Moreover, they tend to have lost KAP1 binding and to be rather linked to active transcription. KZFP binding could have been originally targeted to TE sequences that can no longer be recognized as such, due to accumulation of mutations. Therefore, we would be in presence of a potential gene regulatory network system led by TE presence originally and maintained to now only for the host interest. Of note, KZFPs use only a portion of their zinc fingers to recognize their targets, while the other part could be used to bind other molecules, like RNAs or proteins, widening the regulatory spectrum of these factors.

Repression of TEs by KZFPs is a powerful system: efficient when established, tunable during development, and recyclable for other purposes through evolution. This is another striking example of the ambivalent impact of TEs on host genome and how

their presence drives innovation of systems that will finally serve the host interest, independently of TEs.

## 2. Targeting RNA to control TE expression

### a) The HUSH complex

When KZFPs start targeting TE sequences, these have been integrated in the genome for quite some time and were already able to spread. Although versatile, it is a slowly implemented repression system. To avoid wide spreading and too much damage done to host cells before KZFP “activation”, there is a need for a faster system that acts immediately at new retrotransposon insertions, which is what the Human Silencing Hub (HUSH) partly does. This protein complex was found, at first, to act against retroviral invasions, such as HIV, magnifying the close relationship between new retrotransposon infiltration and retroviruses and the speed at which HUSH counteracts these foreign RNA sequences (L. Liu et al., 2011). As such, HUSH is often seen as an innate immune system.

HUSH is a heterotrimer complex composed of Transcription Activation Suppressor (TASOR or FAM208A), M-phase phosphoprotein 8 (MPP8) and Periphilin (PPHLN1) (Figure 8). The exact role of TASOR is not elucidated yet but some hint comes from the domains it contains: an inactive Poly ADP-Ribosylation Polymerase (PARP) domain, a Spen Paralog and Ortholog C-terminal (SPOC) domain, and two disordered domains previously called DomI and DomII (Douse et al., 2020). DomI is essential to the interaction between TASOR and MPP8, interaction stabilized by the SPOC domain, a type of domain generally involved in protein-protein interactions. The role of the inactive PARP domain remains unclear but it is essential for HUSH-mediated TE repression (Douse et al., 2020). It is noteworthy that deletion of TASOR and therefore abrogation of HUSH function in mice leads to embryonic lethality before gastrulation (Harten et al., 2014). MPP8 can bind to H3K9me3 (or H3 mimicry on other proteins) through its chromodomain and this interaction, together with SETDB1 recruitment, helps propagating H3K9me3 (Figure 8) (Tchasovnikarova et al., 2015). However, this heterochromatin construction or spreading is not the main repression mechanism used by HUSH: deletion of the MPP8 chromodomain does not abolish

HUSH-mediated TE repression. The last protein of the complex, Periphilin, binds to RNA (Figure 8) and only to RNA targeted by HUSH, despite not encoding any recognizable RNA binding motif (Seczynska et al., 2022). Despite not knowing the precise way Periphilin binds to its targets, this protein is thought to be the targeting unit of the complex, allowing to recruit HUSH to its targets, even in absence of H3K9me3 (Tchasovnikarova et al., 2015).

Effector proteins can associate to this tripartite core, such as Microorchidia family CW-type zing finger 2 (MORC2), SETDB1 or Activating transcription factor 7-interacting protein 1 (ATF7IP). Because MORC2 is an ATP-dependent chromatin remodeler that compacts chromatin and SETDB1 is an H3K9 methyltransferase, it was originally thought that this pathway was acting through heterochromatinization, similarly to KZFPs (Tchasovnikarova et al., 2015). However, binding to chromatin is not essential to HUSH silencing, nor the presence of SETDB1 or MORC2. In fact, HUSH only act on targets that are transcribed and most of them are present in “leaky” heterochromatin allowing for some transcriptional activity. Therefore, HUSH targets are mainly young retrotransposons and LINE-1 particularly, although a small number of SINEs are also targeted (Robbez-Masson et al., 2018). It is unclear whether HUSH requires the presence of RNA or RNA Pol II binding, although there is growing evidence that an RNA species may be involved. Indeed, HUSH was showed to be able to silence any and only intronless sequences entering in cells, without the requirement to be integrated in the genome (Seczynska et al., 2022). Notably this recognition of intronless sequences was independent of splicing events and the common denominator of these sequences is proposed to be their speed of transcription, and more precisely, their elongation rate. In fact, long exons (> 1kb) such as full-length LINE-1 elements, tend to have a slow elongation rate, which is accentuated when H3K9me3 is present on these sequences. Another hypothesis would be the recruitment of specific proteins by introns that would prevent binding of Periphilin to its targets. In any case, as intronless sequences trigger the activation of an innate immune system, they can be considered as pathogen-associated molecular patterns (PAMP).

One last piece of evidence to place HUSH as a transcriptional controller, is the physical interaction between MPP8 and ZCCHC8, a member of the Nuclear Exosome Targeting complex (NEXT), which targets non poly-adenylated RNAs for degradation



(Figure 8) (Garland et al., 2022). Although HUSH also represses poly-adenylated transcripts, the NEXT RNA decay is needed for proper HUSH mediated TE-silencing. The role of HUSH role in TE RNA regulation might be completed by other partnering complexes. This fits with the observation that HUSH also plays a role in defense against non-integrated viruses, which are not poly-adenylated and are not subjected to nuclear control.

That being said, ZCCHC8-depleted ESCs are still viable, while MPP8 loss is fatal to them (Garland et al., 2022; Müller et al., 2021), showing the larger importance of the HUSH complex in this particular context. Of note, MPP8 has also been shown to interact with the *de novo* DNA methyltransferase 3A (DNMT3A) (Chang et al., 2011), opening the possibility that DNA methylation is targeted to HUSH-bound loci. Furthermore, HUSH activity is essential in cells with low DNA methylation levels but dispensable for proliferation of more differentiated cells, while being expressed in all cell types (Müller et al., 2021). This can lead to the hypothesis that HUSH acts both on a transcriptional chromatin-based, established and conserved regulation mode, as well as a post-transcriptional RNA degradation-based one (Figure 8). A similar RNA induced transcriptional silencing (RITS) has been described in yeast, which has strong functional similarities with HUSH (Douse et al., 2020).

HUSH can therefore be considered as a system primarily recognizing foreign RNA species obtained from intronless DNA sequences or reverse-transcription, and addressing them for degradation as well as for transcriptional repression through chromatin-based mechanisms similar to the ones used by the KZFP machinery. It particularly targets young L1 elements that are prone to premature transcriptional termination and is essential in lowly methylated cells, and during early embryonic development.

## b) piRNAs

This defense system against TEs has been found in almost all animals. Most of the time, its expression is restricted to germ cells although not all germ cells express it and it has been found in some somatic cells, being there involved in defense against viral infections (Morazzani et al., 2012). Just like HUSH, this system is based on RNA-

directed mediated silencing of TEs, but it differs because it can only target known elements and not newly invasive TE species, acting then more like an adaptive immune system than an innate one.

piRNA stands for PIWI-interacting RNA, PIWI (P-element Induced Wimpy testis) proteins referring to a specific type of Argonautes, a group of proteins that bind small RNAs. Indeed, piRNAs are considered a clade of small RNAs that are ranging from 24 to 35 nucleotides in length (Vagin et al., 2006). They possess other characteristic features allowing them to be recognized once they are processed, such as a 2'-O-methylation at their 3' end established by HENMT1 in mouse (Kirino & Mourelatos, 2007; Lim et al., 2015; Ohara et al., 2007). Despite their peculiar signature, what drives a specific RNA to be processed into a piRNA is not really known although it has been noticed that the promoters of piRNA loci have a low CG content and high DNA methylation levels in somatic cells. piRNA precursor loci were discovered to lie next to each other in big clusters in *Drosophila* (Brennecke et al., 2007). However, in mice, they still bear the name of "clusters" while a high number of piRNAs are generated from small loci with a unique RNA target (Aravin et al., 2006). Their biogenesis varies among species but some common grounds can be found. Two main pathways participate in the production of piRNAs that can be called primary or secondary. Briefly, in mice, primary piRNAs are obtained through the cleavage of long RNA sequences by endonucleolytic activities and then adding the necessary modifications creating piRNAs with a bias for U as a first nucleotide. Secondary piRNAs are obtained when primary piRNAs bind to their complementary RNA targets, generating piRNAs with a bias of A as 10<sup>th</sup> nucleotides (Aravin et al., 2006). This mode of production is often referred to as "ping-pong" amplification.

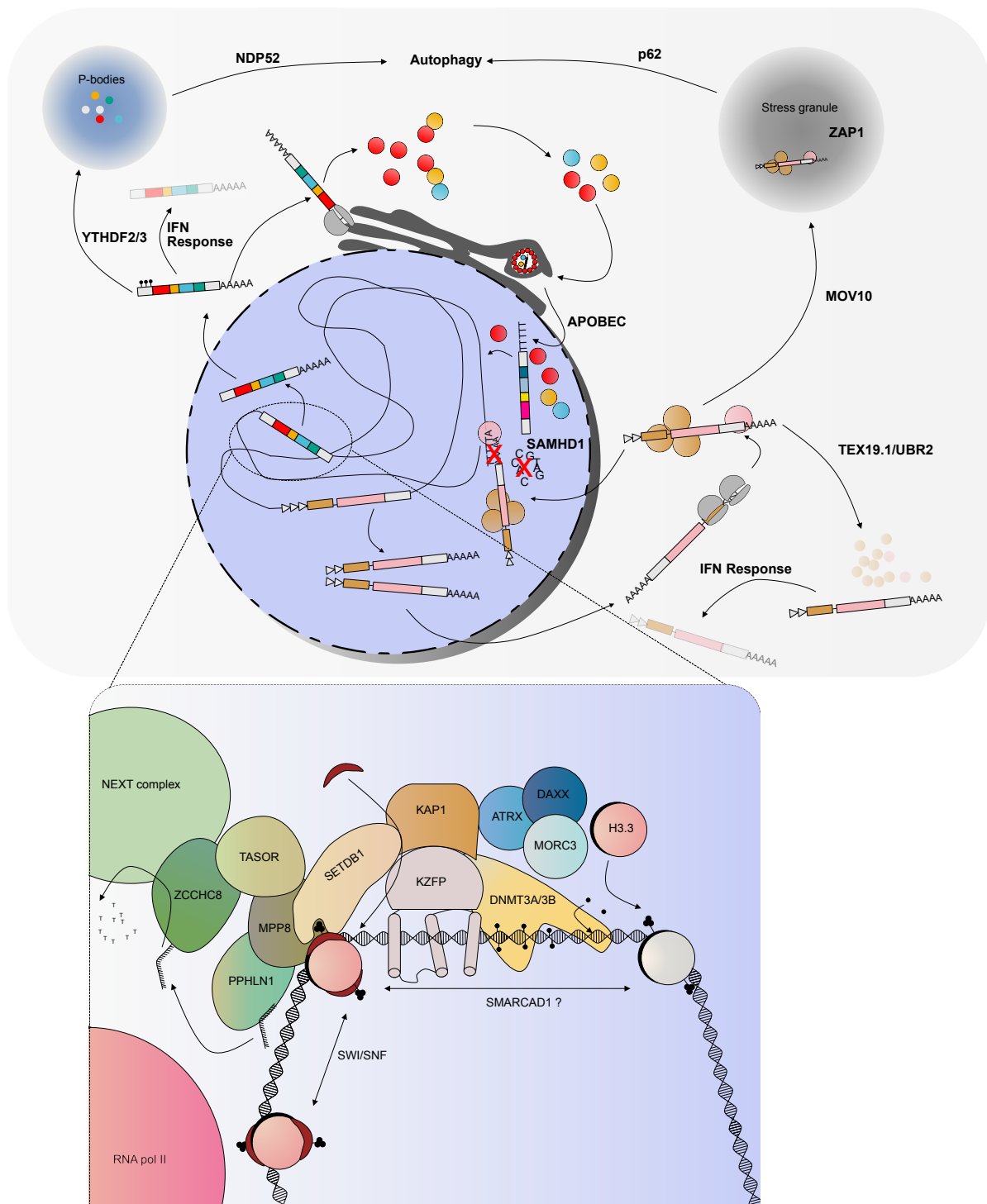
This mode of production participates in the post-transcriptional suppression of undesired RNA species such as TEs, against which piRNAs are targeted. That is the first layer of TE control achieved by the piRNA system. However, piRNAs are the only mammalian small RNA species that can also provide transcriptional control, by promoting DNA methylation to TE sequences (Aravin et al., 2007). In fact, failure to produce piRNAs, through genetic deletion of the proteins essential to their biogenesis, impairs the establishment of DNA methylation at TE sequences in fetal male germ cells in mice, and is linked to a meiotic arrest at the pachytene stage (Chuma et al., 2006;

Shoji et al., 2009; F. Yang et al., 2008). This lack of DNA methylation leads to the expression of otherwise repressed TEs, notably LINE-1 elements for which this transcriptional control is both sufficient and necessary before the onset of meiosis compared to the post-transcriptional control. Whether the meiotic arrest is due solely to the DNA methylation defect, lack of piRNAs or TE expression is still a standing question in the field.

In mice, piRNAs are only produced in male germ cells, which is in contrast with *Drosophila* for example, but also with a closely related hamster species in which piRNAs are also produced in female germ cells (H. Zhang et al., 2021). Therefore, all studies done on piRNAs in mouse are done on male germ cells. In any case, essential proteins to piRNA biogenesis are usually not expressed in ESCs or in somatic cells in mammals. The mouse genome encodes three PIWI proteins: MILI (PIWIL1), MIWI (PIWIL2) and MIWI2 (PIWIL4), which have non-redundant roles in either biogenesis of piRNAs or their downstream effects. Distinct sets of piRNA are expressed across germ line differentiation (Aravin et al., 2007). A first “batch” or set is produced at pre-natal stage and are also referred to as pre-pachytene piRNAs. These are the ones that target TEs and are associated to MILI and MIWI2. The second set is produced in post-natal life, at the pachytene stage, hence their name as pachytene piRNAs. These mainly target intergenic or unique host sequences in genic regions and associate with MILI and MIWI. Pachytene piRNAs are not produced through “ping-pong” amplification and their biological role is to this date quite unknown. Regarding fetal pre-pachytene piRNAs, the primary piRNAs are associated to MILI and are amplified as secondary piRNAs through the “ping pong” amplification cycle. These secondary piRNAs are loaded onto MIWI2 and translocated in nuclei, where they serve to target DNA methylation deposition at TE promoters by sequence complementarity (Shoji et al., 2009; Watanabe et al., 2018). However, MIWI2 does not directly bind to DNA or to DNA methylation machinery. Therefore, several studies have attempted to provide insights into the physical connection between MIWI2 and the DNA methylation machinery. A recent paper succeeded into uncovering the interacting partners of MIWI2 by IP-MS in male germ cells, a quite challenging task, and reported a direct link between a new factor SPOC domain containing 1 (SPOCD1), MIWI2 and the DNA methylation proteins DNMT3A and DNMT3L (Zoch et al., 2020). However, DNMT3C is the DNA methyltransferase responsible for methylating young and active TEs in the male germ

line in mice ([Barau et al., 2016](#)) and it could not be caught by this method, although direct interaction between SPOCD1 and DNMT3C was demonstrated in an ectopic over-expression system. Whether this is a methodology issue or the sign that other factors are key to the piRNA-directed DNA methylation to TEs is still a question.

While the function on TEs of the piRNA pathway is conserved among animals, the piRNA targets, proteins involved in the biogenesis and even PIWI proteins are evolving rapidly. In fact, between, closely related species, the piRNA targets are rarely conserved, which may seem normal under the assumption that they are adapted to the TE load or invasion history ([Ozata et al., 2019](#)). However, the underlying reasons to have proteins that are also under positive selection and rapidly evolving does not fit with this hypothesis and might be linked to other unknown functions of piRNAs. The fact that piRNA proteins have been lost and gained again during evolution also leads the beliefs that their primary function might not be to repress TEs at all.



**Figure 8 | IAP and LINE-1 elements are controlled in ESCs.**

Top panel with grey background represents a summary of all non-targeted pathways acting in the control of IAP and LINE-1; bottom panel with a purple background focuses on targeted control at both types of elements in the nucleus.

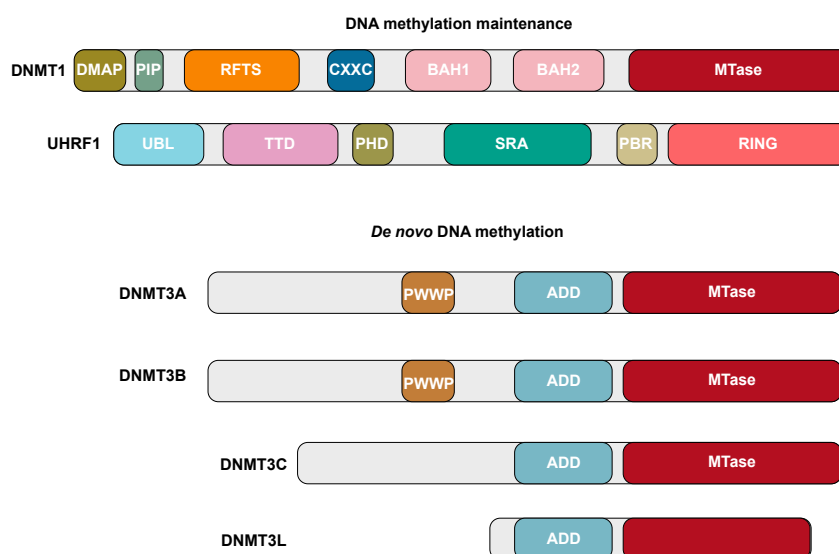
## C. Chromatin determinants of TE targeted control

### 1. DNA methylation

DNA methylation in mammals consists in the addition of a methyl group onto cytosines that are part of CpG di-nucleotides. This DNA modification is not present in all animals, including some common laboratory models (*C.elegans*, *D.melanogaster*, *S.cerevisiae* and *S. pombe*) but it has important roles in transcriptional regulation in mammals. Some CpA dinucleotides are also methylated, at a lower rate and more specifically in the neural lineage (Lister et al., 2013). However, the biological role of CpA methylation is out of the scope of this manuscript and I will focus my interest on CpG methylation that I will term DNA methylation.

In mammalian genomes, CpGs are present at lower rate compared to expectations and are not evenly distributed across the genome. As a whole, the genomic DNA contains a low CpG amount, except for enrichment at specific regions called CpG islands (CGIs), mostly located at gene promoters, including some young ERVs. About 2/3 of CGIs are un- or lowly methylated, mainly at housekeeping or developmental gene promoters (Bird et al., 1985; Gardiner-Garden & Frommer, 1987). Among the methylated CGIs are the germline gene promoters that heavily methylated in somatic cells but remain unmethylated in germ cells and their precursors (Borgel et al., 2010). DNA methylation at promoters is linked to transcriptional repression but requires distinct kind of effectors to have this effect (Stadler et al., 2011; Weber et al., 2007). DNA methylation also defines regions in germ cells that will be mono-allelically expressed depending on the parental origin in somatic cells, referred to as genomic imprints. However, DNA methylation is not sufficient to promote transcriptional silencing, which can be exemplified by its presence in the body of expressed genes. However, this modification changes the transcription factors (TFs) that can bind to promoters: some methylation-sensitive TFs are repulsed or attracted by its presence (Yin et al., 2017). Chromatin modifiers or remodelers can also be attracted or repulsed by DNA methylation or even DNA methyltransferases (DNMTs), the enzymes responsible for its deposition.

DNA methylation is a dynamic modification whose patterning is highly cell-specific. Moreover, as for histone modifications, DNA replication dilutes its presence leaving the newly synthesized strand modification-free. Therefore, different enzymes are needed to establish correct patterning, to maintain this patterning through cell divisions and remove it when cellular fates change. Two enzymes, DNMT3A and DNMT3B, are responsible for the establishment of DNA methylation patterns with different cell specificities and are termed *de novo* DNA methyltransferases (Okano et al., 1999). A third *de novo* enzyme, DNMT3C, is specific to rodents, and its only role is to methylate young TEs promoters in fetal male germ cells, as demonstrated in mice (Barau et al., 2016). DNMT3 enzymes require the presence of a catalytically inactive partner, DNMT3L, for their activity in germ cells (Barau et al., 2016; Bourc'his et al., 2001; Bourc'his & Bestor, 2004). Once established, DNA methylation patterns are maintained by DNMT1 through DNA replication in all dividing cell types (Bostick et al., 2007). Different claims have been made in regards to a potential *de novo* activity of DNMT1 in oocytes (Y. Li et al., 2018) and in ESCs (Haggerty et al., 2021), but no mechanistic model for this activity has been established yet. In contrast, the targeting of DNMT1 to hemi-methylated DNA generated during replication is quite detailed, involving a specific cofactor Ubiquitin-like with PHD and Ring Finger domain 1 (UHRF1) (Xie & Qian, 2018). Finally, DNA methylation can be passively lost with cell divisions or can be actively removed by Ten Eleven Translocation enzymes (TET1, TET2 and TET3).



**Figure 9 | DNA methylation enzymes and co-factors.**

Enzymes and co-factors involved in DNA methylation deposition and maintenance in ESCs, highlighting their functional domains.

### a) DNA methylation maintenance happens through collaboration between DNMT1 and UHRF1

Besides a catalytically active methyltransferase domain, DNMT1 possesses a Zinc Finger Domain of a CXXC type (Figure 9), allowing direct binding to unmethylated CpG. To avoid activity at any CpG, DNMT1 is auto-inhibited due to intra-molecular interactions between different domains (Song et al., 2011). Alleviation of this inhibition involves an interaction between the Replication Foci Targeted Sequence (RFTS) of DNMT1 and H3K14 and H3K18 mono-ubiquitination (Ub) (T. Li et al., 2018). DNMT1 can also interact with Proliferating Cell Nuclear Antigen (PCNA) through its PCNA interacting domain (PIP) (Figure 9), again associating the protein to replication forks (Chuang et al., 1997). Interestingly, PCNA associated Factor 15 (PAF15) also bound during replication contains a H3-like domain that can be ubiquitinated on two close lysine residues and have the same effect on DNMT1 activity than H3K14Ub and H3K18Ub (Nishiyama et al., 2020). Without these ubiquitination, DNMT1 activity is blocked. By operating these modifications on both H3 and PAF15, UHRF1 is essential to DNMT1 activity. Besides these necessary interactions, DNMT1 itself undergoes a number of post-translationally modifications (acetylation, methylation, phosphorylation), which modulates its activity.

UHRF1 activity and presence is hindered by several inhibitory processes, as well as targeting for degradation by the proteasome. Deposition of ubiquitin is ensured through its Really Interesting New Gene (RING) domain, and regulated by or in concert with its Ubiquitin like-domain (UBL) (Figure 9) (Citterio et al., 2004). UBL recruits the ubiquitin-modified E2 enzyme needed to transfer the ubiquitin modification. A Tudor Tandem-like domain (TTD) of UHRF1 (Figure 9) allows binding to H3K9me3, and this interaction to remove UHRF1 auto-inhibitory intra-molecular interactions (Rothbart et al., 2013). This would mean that DNA methylation maintenance can only happen in a chromatin context harboring H3K9me3, which is not what is observed. The explanation for this discrepancy can be explained by the presence of a H3K9me3 mimicry domain in Ligase I (LIG1), also necessary for DNA replication and therefore present on spot (Ferry et al., 2017). Therefore, UHRF1 activity can be ensured by binding to LIG1 or to H3K9me3, which is important in regards to replication timing. In fact, both the speed



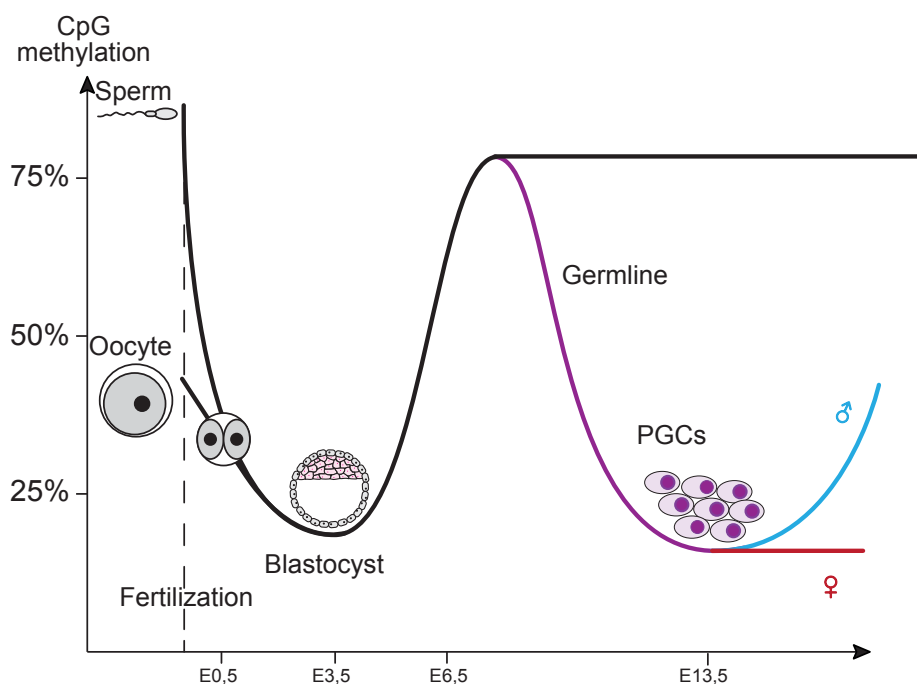
and the number of replication forks are too high for DNMT1 to be able to methylate every hemi-methylated CpGs in the replication bubble and for UHRF1 to target all of them at the same time as well. DNA methylation maintenance will continue for a prolonged timing compared to DNA replication and most of the late methylated regions are enriched for H3K9me3, releasing the need for LIG1 presence. More than allowing DNMT1 activity, UHRF1 is also essential for DNMT1 targeting on hemi-methylated DNA. Indeed, DNMT1 can bind to any unmethylated CpG, while UHRF1 can bind with higher affinity to hemi-methylated CpG compared to other CpG contexts through its SET and RING-finger associated domain (SRA) (Figure 9) (Arita et al., 2008; Hashimoto et al., 2008).

#### b) Defining DNA methylation patterning with de novo DNMTs

DNMT3A and DNMT3B, but also the DNMT3L co-factor, possess an ATRX-DNMT3-DNMT3L (ADD) domain (Figure 9) that is physically attracted by H3K4 devoid of methylation and repulsed by its presence, with forces of repulsion increasing with the number of methyl groups (Ooi et al., 2007; Y. Zhang et al., 2010). In accordance with this role, H3K4 methylation and especially H3K4me3 that mainly covers active (CGIs included) promoters, is mutually exclusive with DNA methylation. DNMT3A and DNMT3B enzymes also harbor a PWWP domain (Pro-Trp-Trp-Pro) (Figure 9) that favors binding to H3K36 methylation (Dhayalan et al., 2010; Sun et al., 2005). This interaction helps to understand the enrichment of DNA methylation on active gene bodies that are covered by this specific nucleosome modification, as a result of interaction between the RNA Pol II and H3K36 methyltransferases. However, the role of DNA methylation on gene bodies is rather discussed between participating in the discrimination of exons and introns or the silencing of transcriptional promoting regions embedded in gene bodies (Greenberg & Bourc'his, 2019; Teissandier & Bourc'his, 2017). DNMT3C lacks this PWWP domain (Figure 9), which might be important for its very specific function on TE promoters (Barau et al., 2016). Mechanistically, DNMT3s can methylate any given cytosines but only methylation at CpGs—due to their symmetric configuration—can be maintained by DNMT1.

DNMT3A and DNMT3B activity and importance is cell type- and genomic context-dependent. DNMT3A is responsible for the large majority of DNA methylation

activity in germ cells, while DNMT3B activity is focused on satellite repeats (Dura et al., 2022; Kaneda et al., 2010). In fact, primordial germ cells (PGCs) that will give rise to differentiated germ cells see their inherited pattern of DNA methylation from differentiating ES cells globally erased actively by TET action and passively through replication and division (Figure 10). The repertoire of expressed genes in germ cells is very peculiar compared to somatic lineages and erasure may be necessary to allow expression of germ cell genes that are otherwise repressed by DNA methylation (Borgel et al., 2010). Therefore, a new patterning of DNA methylation has to be acquired through *de novo* DNA methylation (Monk et al., 1987). After fertilization, the embryo undergoes another combination of active and passive DNA demethylation to allow again a change of cell identity, with the lowest amount of methylation reached in the pluripotent cells from the inner-cell mass of the blastocyst (ICM) (Figure 10), from which ESCs are derived in culture. *In vivo*, PGCs are not derived from this specific stage but from ESCs in the course of their differentiation toward different somatic lineages and therefore with a higher DNA methylation level. However, this transient re-acquirement of DNA methylation is not necessary to allow PGC specification from ESCs *in vitro* (Schulz et al., 2023). During this wave of DNA methylation in ESCs, both DNMT3A and DNMT3B are needed, with DNMT3B playing a larger role in germline gene methylation (Borgel et al., 2010).



**Figure 10 | DNA methylation is dynamic during early development**  
Dynamics of DNA methylation deposition and erasure in early mouse development.

During development, pluripotent embryonic cells can harbor different DNA methylation levels as they go from naïve, to formative to primed pluripotency, which can be recapitulated in a Petri dish by culturing ESCs in different culture conditions (Morgani et al., 2017). The serum-based medium contains Leukemia Inhibitory Factor (LIF) and Fetal Bovine Serum (FBS), necessary to allow continuous renewal without differentiation, respectively (see Methods for the exact composition). FBS provides the necessary inhibitor of Mitogen-Activated protein kinase/Extracellular signal-regulated kinase (MAPK/ERK), Bone Morphogenetic proteins (BMP). ESCs grown in serum/LIF form a transcriptionally heterogeneous population and have a highly methylated genome, close to the one of somatic differentiated cells (Ying et al., 2008). Yet, they are still pluripotent, meaning are able to give rise to the three embryonic lineages, endoderm, mesoderm and ectoderm. Alternatively, ESCs can be cultured in serum-free conditions using two inhibitors of differentiation paths, known as “2i medium” (see Methods). One inhibitor (CHIR99021 or CT99021, shortened as CT) blocks the Wnt pathway through inhibition of its main effector glycogen synthase kinase 3 (GSK3), while the second (PD0032) occupies the same function than BMP and inhibits the MAPK/ERK pathway. Compared to serum conditions, 2i-cultured ESCs form a homogenous population and present a methylation level close to what can be observed in the ICM *in vivo*, meaning around 15-20% of CpG methylation genome-wide.

Despite the wave of DNA methylation loss observed during pre-implantation development, some specific regions retain DNA methylation, among which imprinted regions, satellite DNA and young families of TEs, especially IAP elements (Walter et al., 2016; L. Wang et al., 2014). The exact mechanism that prevents DNA methylation erasure at these sequences is quite unclear and might differ between imprinted regions and TEs, also similar actors could play a role in these processes. In fact, two KZFPs, ZFP57 also helped by ZFP445, bind most imprinted regions (harboring a methylated TGCCGC sequence) in early stages of development (Quenneville et al., 2011; Takahashi et al., 2019). As described previously, KZFPs are bound by KAP1, bringing DNA methylation and H3K9me3 machineries on spot and allowing a proper transmission of the methylation pattern, despite a global erasure (Figure 8). However, ZFP57 binding to methylated TE sequences is not necessary to maintain this DNA modification (Shi, Strogantsev, et al., 2019). Other KZFPs, such as ZFP932 or

Gm15446, could be involved in this process (Ecco et al., 2016). Importantly, DNA methylation is not alone in this process as H3K9me3 is also present at these young TEs and the prevalence between these two repressive modifications is highly context-dependent and will be discussed in following paragraph, after discussing H3K9me3 on its own.

## 2. H3K9me3

Trimethylation of the lysine 9 of the histone 3 (H3K9me3) is a modification associated with transcriptional repression and helps to form compact chromatin regions. These regions are usually called heterochromatin, which can exist in a constitutive form (present in every cell types) or in a facultative form (depending on tissue type). Despite being historically associated to constitutive heterochromatin, there is more and more evidence that H3K9me3 is involved in the regulation of differentiation and tissue-specific gene expression, hence in facultative heterochromatin (Nicetto et al., 2019). Addition of methyl groups is done one after another and trimethylation is the highest methylation state at H3K9. In mammals, three groups of two homologous enzymes are responsible for catalyzing methylation at H3K9 with preferences and redundancy (Montavon et al., 2021). They all possess a catalytic SET domain responsible for their activity. Of note, in *Drosophila*, these groups of enzymes are also represented albeit with only one member per group (Setdb1, Su(var)3-9, and G9a). Importantly, H3K9me3 and their writers have been mainly studied in ESCs. Although the reported roles and relative importance might change with cell types, I will focus on findings in ESCs.

### a) SETDB1 and SETDB2

Previously, I already mentioned SET domain bifurcated 1 (SETDB1)—also named ERG-associated protein with SET domain (ESET)—which has an understudied homolog SETDB2, on which not much can be said apart from its necessity to establish complete H3K9me3 patterning in mammals, and its recent involvement in regulation of inflammatory pathways (Kimball et al., 2019; L. Yang et al., 2002). The catalytic domain of SETDB1/SETDB2 is modified compared to the canonical SET domain, with a 300bp-insertion that preserves its activity. Most of H3K9me1 is deposited by

SETDB1, which can also add further methyl groups. Besides the SET domain, SETDB1 contains a Methyl-CpG Binding domain (MBD), a Tudor domain allowing binding to bivalent H3K9me1 and H3K4ac chromatin, and interacts with Activating Transcription Factor 7-interacting protein (ATF7IP), Heterochromatin protein 1 (HP1), and KAP1 (Kimball et al., 2019; Schultz et al., 2002; Timms et al., 2016). These interactions determine its targeting and activity. The specific use of MBD is rather unclear. While it seems able to bind methylated DNA, this domain is also present in *Drosophila* and *C.elegans*, two organisms devoid of DNA methylation. The ability to bind H3K9me1/H3K4ac domains is quite interesting, although this combination of modifications has not yet been studied in depth. Relationship between ATF7IP and SETDB1 has been more extensively studied and this protein can be considered as a co-factor of SETDB1 (Timms et al., 2016; Tsusaka et al., 2019). Indeed, their interaction promotes mono-ubiquitination at the K867 residue of SETDB1, increasing its processivity, stability and localization in the nucleus. Regulation of SETDB1 activity can be easily regulated by its localization because it harbors both nuclear localization and nuclear export signals. Hiding one or the other signal by binding of partners can be quite dynamic and may be used in response to environmental stresses (Beyer et al., 2016; Delaney et al., 2019).

Notably, the interactions of SETDB1 with HP1 or ATF7IP are mutually exclusive and lead to different outcomes. Association to HP1 is more likely to target SETDB1 to pericentromeric chromatin and to promote the deposition of one methyl group on naked H3K9 (Loyola et al., 2009; Tsusaka et al., 2019), while interaction with ATF7IP directs SETDB1 toward other genomic targets, notably TEs, and facilitates enhanced activity toward trimethylation of H3K9 (Tsusaka et al., 2019). Two comments emerge from these observations: although SETDB1 activity has been mainly reported at young ERVs (Figure 11), it is also involved in formation of constitutive heterochromatin at telomeric repeats, as well as pericentromeric and centromeric repeats by providing a substrate used by other H3K9 methyltransferases (like SUV39H1/H2). Therefore, depending on the abundance or availability of ATF7IP and HP1, which can vary during the cell cycle, the main activity of SETDB1 will change, with biological different impacts. That being said, at least 40% of SETDB1 binding sites are found in close proximity (at max. 100bp) from ERVs, which can be explained by its interaction with KAP1 that is itself targeted to these elements by KZFPs (Karimi et al., 2011). Moreover, ATF7IP is

part of the HUSH complex that targets not only ERVs but also other types of TEs, like LINEs, and links even more SETDB1 to the control of TEs. Although being involved in H3K9me3 deposition at other repeats, SETDB1 is still the main enzyme responsible for H3K9me3 at ERVs and notably, IAPs. Source of gene mis-regulation in SETDB1 mutants is even though to emerge from expression of chimeric transcripts using ERV LTRs as an available promoter or TSS (Karimi et al., 2011).

## b) SUV39H1 and SUV39H2

Discovered primarily in *Drosophila*, Suppressor or Variegation (3-9) homologue (SUV39H1) and its homolog SUV39H2 in mammals are able to add mono-, di- and tri-methylation at H3K9 through their chromodomain. Although being able to deposit methylation on naked H3K9 *in vitro*, this activity is not essential *in vivo*, contrary to its involvement in di and tri-methylation at centromeres and telomeres that are part of the constitutive heterochromatin (Figure 11) (Lehnertz et al., 2003). These binding abilities suggest that SUV39H1/H2 target regions that are already H3K9 methylated, further enhancing the level of methylation and its propagation to neighboring regions. These already methylated regions could have been targeted by other H3K9 methyltransferases, such as SETDB1. Indeed, a collaboration between SETDB1 and SUV39H1/H2 has been reported at telomeric regions (García-Cao et al., 2003; Gauchier et al., 2019). In addition, binding to HP1 also promotes spreading of H3K9 methylation (Bannister et al., 2001; Lachner et al., 2001).

The SUV39H1/H2 chromodomain can also bind RNA, which is used to enhance its targeting to expressed satellite repeats as well as telomeric repeats. Indeed, a specific telomere repeat-containing RNA (TERRA) emerges from telomeres and is bound by SUV39H1/H2 helping its recruitment (Porro et al., 2014). Moreover, these RNA-mediated recruitment of SUV39 enzymes have also been described in the embryo at the 2-cell stage, with a long non-coding RNA (lncRNA), *oct4/4*, directed to *Oct4* and leading to deposition of H3K9 at *Oct4* promoter (Scarola et al., 2015). However, the specificity of the chromodomain toward RNA is not known. Other RNAs could be recognized by SUV39H1/H2. Their identification would allow to understand new modes of targeting, for example at rDNA loci, and the relative dependencies toward these two homologs for H3K9 patterning.

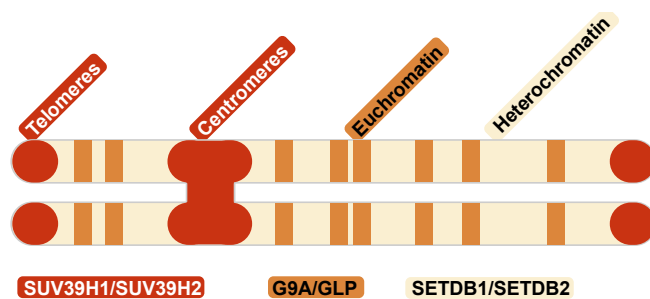
Importantly, SUV39H1/H2 are also essential for H3K9me3 deposition at several ERV subfamilies (MMETn, ETnERV, IAPEz, IAPEy) and LINE-1 elements, and other genomic regions could depend on SUV39H1/H2 to establish their correct chromatin features (Bulut-Karslioglu et al., 2014). Redundancy between SETDB1 and SUV39H1/H2 at TEs explain why the loss of SUV39H1 does not have much impact on TE repression (Bulut-Karslioglu et al., 2014), compared to satellite and centromeric repeats. Indeed, complete knock-out of SUV39H1 and SUV39H2 in mice revealed no developmental delay or defect but a high post-natal lethality linked to genome instability (Peters et al., 2001).

Finally, the activity of SUV39H1/H2 is regulated through ubiquitination. SIRT1, an H3 and H4 deacetylase, is also responsible for the de-acetylation of a specific lysine in the SET domain of SUV39H1, which prevents the polyubiquitination of another lysine residue and further protein degradation by the proteasome (Bosch-Presegué et al., 2011). SIRT6, another deacetylase that also interacts with KAP1, acts on other sites of the SUV39H1 SET domain, allowing mono-ubiquitination by SKP2 (a E3 Ubiquitin ligase like UHRF1), which promotes SUV39H1 activity (Santos-Barriopedro et al., 2018).

### c) G9a/EHMT2 and GLP/EHMT1

Euchromatin histone-lysine methyltransferases 1 and 2 (EHMT1/EHMT2), also named GLP and G9a, respectively, are the last enzymes needed to form the entire patterning of H3K9me3 in the mouse genome. Contrary to SUV39H1/H2, G9a and GLP can add one to two methyl groups on naked H3K9, but not 3. These two enzymes might be responsible for the formation of large H3K9me2 domains observed in ESCs, but could also provide substrate for the other aforementioned H3K9 methyltransferases. These enzymes are also able to add methyl groups to H3K27 residues, although a low number of methylated H3K27 results from their actions (Mozzetta et al., 2014; Wu et al., 2010). As their name suggests, these enzymes are responsible for the methylation at H3K9 in euchromatin (Figure 11), meaning chromatin regions that are poorly condensed and could be transcribed. Therefore, their action is more linked to gene regulation and is essential during development with their deletion causing an early

embryonic lethality (at embryonic day 9.5, E9.5) (Tachibana et al., 2002). It is noteworthy that they also target some TE families, including MERVL and IAP, which may be linked to the observed embryonic lethality, at least to some extent. In the mouse testis, both G9a and SETDB1 are needed to reach full H3K9me3 levels at IAPs (S. Liu et al., 2014). This raises the question as to whether this is due to an interdependency with the third methyl group being only deposited by SETDB1, or whether it reflects interindividual differences, with different copies of IAP being targeted by the different H3K9 methyltransferases. In fact, loss of G9a or GLP in ESCs leads to a decrease in H3K9me2 and DNA methylation at ERVK and ERVL, while H3K9me3 levels are unchanged at these sequences and no up-regulation is observed (Dong et al., 2008).



**Figure 11 | H3K9me3 methyltransferases have preferences over genomic compartments**  
Schematic chromosome displaying preferences of the three groups of H3K9me3 methyltransferases for different genomic compartments.

### 3. Relative dependency over H3K9me3 or DNA methylation for TE repression

In ESCs, DNA methylation and H3K9me3 are both present at a variety of sequences, including ERVs, and in particular IAP elements. The role played by each modification in the transcriptional control of these elements has been under close scrutiny even though some questions are still unanswered. I will try to summarize the knowledge acquired about the interplay between DNA methylation and H3K9 methylation at IAP elements into a comprehensive yet precise overview.

What has been observed in several studies and has not been contradicted so far is the independence of enzymes responsible for H3K9 methylation toward DNA methylation deposition. In other words, in absence of DNA methylation, H3K9 methylation is mainly unchanged. This has been observed in the context of ESCs



deficient for DNMT1, DNMT3A and DNMT3B, single-handedly or in combination (Dong et al., 2008; Karimi et al., 2011; Lehnertz et al., 2003). However, several pieces of evidence point toward a dependency of DNA methylation toward the H3K9me3 mark or H3K9 methyltransferases. Loss of G9A/GLP leads to loss of DNA methylation at IAP elements, LINE-1 elements and pericentromeric repeats, while DNA methylation enzymes are still expressed (Dong et al., 2008). Reintroduction of catalytically inactive mutants leads to a partial rescue of DNA methylation with a better recruitment at IAP elements of DNMT3A than in *G9a*-KO, but the rescue is not complete compared to WT. SUV39H1/H2 seems also necessary for proper localization of DNMT3B at pericentromeric repeats, which is linked to a decrease in DNA methylation in absence of SUV39H1/H2 (Lehnertz et al., 2003).

The dependency of DNA methylation toward H3K9 methylation or its methyltransferases might be explained by two observations. As said previously, G9A/GLP are able to methylate the H3 mimic domain present on LIG1, enhancing recruitment of DNMT1 at replication sites (Ferry et al., 2017). In its absence, DNA methylation maintenance efficiency is lower, although not completely abrogated. We can also assume that catalytically dead G9A/GLP are not able to methylate LIG1 and therefore, not able to recruit efficiently UHRF1-DNMT1. However, they could attract the *de novo* DNMTs through an interaction with HP1, to which G9A/GLP and DNMT3A/DNMT3B bind (Dong et al., 2008). Recruitment of DNMT3B at SUV39H1/H2 target sites is also thought to happen through HP1a and HP1b acting as scaffold proteins (Lehnertz et al., 2003). Despite these assumptions are in agreement with current knowledge and could explain these observations, they have not been properly tested and should be taken with caution.

Importantly, SUV39H1/H2 deletion only leads to minor reduction of H3K9me3 at IAPs with no impact on their expression, while *G9a/Glp* mutant cells present a decrease in H3K9me2 and H3K9me3 at IAPs, accompanied by increased transcription of these elements (Dong et al., 2008; Lehnertz et al., 2003). As both mutants suffer from a DNA methylation defect, the necessity for one or the other on IAP repression is unclear. While IAP reactivation is not drastic in *Dnmt1*-KO or *Dnmt1-Dnmt3a-Dnmt3b* (tKO) mutant ESCs, loss of SETDB1—the H3K9 methyltransferase that is more specifically targeted toward TEs—leads to a major IAP de-repression, in the context of

only a minor DNA methylation loss and an expected H3K9me3 loss. These observations argue that, in ESCs, H3K9me3 is more important for IAP repression than DNA methylation (Karimi et al., 2011; Matsui et al., 2010). It is noteworthy that DNA methylation behaves differently toward SETDB1-dependent H3K9me3 compared to G9A/GLP and SUV39H1/H2-dependent H3K9 methylation, being more dependent on the latter. This suggests a dependency toward these specific enzymes compared to a dependency toward the H3K9 methylation mark *per se*.

Some studies brought some light onto these unexpected observations. First, while long-term loss of DNMT1 in ESCs leads to only a mild IAP transcriptional reactivation, acute loss of DNMT1 leads to major IAP up-regulation, transiently peaking at 4 days after deletion before expression declines again (Sharif et al., 2016). This observation was also reported during the transition of ESCs from high (serum-based) to low DNA methylation (2i-based) culture conditions (Walter et al., 2016). Second, UHRF1 depletion, which leads to the same DNA methylation loss than measured in *Dnmt1*-KO, triggers lower re-expression of IAP elements compared to *Dnmt1*-KO ESCs (Sharif et al., 2016). It seems then clear that the absence of DNA methylation is not what triggers IAP up-regulation right after DNMT1 loss, or at least is not sufficient to explain it. The hypothesis argued by the authors is indeed not linked to DNA methylation *per se* but to the recruitment of UHRF1 to accumulated hemi-methylated CpGs that might block the recruitment of H3K9 methyltransferases, causing then IAP de-repression. After a few division cycles, the abundance of hemi-methylated CpGs decreases and the recruitment of UHRF1 with it, leaving space for H3K9me3 deposition. This last study confirms the major importance of H3K9me3 at IAP elements for their control in ESCs compared to DNA methylation, although there could be some synergistic effects of DNA methylation loss in H3K9me3-depleted sequences.

Importantly, this relative importance between H3K9me3 and DNA methylation for IAP control is cell type-dependent. In mouse embryonic fibroblasts (MEFs), H3K9me3 is still important for IAP control but is maintained after SETDB1 depletion if correct methylation of H3K9 is ensured in precursor cells, namely ESCs (Matsui et al., 2010). In endoderm lineages, H3K9me3 is not sufficient to secure correct IAP repression, whereas DNA methylation is essential (Z. Wang et al., 2022).

## IV. New factors involved in TE regulation

Despite all the scientific knowledge related to TE regulation accumulated in the past 20 years, several outstanding questions are still pending. Among them, why some TE sequences or some cell types are more dependent over H3K9me3-dependent control, while others more largely rely on DNA methylation? Both H3K9me3 and DNA methylation correlate with transcriptional repression but variations of their relative importance on TE repression might reflect an association with repressive effectors. Moreover, during the first wave of DNA demethylation occurring in early embryogenesis, only some TE sequences maintain high levels of DNA methylation and mechanisms at stake are still unclear. Maintenance of H3K9me3 at these sequences seems to matter more than DNA methylation in ESCs (Karimi et al., 2011). However, acute loss of DNA methylation in ESC cultures leads first to increased TE transcription, including IAP elements or LINE-1, followed by a secondary DNA methylation-independent repression phase (Walter et al., 2016). Three types of chromatin responses were observed upon adaptation to the loss of DNA methylation, reflecting different TE types: a switch from H3K9me3 to H3K27me3 for LINE-1, further enrichment of H3K27me3 for MERVL and further enrichment of H3K9me3 for IAP. How these chromatin responses are initiated is unknown but this might play a role in regulating TEs *in vivo* during demethylation waves. Overall, a lack of knowledge of the exact mechanisms leading to repression associated with heterochromatin formation at TEs needs to be challenged. The development of CRISPR-Cas9 targeting technologies and silencing RNAs during the past decades have allowed the development of various screening methods for factors involved in controlling specific TE families or to identify factors present in their vicinity.

### A. Identification of TE regulators through loss-of-function screens

Several kinds of strategies have been applied to find new TE regulators and four of these previous studies will be more thoroughly discussed here: (A) B. X. Yang et al., 2015b, (B) N. Liu et al., 2018, (C) Groh et al., 2021, and (D) Chelmicki et al., 2021 to which I actively participated during my master internship and that initiated my PhD

project. These screens differ in some key parameters: the targeting technique applied to deplete a library of factors, the class and family of TE used as reporter sequence, the measurement of TE expression, and the cell lines used to perform the experiment. Each of this parameter plays an important part in obtained results. In A, loss-of-function was achieved using a small interfering RNA (siRNA) strategy, implying transient degradation of transcripts emanating from targeted genes, while the three others used CRISPR-Cas9 deletion, meaning a permanent genetic deletion. Both approaches have their own obstacles. In the siRNA approach, some siRNAs may be not efficient or some portion of siRNA-targeted transcripts might still be present. In the CRISPR-Cas9 approach, sgRNAs might also have variable recognition efficiencies (which is mitigated by having multiple sgRNAs per gene target), and the DNA cut might not impact the expression of the RNA or the function of produced protein. Moreover, the use of the CRISPR-Cas9 system implies further genetic manipulation of cells for stable Cas9 expression. Nevertheless, these different loss-of-function systems might finally bring little disparities compared to other parameters.

In fact, the four screens all relied on reactivation of a transgenic TE reporter and the design of this reporter has an important impact on the screen readout. The rationale behind reporter assays lies in the hypothesis that the DNA sequence of a specific TE element, family or class is sufficient to acquire similar regulation compared to endogenous elements. In A, a full sequence of one MuLV element linked to GFP was used as a reporter, and GFP expression was measured by fluorescence intensity. Then in this experience, the potential of the whole element to trigger repression is tested. In C, only a small part of the *gag* sequence of IAPEz elements was used, the short heterochromatin inducing sequence (SHIN), formerly defined as sufficient to induce heterochromatin formation at IAP elements ([Sadic et al., 2015](#)). This sequence is identical in hundreds of different IAPEz elements across the mouse genome. In D, we used a consensus sequence of 694 bp starting from the transcription start site of IAPEz elements to 60 nucleotides inside the *gag* sequence. This fragment does not encompass the full SHIN sequence but it does include the 5'LTR, the 5'UTR and the PBS ([Figure 4](#)). Strategies used in C and D are otherwise strikingly similar, while the top obtained candidates are mainly different between the two, hence the hypothesis that the repression potential of SHIN and LTR + PBS might be very different.

In B, LINE-1 elements were at the center of the screen, and moreover, the final retotransposition potential was assayed using a neomycin-based retrotransposition cassette. Compared to previous strategies, this one allows to detect factors that might be involved in repressing LINE-1 beyond transcriptional control. However, like C and D, selection of factors whose deletion enhance TE expression was achieved through antibiotic resistance selection, meaning a survival screen: only sgRNAs in cells that survive the selection can be enriched. It implies that deletion of factors that are essential to cell survival (like KAP1 or SETDB1 in mESCs) will never be enriched in this kind of screens, despite their possible importance in TE regulation.

C and D were directly performed in mouse ESCs, while in A, screening was performed in mouse Embryonic Carcinoma cells (ECCs), and then applied to mESCs. In contrast, B was performed in human chromatic leukaemia K562 cells. All screens performed in ESCs were done using 'serum'-based medium, which favors cellular state heterogeneity and high genomic methylation levels. Therefore, factors needed to compensate a lack of DNA methylation will not be particularly enriched in this condition.

Despite methodological divergence, each of these techniques was successful in identifying pathways involved in TE regulation, notably in heterochromatin formation at TE sequences. In A, Chromatin assembly factor 1 subunit 1 (CAF1) and small ubiquitin related modifier 2 (SUMO2) were found as top enriched factors. In fact, CAF1 is associated with enhanced nucleosome renewal at MuLV that decreases H3K4me3 or H3 acetylation, while SUMO2 participates via deposition of SUMO on KAP1, promoting its recruitment to KZFP-bound DNA. Action of CAF1 and SUMO2 is then not limited to MuLV but extends to other KAP1-bound ERVs. In the screen developed in C, formation of heterochromatin at IAPs by DAXX-ATRAX-mediated incorporation of H3.3 was demonstrated, with necessity of the Microorchidia family protein 3 (MORC3). Notably, MORC3 was also part of the top enriched factors in D. Targeted to LINE-1, the study led in B revealed the implication of the HUSH complex in TE repression, a finding that was confirmed by another publication ([Robbez-Masson et al., 2018](#)). In D, we focused on four proteins (METTL3, METTL14, Wilms Tumor Associated Protein WTAP and Zinc finger containing CCCH motif 13 ZC3H13) involved in RNA modification by m<sup>6</sup>A, a pathway not linked to TE repression before but which was confirmed by several independent studies ([J. Liu et al., 2021](#); [Xu et al., 2021](#)). However, other proteins were

ranked higher in our screen, like SPINDLIN1 (SPIN1) and AI846148, also named C11ORF84 or SPINDOC. Resolving the role of these proteins in TE control was at the basis of my PhD project.

## B. SPIN1

SPIN1 is expressed from the *Spindlin1* gene and was first described as a highly abundant maternal transcript and protein transmitted after fertilization (Oh et al., 1997). In oocytes, it was found to associate to the meiotic spindle, hence its name, and its association was shown to be both cell cycle- and phosphorylation-dependent. Later on, two transcriptional *Spin1* isoforms were described, one expressed solely in oocytes, and another ubiquitous form, also expressed in the testis (K.-M. Zhang et al., 2008). *Spin1* is the most expressed and most studied member of a gene family also comprising *Spin2a*, *Spin2b*, *Spin3* and *Spin4* in humans, and *Spin2c* and *Spin4* in the mouse, and other *Spin*-related pseudogenes. SPIN proteins contain functional domains called either Spin/Ssty—for their presence in SPIN family and Spermiogenesis Specific Transcript on the Y (SSTY) proteins—or Tudor domains for their ability to read chromatin. Moreover, few years after its discovery in oocytes, *Spin1* over-expression was associated with deleterious phenotypes such as multinucleation and enhanced cell proliferation in human cells (Gao et al., 2005). Due to these multiple aspects, research on this protein is articulated in distinct axes: *i*) its importance in germ cells, sex-determination and early development, *ii*) the genomic regions it binds and the functional impact of this binding and *iii*) its multi-faceted role in cancer development and how it can be targeted. Combination of these different axes could support an implication in TE regulation and explains our choice for this candidate from our screen. Moreover, as far as I could search, a role for SPIN1 in ESCs had never been reported, further highlighting the originality and potential impact of my study. Finally, most of the knowledge on SPIN1 comes from studies in human cancer cells or *in vitro*, whose results may not apply to ESCs.

## 1. SPIN1 reads several chromatin modifications

SPIN1 is a rather small protein (around 29 kDa) encoding three distinct Tudor (or Spin/Ssty) domains with different binding abilities (Figure 12). Despite their structure being very similar, each of these domains varies in sequence with only 25 to 50% homology between them (N. Yang et al., 2012). The second domain has strong affinity for H3K4me3 binding *in vitro*, in a structure that is quite different from other known H3K4me3 readers (N. Yang et al., 2012). This observation has been strengthened by ChIP-seq experiments in different cell types, showing a quasi-systematic presence of H3K4me3-modified nucleosomes at sites of SPIN1 binding (Franz et al., 2015; Greschik et al., 2017). H3K4me3 is strongly associated with promoters of actively transcribed genes, hence a proposed role of SPIN1 as a transcriptional activator. Consistent with a role in transcription, SPIN1 localizes to nucleoli in somatic cells, and it was also shown to bind active ribosomal DNA loci (rDNA), enriched in H3K4me3, leading to transcription of ribosomal RNA (rRNA) (W. Wang et al., 2011). Conversely, knock-down of *Spin1* led in this study to decreased rDNA transcription. However, despite its presence at numerous gene promoters, loss of SPIN1 was associated with down-regulation (or even up-regulation) of very few genes (Franz et al., 2015; Greschik et al., 2017). The biological importance of SPIN1 as a transcription activator is therefore not completely understood.

The first domain of SPIN1 has specific binding abilities toward asymmetrically di-methylated arginine 8 of H3 (H3R8me2a) (Figure 12A, 12C), which is notably much better fitted for asymmetric versus symmetric H3R8me2 (Su et al., 2014). Importantly, using different kinds of modified histone peptides *in vitro*, the primary binding of SPIN1 is toward H3K4me3 and the presence of H3R8me2a led to an even stronger interaction. *In vivo*, H3K4me3 and H3R8me2a modifications are found at Wnt target genes (Su et al., 2014). The Wnt pathway is very conserved among animals and is particularly involved in self-renewal and survival of adult stem cells, as well as differentiation and overall body plan definition. In agreement with its role in self-renewal, the Wnt pathway is ectopically activated in several types of cancer, in association with increased proliferation and migration. The association between Wnt signaling and SPIN1 together with observation that SPIN1 overexpression leads to abnormal mitosis and decreased apoptosis opened a new avenue of research: the

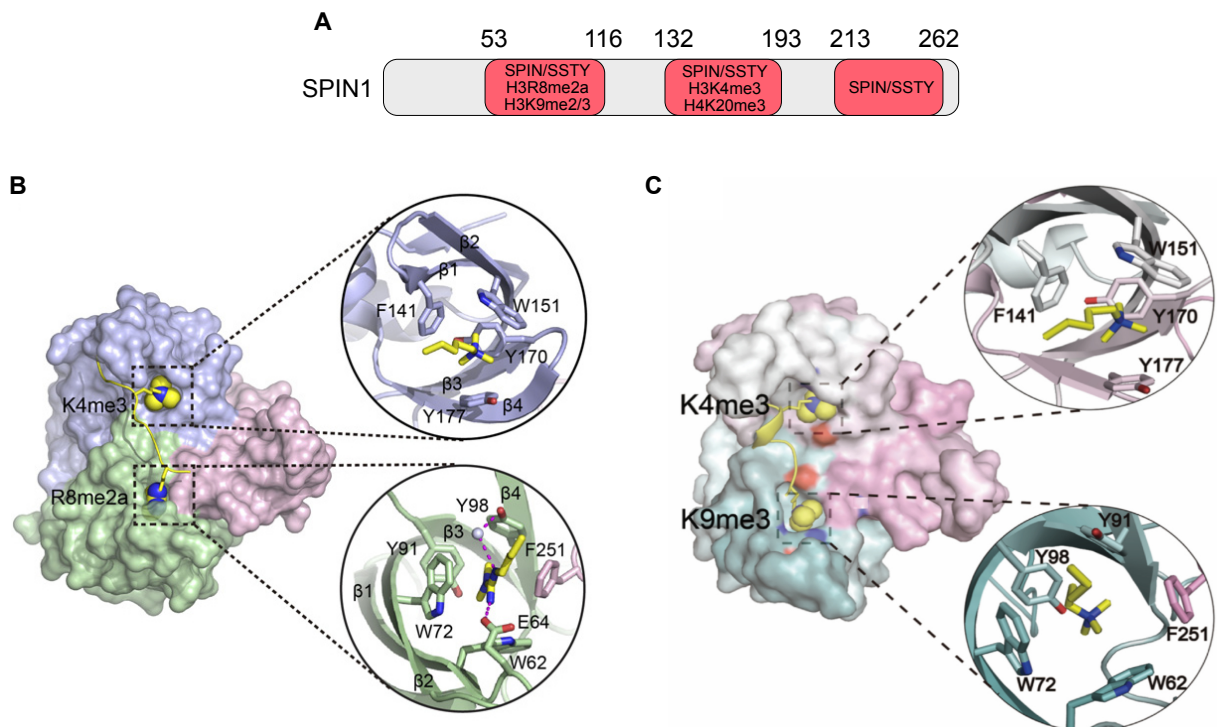
involvement of SPIN1 in cancer development and its use as a potential drug target for treatment.

Interestingly, this first domain can also bind a distinct chromatin mark: H3K9me3 (Figure 12A, 12B) (Zhao et al., 2020). In fact, H3K4me3 and H3K9me3 are usually found in non-overlapping regions, as these marks are linked to transcriptional activation and repression, respectively. The existence of domains harboring both marks is still a question. They have been found co-occurring at imprinted regions but on distinct alleles (Mikkelsen et al., 2007). Co-marking by H3K4-H3K9me3 has although been described in extra-embryonic stem cells (Rugg-Gunn et al., 2010) and in mesenchymal stem cells (Matsumura et al., 2015), in association with a small number of lowly expressed genes. Moreover, the existence of an H3K4me3-dependent H3K9 demethylase, JMJD2A, points toward a possible transient co-occurrence of these marks at promoters (Pedersen et al., 2016). However, their co-occurrence was deduced from chromatin immunoprecipitation experiments on cell populations, and could rather reflect DNA loci that present heterogenous chromatin states across cells. The estimated low number of these potential H3K4-K9me3 dual regions and their likely transient state may prevent them from being confirmed through conventional sequential ChIP-seq. However, with the development of new genome-wide bimodal chromatin profiling methods at the single cell level (Bartosovic & Castelo-Branco, 2022), this question could be answered. Nonetheless, it is interesting that *in vitro*, as for H3R8me2a, SPIN1 binding to H3K4me3 is further enhanced by H3K9me3 (Figure 12B) (Zhao et al., 2020). This would make SPIN1 the only H3K4me3 reader for which the presence of H3K9me3 does not lower but rather strengthen its interaction with H3K4me3. While the strong association of SPIN1 with open and transcriptionally active chromatin is difficult to reconcile with a potential role on TE repression (which our screen suggests), its ability to bind putative non-canonical H3K4-K9me3 domains is more appealing and could involve a role in reducing TEs to complete silencing.

Finally, a peptide array study of Tudor domain-containing proteins revealed the ability for the second domain of SPIN1 to also interact with H4K20me3 (Figure 12A) (R. Lu & Wang, 2014), a repressive mark (Martens et al., 2005). This interaction has not been studied in details yet, probably because binding to H3K4me3 uses the same domain and the strength of this interaction is much higher. The panel of marks that



SPIN1 can read or bind (H3K4me3, H3K9me3, H3R8me2a and H4K20me3, [Figure 12A](#)) together with its localization to the nucleolus (notably in MEFs) strengthen its link to rRNA expression or rDNA regulation. In fact, all of these chromatin modifications have been found to exist at rDNA loci, which are either active (H3K4me3 and H3R8me2a) or inactive (H3K9me3 and H4K20me3) and can switch between these two states. Decreased *Spin1* expression has been shown to dampen rRNA levels in different human cell lines (HEK293T ([Y. Du et al., 2021](#); [X. Zhang et al., 2018](#)); HeLa ([W. Wang et al., 2011](#))), which could be secondary to lower proliferation in *Spin1*-KD experiments. Intriguingly, a large number of factors involved in rDNA loci regulation are also involved in TE regulation such as ATRX, DNMT3A, DNMT1, SUV39H1 and G9A ([Grummt & Längst, 2013](#)).



**Figure 12 | SPIN1 binds to distinct chromatin modifications.**

**A**, Schematic representation of SPIN1 functional domain and interaction affinities. Numbers indicate the amino-acids position. **B**, Extracted from [Su et al, 2014](#). SPIN1 is shown in surface mode with Tudor domains 1,2 and 3 colored in green, purple and pink respectively. R8me2a and K4me3 in yellow are shown in their binding pockets respectively in Tudor domain 1 and Tudor domain 2 as determined after crystallization with peptides. **C**, Extracted from [Zhao et al., 2020](#). SPIN1 is shown in surface mode with Tudor domains 1, 2 and 3 colored in green, gray and pink respectively. K9me3 and K4me3 in yellow are shown in their binding pockets respectively in Tudor domain 1 and Tudor domain 2 as determined after crystallization with peptides.

## 2. SPIN1 is involved in distinct biological processes

### a) SPIN1 is an essential protein

Complete *Spin1* knock-out (KO) mouse models die at 2 days after birth, with some reported *in utero* lethality (Chew et al., 2013; Greschik et al., 2017). In the first study, this phenotype was linked to its abundance in oocytes and to the role it could have in early embryogenesis as a maternally inherited transcript. Indeed, while the lack of SPIN1 does not impair oocyte growth and maturation, more than 50% of *Spin1*-KO oocytes (upon grafting of *Spin1*-KO ovaries in wild-type females) were not able to resume meiosis. The exact mechanism involving SPIN1 in meiosis completion has not been unveiled but was proposed to involve the regulation of specific maternal transcripts, through interaction with Serpin 1 mRNA binding protein (SERBP1), and/or localization to the meiotic spindle. In the second study, in 80% of the cases, conditional *Spin1* mutations in myogenic precursors led to the same early post-natal death linked to neuromuscular defects than constitutive KO. Surviving pups still presented limb defect with necrotic muscular fibers (Greschik et al., 2017). These two studies point towards the necessity of SPIN1 in germ cells and during embryogenesis. To finish on its implication in germ cells, SPIN1 was found to be expressed in the nucleus from the spermatocyte stage onward and to be localized to both the nucleus and flagellum of spermatozoa, although no involvement in male meiosis has been shown (K.-M. Zhang et al., 2008). However, the SSTY proteins, encoded by multi-copy genes (approx. 150) only present on the Y chromosome, contain homologous domains to SPIN1 and these are involved in correct spermatozoa morphology and fertility (Touré et al., 2004). The strong deleterious phenotypes in SPIN1-depleted animals or germ cells may obscure the exact mechanisms leading to this defect. Therefore, the role of SPIN1 in specific biological processes has mostly been investigated in cultured cells.

### b) SPIN1 is involved in viral defense

A stand-alone study has linked SPIN1 to defense against viral infections of Hepatitis B Virus (HBV) and Herpes Simplex Virus type 1 (HSV-1) (Ducroux et al., 2014). Both are DNA viruses whose DNA is circularized in the nucleus and is treated as host DNA (*i.e.* chromatin modified) and serve for expression of the viral proteins. SPIN1 was shown to interact with HBX, an essential factor of HBV transcription. It was also shown to

interact with the circularized DNA itself through its second Tudor domain. *Spin1* over-expression decreased transcription from the circularized DNA, while knock-down increased it. However, contradictory to other reports, the presence of SPIN1 on the circular DNA was anti-correlated to H3K4me3. Recruitment of SPIN1 to this viral integrated DNA must therefore be H3K4me3-independent and might even rely on direct interaction with the DNA itself. A further hypothesis would be that HBX is needed to counteract SPIN1 transcriptional blockade. This is the first (and only) example of SPIN1 being reported as a transcriptional down-regulator, and moreover towards foreign DNA. This particular role of SPIN1 is of interest to our hypothesis linking SPIN1 to the transcriptional repression of TEs. As mentioned before, several regulators of TEs are also involved in viral repression and it is reasonable to think that SPIN1 could be one of them. However, in this case, repression occurs through H3K4me3 binding, which TEs are not really enriched for, at least when they are repressed.

### c) SPIN1 is linked to cancer development

*Spin1* was found to be highly expressed in several cancer tissues compared to healthy ones, notably ovarian, breast, gastric, brain cancers as well as in liposarcoma and seminoma (X. Chen et al., 2016, 2018; Franz et al., 2015; Janecki et al., 2018; Lv et al., 2020; J.-X. Wang et al., 2012). Notably, high *Spin1* expression was linked to increased tumor aggressiveness, proliferation, migration, drug-resistance and poor survival outcome. One proposed explanation for these oncogenic traits is the potential role of SPIN1 as transcriptional activator of a p53 regulator named MDM2 (Fang et al., 2018; Lv et al., 2020). p53-dependent protection of genome stability is often (in more than 50%) compromised in cancer development. Although SPIN1 is not the only p53 regulator, this new p53-related function combined with the observation that *Spin1* over-expressing cells show multinucleation and higher proliferation makes it an interesting target in cancer research. Moreover, *Spin1* over-expression was linked to increased expression of Wnt target genes, a pathway notably involved in self-renewal and whose mis-regulation is also associated with cancer development (J.-X. Wang et al., 2012). Therefore, several efforts have been made in designing drugs or chemical probes that would target SPIN1 specifically, and especially its second Tudor domain (H3K4me3 binding) that has been linked to its transcriptional activator role in multiple

situations (Bae, Viviano, et al., 2017; Fagan et al., 2019; Wagner et al., 2016; Y. Xiong et al., 2019).

So far, one molecule (A366, (Wagner et al., 2016))—initially targeted toward the H3K9 dimethyltransferase G9a—has shown abilities to bind the second Tudor domain of SPIN1. A second one (EML405, (Bae, Viviano, et al., 2017)) has abilities to bind the first and the second Tudor domains. That said, EML405 selectivity toward SPIN1 is higher compared to A366, which has other targets, likely because of this interaction with several domains instead of one. In a third approach, another potent binder of G9a was used as a basis for targeting again the second Tudor domain and then modified to improve its selectivity toward SPIN1 over G9A (Y. Xiong et al., 2019). The rationale was to create an inhibitor that would target this domain instead of two, a lower molecular weight compound having usually less interactions and thus less secondary unwanted effects. From this work, a last compound, MS31, emerged with a high selectivity toward the second Tudor domain of SPIN1, low to no cell toxicity and potent inhibition of SPIN1-H3K4me3 interaction. In this chase for SPIN1 inhibitors, SPIN1 interaction partners were also interrogated, leading to the discovery of a protein named SPINDOC for its ability to bind SPIN1 (SPIN1 DOCKing protein) (Figure 12A, Figure 13A,B,C), and which was originally described as a potential endogenously produced inhibitor of SPIN1 (Bae, Gao, et al., 2017).

### c. SPINDOC

SPINDOC, also named with the aliases C11orf84 or AI846148, is a protein of approximately 41 kDa which has only been studied since the discovery of its interaction with SPIN1 in 2017. From the predicted structure of SPINDOC by AlphaFold and homologous domains in other proteins, only a putative Zinc Finger domain can be inferred. However, SPINDOC has not been shown to bind chromatin or DNA on its own.

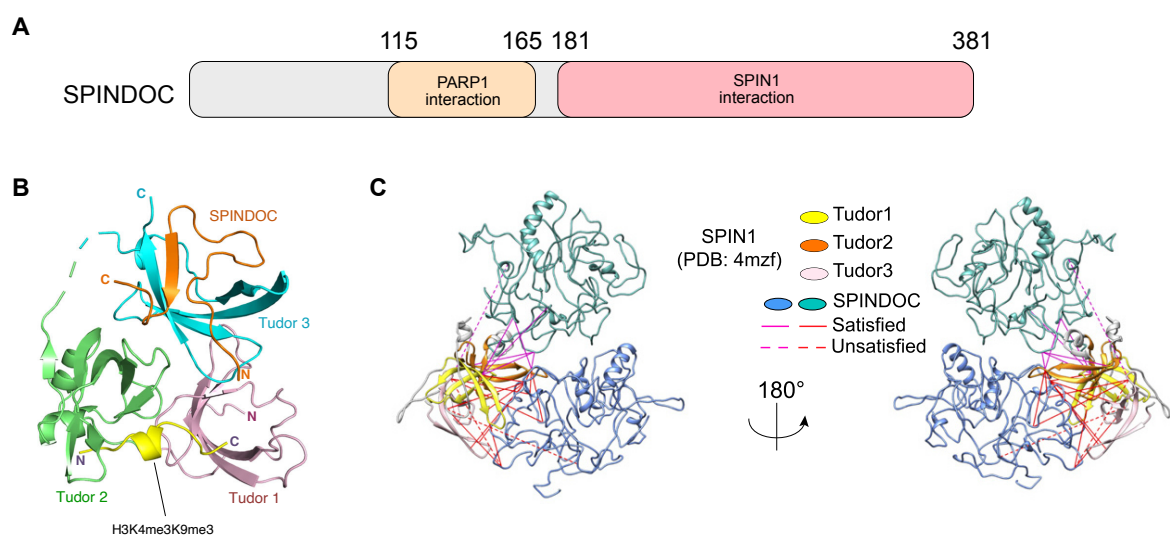
## 1. SPINDOC interacts with SPIN1

Immunoprecipitation followed by mass spectrometry (IP-MS) experiments using over-expressed and tagged SPIN1 in HEK293T cells revealed SPINDOC as the most abundant SPIN1 interacting partner (Bae, Gao, et al., 2017). Immunoprecipitation of tagged SPIN1 and SPINDOC also led to the precipitation of the other partners in another transgenic approach in HeLa cells (Devi et al., 2019). Interaction between SPIN1 and SPINDOC has then been studied on its own and revealed to be strong (resisting high salt concentration) (X. Liu et al., 2020). Of note, previous experiments trying to identify SPIN1 interaction partners in MEFs failed (W. Wang et al., 2011).

Assays were rapidly developed to understand whether interactions between SPIN1 and SPINDOC either disrupted or promoted SPIN1 binding to chromatin. First, over-expression of SPINDOC in HEK293T cells changed the localization of exogenous SPIN1 from chromatin towards to the nuclear soluble fraction, interpreted as SPINDOC antagonizing SPIN1 binding to chromatin (Bae, Gao, et al., 2017). However, this observation has been contradicted by several papers, first with the fact that SPINDOC over-expression increased endogenous SPIN1 localization to chromatin (Devi et al., 2019), and then by crystallizing the structure of SPIN1-SPINDOC bound to H3K4me3-K9me3 peptides (Figure 13B) (Y. Du et al., 2021). In this context, the first Tudor domain of SPIN1 is bound to K9me3 residue of the peptide, the second domain to K4me3 and SPINDOC to the third Tudor domain, whose deletion impaired SPINDOC interaction in previous studies (Figure 13B) (Devi et al., 2019). Despite being stable, this form of interaction might not be the one preferred in cells, depending on the relative expression levels of SPIN1 and SPINDOC and in presence of a more complex chromatin context. However, another crystal structure revealed that SPIN1 can interact alone with H3K4me3-K9me3 peptides (Figure 12B) (Zhao et al., 2020). Then, a new technique aiming at unraveling the interaction of two partners in cells with tagged-exogenous expression was used on SPIN1 and SPINDOC and revealed that *i*) their interaction does not solely depend on the third Tudor domain of SPIN1 and can involve other SPIN1 Tudor domains (Figure 13C), and *ii*) two SPINDOC molecules might be bound to SPIN1, forming a hetero-trimer complex (Figure 13C) (X. Liu et al., 2020). In presence of SPINDOC, SPIN1 would therefore not be able to bind chromatin anymore,

which would rather confirm the initial observation whereby over-expression of SPINDOC could re-locate SPIN1 away from chromatin.

Interaction between H3K4me3 (or chromatin in general) and SPIN1 might also be necessary for other observed biological roles of SPIN1, such as the transcriptional activation of Wnt target genes and ribosomal DNA. For a first study, over-expressing *Spindoc* had the same overall effect than using a SPIN1 inhibitor (EML631) (Wagner et al., 2016) on over-expressed *Spin1*: a reduction of Wnt signaling and a reduction of rRNA production, as measured by reverse transcription followed by quantitative PCR (RT-qPCR) (Bae, Gao, et al., 2017). While the repressive role of SPINDOC over-expression on Wnt signaling was confirmed in another publication (Devi et al., 2019), this role for rRNA expression is disputed, notably by showing that decreased level of SPINDOC led to less expression of rRNAs, similar to *Spin1*-KD (Y. Du et al., 2021). In conclusion, while the interaction between SPIN1 and SPINDOC is certain, the exact outcome of this interaction on SPIN1 function needs to be further dissected in physiological contexts.



**Figure 13 | SPINDOC interacts with SPIN1.**

**A**, Schematic representation of SPIN1 functional domain and interaction affinities. Numbers indicate the amino-acids position. **B**, Extracted from Du et al., 2021. Cartoon representation of SPINDOC/SPIN1 complex bound to H3K4me3K9me3 peptides as determined after crystallization of the complex. SPIN1 Tudor domains 1, 2 and 3 are represented in pink, green and blue respectively, a fragment of SPINDOC in orange and H3K4me3K9me3 peptide in yellow. **C**, Extracted from Liu et al., 2020. Model of interaction between SPIN1 and two SPINDOC molecules satisfying the maximum numbers of intermolecular cross-links. SPIN1 is engaging interaction with SPINDOC all its Tudor domains.

## 2. SPINDOC interacts with PARP1

Recent SPINDOC-centered IP-MS have revealed a new interactor of SPINDOC beside SPIN1, the Poly ADP-Ribose polymerase 1 protein (PARP1) (Figure 13A) (F. Yang et al., 2021). The PARP family of proteins, which contains also PARP2 and PARP3, functions in repair of single and double strand DNA breaks. This role might be achieved by three distinct means: *i*) by recruiting DNA repair factors such as XRCC1 to the site of damage through poly-ADP ribosylation of itself and neighboring proteins on site, *ii*) by serving as a scaffold for DNA repair factors, or *iii*) by acting as a transcriptional activator for genes involved in repair pathways (Kamaletdinova et al., 2019). SPINDOC could therefore interact with either SPIN1 or PARP1, forming apparently two distinct complexes. SPINDOC seems to bind to the DNA binding domain of PARP, therefore creating a competition between SPINDOC and DNA for PARP binding. Furthermore, according to this same study, DNA damage would enhance *Spindoc* expression through the action of the transcription factor KLF4. Deletion of endogenous *Spindoc* in HEK293T and Hela cells led to decreased DNA damage repair: less deposition of ADP-ribose, less clearing of PARP (which is necessary to access the DNA damage), and less expression of genes involved in DNA repair regulated in part by PARP. This observation was further investigated *in vivo* in a *Spindoc*-KO mouse, which showed increased sensitivity toward DNA damage in irradiated thymus.

## 3. SPINDOC is highly expressed in the testis

Few more observations were made on *Spindoc*-KO mice, which were coincidentally developed by two independent groups. Although these mice are viable and fertile, a smaller size was observed in KO compared to WT littermates (F. Yang et al., 2021). The tissue expressing *Spindoc* at the highest level is the testis and loss of SPINDOC was associated with morphological defects of spermatozoa (Jiang et al., 2021). A mild-decreased fertility was also reported in *Spindoc*-KO males. However, the absence of meiotic defects somehow excludes the eventuality that SPINDOC plays a role in piRNA pathway-directed TE silencing, despite SPIN1 interacting with SPOCD1, a protein bridging the piRNA machinery and DNA methylation establishment (Zoch et al., 2020). In our hypothesis, SPINDOC would be involved in TE repression (and especially IAP elements) in ESCs, which does not necessary implicate that SPINDOC is required for

the piRNA pathway and IAP regulation in male germ cells. Nonetheless, the expression pattern of SPINDOC is intriguing and deserves (or deserved) further investigation and all of these data were not known when I started the work presented in this manuscript. In consequence, our initial hypothesis was that SPINDOC might play a role in regulating TEs in germ cells, a question that I tried to answer by deriving a *Spindoc*-KO mouse model, as none of them were available at that time.



# RESULTS



## I. Involvement of SPIN1 and SPINDOC in TE control in mESCs

The major aim of my PhD project was to confirm and further elucidate the role of SPIN1 and SPINDOC, as a complex or individually, in the regulation of TEs, and especially of IAP elements, in mouse ESCs. This question emerged from a previous project in the lab, led by Tomasz Chelmicki (Post-doc), who developed a CRISPR-Cas9 loss-of-function screen to find new IAP regulators in mESCs (Chelmicki et al., 2021). To address this question, I generated two cell lines where *Spin1* or *Spindoc* were genetically edited and Alexis Cornec (a Master student under my supervision) generated a cell line where both genes were deleted (double KO). Genome-wide assays at DNA, RNA and chromatin levels constitute the heart of the study developed hereafter.

First, I confirmed the results of the screen by showing that loss of SPIN1 or SPINDOC or both indeed induces IAP up-regulation, while extending their role on other retrotransposon types too. Notably, only SPINDOC loss results in LINE-1 de-repression. This discrepancy between loss of SPIN1 and loss of SPINDOC indicates they could have distinct roles in regulating TEs in ESCs. Loss of SPIN1 or SPINDOC also triggers a few gene transcriptional changes, but expression of known TE regulators or genes involved in the pluripotency of ESCs are globally unaltered. Then, I tried to understand the mechanism behind TE up-regulation in absence of SPINDOC or SPIN1, starting with chromatin profiling of H3K4me3 and H3K9me3, both of which can be bound by SPIN1 *in vitro* and could reflect the observed transcriptional changes. However, no correlation could be made between the few observed chromatin changes and TE transcriptional mis-regulation. Finally, I measured DNA methylation levels by several orthogonal methods and was able to link the loss of SPIN1 or SPINDOC to a genome-wide decrease in DNA methylation, accompanied by a destabilization of UHRF1. However, DNA methylation loss is not enough to explain the level of observed TE up-regulation. Finally, SPIN1-centered mass spectrometry allowed me to identify a network of interactors of SPIN1 in ESCs, which had not been found before in other cell types and could explain the molecular phenotype of *Spin1*- or *Spindoc*-KO ESCs.

Following is a first draft of a publication that could arise from my PhD work, which will be completed with several ongoing and follow-up experiments. I performed

most of the experiments and Alexis Cornec punctually took part in the characterization of mutant ESCs under my supervision. Bioinformatic directions, notably related to repeated TE sequences, were discussed in close proximity with Aurélie Teissandier, who performed most of the analyses. For my part on this aspect, after mapping and annotation of RNA-seq reads, I performed differential expression analyses and generated all figures based on this data.

## **The SPIN1-SPINDOC chromatin reading complex controls retrotransposons partially through stabilization of DNA methylation maintenance**

---

Emeline ROGER<sup>1</sup>, Aurélie TEISSANDIER<sup>1</sup>, Alexis CORNEC<sup>1,2</sup>, Florent DINGLI<sup>3</sup>, Holger GRESCHIK<sup>4</sup>, Damarys LOEW<sup>3</sup>, Roland SCHÜLE<sup>4</sup>, Tomasz CHELMICKI<sup>1,5</sup>, Deborah BOURC'HIS<sup>1</sup>

1: Institut Curie, PSL Research University, INSERM U934, CNRS UMR 3215, Paris, France

2: Now at Institut Curie, PSL Research University, INSERM U932, Paris, France

3: Institut Curie, PSL Research University, Laboratoire de spectrométrie de masse protéomique, Paris, France

4: Urologische Klinik und Zentrale Klinische Forschung, Klinikum der Universität Freiburg, Freiburg, Germany

5: Now working at Roche, Basel, Switzerland.

Corresponding author: [Deborah.bourchis@curie.fr](mailto:Deborah.bourchis@curie.fr)

## ABSTRACT

Transposable elements (TEs) are mobile genetic elements present in every living species. To limit their harmful effects, species have developed over evolutionary time-scales several mechanisms to control TE expression and mobilization. These are of utmost importance to protect genomic integrity in germ cells and pluripotent cells, notably embryonic stem cells (ESCs). By suppressing their initial transcription, DNA methylation is a potent TE repressor. However, during the epigenetic reprogramming that accompanies pre-implantation development, ESCs lose most of their genomic methylation, except at a few regions including some TEs. Mechanisms behind TE control in this context are not completely resolved. We report here the involvement of SPIN1-SPINDOC, a peculiar chromatin reading complex with abilities to bind both H3K4me3 and H3K9me3 *in vitro*, in the control of TEs in mouse ESCs. Presence of both factors also proved to be essential for proper DNA methylation maintenance with an influence on UHRF1 stability. However, DNA methylation is not the only mechanism by which SPIN1 and SPINDOC exert control over TEs in ESCs. SPIN1-SPINDOC-dependent control of TEs re-opens the question of the relevance of DNA methylation in naive ESCs.



## INTRODUCTION

Transposable elements (TEs) are mobile genetic elements present in all genomes. In mammals, only retrotransposons, which represent approximately 40% of the DNA content, are still able to jump with few elements responsible for the majority of transposition activity (Goodier, 2016). In humans, only a subgroup of Long Interspersed Nuclear Element (LINE-1) fall into this category (L1 Human Specific, L1Hs), while several families of LINE-1 and endogenous retroviruses (ERVs) are still transposition-competent in the mouse genome (Kazazian & Moran, 2017). Among mouse ERVs, the Intracisternal-A-particle family (IAP) is particularly active and is responsible for half of the deleterious mutations observed in laboratory mouse strains (Gagnier et al., 2019). Moreover, deleterious effects not only come from TE movements but also from their transcription, which can be pervasive and affect neighboring gene expression (Modzelewski et al., 2022). While TE over-expression has impacts in all cell types and can be linked to cancer and neurodevelopmental disorders (Burns, 2017; Popova et al., 2022), control of TEs is of utmost importance in germ cells and their progenitors—including pluripotent embryonic stem cells—that transmit the genetic information across generations. In particular, lack of TE control in mouse male germ cells typically leads to meiotic arrest and infertility (Aravin et al., 2008; Barau et al., 2016; Bourc'his & Bestor, 2004; Zoch et al., 2020).

Transcriptional silencing of TEs is ensured by their heterochromatinization, a targeted process requiring sequence recognition to distinguish them from host genes (Groh & Schotta, 2017). Several Krüppel associated box domain zinc finger proteins (KRAB-ZFPs) are involved in this process (Yang et al., 2017), warranting binding of KRAB-associated protein 1 (KAP1), a scaffolding protein (Rowe et al., 2010). KAP1-recruited partners include the DNA methylation machinery, SET domain bifurcated 1 (SETDB1), a methyltransferase specific to lysine 9 of histone 3 (H3K9), and Heterochromatin protein 1 (HP1) (Schultz et al., 2002). DNA CpG methylation, H3K9-methylation and HP1 license the access of targeted regions to other effectors, and prevent binding of transcription factors or machinery. Among other factors recruited to TE loci are chromatin remodelers, such as ATRX and DAXX (Lewis et al., 2010).

The early mammalian development is characterized by two waves of DNA demethylation, a first one linked to the acquisition of pluripotency by the embryonic stem cells in the inner cell mass of the blastocyst and a second one to the acquisition of a



germ cell identity (Greenberg & Bourc'his, 2019). In cultured mouse embryonic stem cells (ESCs), acute DNA methylation withdrawal leads to transient TE up-regulation, followed by chromatin-based mechanisms that restore TE silencing, although the origin of their recruitment to TEs stays unknown (Walter et al., 2016). To find new suppressors of IAP elements, we previously designed a CRISPR-Cas9-based loss-of-function screen in mouse ESCs and notably revealed the implication of RNA modification in IAP control that could provide a layer of control in absence of DNA methylation-based transcription silencing (Chelmicki et al., 2021), a finding corroborated by several publications (Liu et al., 2021; Xu et al., 2021). The top enriched factors of the screen also included two factors involved in a common protein complex, Spindlin1 (SPIN1) and SPIN1 DOCKing protein (SPINDOC). SPIN1 is a chromatin reader with Tudor domains, which has the ability to bind H3K4me3 (Lu & Wang, 2014; Shanle et al., 2017; Wang et al., 2011; Yang et al., 2012), a histone modification usually associated with active transcription, as well as H3K9me3 concomitantly, at least *in vitro* (Du & Qian, 2022). *Spin1* is highly expressed in oocytes where it binds to the meiotic spindle, and in the early embryo as a maternally provided factor (Oh et al., 1997), and is moreover essential for post-natal viability in mice (Greschik et al., 2017). However, SPIN1 is currently known mostly for its role as a transcriptional activator of ribosomal RNA (rRNA) genes and Wnt target genes, both of which relate to cell proliferation and migration (Wang et al., 2012; Wang et al., 2011). The transcription activating function and oncogenic potential of SPIN1 are thought to be attenuated by its cognate repressor, SPINDOC, identified on the basis of its interaction with SPIN1 in human cell lines (Bae et al., 2017; Devi et al., 2019; Du & Qian, 2022).

The unknown biological relevance of SPIN1 and SPINDOC in ESCs and their possible link to IAP regulation spurred us to characterize mutant ESCs lacking these proteins. Here, we reveal that both factors are essential for IAP repression and DNA methylation maintenance through UHRF1 stabilization in naive ESCs. Moreover, proteomic analyses indicate that SPIN1-SPINDOC role on IAP control is not restricted to DNA methylation maintenance and might also involve the recruitment of chromatin remodelers.

## RESULTS

### **SPINDOC and SPIN1 are involved in the control of common and distinct TE families**

We previously identified SPIN1 and SPINDOC as highly scored candidates for IAP repression (Figure 1A), using a CRISPR-Cas9-based genetic screening based on re-expression of transgene that contained the regulatory region of IAPs fused to a blasticidin resistance gene (Chelmicki et al., 2021). To confirm the screen results and further investigate the role of SPIN1 and SPINDOC in mouse ESCs, we generated two *Spindoc*-KO clones, two *Spin1*-KO clones and one *Spin1*-*Spindoc*-KO (dKO) clone. All cell lines were obtained using CRISPR-Cas9 targeting, with two different single-guide RNA (sgRNAs) for *Spindoc*-KO leading to a 4 exon-deletion, or one sgRNA for *Spin1*-KO targeting exon 4 common to all *Spin1* transcript isoforms (Extended Figure 1A and 1B). The dKO deletion was built upon *Spindoc* sgRNA transfection of the single *Spin1*-KO background. Cell growth and morphology of mutant cell lines was comparable to WT ESCs, although dKO may show slightly lower proliferation rate (Extended Figures 1C and 1D). Immunoblot analyses confirmed protein depletion, and further revealed that SPINDOC is highly destabilized in absence of SPIN1, although not completely absent (Figure 1B).

RT-qPCR confirmed increased levels of IAP expression in all mutant cell lines, validating the screen results (Figure 1C). Transcriptomic features were further investigated by short-read bulk RNA sequencing (RNA-seq), in which differential expression analysis ( $|\log_2FC| > 1$  and  $FDR < 0.01$ ) of *Spindoc*-KO, *Spin1*-KO and dKO ESC clones was done relative to the specific wildtype (WT) ESC batch from which they independently derived (Extended Figure 1E). A first differential analysis at the TE family level using random mapping of repeated sequences revealed increased expression of several IAP subfamilies in all mutant cells, including IAP-d, IAPEy, IAPEY3, and IAPEz, whose consensus sequence was used as a reporter in the original screen (Figure 1D-F). It is noteworthy that while IAPEz copies are distributed equally between autosomes and sex chromosomes, the other reactivated IAP subfamilies tend to be enriched or almost exclusively present on the Y chromosome. Other ERVK families were consistently more expressed in all mutant ESC lines: MMERVK10D3, MMERVK9C and MMERVK9E, which are close relative of IAP family, or RLTR45 (Extended Figure

1E). In contrast, increased expression of several evolutionarily young LINE-1 families (Tf in particular) was only observed in *Spindoc*-KO mutants, not in single *Spin1*-KO, both by RT-qPCR (Figure 1C) and RNA-seq (Figures 1D-F and Extended Figure 1F). In *Spindoc*-KO ESCs, SPIN1 is left without its potential SPINDOC repressor, which could have a different impact than the loss of SPIN1 itself. Similarly, LINE-1 subfamilies were also not up-regulated in dKO (Figures 1C and 1F, Extended Figure 1F), which may relate to the fact that both *Spin1*-KO and dKO cells have in common to share both SPIN1 loss and SPINDOC loss, through genetic deletion or destabilization for the latter.

To decipher whether the reactivation of several TE subfamilies emanated from one or several individual copies within these families, we performed a second differential analysis focusing on reads that mapped to unique loci only. *Spindoc*-KO ESC lines had the highest number of individual up-regulated TE copies compared to other mutants, and the highest proportion of elements commonly up-regulated in all mutants were from the ERVK subclass (Figure 1G). We were able to reveal expression from several copies of IAP and MMERVK families in each mutant (Figure 1H-J). Expression from uniquely mapped L1 copies was detected with the highest number for *Spindoc*-KO compared to *Spin1*-KO or dKO, mirroring results obtained from random mapping analysis. Finally, individual TE copies that were scored as up-regulated in mutants over WT ESCs tended to be lowly expressed in WT, indicative of a loss of control over these elements rather than increased transcription of already expressed elements (Extended Figure 1G). Overall, we report that both SPINDOC and SPIN1 are involved in the control of IAP elements, and that SPINDOC also impacts L1 expression.

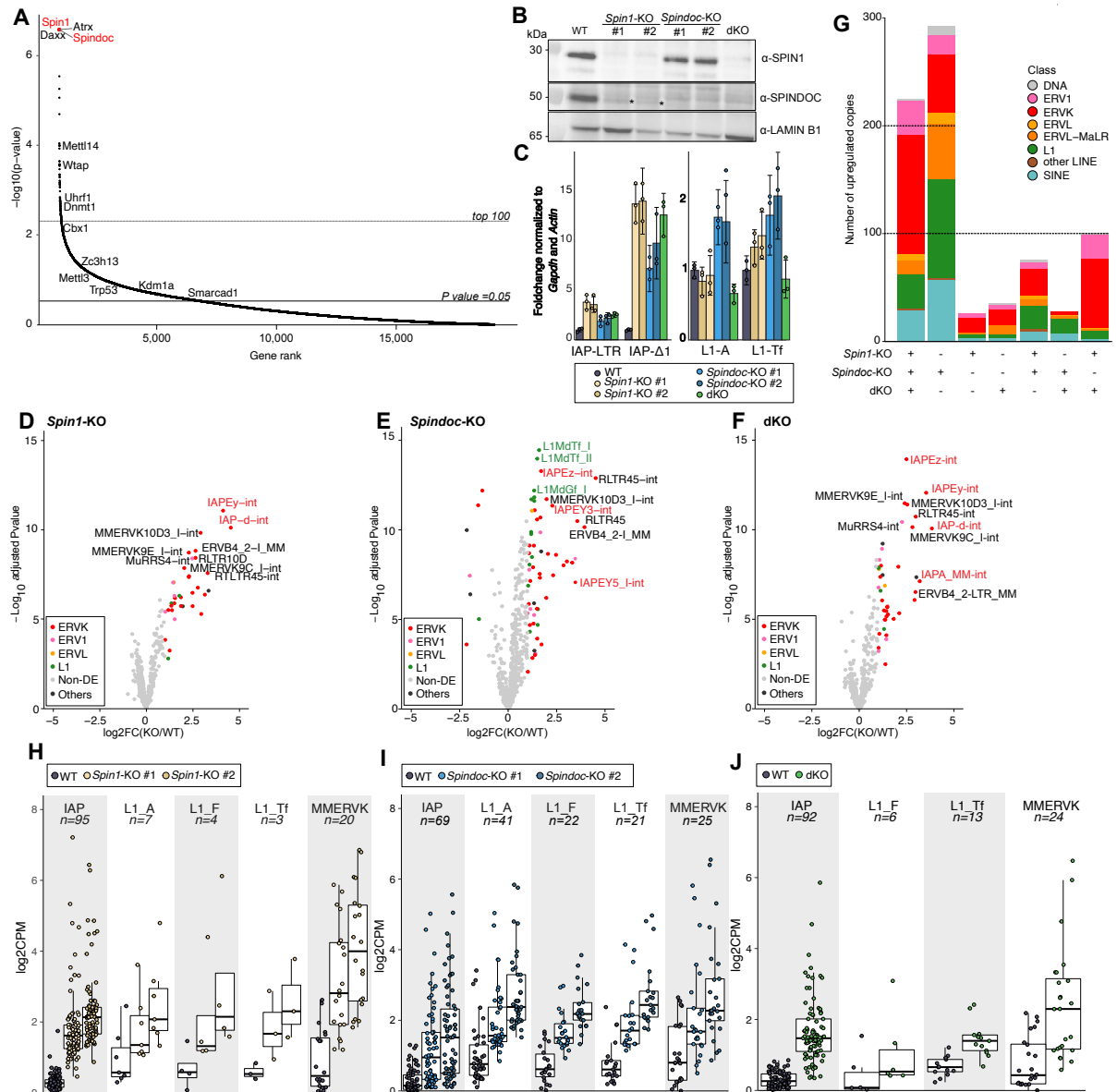
### **Loss of SPIN1 or SPINDOC leads to minor gene transcriptional changes**

We next investigated the gene transcriptional changes induced by loss of SPIN1 and/or SPINDOC. A total of 701 and 550 genes were considered as mis-expressed in *Spin1*-KO or dKO ESCs, respectively, compared to their WT counterpart, with as much up- than down-regulation (Figure 2A). In contrast, 1663 genes were mis-expressed in *Spindoc*-KO versus WT ESCs, and two thirds were up-regulated (Figure 2A), consistent with the role of SPINDOC as a repressor of SPIN1, being itself considered as a transcriptional activator. There was little overlap of gene changes between the three mutants, even between *Spin1*-KO and dKO, with 43% and 56% of commonly

mis-regulated genes, respectively (Figure 2B). In particular, approximately 13% of the dKO-specific mis-regulated genes have not yet acquired their canonical name and could be pseudogenes, long non-coding RNA or antisense transcripts (transcripts emanating from RIKEN or Gm projects). These transcriptomic differences between dKO and *Spin1*-KO—although SPIN1 and SPINDOC are present at similar levels in both conditions—could be explained by either SPINDOC leftover in *Spin1*-KO, or by the *Spindoc* RNA being still present in *Spin1*-KO and not differentially expressed compared to WT (Extended Figure 1B). Globally, there was therefore a discrete gene mis-regulation in mutant cells and we were unable to find GO terms significantly affected using Gene Set Enrichment Analysis (GSEA) (Extended Figure 2A).

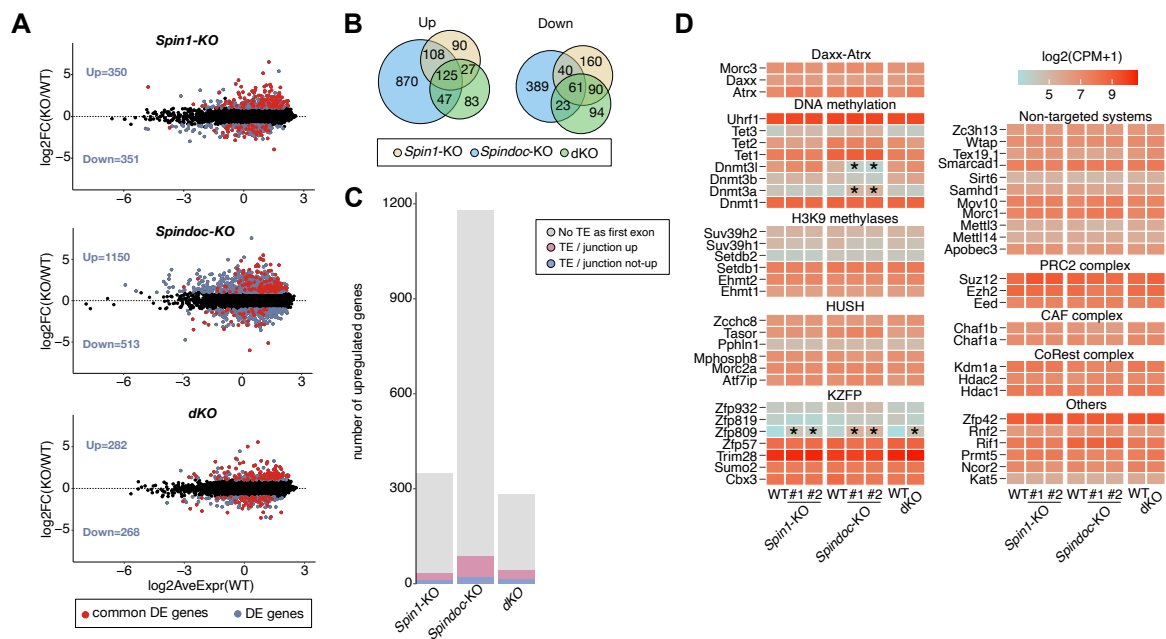
Gene mis-regulation in *Spin1*- and/or *Spindoc*-KO ESCs, and in particular up-regulation, could reflect induction from nearby de-repressed TEs acting as proximal promoters. Chimeric transcripts initiating from a uniquely mapped element and containing at least the first exon of an up-regulated gene were quantified and represented only 6% to 10% of up-regulated genes among the *Spin1*-KO, *Spindoc*-KO and dKO transcriptomes (Figure 2C). Several of these cases were previously reported, such as *Cyp2b23* (Groh et al., 2021) (Extended Figure 2B) and *Aass* (Elsässer et al., 2015). Overall, our analysis revealed that de-repressed TEs participated only very minimally to gene up-regulation in these mutants.

When focusing on genes related to different states of pluripotency, the expression of markers of core and naive pluripotency were within WT levels among the different mutant backgrounds (Extended Figure 2C). We noticed a slight up-regulation of a minority of markers related to formative/primed pluripotency, more especially in *Spindoc*-KO ESCs, indicating that these mutants may stand in a more advanced state along the pluripotency continuum. Nonetheless, all mutant cells, including *Spindoc*-KO, were able to form and survive embryoid body induction similarly to WT controls (Extended Figure 2D), which illustrates their ability to undergo multi-lineage somatic differentiation.



**Figure 1 | SPIN1 or/and SPINDOC loss triggers TE up-regulation.** **A**, Ranked  $P$  values ( $-\log_{10}$  scale, permutation tests by MAGECK) for enriched genes upon IAP-centered loss-of-function screen in ESCs (Chelmicki et al., 2021) after 9 days of blasticidin treatment. Dashed lines indicate top 100 ranked genes and  $P$  value= 0.05. The name of known IAP repressors is reported in black. **B**, Immunoblotting on whole cell protein extracts in 2i-grown ESCs against SPIN1, SPINDOC, and LAMINB1 as loading control. Asterix: faint SPINDOC band in *Spin1*-KO ESCs. **C**, Steady-state levels of IAP and LINE-1 transcripts as assessed by RT-qPCR using primer pairs targeting the LTR of IAPs (IAP-LTR), a specific class of IAPs with an internal deletion (IAP- $\Delta$ 1), and two young subfamilies of LINE-1 (A and Tf). Data are shown as  $\Delta$ CT mean normalized to *Actin* and *Gapdh* over three biological replicates  $\pm$  standard deviation. **D**, **E**, **F** Volcano plots of RNA-seq showing differential expression ( $|\log_2FC| > 1$ ,  $FDR < 0.01$ ) of TE families obtained from random mapping in mutant over WT ESCs in  $-\log_{10}(\text{adjusted } P\text{-value})$  versus  $|\log_2FC|$  in *Spin1*-KO (**D**), *Spindoc*-KO (**E**) and dKO (**F**). For *Spin1*-KO and *Spindoc*-KO, sequencing of both clones were used as replicates for differential analysis. Names of the 10 most up-regulated families are displayed, with IAPs in red and LINE-1 in green. Up-regulated family dots are colored by classes with ERVK: red; ERV1: pink; ERVL: orange; L: green; others: black. **G**, Upset plot showing numbers of uniquely mapped up-regulated TE copies ( $|\log_2FC| > 1$  and  $FDR < 0.01$ ) assigned to their corresponding TE-class in mutant ESCs. DNA transposons: grey; ERVK: red; ERV1: pink; ERVL: orange; L1: green; other LINES: brown; SINEs: blue. **H**, **I**, **J**, Boxplot showing expression in  $\log_2\text{CPM}$  of uniquely mapped copies in several TE families, in *Spin1*-KO (**H**), *Spindoc*-KO (**I**) and dKO (**J**). Differential analysis was performed from the same table of uniquely mapped elements in all mutants. Each dot represents a uniquely mapped up-regulated TE copy ( $|\log_2FC| > 1$ ,  $FDR < 0.01$ ) in the displayed genetic background.

SPIN1 has been reported to activate rRNA transcription (Wang et al., 2011), a function that may be further enhanced by SPINDOC association (Du et al., 2021). As repeated rRNA transcripts are poorly accessible in RNA-seq datasets, we used RT-qPCR to measure the expression of the 45S pre-ribosomal precursor using several primer pairs (Extended Figure 2E). While rRNA levels were globally unaffected by the single loss of SPIN1 or SPINDOC, dKO ESCs presented decreased 45S expression. This result stresses the distinct nature of *Spin1*-KO with SPINDOC destabilization and dKO mutants with both SPIN1 and SPINDOC ablation. SPIN1 is also a known activator of the Wnt pathway, but the use GSK3 inhibitor in 2i conditions of ESCs culture prevented us from assessing gene changes related Wnt signaling in absence of SPIN1 and SPINDOC.



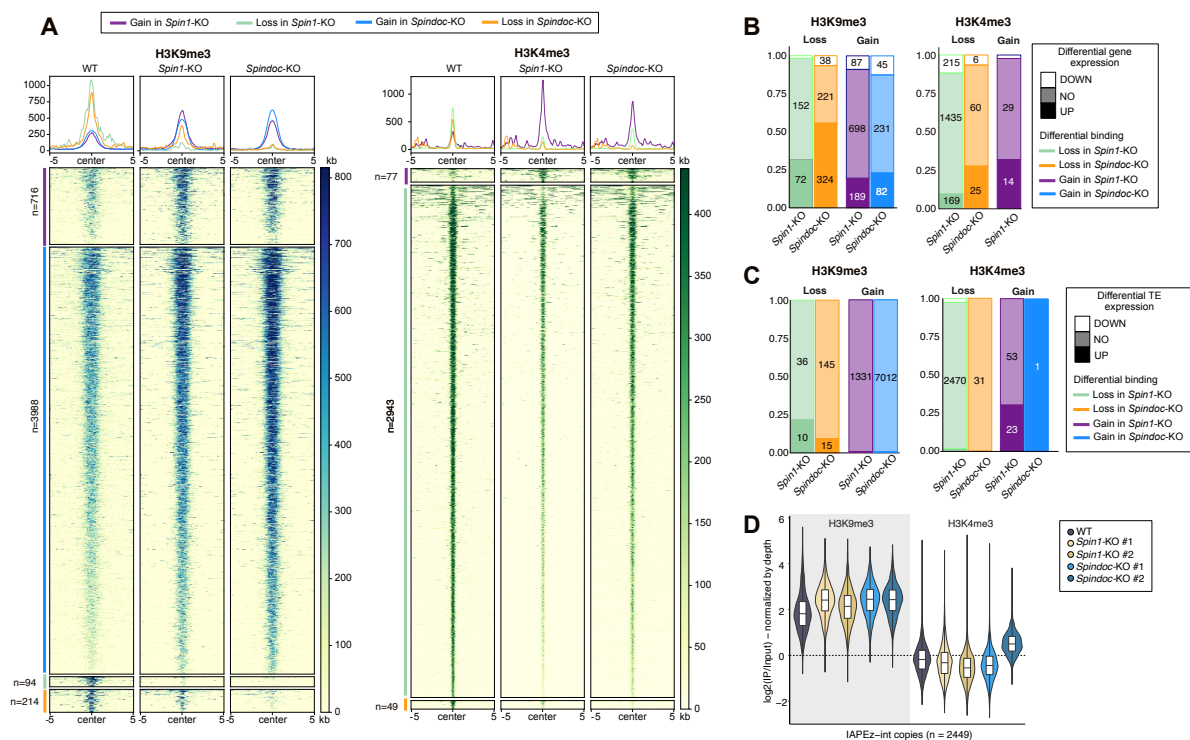
**Figure 2 | Genic transcriptional changes upon SPIN1 or SPINDOC loss.** **A**, Bland-Altman plot depicting log<sub>2</sub>FC values of genes in the different mutants in relation to their average expression in WT ESCs. For *Spindoc*-KO and *Spin1*-KO, both clones were considered as replicates for this analysis. Numbers of up- and down-regulated genes in the appropriate background are displayed, mis-expressed genes being considered if |log<sub>2</sub>FC| > 1 and FDR < 0.01. Dotted lines indicate no foldchange. Grey dots depict differentially expressed genes (DE) and red dots DE genes commonly found in *Spin1*-KO, *Spindoc*-KO and dKO. TEs are excluded from this analysis. **B**, Venn diagram with numbers of up- and down-regulated genes in all possible groups formed by *Spindoc*-KO, *Spin1*-KO and dKO. Only genes were considered in these numbers. **C**, Barplot of the number of up-regulated genes belonging to three distinct categories: no TE found as first exon of transcripts containing this gene (grey), TEs found as first exon of transcripts containing this gene, with this transcript being either up-regulated (pink) or not (purple). **D**, Heatmap displaying expression in log<sub>2</sub>(CPM+1) of the main known regulators of TEs in ESCs in each genetic background. Stars indicate a |log<sub>2</sub>FC| > 1 in KO ESCs relative to their respective WT.

Finally, we looked more precisely at expression of known suppressors of TE expression, as an indication of potential indirect effects of SPIN1 or SPINDOC deletion on TE regulation. Specifically in *Spindoc*-KO ESCs, *Dnmt3L* was down-regulated (Figure 2D), but while this co-factor is essential for *de novo* methylation in germ cells (Bourc'his et al., 2001), it is not required in embryonic cells and lowly expressed in 2i-grown ESCs (von Meyenn et al., 2016). Moreover, it was accompanied by an up-regulation of *Dnmt3A*. In all mutants, one gene encoding a KZFP that is known to repress TEs, *Zfp809* (Wolf & Goff, 2009), was consistently up-regulated (Figure 2D). All considered, the observed changes in gene expression are unlikely to account for TE up-regulation, and *vice versa*, in *Spin1*-KO and/or *Spindoc*-KO ESCs.

### **Loss of SPINDOC and SPIN1 leads to minor changes in H3K4me3 and H3K9me3 patterns**

SPIN1 can read different histone marks, among which H3K4me3 and H3K9me3 (Zhao et al., 2020). TE sequences and in particular IAP elements are enriched in H3K9me3, associated with a transcriptionally repressed chromatin, and deprived in H3K4me3, associated with transcriptionally active chromatin. Therefore, the observed IAP up-regulation upon *Spin1* and/or *Spindoc* KO could be linked to decreased H3K9me3 and/or increased H3K4me3 enrichment. To test this hypothesis, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) against these two marks in WT, *Spin1*-KO and *Spindoc*-KO ESCs. The majority of peaks (representative of significant enrichment at a specific region compared to input level) defined in WT were unchanged in mutants (Extended Figure 3A, 3B). Among detected enrichment changes (Figure 3A), we observed a strong bias toward gain of H3K9me3 with a majority of gained peaks in *Spindoc*-KO (approximately 6% of the peaks defined in WT), and loss of H3K4me3, especially in *Spin1*-KO (19% of peaks defined in WT). Both histone modification changes were not in accordance with a more transcriptionally active state. Moreover, acquired or lost peaks overlapping with genes were not linked to expected transcriptional changes. The majority of concerned genes and TEs were not differentially expressed, while we would expect a down-regulation in associated with the globally observed H3K9me3 gain or H3K4me3 loss (Figure 3B, 3C). This analysis stresses the fact that these marks are not causative but correlative with transcriptional changes. To complete this trend, in *Spin1*-KO and *Spindoc*-KO ESCs we observed a minor gain of H3K9me3 and no further enrichment of H3K4me3 on

IAPEz elements, one of the commonly up-regulated IAP subfamilies (Figure 3D). In addition, SPIN1 or SPINDOC loss did not lead to a lower H3K9me3 or higher H3K4me3 enrichment at all up-regulated TE subfamilies (Extended Figure 3C). Despite a multi-class TE up-regulation that can be seen at the unique copy-level, ESCs deprived of SPIN1 or SPINDOC did not present major H3K9me3 loss, or H3K4me3 gain at these elements or genome-wide. Therefore, we conclude that up-regulation of TEs is not linked to changes in chromatin states.



**Figure 3 | H3K9me3 and H3K9me3 changes in absence of SPIN1 or SPINDOC are not associated with transcriptional changes.** **A**, Heatmaps and metaplots showing H3K9me3 (left) and H3K4me3 (right) differential peak enrichment in RPKM-normalized ChIP-seq signal from *Spin1*-KO or *Spindoc*-KO compared to WT ESCs. Differentially enriched peaks are divided into four categories, with numbers of each peak category provided on the heatmap right side and ordered from top to bottom as: *Spin1*-KO>WT (purple), *Spindoc*-KO>WT (blue), *Spin1*-KO<WT (light green) and *Spindoc*-KO<WT (orange). Enrichment is assessed +/- 5kb from the center of the peak. **B**, Barplot showing the proportion of genes overlapping with differentially enriched ChIP-seq peaks shown on **A**, as arranged in three categories: overlapping down-regulated genes (DOWN), overlapping non-differentially expressed genes (NO) or overlapping up-regulated genes (UP). Numbers of genes in each category are provided on the barplot. NB: one peak can overlap several genes. **C**, Same as in **B**, but for TEs. **D**, Violin plot representing  $\log_2(\text{Input})$  ChIP-seq enrichment of H3K9me3 and H3K4me3 at all IAPEz elements mapped uniquely in all conditions. Dotted lines represent no enrichment compared to Input.



### **Loss of SPIN1 and SPINDOC leads to genome-wide DNA methylation loss**

During our study of H3K4me3 and H3K9me3 patterns, we noticed that the up-regulated TE copies (uniquely mapped) observed in *Spin1*-KO and *Spindoc*-KO ESCs were quite lowly enriched in H3K9me3 in WT ESCs. Using publicly available KAP1 ChIP-seq datasets from ESCs (Castro-Diaz et al., 2014; De Iaco et al., 2017), we further found that SPIN1- and SPINDOC-specific TE targets were also not particularly enriched for KAP1 binding, compared to all IAPEZ elements (Extended Figure 4A, 4B, 4C), suggesting an alternative mode of regulation than KAP1-dependent H3K9me3 deposition for these TEs. This prompted us to consider that TEs that are up-regulated upon SPIN1-SPINDOC loss may rely on DNA methylation-dependent control. We measured CpG methylation levels globally in WT, *Spin1*-KO and *Spindoc*-KO ESCs by Lumometric Assay (LUMA) (Figure 4A), and targeted at promoter sequences of IAP, L1\_A and L1\_Tf elements by bisulfite conversion followed by PCR-pyrosequencing (Figure 4B). In WT ESCs, due to culture in 2i-based medium, CpG methylation levels were low, between 20 to 30%. Nevertheless, we consistently observed further decreased CpG methylation levels in both LUMA and TE-targeted pyrosequencing assays in all tested mutants.

Wondering if specific regions were especially losing DNA methylation in mutants compared to WT, we performed genome-wide single-nucleotide-based profiling using Enzymatic Methyl Sequencing (EM-seq). We found that DNA methylation loss was equally distributed in all genomic compartments, with an average CpG methylation between below 10% for mutants compared to between 20 and 30% for WT cells (Figure 4C). As DNA methylation loss was equally distributed genome-wide, few differentially methylated regions (DMRs) could be called (*Spin1*-KO #1 vs WT= 10,858 DMRs, *Spindoc*-KO #1 vs WT DMRs= 17,494) (Figure 4D), as their identification pre-supposes a difference compared to neighboring regions. Although DNA methylation decreased across the entire genome in *Spin1*-KO and *Spindoc*-KO ESCs, levels were still above the ones from ESCs cultured in 2i medium complemented with Vitamin C (Figure 4D), a condition where CpG methylation is almost completely removed due to enhanced activity of ten-eleven translocation (TET) enzymes (Blaschke et al., 2013). Loss of SPIN1 or SPINDOC therefore generates an intermediate DNA methylation level for ESCs, between ESCs cultures in 2i only and in 2i+Vitamin C medium. Comparing the relative effects of SPIN1 or SPINDOC loss on DNA methylation levels, we noticed a

consistent higher decrease when SPINDOC was absent (Figure 4C), except when focusing on IAPeZ elements for which greater decrease was observed in *Spin1*-KO ESCs (Figure 4E, Extended Figure 5A).

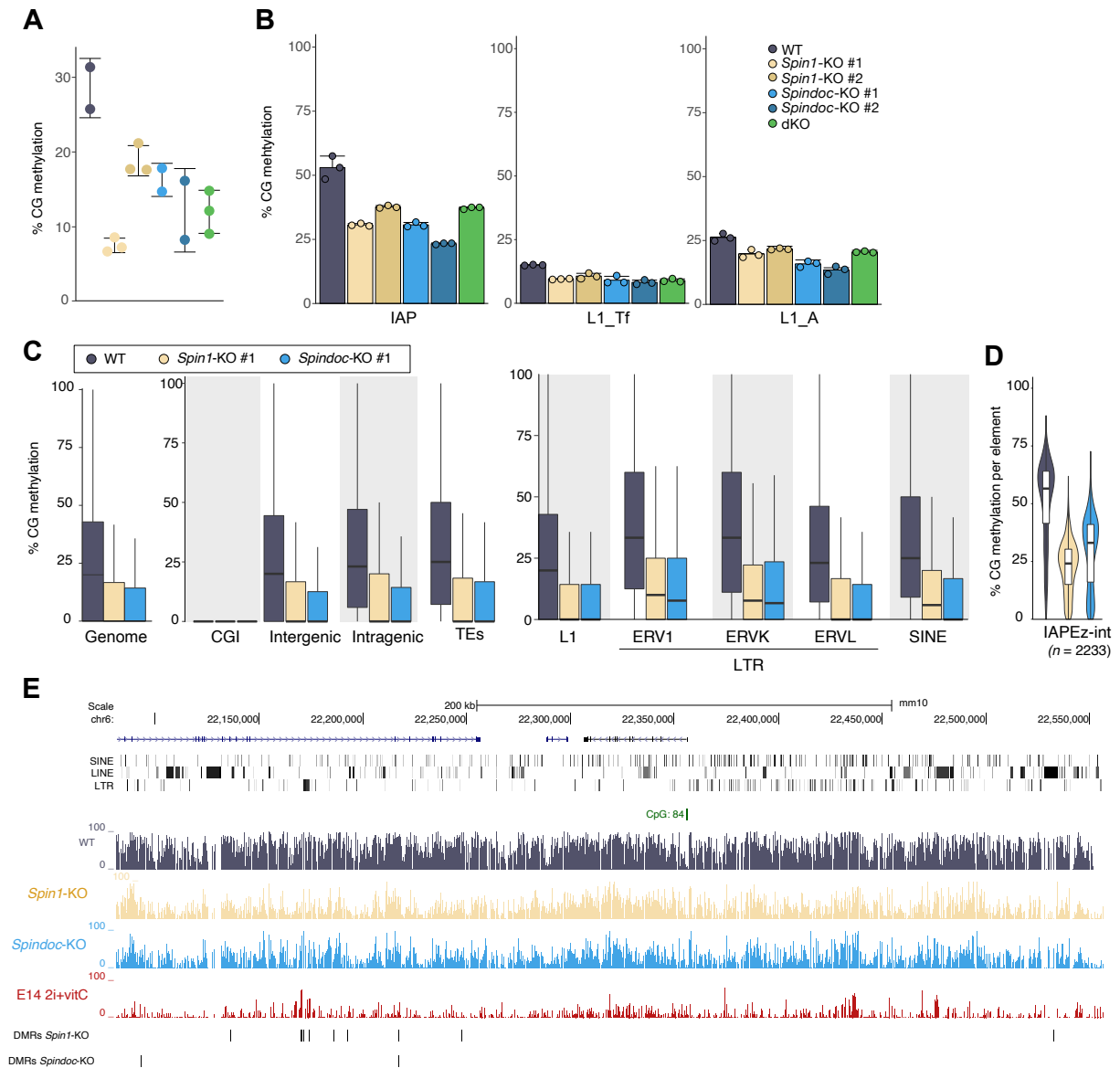
### **Loss of DNA methylation is accessory to TE up-regulation in absence of SPIN1 and SPINDOC**

DNA methylation loss might be either “causative”, meaning upstream of the path leading to TE up-regulation in *Spin1*-KO and *Spindoc*-KO ESCs, or “reflective” as a downstream consequence of their increased expression. However, DNA methylation loss is observed genome-wide, while transcriptional changes occur over discrete genomic regions, the last hypothesis seems therefore unlikely. Strikingly, we observed that when cultured in serum-based medium and therefore with highly methylated genome than in 2i (70-80% in serum-based medium compared to 20-30% in 2i), the absence of SPINDOC or SPIN1 had no effect on CpG methylation levels, as assessed by LUMA and TE-centered pyrosequencing (Extended Fig 5B, 5C). Despite displaying normal DNA methylation levels, serum-grown *Spin1*-KO and *Spindoc*-KO ESCs still showed IAP up-regulation, albeit to a lesser extent, as shown by targeted RT-qPCR (Figure 5A). For instance, IAP- $\Delta$ 1 transcripts were increased by 15-fold in *Spin1*-KO compared to WT control ESCs grown in 2i, and increased by 7-fold in serum-grown conditions. Therefore, while DNA methylation loss might play a role in TE de-repression, it is not the only determinant of IAP reactivation in ESCs lacking SPIN1 or SPINDOC. Moreover, while LINE-1 expression was up-regulated in *Spindoc*-KO but not *Spin1*-KO, LINE-1 DNA methylation was equally decreased in *Spindoc*-KO and *Spin1*-KO compared to WT 2i-grown ESCs (Extended Figure 5A). Collectively, these results provide indirect evidence that SPIN1 and SPINDOC role in repressing TEs is partly but not completely linked to DNA methylation-dependent regulation.

### **SPIN1 and SPINDOC are essential for UHRF1 stability in 2i-grown ESCs**

Reduction in global genome methylation levels is reminiscent to a defect in maintenance of DNA methylation. The transcript level of the maintenance DNA methyltransferase *Dnmt1* and its co-factor Ubiquitin-like with PHD and ring finger domain 1 (*Uhrf1*) were unchanged in RNA-seq of 2i-grown *Spin1*-KO and *Spindoc*-KO compared to WT ESCs (Figure 2D). However, immunoblot analysis revealed a strong decrease in UHRF1 protein in all single mutants, while DNMT1 was normally present

(Figure 5B, Extended Figure 6A). Importantly, UHRF1 destabilization did not occur when *Spin1*-KO and *Spindoc*-KO were cultured in serum-based medium, in agreement with unaltered DNA methylation levels in this condition (Figure 5C, Extended Figure 5B, 5C). In ESCs, decrease of DNA methylation during serum to 2i medium conversion is accompanied by a dynamic control of UHRF1 stability, in part through proteasome targeting by PRAMEL7 (Graf et al., 2017). Notably, transcript levels of *Pramel7* was unchanged in our mutants, as assessed by RNA-seq (*data not shown*). A recent shRNA screen identified new kinases whose expression is necessary to achieve UHRF1 degradation in ESCs during serum to 2i transition specifically (Rushton et al., 2022). However, the mechanism by which by these kinases induce UHRF1 destabilization has not been studied yet. Nonetheless, as SPIN1- SPINDOC loss only affects DNA methylation in 2i conditions, we reasoned that our proteins of interest might regulate such kinases in this system. In our RNA-seq data, we found that two kinases were up-regulated in both *Spin1*-KO and *Spindoc*-KO ESCs: Mitogen activated protein kinase 12 (*Mapk12*) and Potassium voltage-gated channel subfamily H member 2 (*Kcnh2*) (Extended Figure 6B). This may be of interest to further study these proteins as involved in UHRF1 destabilization in 2i-grown ESCs.

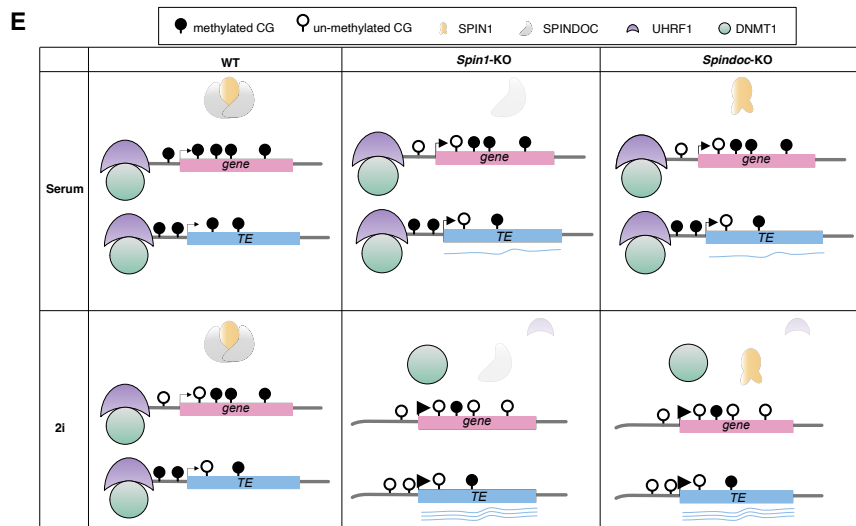
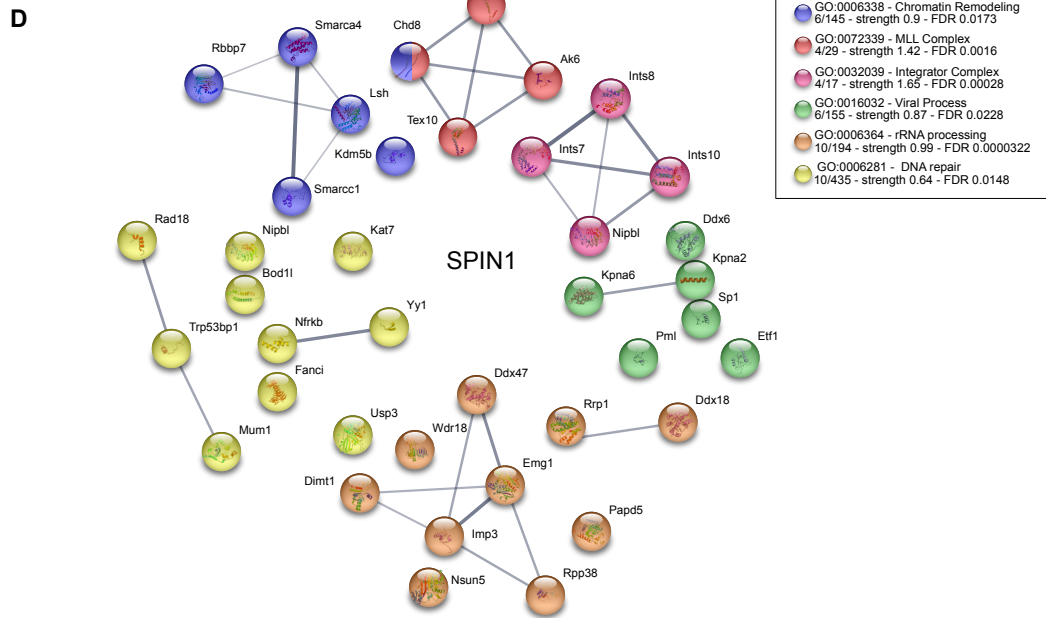
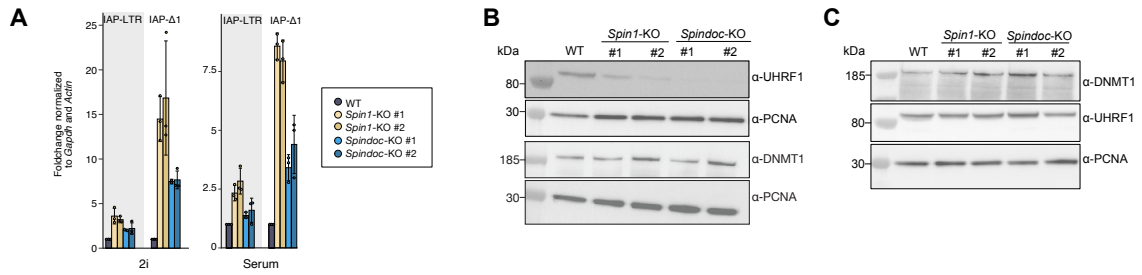


**Figure 4 | SPIN1 or SPINDOC deletions induce a genome-wide loss of DNA methylation.** **A**, Dotplot showing the percentage of global CpG methylation levels in 2i-grown WT and mutant ESCs, determined with Lumometric Assay (LUMA). Each dot represents a biological replicate of cells grown independently. Vertical bars show mean  $\pm$  standard deviation. **B**, Barplot representation of % CG methylation in WT and mutant ESCs at the promoter region of IAP (4 CpGs), L1\_Tf (6 CpGs), and L1\_A (7 CpGs) elements determined by bisulfite conversion followed by PCR-pyrosequencing. Each dot represents a biological replicate of cells grown independently. Vertical bars show mean + standard deviation. **C**, Barplot representing mean % CG methylation + standard deviation on whole genome/ CpG islands (CGIs), Intergenic and, Intragenic sequences and TEs / LINE-1, ERV1, ERVK, ERVL, SINEs, as determined by EM-seq in 2i-grown WT (grey), *Spin1*-KO #1 (yellow) and *Spindoc*-KO #1 (blue) ESCs. **D**, Violin plot of % CG methylation at IAPEZ-int elements with at least 10 reads uniquely mapped in WT, *Spin1*-KO #1 and *Spindoc*-KO #1 ESCs, as determined by EM-seq. **E**, Genomic tracks extracted from UCSC showing genes (dark blue), repeat masker (black), CGIs (green), % CG methylation at WT (grey), *Spin1*-KO #1 (yellow), *Spindoc*-KO #1 (blue) ESCs, as well as ESCs grown in 2i + Vitamin C (2i+VitC) from published WGBS datasets (Walter et al., 2016).

### **SPIN1 interacts with chromatin remodelers**

To gain insights into the mechanisms behind the control that SPIN1 and SPINDOC exert on TE expression and on DNA methylation, we performed SPIN1 immunoprecipitation followed by liquid chromatography and mass spectrometry (IP-LC/MS) on WT and *Spin1*-KO ESCs (clone #1). As expected, SPIN1 and SPINDOC were found only in WT IP, together with a set of 19 proteins not known to bind SPIN1 in other cell types ([Extended Table 1](#)). Among the list of proteins enriched in WT IP compared to *Spin1*-KO IP, we found several chromatin interactors among which several chromatin remodelers (CHD8, CHD9 and SMARCC1) and several transcription factors (SP1, YY1) ([Figure 5D](#), [Extended Table 1](#)). Consistent with previous publications linking SPIN1 to rRNA transcription and Wnt signaling (Wang et al., 2012; Wang et al., 2011), we also found several proteins involved in the same process and GSK3B, one of the main effectors of Wnt signaling. Interestingly, factors involved in DNA repair were enriched among SPIN1 interaction partners. Recently, SPINDOC was revealed to interact with Poly-ADP Ribosylation Protein (PARP) involved in DNA damage repair (Yang et al., 2021). However, a complex encompassing SPIN1, SPINDOC and PARP was not found, and SPINDOC seems to interact either with SPIN1 or PARP. Links between SPIN1 and DNA repair strengthens the biological relevance of both partners in this essential process.

Among enriched proteins, two caught our attention: Promyelocytic leukemia protein (PML) and Lymphocyte-Specific Helicase (LSH). On one hand, PML associates with ATRX-DAXX and H3.3-enriched heterochromatin (Delbarre et al., 2017), such as the one found at IAP sequences, both of which are linked to TE repression (Lewis et al., 2010). On the other hand, LSH is involved in regulation of *de novo* and maintenance DNA methylation, with a role in UHRF1 chromatin binding (Han et al., 2020). SPIN1-LSH and SPIN1-PML association could explain our observations and require further investigations.



(See next page for legends)

**Figure 5 | DNA methylation loss is associated with UHRF1 destabilization in 2i- but not serum-grown ESCs** **A**, Barplot showing expression in foldchange of mutant over WT ESCs of CT normalized to *Actin* and *Gapdh* of IAP with two different primer pairs targeting IAP-LTR or IAP- $\Delta$ 1 in WT and *Spin1*-KO or *Spindoc*-KO ESCs grown in 2i (left panel) or serum medium (right panel). Folchange is shown as mean + standard deviation over three biological replicates of cells grown independently represented by each dot. **B, C**, Immunoblot in WT, *Spin1*-KO and *Spindoc*-KO ESCs grown in 2i-based medium (**B**) or in serum-based medium (**C**) against UHRF1, DNMT1, and PCNA as loading control. **D**, Selected factors interacting with SPIN1 in 2i-grown ESCs, as determined by IP/LC-MS, chosen to represent the diversity of biological processes statistically enriched in SPIN1-interacting partners. Links between two proteins represent a physical interaction, with strength of links as a function of confidence in their interaction. Each color represents a GO term: Chromatin remodeling (blue), MLL Complex (red), Integrator complex (pink), Viral process (green), rRNA process (orange) and DNA repair (yellow). Network visualization was obtained via the STRING database. **E**, Schematic summary of major findings of the study. Loss of SPIN1 or SPINDOC triggers a destabilization of UHRF1, associated with a loss of DNA methylation and up-regulation of TEs in 2i-grown ESCs. TE up-regulation is not restricted to 2i-grown ESCs but also extends to serum-grown ESCs, as opposed to other downstream effects of SPIN1 or SPINDOC loss (DNA methylation and UHRF1 destabilization).

## DISCUSSION

In this study, we revealed the involvement of SPIN1 and SPINDOC in TE repression and in particular of IAP elements in mouse ESCs. SPIN1 and SPINDOC also proved to be essential for UHRF1 stability in 2i culture conditions, linking both factors to a role in DNA methylation maintenance. Despite DNA methylation being known as a potent repressor of TEs, its loss does not solely explain the level of TE up-regulation observed in *Spin1*- or *Spindoc*-deficient cells. Moreover, TE up-regulation was not linked to a loss of H3K9me3 at TE regulatory sequences. All of these findings are completely novel, for both the fields of DNA methylation and TE biology.

In addition, this study constitutes the first of its kind on the role of SPIN1 and SPINDOC in ESCs. SPIN1 and SPINDOC were mainly studied in human cancer cell lines for their involvement in rRNA transcription, chromatin interaction and Wnt signaling. According to this, we found that *Spin1* deletion leads to SPINDOC destabilization, and that SPIN1 physically interacts with GSK3, a central effector in Wnt signaling. Although SPIN1 is thought to act as a transcriptional activator at rRNA in human somatic cell lines (HeLa, HEK293T, SW480) (W. Wang et al., 2011; Zhang et al., 2018), its loss had no impact on rRNA expression in ESCs. Similarly, *Spindoc* loss of expression had previously been linked both to rRNA up-regulation and down-regulation in HEK293T (Bae et al., 2017; Du et al., 2021), while we observe no perturbation of rRNA gene expression in *Spindoc*-KO ESCs. However, the deletion of both *Spin1* and *Spindoc* in ESCs provokes a down-regulation of rRNA expression, with a slight decrease in cell proliferation. Exploration of factors found to interact with SPIN1

and involved in rRNA regulation found in this study may provide more clarity about exact mechanism linking SPIN1-SPINDOC to rRNA expression.

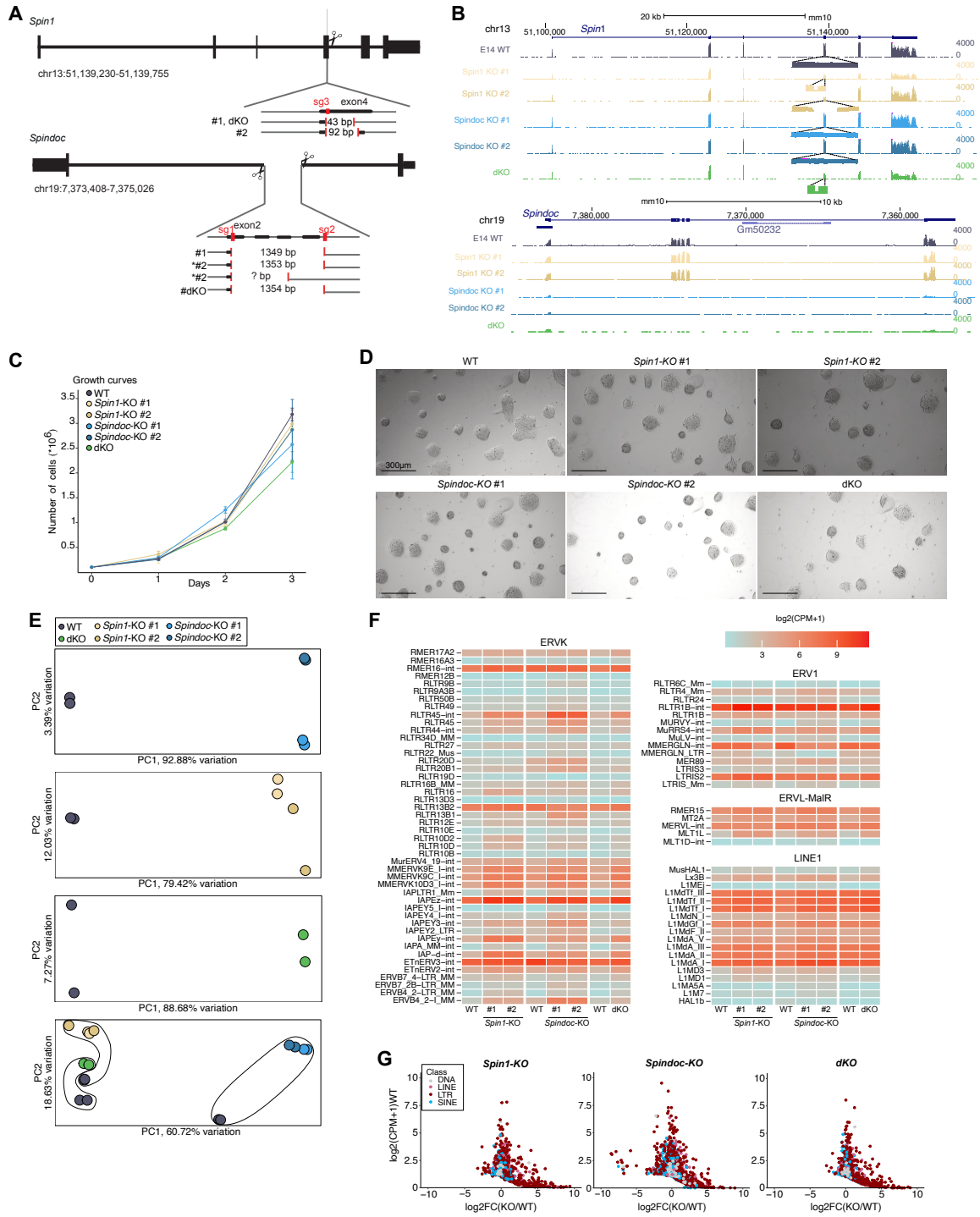
Despite SPIN1 being an H3K4me3 reader and thought to act as a transcriptional activator, the transcriptome of *Spin1*-KO ESCs was only minimally impacted, and did not show pronounced trend towards down-regulation. In contrast, loss of SPINDOC—described as a SPIN1 repressor—promoted up-regulation of more than 1000 genes, corroborating this role. However, genetic deletion of both *Spin1* and *Spindoc* was more similar to single *Spin1* deletion, with again a low transcriptional impact. Therefore, the absence of SPINDOC might be seen as a condition whereby SPIN1 has increased availability—like an over-expression system—compared to WT cells where most SPIN1 interacts with SPINDOC. Regulation of SPINDOC level might therefore be key to SPIN1 availability on chromatin and its enhancing effect on transcription. However, we noted a significant loss of H3K4me3 (not accompanied by transcriptional changes) in absence of SPIN1. It is possible that H3K4me3 is protected by SPIN1 binding, an hypothesis that would be confirmed by a genome-wide binding analysis of SPIN1 in ESCs.

Loss of TE transcriptional control, in particular at IAP elements, is often associated to H3K9me3 and/or DNA methylation loss (Karimi et al., 2011; Sharif et al., 2016; Walter et al., 2016). Despite their strong up-regulation, IAPs did not substantially lose H3K9me3 in *Spin1*-KO or *Spindoc*-KO ESCs, and may even gain enrichment for this mark compared to WT levels. However, the absence of SPIN1 or SPINDOC absence led to massive, genome-wide DNA methylation loss in ESCs grown in 2i conditions. Maintenance of IAP up-regulation in serum medium, where DNA methylation levels are normal in *Spin1*-KO or *Spindoc*-KO ESCs, showed that this DNA methylation loss is neither an up- nor a down-stream effect of TE-up regulation. It may rather be linked to the requirement of SPIN1 and SPINDOC for UHRF1 stability, which may happen through an indirect mechanism via which the absence of the SPIN1-SPINDOC complex allows higher expression or availability of one or several proteins that destabilize UHRF1. In sum, our results show that the presence of both H3K9me3 and DNA methylation at IAPs might not be sufficient to ensure full transcriptional control of these elements. Therefore, despite H3K9me3 and DNA methylation being hallmarks of TE control, we challenge this view by finding conditions in which none are lost and TE up-regulation is still observed.



The fact that SPIN1 and SPINDOC are both necessary for proper TE control in ESCs does not fit with the hypothesis that SPIN1 acts as a transcriptional activator of TEs, in which case loss of SPIN1 should lead to TE expression levels similar to WT or to their further repression. However, we noted that *Spindoc*-KO ESCs, in which SPIN1 is deprived of SPINDOC control, de-repressed of a wide variety of TEs, included LINE-1, which are not affected by the combined loss of SPIN1 and SPINDOC, as in dKO ESCs.

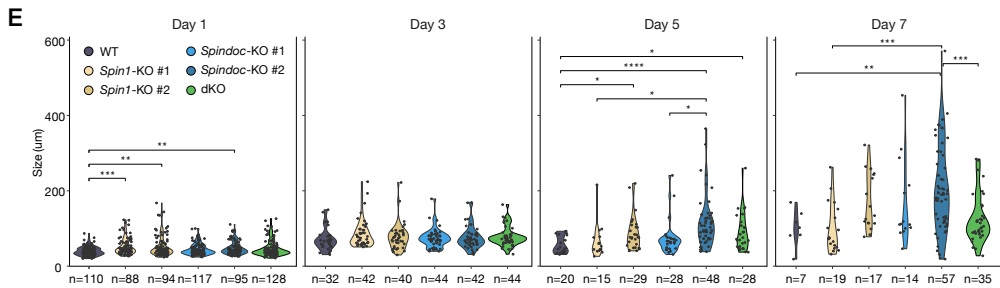
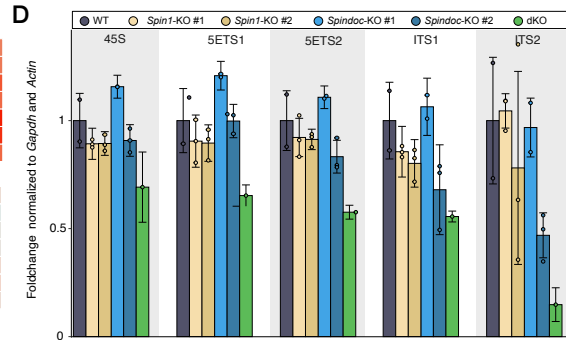
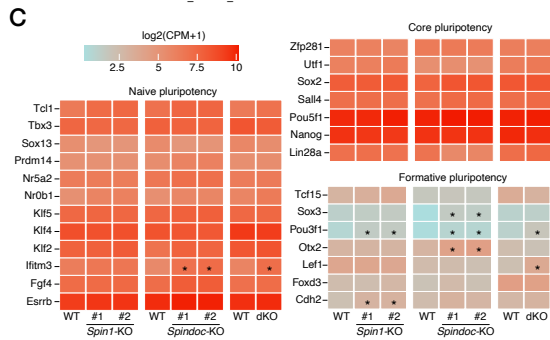
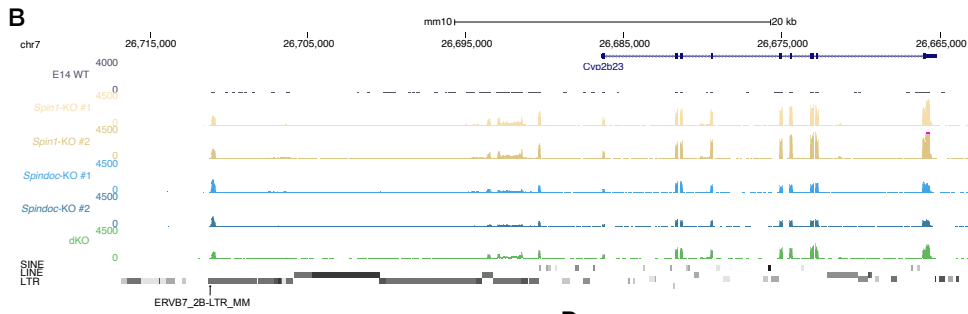
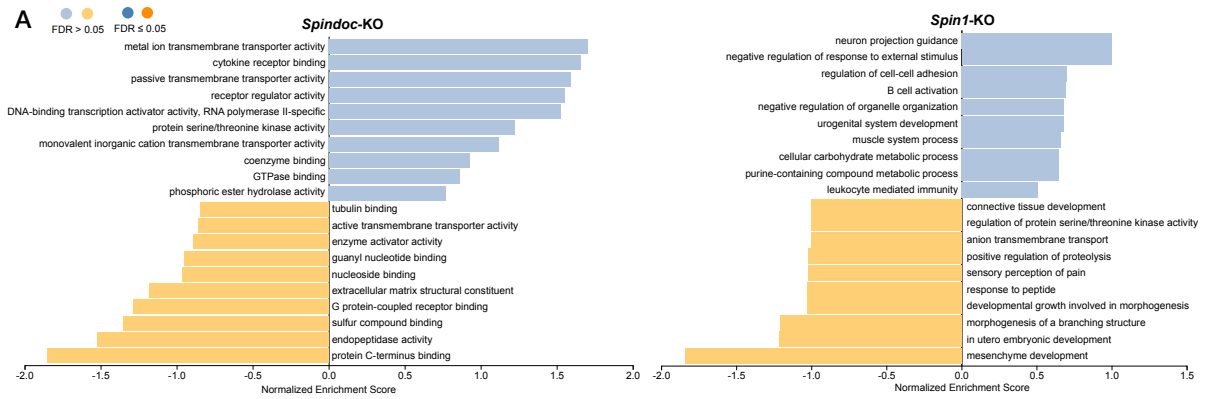
Finally, thanks to the identification of SPIN1 interaction partners in ESCs, we were able to link the SPIN1-SPINDOC complex to chromatin remodeling, notably via PML interaction. We propose a final working model in which SPIN1 alone acts as a transcriptional activator of TEs, while SPIN1-SPINDOC would work as a complex to allow proper heterochromatinization at specific TE sequences. This model highlights the importance of SPINDOC for TE control and the unwarranted effects of SPIN1 expression in ESCs.



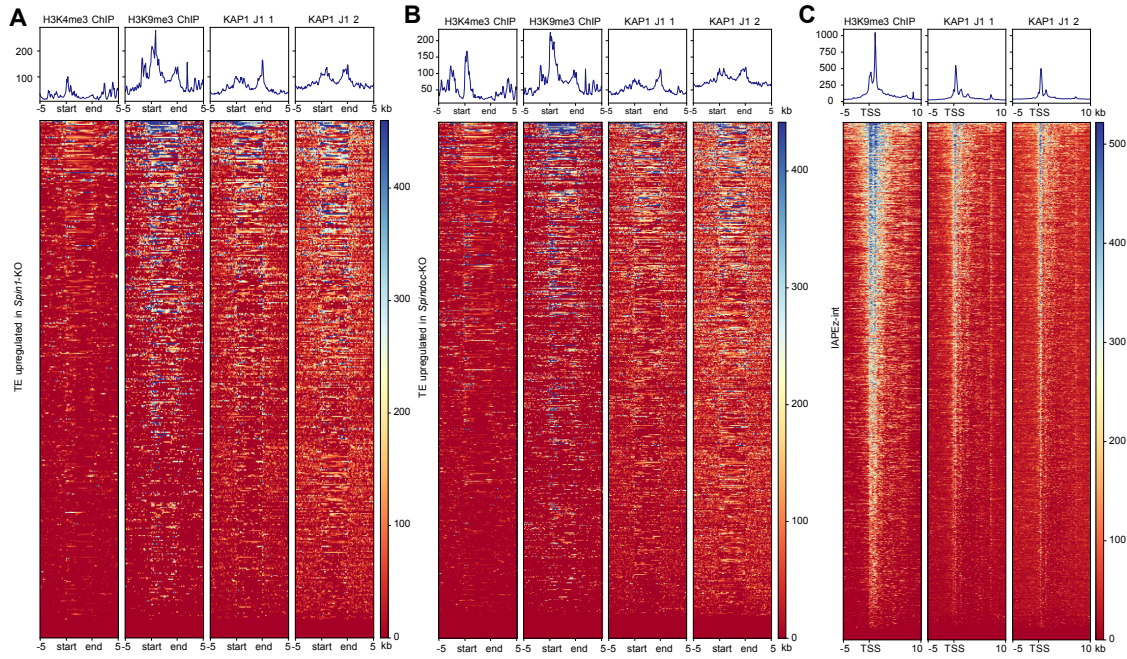
(See next page for legends)

**Extended Fig 1 | *Spin1*-KO, *Spindoc*-KO and dKO ESCs growth and morphology is not affected despite up-regulation of multiple TE families.** **A**, Scheme of *Spin1* and *Spindoc* loci with sgRNA binding regions in red, cut sites notified with scissors and size of sequenced deletions provided for each clone. *Spindoc*-KO #2 has a heterozygous deletion with one unknown boundary on one allele. **B**, Genomic tracks from UCSC Genome browser showing deletions at the transcript level at *Spin1* and *Spindoc* loci. Expression is shown as number of reads normalized to sequencing depth. Small exonic deletions are magnified for the *Spin1* locus. *Spin1*-KO clones are depicted in yellow, *Spindoc*-KO clones in blue and dKO in green. **C**, Growth curves in million cells per day of 2i-cultured WT and mutant ESCs. Each dot represents an average over three biological replicates of ESCs grown independently  $\pm$  standard deviation. **D**, Pictures of 2i-grown ESCs of all genotypes, cultured at comparable confluency. **E**, Principal Component Analysis (PCA) of *Spin1* sequencing batch, *Spindoc* sequencing batch, dKO sequencing batch and all sequencing batches from top to bottom. Raw CPM data after random mapping were used to compute PCA **F**, Heatmap of TE families up-regulated in *Spin1*-KO, *Spindoc*-KO or dKO ESCs defined with random alignment. Expression is in  $\log_2(\text{CPM}+1)$ . **G**, Log2FC of up-regulated TE copies determined by unique mapping in *Spin1*-KO, *Spindoc*-KO and dKO (from left to right) ESCs, in relation to their expression in  $\log_2(\text{CPM}+1)$  in WT ESCs. For *Spin1*-KO and *Spindoc*-KO, both clones were considered as replicates for this analysis.

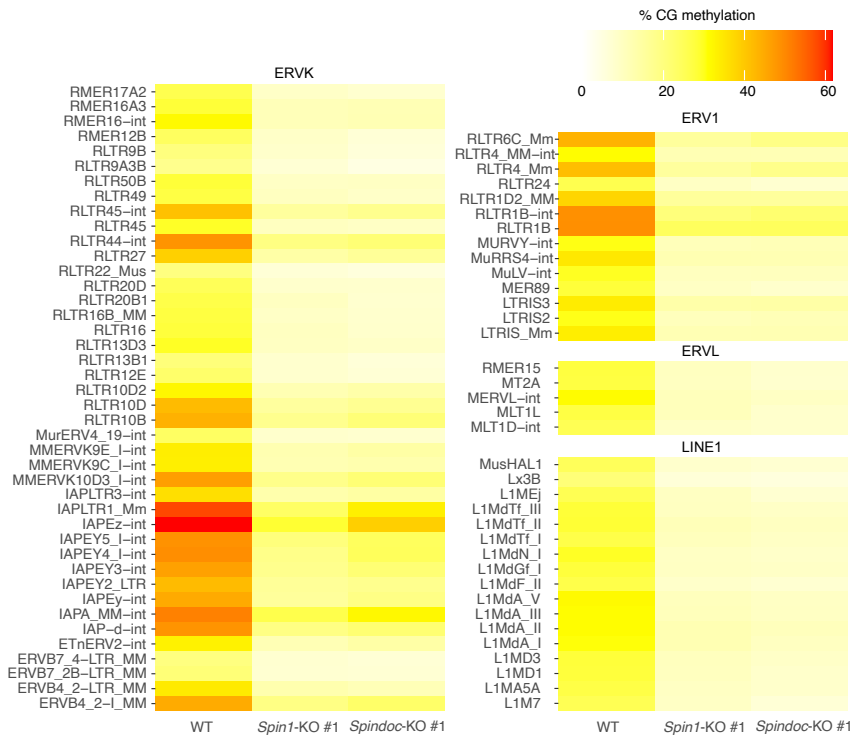
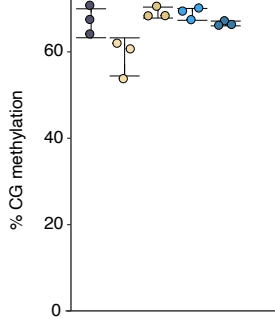
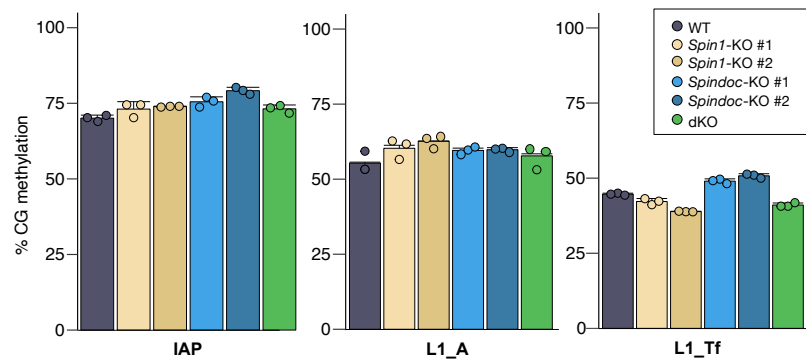
**►Extended Fig 2 | The loss of SPINDOC and/or SPIN1 minimal impacts the ESC gene transcriptional program.** **A**, GO terms enriched after Gene-Set enrichment analysis performed with WEB-based Gene Set Analysis Toolkit (WEB Gestalt) with FDR > 0.05 in yellow and light blue; and FDR < 0.05 in orange and dark blue. Vertical bar size is function of Normalized Enrichment Score (NES). **B**, Genomic tracks adapted from UCSC Genome browser showing expression of *Cyp2b23* with a TE (ERVB7) fused as exon1 consistently found in *Spin1*-KO (yellow), *Spindoc*-KO (blue) and dKO (green) ESCs but not WT ESCs. Repeat masker annotation is shown in black. **C**, Barplot of rRNA expression determined by RT-qPCR with several primer pairs across the 45S rDNA sequence (Fulka & Langerova, 2014). Expression is in Foldchange of CT normalized by *Actin* and *Gapdh* in mutants over WT. Average + standard deviation of one to three independent biological replicates with three technical replicates is depicted, with each dot representing a biological replicate. **D**, Heatmap of pluripotency marker expression showing normalized  $\log_2(\text{CPM}+1)$  counts in WT and *Spin1*-KO, *Spindoc*-KO and dKO ESCs from RNA-seq. Stars indicate differential expression of mutants over WT, as  $|\log_2\text{FC}|>1$  and FDR<0.01. **E**, Violin plots at day 1, day 3, day 5, and day 7 of EB differentiation showing EB size for WT and mutant ESCs, with numbers of counted EBs at the bottom (see Methods). Statistical analysis was done with ANOVA Welsh, and p-value adjusted with Tukey method. \*, \*\*, \*\*\*, \*\*\*\* indicates p-value >0.05;0.01;0.001;0.0001, respectively.



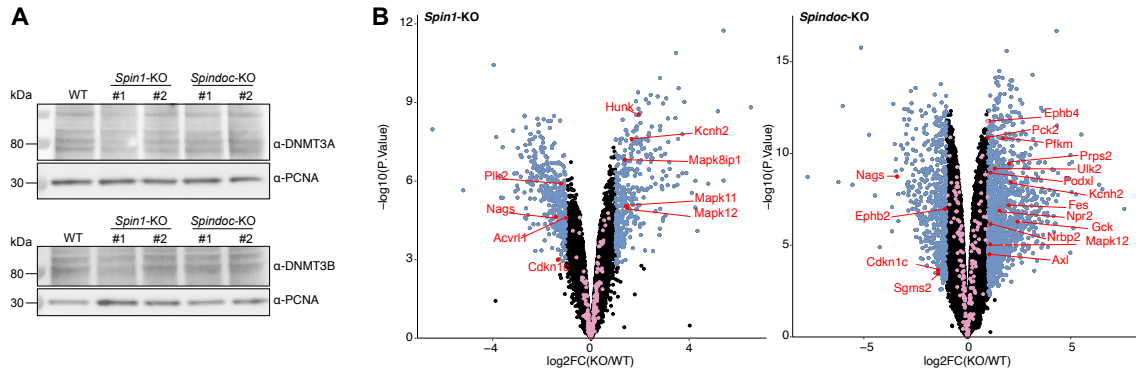




**Extended Figure 4 | TEs up-regulated in *Spin1*-KO or *Spindoc*-KO ESCs are not enriched for KAP1 binding. A, B,** Heatmaps and metaplots of H3K4me3 and H3K9me3 from this study, and KAP1 peaks from Castro-Diaz et al., 2014 and De Iaco et al., 2017. Enrichment is represented in RPKM in WT ESCs at TE copies defined as up-regulated with unique alignment in *Spin1*-KO (A) and *Spindoc*-KO (B). **C,** Heatmaps and metaplots of H3K9me3 and KAP1 peaks in RPKM in WT ESCs at all uniquely mapped IAPEZ-int copies.

**A****B****C**

**Extended Figure 5 | DNA methylation observed in 2i-grown *Spin1*-KO and *Spindoc*-KO ESCs is not observed when grown in serum-based medium. A**, Heatmap of % CG methylation at up-regulated TE families in at least one clone in 2i-grown ESCs of WT, *Spin1*-KO #1 and *Spindoc*-KO #1 genotypes, as defined by EM-seq. **B**, % CG methylation in WT (grey), *Spin1*-KO #1 (light yellow), *Spin1*-KO #2 (dark yellow), *Spindoc*-KO #1 (light blue), *Spindoc*-KO #2 (dark blue) as defined by LUMA for ESCs grown in serum-based medium. Vertical lines display  $\pm$  standard deviation from the mean and each dot is a biological replicate of ESCs grown independently. **C**, Barplot of % CG methylation at IAP, L1\_A and L1\_Tf elements (from left to right) in WT (grey), *Spin1*-KO #1 (light yellow), *Spin1*-KO #2 (dark yellow), *Spindoc*-KO #1 (light blue), *Spindoc*-KO #2 (dark blue) and dKO (green) as defined by pyrosequencing for serum-grown ESCs. Shown is mean  $\pm$  standard and each dot a biological replicate of ESCs grown independently.



**Extended Figure 6 | DNA methylation loss in *Spin1*-KO and *Spindoc*-KO ESCs is not due to mis-regulation of *de novo* DNMTs or known factors involved in UHRF1 stability.** **A**, Immunoblot on 2i-grown WT, *Spin1*-KO and *Spindoc*-KO ESCs against DNMT3A, DNMT3B, and PCNA as loading control. **B**, Volcano plot showing differential gene expression in *Spin1*-KO (left) and *Spindoc*-KO (right) over WT ESCs. For *Spin1*-KO and *Spindoc*-KO, sequencing of both clones were used as replicates for differential analysis. DE genes: grey; non-DE genes: black; genes encoding kinases found to be involved in UHRF1 stability in 2i medium (Rushton et al., 2022): red when mis-regulated and pink when non-differentially expressed in mutants.



## METHODS

### Cell Culture

E14 mouse embryonic stem cells were cultured on 0.2% gelatin coated-plates at 37°C with 8% CO<sub>2</sub>. For all experiments except stated otherwise, cells were grown in 2i medium consisting of 50% of Neurobasal medium (Gibco) with 50% of DMEM-F12 (Gibco) complemented with 2mM L-Glutamine, 0.1mM β-Mercaptoethanol, Ndiff Neuro-2 medium supplement, B-27 medium supplement, 1000U/mL of Leukemia Inhibitory Factor (LIF), 3μM of GSK3 inhibitor (CT-99021), 1μM MEK inhibitor (PD0325901). When grown in serum-based medium, ESCs were grown in Glasgow Minimum Essential Media complemented with 15% Fetal Bovine Serum, 0.1nM Essential Amino Acids, 1nM Sodium Pyruvate, 2mM L-Glutamine, 0.1mM β-Mercaptoethanol and 1000U/mL of LIF. Cells were split every 2 to 3 days using Accutase (Invitrogen) when cultured in 2i, or TrypLE (Gibco) when cultured in serum. For embryoid body (EB) differentiation, ESCs were dissociated with TrypLE and  $2.5 \times 10^4$  were seeded in uncoated B10 plates in serum without LIF to allow the formation of cell aggregates. Medium was changed every 2 days. At least ten pictures were taken every day for 7 days for each genotype, in order to count at least 30 EBs. A  $n < 30$  means that no other EB were found on the plate.

### Plasmid construction

To generate KO cell lines, targeting sequences (Table 1) were inserted in plasmid pX459 (pSpCas9(BB)-2A-Puro, Addgene #62988) after treatment with *BbsI* enzyme (NEB, R3539). Single stranded DNA sequences were annealed with their reverse complement 5 min at 95°C and cooled down slowly to room temperature (RT). Annealing products were then ligated to digested plasmids with Ligase T4 (NEB, M0202) and transformed in NEB Turbo competent cells (NEB, C2984) according to manufacturer's instructions.

### Transfection and clone selection

Insertion of plasmids in ESCs and subsequent genome editing was done using  $3 \times 10^6$  cells and 2μg of each plasmid through nucleofection with a 4D-Nucleofector device, P3 primary cell kit 4D-Nucleofector kit and appropriate protocol for ESCs provided by Lonza. All cells were plated with four different dilution factors. After 16 to 24 hours of recovery, puromycin selection (1 μg/mL) for 2-3 days and 4-5 days of growth, 96 single colonies were isolated and genotyped by PCR. For selected clones, (i.e *Spin1*-KO #1, *Spin1*-KO #2, *Spindoc*-KO #1, *Spindoc*-KO #2, dKO) PCR products were sequenced (Sanger sequencing, service proposed by Eurofins) to determine precisely genomic deletions.

### **Protein extraction and Western blotting**

Whole cell protein extraction was performed using RIPA buffer (0.5% Sodium Deoxycholate, 0.1% SDS and 1% NP-40 in PBS) with 20 min incubation on ice and centrifugation at 25,000g for 20 min. Nuclear extraction used for immunoprecipitation followed by mass spectrometry was performed by a first step of hypotonic lysis (10mM HEPES pH7.9, 5mM MgCl<sub>2</sub>, 250mM Saccharose, 0.1% NP-40) followed by nuclei swelling (25 mM HEPES pH7.9, 1.5mM MgCl<sub>2</sub>, 0.1mM EDTA, 20% Glycerol, 700mM NaCl, 1mM β-Mercaptoethanol) and sonication (6x 15'ON/30'OFF – Ultra High Frequency with a BioRuptor Pico, Dlagenode). All buffers were complemented with protease inhibitors (either 1µg/mL Aprotinin, 1µg/mL Leupeptin, 1µg/mL Pestat A and 1mM PMSF, or complete EDTA-Free, Roche). For Western blotting, after equilibration of concentration, protein extracts were denatured with LDS (NuPAGE, Thermofisher) and 10mM DTT for 5 min at 95°C and loaded on pre-cast 4-12% Bis-Tris gels (NuPAGE, Thermofisher) for migration with MOPS running buffer. Content of the gel was transferred on a 0.45µm Nitrocellulose membrane for 1h10 at 4°C in 200mM Glycine, 25mM Trizma Base, 10% Methanol. Membrane was blocked in 5% milk diluted in 0.3% NP40 in PBS (PBS-NP40). Primary antibodies ([Table 2](#)) were incubated O/N at 4°C and secondary antibodies for 45min at RT with 1% milk in PBS-NP40. PBS-NP40 also served as washing buffer. Revelation thanks to LumiLight<sup>Plus</sup> substrate (Roche) were imaged with a ChemiDoc MP instrument (BioRad).

### **RNA extraction, reverse transcription, RT-qPCR and RNA-seq**

RNA extraction was performed with the RNeasy Mini Extraction Kit (Qiagen) following the manufacturer's instructions, with extended DNase treatment (RNase-free DNase set, Qiagen) of 30 min to 1 hour. RNA concentrations measured with Nanodrop 2000 were used to perform reverse transcription with equal amount of RNA (200ng or 500ng, depending on RNA availability). RNA annealed with 20ng/µL random primers was reverse transcribed by SuperScript-III (Invitrogen) in presence of RNase inhibitor (Promega), 500nM dNTPs and 0.1M DTT at 50°C for 1h. 1µL of 5-fold-diluted reaction mix was used for qPCR with SYBR Green Master Mix (Invitrogen) on a Viia7 instrument (Applied Biosystems). All primers used in qPCR experiments are listed in [Table 1](#). Normalization of RT-qPCR analysis was done over a geometric mean of *Actin* and *Gapdh* and foldchange over WT as such:  $2^{-CT\_Sample}/\sqrt{[(2^{-CT\_Gapdh}) \times (2^{-CT\_Actin})]} = \text{Norm\_Sample}$  and  $\text{Norm\_Sample}/\text{Norm\_WT} = \text{Foldchange}$ . RNA-seq libraries were prepared from a minimum of 200 ng of DNase-treated total RNA with the TRuSeq Stranded mRNA protocol (Illumina) for *Spindoc*-KO ESC and associated WT, and with the Stranded mRNA Prep Ligation for other samples. Between 110M and 160M reads were sequenced per sample for in a 100bp paired-end format using a NovaSeq 6000 (Illumina) instrument. Libraries and sequencing were performed by the Institut Curie Next Generation Sequencing platform.

### **DNA extraction, LUMA, pyrosequencing assay and EM-seq**

DNA extraction was performed either with GenElute kit (Sigma) or PureLink kit (Invitrogen) following manufacturer's protocol for mammalian cells. For LUMA, 500ng

of each samples tested DNA was digested either with *HpaII* (NEB, R0171) and *EcoRI* (NEB, R3101) (mixA) or *MspI* (R0106) and *EcoRI* (mixB) for 1 hour at 37°C in 20uL. 6µL of each reaction was used on PyroMark Q48 Pyrosequencer (Qiagen) to determine frequency of fragmentation as a proxy for genome wide DNA methylation level (with dispensation order of nucleotides as ACTCGA). For pyrosequencing, 200ng of DNA were subjected to bisulfite treatment with the EZ-DNA methylation-Lightning kit (Zymo) following manufacturer's instructions. For targeted amplification, the PyroMark PCR kit (Qiagen) was used with standard proposed PCR mix, 35 cycles and primers listed in Table 1 together with their sequencing primers. Pyrosequencing on Q48 instrument followed manufacturer's instructions with appropriate PyroMark Q48 Advanced reagents. For EM-seq, 100ng of DNA were used for enzymatic conversion and library preparation accordingly to NEB NExt Enzymatic Methyl-seq (E7120). A pool containing 5nM of each sample (WT, *Spin1*-KO #1, *Spindoc*-KO#1) was formed and between 750M and 860M reads were sequenced per sample in a 100bp paired-end format using a NovaSeq 6000 (Illumina) instrument at the Institut Curie Next Generation Sequencing (NGS) Platform.

### ChIP-seq

For each replicate,  $2 \times 10^7$  cells were cross-linked with 1% formaldehyde for 9 minutes (H3K4me3) or 4 minutes (H3K9me3) and quenched with 0.125M Glycine. Cells were collected in Farnham Lysis Buffer (5mM PIPES pH8.0, 85mM KCl, 0.5%NP-40) and nuclei separated using Kountes dounce size B. Nuclei were re-suspended in RIPA (1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS in PBS) and sonicated for 30 cycles of 30sON/30sOFF at Ultra-High Frequency with a BioRuptor Pico (Diagenode). The soluble fraction was recovered and 10% was set aside as input before incubation with 5µg of antibody for 16 hours at 4°C under agitation. Antibodies were then bound to 1% BSA-PBS washed Dynabeads M-280 Sheep Anti-Rabbit IgG for 2 hours at 4°C under agitation. Beads were then washed in 5x in RIPA Buffer for 5 min, 1x in LiCl solution (100mM Tris-HCl pH7.5, 500mM LiCl, 1% NP-40, 1% Sodium deoxycholate), and 2x in TE buffer (10mM Tris-HCl pH7.5, 0.1mM EDTA) before elution in 1%SDS, 0.1M NaHCO<sub>3</sub>. Input and immunoprecipitated samples were subjected to RNase A (Invitrogen) treatment for 1hour at 37°C and Proteinase K (Invitrogen) for 20 minutes at RT before de-crosslinking at 65°C O/N. DNA was recovered using MinElute purification kit (Qiagen) according to manufacturer's instructions. Sequencing libraries were prepared using Illumina ChIP-seq preparation kit between 30M and 120M reads were sequenced per sample in a 100bp paired-end format using a NovaSeq 6000 (Illumina) instrument.

### Annotation of transposable elements

Repeat annotation was downloaded from repeatmasker (<http://www.repeatmasker.org/>). To reconstruct full-length LTR copies, we used the same strategy as done previously (Walter et al., 2016) using the perl tool “*One code to find them all*” (Bailly-Bechet et al., 2014). Reconstructed transposon annotation was used in the following analyses.

### **RNA-seq analysis**

Paired-end reads were trimmed with Trim Galore (v0.4.4, <https://github.com/FelixKrueger/TrimGalore>) to remove adapter sequences and Ns nucleotides from both side of the read. Cleaned reads were aligned onto the Mouse reference genome mm10 with STAR (v2.7.0a) (Dobin et al., 2013) allowing at most 6% of mismatches (`--outFilterMismatchNoverLmax 0.06`), reporting randomly one position among at most 5000 positions for a read (`--outFilterMultimapNmax 5000 --outSAMmultNmax 1`). Reconstructed transposon annotations and evidence-based gene annotation from GENCODE v25 were merged and used as input for quantification. Family-based transposon and gene expression were quantified using the previous mapped reads with FeatureCounts (v.1.5.1, parameters: `-M -p -C -s 2`) (Liao et al., 2014). Copy-based transposon quantification was performed using only uniquely-mapped reads (filtering keeping NH tag equal to 1) with FeatureCounts and the same parameters as previously. Bigwig files were produced with deepTools (v2.4.1) (Ramírez et al., 2016) using the option `--normalizeUsingRPKM`. Differential analysis was performed between samples sequenced on the same flow cell. For random alignment data, transcripts with more than 1 CPM in at least one sample were selected for EdgeR CPM normalization, followed by differential expression analysis using Voom transformation from LimmaR package (Robinson et al., 2010). We considered differentially expressed transcripts with  $|\log_2FC| > 1$  and  $FDR < 0.01$ . Heatmap displaying  $\log_2CPM$  or  $\log_2(CPM+1)$  uses non-normalized CPM data. For unique alignment data, expression threshold was raised to 1.5 CPM. Selected transcripts for each sequencing run were gathered and formed the starting point for differential analysis, performed then on the same amount and transcript diversity for each mutant.

### **Chimeric TE-gene transcript analysis**

To increase the detection of spliced RNA-seq reads, we applied the two-pass alignment strategy with STAR using parameters and methods developed in Modzelewski et al., 2021. The second pass mapping was used to reconstruct the transcript with StringTie2 (Kovaka et al., 2019) for each sample. A unified non-redundant set of isoforms was generated using all samples with StringTie2 (option `--merge`). Mono-exonic transcripts were removed. First exons were overlapped with repeat annotations. TE-overlapping first and second exons were used as input for quantification with FeatureCounts. Differential expression analysis was performed using edgeR's normalization combined with voom transformation from limma R package (Robinson et al., 2010). Chimeric transcripts were declared as differentially expressed if  $FDR < 0.01$  and  $\log_2FC > 1$ .

### **ChIP-seq analysis**

Paired-end reads were trimmed with Trim Galore (v0.4.4) to remove adapter sequences and Ns nucleotides from both side of the read. Cleaned reads were aligned onto the Mouse reference genome mm10 with STAR (v2.7.0a, parameters: `--outFilterMultimapNmax 1 --alignEndsType Local --outFilterMismatchNmax 999 --`

outFilterMismatchNoverLmax 0.06 --alignIntronMax 1 --alignMatesGapMax 2000) reporting only uniquely mapped reads. Duplicated reads were removed using Picard (v2.6.0). Bigwig files for UCSC genome browser and heatmaps were created with deepTools (v2.5.3). Peak calling was performed using MACS2 (v2.2.7.1) with a q-value threshold of 0.05 and broad option for H3K9me3 samples, discarding the ones that overlap with blacklisted regions (Amemiya et al., 2019). Common peaks between biological replicates were used for the following analyses. Peaks were quantified using the *regionCounts* function with mapping quality threshold of 20. Normalisation factors were calculated using 10kb-long bins to remove composition biases. Differential binding analysis was performed with csaw (Lun & Smyth, 2016). Peaks with a FDR<0.05 were declared as differentially bound. KAP1 ChIP-seq samples were downloaded from GEO database: GSM1406445 (Castro-Diaz et al., 2014), GSM2473148 and GSM2473150 (De Iaco et al., 2017) and analyzed as previously mentioned. The two last samples belonged to the same study so they were averaged. Cleaned reads were also aligned onto the Mouse (Langmead & Salzberg, 2012) in end-to-end and very sensitive mode in order to report randomly one position per read. After removing PCR duplicates as done previously, family-based transposons were quantified with FeatureCounts.

### **EM-seq analysis**

Paired-end reads were trimmed with Trim Galore (v0.6.5) with the following parameters: --clip\_R1 10 --clip\_R2 10 --three\_prime\_clip\_R1 10 --three\_prime\_clip\_R2 10 --length 15, as recommended in the user guide of Bismark. Cleaned reads were aligned onto the Mouse reference genome (mm10) using Bismark v0.23.0 (Krueger & Andrews, 2011) with Bowtie2-2.2.9 and allowing one mismatch in the seed. Methylation calls were extracted after duplicate removal. Biological replicates were merged. Only CG dinucleotides covered by a minimum of 5 reads were conserved for the rest of the analysis. CpG islands (“CGIs”) were defined previously (Illingworth et al., 2010). Intragenic compartments were defined as non-CGI promoters, exonic and intronic annotations from GENCODE v25. Intergenic partitions were defined as genomic regions that did not overlap with CGIs, intragenic regions or repeats. Differentially Methylated Region (DMR) calling was performed using the bioconductor package DSS (Feng et al., 2014) with the following parameters: a CpG methylation level difference of at least 20%, at least five CpGs called, minimum length of 200 bp, and at least 500 bp between two DMRs.

### **Protein immunoprecipitation**

Equal amount of nuclear protein extracts (1mg for mass spectrometry) were diluted to 250mM cations (Na<sup>+</sup>/K<sup>+</sup>) with BC0 (20mM Tris-HCl pH7.9, 2mM EDTA, 10% Glycerol, 1mM DTT) and pre-cleared for 4 hours at 4°C with 1%BSA-PBS blocked Dynabeads Protein A (Invitrogen). The SPIN1 antibody (Franz et al., 2015) was bound to blocked beads for 4 hours at 4°C on a rotating wheel. After pre-clearing, 10% of volume was set aside as Input and the remaining protein extract was incubated O/N with SPIN1-ProteinA beads at 4°C. After 4x washes with BC250 (BC0 + 250mM KCl), proteins

were either eluted in 0.2M Glycine pH2.6 for 10 min on ice or went through sample preparation for liquid chromatography coupled to mass spectrometry (LC-MS/MS). When eluted, reaction was stopped with 1.5M Tris-HCl pH8.8 and denatured at 95°C in Laemmli buffer.

### **LC-MS/MS sample preparation**

Beads were washed three times with 100 µL of 100 mM NH<sub>4</sub>HCO<sub>3</sub>, were resuspended in 100 µL of 100 mM NH<sub>4</sub>HCO<sub>3</sub> and digested directly by adding 0.2 µg of trypsin/LysC (Promega) for 1 h at 37°C. Samples were then loaded onto custom-made C18 StageTips packed by stacking three AttractSPE® disk (#SPE-Disks-Bio-C18-100.47.20 Affinisep) into a 200 µL micropipette tip for desalting. Peptides were eluted using a ratio of 40:60 CH<sub>3</sub>CN:H<sub>2</sub>O + 0.1% formic acid and vacuum concentrated to dryness. Peptides were reconstituted in injection buffer (10µl of 0.3% trifluoroacetic acid (TFA) [vol/vol]) before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

### **LC-MS/MS analysis**

Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Scientific). Peptides were trapped on a C18 column (i.d. 75 µm x 2 cm; nanoViper Acclaim PepMap™ 100, Thermo Scientific) at a flow rate of 3.0 µL/min in buffer A (2/98 MeCN/H<sub>2</sub>O in 0.1% formic acid) over 4 min for desalting and concentration of the samples. Separation was performed on a 50 cm x 75 µm C18 column (nanoViper Acclaim PepMap™ RSLC, 2 µm, 100Å, Thermo Scientific) regulated to a temperature of 50°C with a linear gradient from 2% to 25% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nl/min over 91 min. MS1 data were collected in the Orbitrap (120,000 resolution at *m/z* 200; maximum IT 60 ms; AGC target 4 x 10<sup>5</sup>). Charge states between 2 and 7 were required for MS2 analysis, and a 60 sec dynamic exclusion window was used. MS2 scan were performed in the ion trap in rapid mode with HCD fragmentation (isolation window 1.2 Da; NCE 30%; maximum IT 60 ms; AGC target 10<sup>4</sup>)

### **Data processing**

For identification, the data were searched against the *Mus Musculus* (UP000000589\_10090) UniProt databases using Sequest HT through proteome discoverer (version 2.4). Enzyme specificity was set to trypsin and a maximum of two miss cleavages sites were allowed. Oxidized methionine, Met-loss, Met-loss-Acetyl and N-terminal acetylation were set as variable modifications. Carbamidomethylation of cysteines were set as fixed modification. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da for MS/MS peaks. The resulting files were further processed using myProMS v3.9.3 (<https://github.com/bioinfo-pf-curie/myproms>) (Pouillet et al., 2007). FDR calculation used Percolator (The et al., 2016) and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs),

reextracted across all conditions and computed with MassChroQ version 2.2.21 (Valot et al., 2011). For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) and missed cleavages were allowed. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate (N=5 in each condition). To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed, and p-values were adjusted using the Benjamini–Hochberg FDR procedure. Proteins with at least 2 distinct peptides in 3 replicates of a same state, a 2-fold enrichment and an adjusted p-value  $\leq 0.05$  were considered significantly enriched in sample comparisons. Proteins uniquely present in WT samples were also considered if they matched the peptides criteria.

TARGET	FORWARD	REVERSE	APPLICATION
<i>sgRNA Spin1</i>	CACCGATCCAGCATGGATGGAGAGA	AAACTCTCTCCATCCATGCTGGATC	CRISPR-Cas9 Spin1-KO
<i>sgRNA Spindoc_ex2</i>	CACCGCCACCTCGGCCTAACCTGC	AAACGCAGGTTAGGCCGAGGTGGC	CRISPR-Cas9 Spindoc-KO
<i>sgRNA Spindoc_ex4</i>	CACCGTTGCTAGGCCTTCCAACGCT	AAACAGCGTTGGAAGGCCTAGCAAC	CRISPR-Cas9 Spindoc-KO
<i>Spin1</i>	ACATCTGCCTGGATCAGTTC	CTGCCTACCAGTAGGAGTT	Genotyping PCR
<i>Spindoc</i>	GGGGTGGGACCATGACTAGA	CACCAGAGTGACGGAAGCTC ACATCTGCCTGGATCAGTTC	Genotyping PCR 3 primers
<i>IAP_seq</i>	TTTTTATTTTATGTGTTTTGTTTT		Pyrosequencing
<i>L1-A_seq</i>	GGTATATAGGGAAGTAGGTTA		Pyrosequencing
<i>L1-Tf_seq</i>	GGTATATAGGGAAGTAGGTTA		Pyrosequencing
<i>IAP</i>	GAGGGTGGTTTTTATTTTATGTGT	ATCACTCCCTAATTAACAACCC	Pyrosequencing PCR
<i>L1-A</i>	AGATTGAGGTATATAGGGAAGTAGGTT	ATCCACTCACCAAAAATCTTAAAAT	Pyrosequencing PCR
<i>L1-Tf</i>	GGTTGGGGAGGAGGTTAAGTTATA	CTACCTATCCAAAACTATCAAATTCTCT	Pyrosequencing PCR
<i>45S</i>	CTCTTAGATCGATGTGGTGCTC	GCCCGCTGGCAGAACGAGAAG	RT-qPCR
<i>5'-ETS1</i>	TGTTTCACTTTGGTCGTGTCTC	TCGACGTTACAAGAAACAGC	RT-qPCR
<i>5'-ETS2</i>	GTCTTCTGGTTCCCTGTGTG	GCTAGAGAAGGAACTTTCTCACTG	RT-qPCR
<i>Actin</i>	AAGTGACGTTGACATCCG	GATCCACATCTGCTGGAAGG	RT-qPCR
<i>Gapdh</i>	TCACCACCATGGAGAAGGC	GCTAAGCAGTTGGTGGTGCA	RT-qPCR
<i>IAP-LTR</i>	CGCATGTGCCAAGGGTATCT	GAGCATTCTCCTTCGCCTCG	RT-qPCR
<i>IAPEZd1</i>	AACGCTGCTGCTTTAACTCC	TGCACATAAAGCTGGCACA	RT-qPCR
<i>ITS1</i>	TCTCGTTTCGTTCCCTGCTGG	GATCCACCGCTAAGAGTCGTATC	RT-qPCR
<i>ITS2</i>	CGTGTGAGTAAGATCCTCCAC	GTTACTGAGGGAATCCTGGTTAG	RT-qPCR
<i>L1-A</i>	GGATTCCACACGTGATCCTAA	TCCTCTATGAGCAGACCTGGA	RT-qPCR
<i>L1-Tf</i>	CAGCGTTCGCCATCTTG	CACCCTCTACCTGTTTACAGACTAA	RT-qPCR

**Table 1 | Oligonucleotides used and their applications**

TARGET	REFERENCE	FEATURES	APPLICATION	DILUTION
<b>SPIN1</b>	Spin1-5867 (Franz et al., 2015)	Rabbit polyclonal	Western blot Immunoprecipitation	1/1000, 1/8
<b>SPINDOC</b>	HPA040128	Rabbit polyclonal	Western blot	1/250
<b>LAMINB1</b>	ab16048	Rabbit polyclonal	Western blot, Immunofluorescence	1/10000, 1/2000
<b>PCNA</b>	CST 2586	Mouse polyclonal	Western blot	1/2000
<b>UHRF1</b>	sc373750	Rabbit polyclonal	Western blot	1/1000
<b>DNMT3A</b>	ab188470	Rabbit polyclonal	Western blot	1/2000
<b>DNMT3B</b>	ab122932	Rabbit polyclonal	Western blot	1/1000
<b>DNMT1</b>	CST 50325	Rabbit polyclonal	Western blot	1/1000
<b>H3K9ME3</b>	ab176916	Rabbit monoclonal	Western blot, ChIP	1/1000, 5µg
<b>H3K4ME3</b>	active motif 39159	Rabbit polyclonal	Western blot, ChIP	1/1000, 5µg
<b>H3</b>	ab1791	Rabbit polyclonal	Western blot	1/1000
<b>FLAG-M2</b>	F1804	Mouse monoclonal	Western blot	1/1000

**Table 2 | Antibodies used with their application and working conditions.**

ID	RATIO	LOG2(RATIO)	ADJ. P-VALUE
Q61142	1000	1000	
Q05AH6	1000	1000	
Q8BG79	1000	1000	
Q64368	1000	1000	
Q35144	1000	1000	
Q80XP8	1000	1000	
P58462	1000	1000	
E9Q6J5	1000	1000	
Q80Y84	1000	1000	
Q9CWK3	1000	1000	
Q9QXK2	1000	1000	
E9Q7E2	1000	1000	
Q7TQK1	1000	1000	
Q91231	1000	1000	
A2A791	1000	1000	
Q8BPM2	1000	1000	
Q8C3X4	1000	1000	
Q8VI33	1000	1000	
Q68ED3	1000	1000	
Q9D4H9	1000	1000	
G5E8V9	1000	1000	
Q61624	3,673411	1,87712	6,93E-05
P58501	3,5854	1,842134	7,32E-06
A2RTL5	3,520052	1,815597	1,27E-06
A2AU4	3,239504	1,695773	0,000173
P70399	3,202407	1,679156	2,49E-08
Q91W36	3,187411	1,672385	9,96E-06
A0A1D9BZF0	3,051393	1,609468	2,09E-08
Q8C0T5	2,94583	1,558674	0,000651
Q09XV5	2,886061	1,529102	1,09E-13
Q9CWL8	2,867962	1,520026	0,017613
Q3U319	2,841943	1,506878	0,000195
P61924	2,728555	1,448137	8,76E-06
Q8CCG4	2,728248	1,447975	2,09E-06
Q9D7N3	2,712037	1,439377	1,18E-05
Q60953	2,700026	1,432973	7,83E-08
Q920Q8	2,693086	1,42926	0,001086
Q8K284	2,654351	1,408359	6,36E-05
Q9QWV9	2,650345	1,40618	8,37E-05
Q921Y2	2,620764	1,389988	0,003444
Q9JKY0	2,618886	1,388953	0,004089
Q8K368	2,61574	1,387219	0,002448
O54774	2,578697	1,366642	0,009087
Q8BYH8	2,576822	1,365593	0,024286
Q6PDL0	2,575712	1,364971	0,001867
P56183	2,565506	1,359244	0,000101
Q9DBZ5	2,509144	1,327195	0,017275
Q9QXK3	2,485598	1,313593	3,99E-10
Q8K205	2,484538	1,312977	0,000475
Q61879	2,476915	1,308544	6,33E-09
Q8K2A7	2,466174	1,302275	0,000866
O89090	2,464162	1,301097	0,005648
B9EKI3	2,430426	1,281209	0,002635
O35345	2,398409	1,262078	0,001754
P55194	2,389658	1,256804	0,00479
A2AF47	2,383802	1,253264	0,01552
Q9Z2I9	2,381168	1,25167	0,030186
P51660	2,378048	1,249778	8,74E-07
Q9WV60	2,361346	1,23961	0,043186
Q6ZQ29	2,359629	1,23856	0,017376
Q9DBL1	2,353226	1,23464	0,014763
Q9CYI4	2,336514	1,224358	1,07E-06
Q9DBU6	2,312898	1,209701	0,002267
Q8R0L9	2,307694	1,206452	6,87E-05
Q8QZT1	2,283248	1,191088	0,002723
P62827	2,2689	1,181993	6,9E-10
Q8VE97	2,268872	1,181975	0,001469
Q80V86	2,266766	1,180636	0,001247
Q922E6	2,256692	1,17421	0,009517
Q3URQ0	2,255194	1,173251	0,000887
Q9Z0W3	2,239396	1,16311	8,13E-05
Q4VBE8	2,22675	1,154939	0,011735
Q8K4F6	2,222742	1,152341	0,005376
Q6KCD5	2,220657	1,150987	2,77E-08
Q62419	2,219991	1,150554	7,51E-05
Q6ZPL9	2,201236	1,138314	0,045369
Q8BWY3	2,19626	1,135049	0,002053
Q9CWX9	2,194563	1,133934	9,14E-06
Q9WTX8	2,186677	1,12874	0,016989
Q91YJ2	2,177328	1,122559	0,001343
Q6DID5	2,175864	1,121588	0,00013
Q8K202	2,164567	1,114078	0,00633
Q3TKT4	2,164175	1,113817	3,19E-15
Q9DBR3	2,147981	1,102981	0,000898
Q60848	2,142646	1,099393	1,33E-12
Q9EQM6	2,142108	1,099031	0,000575
Q5SVQ0	2,135162	1,094346	5,32E-06
O35130	2,133687	1,093348	0,000548
Q9JIF7	2,133672	1,093338	4,04E-16
Q4FK66	2,129439	1,090474	0,00875
Q00899	2,120734	1,084564	0,000417
Q6PIJ4	2,116748	1,08185	5,54E-08
Q61103	2,11542	1,080944	1,86E-07
O54950	2,1132	1,079429	0,000665
P54823	2,093656	1,066024	1,35E-05
B2RXQ2	2,087365	1,061683	0,000916
Q9ESZ8	2,079936	1,056539	3,2E-13
Q99NA9	2,072941	1,051679	0,003601
Q8CIN4	2,06738	1,047803	0,038537
P52293	2,06176	1,043877	4,08E-07
A6PWY4	2,059306	1,042158	3,68E-09
Q8K363	2,053025	1,037751	1,38E-19
Q99PP7	2,04578	1,032651	0,00097
Q02614	2,04578	1,032651	0,000753
Q80UU2	2,043967	1,031372	0,00245
A2ANY6	2,038711	1,027657	6,24E-21
A2AWL7	2,031114	1,022271	2,09E-08
Q5SSK3	2,025891	1,018556	0,036562
Q9D0D4	2,021653	1,015535	0,02894
P32233	2,021592	1,015492	4,12E-13
Q921G8	2,020029	1,014376	0,018545
P30276	2,017845	1,012815	0,015441
Q71LX4	2,016076	1,01155	0,002139
Q91YT7	2,014798	1,010635	0,000103
P97496	2,010614	1,007636	1,9E-11
Q60973	2,008136	1,005857	0,007207
Q8BYK6	2,000714	1,000515	1,96E-06

**Extended Table 1** | List of proteins found enriched in WT cells compared to *Spin1*-KO after SPIN1-IP followed by LC/MS-MS ranked by enrichment ratio.



## REFERENCES

- Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports* 2019 9:1, 9(1), 1–5. <https://doi.org/10.1038/s41598-019-45839-z>
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., & Hannon, G. J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, 31(6), 785–799. <https://doi.org/10.1016/J.MOLCEL.2008.09.003>
- Bae, N., Gao, M., Li, X., Premkumar, T., Sbardella, G., Chen, J., & Bedford, M. T. (2017). A transcriptional coregulator, SPIN·DOC, attenuates the coactivator activity of Spindlin1. *Journal of Biological Chemistry*, 292(51), 20808–20817. <https://doi.org/10.1074/jbc.M117.814913>
- Bailly-Bechet, M., Haudry, A., & Lerat, E. (2014). “One code to find them all”: A perl tool to conveniently parse RepeatMasker output files. *Mobile DNA*, 5(1), 1–15. <https://doi.org/10.1186/1759-8753-5-13/TABLES/3>
- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, 354(6314), 909–912. <https://doi.org/10.1126/science.aah5143>
- Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martínez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., & Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, 500(7461), 222–226. <https://doi.org/10.1038/nature12362>
- Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004), 96–99. <https://doi.org/10.1038/nature02886>
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, 294(5551), 2536–2539. <https://doi.org/10.1126/SCIENCE.1065848>
- Burns, K. H. (2017). Transposable elements in cancer. In *Nature Reviews Cancer* (Vol. 17, Issue 7, pp. 415–424). <https://doi.org/10.1038/nrc.2017.35>
- Castro-Díaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., & Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes and Development*, 28(13), 1397–1409. <https://doi.org/10.1101/gad.241661.114>
- 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>
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics* 2017 49:6, 49(6), 941–945. <https://doi.org/10.1038/NG.3858>
- Delbarre, E., Ivanauskienė, K., Spirkoski, J., Shah, A., Vekterud, K., Moskaug, J., Bøe, S. O., Wong, L. H., Küntziger, T., & Collas, P. (2017). PML protein organizes heterochromatin domains where it regulates histone H3.3 deposition by ATRX/DAXX. *Genome Research*, 27(6), 913–921. <https://doi.org/10.1101/gr.215830.116>
- Devi, M. S., Meiguilungpou, R., Sharma, A. L., Anjali, C., Devi, K. M., Singh, L. S., & Singh, T. R. (2019). Spindlin docking protein (SPIN·DOC) interaction with SPIN1 (a histone code reader) regulates Wnt signaling. *Biochemical and Biophysical Research Communications*, 511(3), 498–503. <https://doi.org/10.1016/j.bbrc.2019.02.096>
- 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*, 29(1), 15–21. <https://doi.org/10.1093/BIOINFORMATICS/BTS635>
- Du, Y., & Qian, C. (2022). Non-canonical bivalent H3K4me3K9me3 recognition by Spindlin1/C11orf84 complex. <https://doi.org/10.1002/bies.202100229>
- Du, Y., Yan, Y., Xie, S., Huang, H., Wang, X., Ng, R. K., Zhou, M. M., & Qian, C. (2021). Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-21236-x>
- 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* 2015 522:7555, 522(7555), 240–244. <https://doi.org/10.1038/NATURE14345>
- Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, 42(8), e69–e69. <https://doi.org/10.1093/NAR/GKU154>

- Franz, H., Greschik, H., Willmann, D., Ozretić, L., Jilg, C. A., Wardelmann, E., Jung, M., Buettner, R., & Schüle, R. (2015). The histone code reader SPIN1 controls RET signaling in liposarcoma. *Oncotarget*, *6*(7). <https://doi.org/10.18632/oncotarget.3000>
- Fulka, H., & Langerova, A. (2014). The maternal nucleolus plays a key role in centromere satellite maintenance during the oocyte to embryo transition. *Development*, *141*(8), 1694–1704. <https://doi.org/10.1242/DEV.105940>
- Gagnier, L., Belancio, V. P., & Mager, D. L. (2019). Mouse germ line mutations due to retrotransposon insertions. In *Mobile DNA* (Vol. 10, Issue 1, pp. 1–22). Mobile DNA. <https://doi.org/10.1186/s13100-019-0157-4>
- Goodier, J. L. (2016). Restricting retrotransposons: a review. *Mobile DNA*, *7*(1). <https://doi.org/10.1186/S13100-016-0070-Z>
- Graf, U., Casanova, E. A., Wyck, S., Dalcher, D., Gatti, M., Vollenweider, E., Okoniewski, M. J., Weber, F. A., Patel, S. S., Schmid, M. W., Li, J., Sharif, J., Wanner, G. A., Koseki, H., Wong, J., Pelczar, P., Penengo, L., Santoro, R., & Cinelli, P. (2017). Pramel7 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. *Nature Cell Biology*, *19*(7), 763–773. <https://doi.org/10.1038/ncb3554>
- Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(October). <https://doi.org/10.1038/s41580-019-0159-6>
- Greschik, H., Duteil, D., Messaddeq, N., Willmann, D., Arrigoni, L., Sum, M., Jung, M., Metzger, D., Manke, T., Günther, T., & Schüle, R. (2017). The histone code reader Spin1 controls skeletal muscle development. *Cell Death & Disease*, *8*(1), 1–15. <https://doi.org/10.1038/cddis.2017.468>
- Groh, S., Milton, A. V., Marinelli, L. K., Sickinger, C. V., Russo, A., Bollig, H., de Almeida, G. P., Schmidt, A., Forné, I., Imhof, A., & Schotta, G. (2021). Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nature Communications* *2021* *12*:1, *12*(1), 1–18. <https://doi.org/10.1038/S41467-021-26288-7>
- Groh, S., & Schotta, G. (2017). Silencing of endogenous retroviruses by heterochromatin. *Cellular and Molecular Life Sciences*, *74*(11), 2055–2065. <https://doi.org/10.1007/s00018-017-2454-8>
- Han, M., Li, J., Cao, Y., Huang, Y., Li, W., Zhu, H., Zhao, Q., Han, J. D. J., Wu, Q., Li, J., Feng, J., & Wong, J. (2020). A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. *Nucleic Acids Research*, *48*(21), 12116–12134. <https://doi.org/10.1093/NAR/GKAA1003>
- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLOS Genetics*, *6*(9), e1001134. <https://doi.org/10.1371/JOURNAL.PGEN.1001134>
- Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., & Lorincz, M. C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs. *Cell Stem Cell*, *8*(6), 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>
- Kazazian, H. H., & Moran, J. V. (2017). Mobile DNA in Health and Disease. *New England Journal of Medicine*, *377*(4), 361–370. [https://doi.org/10.1056/NEJMRA1510092/SUPPL\\_FILE/NEJMRA1510092\\_DISCLOSURES.PDF](https://doi.org/10.1056/NEJMRA1510092/SUPPL_FILE/NEJMRA1510092_DISCLOSURES.PDF)
- Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, *20*(1), 1–13. <https://doi.org/10.1186/S13059-019-1910-1/FIGURES/6>
- Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572. <https://doi.org/10.1093/BIOINFORMATICS/BTR167>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., & Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14075–14080. <https://doi.org/10.1073/PNAS.1008850107/-/DCSUPPLEMENTAL>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/BIOINFORMATICS/BTT656>
- 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* 2021 591:7849, 591(7849), 322–326. <https://doi.org/10.1038/S41586-021-03313-9>
- Lu, R., & Wang, G. G. (2014). Tudor: a versatile family of histone methylation “readers” Histone modification and its “reader” proteins in gene regulation. *38*(11), 1–18. <https://doi.org/10.1016/j.tibs.2013.08.002>
- Lun, A. T. L., & Smyth, G. K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Research*, 44(5), e45–e45. <https://doi.org/10.1093/NAR/GKV1191>
- Modzelewski, A. J., Gan Chong, J., Wang, T., & He, L. (2022). Mammalian genome innovation through transposon domestication. *Nature Cell Biology* 2022 24:9, 24(9), 1332–1340. <https://doi.org/10.1038/S41556-022-00970-4>
- Modzelewski, A. J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., Tjokro, K., Sales, G., Biton, A., Anand, A., Speed, T. P., Xuan, Z., Wang, T., Risso, D., & He, L. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*, 184(22), 5541–5558.e22. <https://doi.org/10.1016/J.CELL.2021.09.021>
- Oh, B., Hwang, S.-Y., Solter, D., & Knowles, B. B. (1997). Spindlin, a major maternal transcript expressed in the mouse during transition from oocyte to embryo. *Development*, 124, 493–503. <http://dev.biologists.org.ezproxyhost.library.tmc.edu/content/development/124/2/493.full.pdf>
- Popova, K., Rybnikova, E., Chesnokova, E., Beletskiy, A., & Kolosov, P. (2022). *The Role of Transposable Elements of the Human Genome in Neuronal Function and Pathology*. <https://doi.org/10.3390/ijms23105847>
- Pouillet, P., Carpentier, S., & Barillot, E. (2007). myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *PROTEOMICS*, 7(15), 2553–2556. <https://doi.org/10.1002/PMIC.200600784>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/NAR/GKW257>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/BIOINFORMATICS/BTP616>
- 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), 237–240. <https://doi.org/10.1038/nature08674>
- Rushton, M. D., Saunderson, E. A., Patani, H., Green, M. R., & Ficzi, G. (2022). An shRNA kinase screen identifies regulators of UHRF1 stability and activity in mouse embryonic stem cells. *Epigenetics*, 17(12), 1590–1607. <https://doi.org/10.1080/15592294.2022.2044126>
- Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, 16(8), 919–932. <https://doi.org/10.1101/GAD.973302>
- Shanle, E. K., Shinsky, S. A., Bridgers, J. B., Bae, N., Sagum, C., Krajewski, K., Rothbart, S. B., Bedford, M. T., & Strahl, B. D. (2017). Histone peptide microarray screen of chromo and Tudor domains defines new histone lysine methylation interactions. *Epigenetics & Chromatin*, 10(1), 12. <https://doi.org/10.1186/s13072-017-0117-5>
- Sharif, J., Endo, T. A., Nakayama, M., Karimi, M. M., Shimada, M., Katsuyama, K., Goyal, P., Brind’Amour, J., Sun, M. A., Sun, Z., Ishikura, T., Mizutani-Koseki, Y., Ohara, O., Shinkai, Y., Nakanishi, M., Xie, H., Lorincz, M. C., & Koseki, H. (2016). Activation of Endogenous Retroviruses in Dnmt1<sup>-/-</sup> ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell*, 19(1), 81–94. <https://doi.org/10.1016/j.stem.2016.03.013>
- The, M., MacCoss, M. J., Noble, W. S., & Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *Journal of the American Society for Mass Spectrometry*, 27(11), 1719. <https://doi.org/10.1007/S13361-016-1460-7>
- Valot, B., Langella, O., Nano, E., & Zivy, M. (2011). MassChroQ: A versatile tool for mass spectrometry quantification. *PROTEOMICS*, 11(17), 3572–3577. <https://doi.org/10.1002/PMIC.201100120>
- von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., Kroeze, L. I., Nesterova, T. B., Jansen, J. H., Xie, H., He, C., Reik, W., & Stunnenberg, H. G. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of

- Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, 62(6), 848–861. <https://doi.org/10.1016/j.molcel.2016.04.025>
- Walter, M., Teissandier, A., Pérez-Palacios, R., Bourc'His, D., & Pe, R. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, 5(JANUARY2016), 1–30. <https://doi.org/10.7554/eLife.11418.001>
- Wang, J.-X., Zeng, Q., Chen, L., Du, J.-C., Yan, X.-L., Yuan, H.-F., Zhai, C., Zhou, J.-N., Jia, Y.-L., Yue, W., & Pei, X.-T. (2012). SPINDLIN1 Promotes Cancer Cell Proliferation through Activation of WNT/TCF-4 Signaling. *Molecular Cancer Research*, 10(3), 326–335. <https://doi.org/10.1158/1541-7786.MCR-11-0440>
- Wang, W., Chen, Z., Mao, Z., Zhang, H., Ding, X., Chen, S., Zhang, X., Xu, R., & Zhu, B. (2011). Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. *EMBO Reports*, 12(11), 1160–1166. <https://doi.org/10.1038/embor.2011.184>
- Wolf, D., & Goff, S. P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242), 1201–1204. <https://doi.org/10.1038/nature07844>
- 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* 2021 591:7849, 591(7849), 317–321. <https://doi.org/10.1038/S41586-021-03210-1>
- Yang, F., Chen, J., Liu, B., Gao, G., Sebastian, M., Jeter, C., Shen, J., Person, M. D., & Bedford, M. T. (2021). SPINDOC binds PARP1 to facilitate PARylation. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-26588-y>
- Yang, N., Wang, W., Wang, Y., Wang, M., Zhao, Q., Rao, Z., Zhu, B., & Xu, R. M. (2012). Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44). <https://doi.org/10.1073/pnas.1208517109>
- Yang, P., Wang, Y., & Macfarlan, T. S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. In *Trends in Genetics* (Vol. 33, Issue 11, pp. 871–881). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2017.08.006>
- Zhang, X., Zhu, G., Su, X., Li, H., & Wu, W. (2018). Nucleolar localization signal and histone methylation reader function is required for SPIN1 to promote rRNA gene expression. *Biochemical and Biophysical Research Communications*, 505(1), 325–332. <https://doi.org/10.1016/j.bbrc.2018.09.098>
- Zhao, F., Liu, Y., Su, X., Lee, J.-E., Song, Y., Wang, D., Ge, K., Gao, J., Zhang, M. Q., & Li, H. (2020). Molecular basis for histone H3 “K4me3-K9me3/2” methylation pattern readout by Spindlin1. *Journal of Biological Chemistry*, 295, 16877–16887. <https://doi.org/10.1074/jbc.ra120.013649>
- Zoch, A., Auchynnikava, T., Berrens, R. V., Kabayama, Y., Schöpp, T., Heep, M., Vasiliauskaitė, L., Pérez-Rico, Y. A., Cook, A. G., Shkumatava, A., Rappsilber, J., Allshire, R. C., & O'Carroll, D. (2020). SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature*, 584(7822), 635–639. <https://doi.org/10.1038/s41586-020-2557-5>

## II. The biological relevance of SPINDOC in the mouse

At the beginning of my PhD, the effects of *Spindoc* deletion had not been studied *in vivo*, and in particular, in a mouse model. Due to the role of SPINDOC in TE regulation in ESCs and its reported high expression in the testis, I aimed at investigating the involvement of SPINDOC in TE regulation in male germ cells. I was particularly interested by a potential meiotic phenotype, in regards to the systematic meiotic arrest that results from failure to silence TEs during spermatogenesis (Zamudio & Bourc'his, 2010). This project started with the generation of a *Spindoc*-KO mouse line with the help of the transgenesis platform of Institut Curie, led by Fatima El Marjou. Here after follows a characterization of this mouse model, as a first step to understand the biological relevance of SPINDOC, which might be independent of SPIN1. *Spindoc*<sup>-/-</sup> mice showed no major deleterious phenotype, although I quantified a male-specific weight defect (helped by Mélanie Armand, engineer) and a minor sub-fertility for both sexes. Accordingly, IAP and LINE-1 expression was comparable in the testis of *Spindoc*<sup>+/+</sup> and *Spindoc*<sup>-/-</sup> littermates. In the meantime, two publications came out describing the phenotypic impact of SPINDOC (Jiang et al., 2021; F. Yang et al., 2021). Their observations globally corroborate mine, although some differences will be discussed in the following pages.

## RESULTS

### 1. SPIN1 and SPINDOC expression pattern

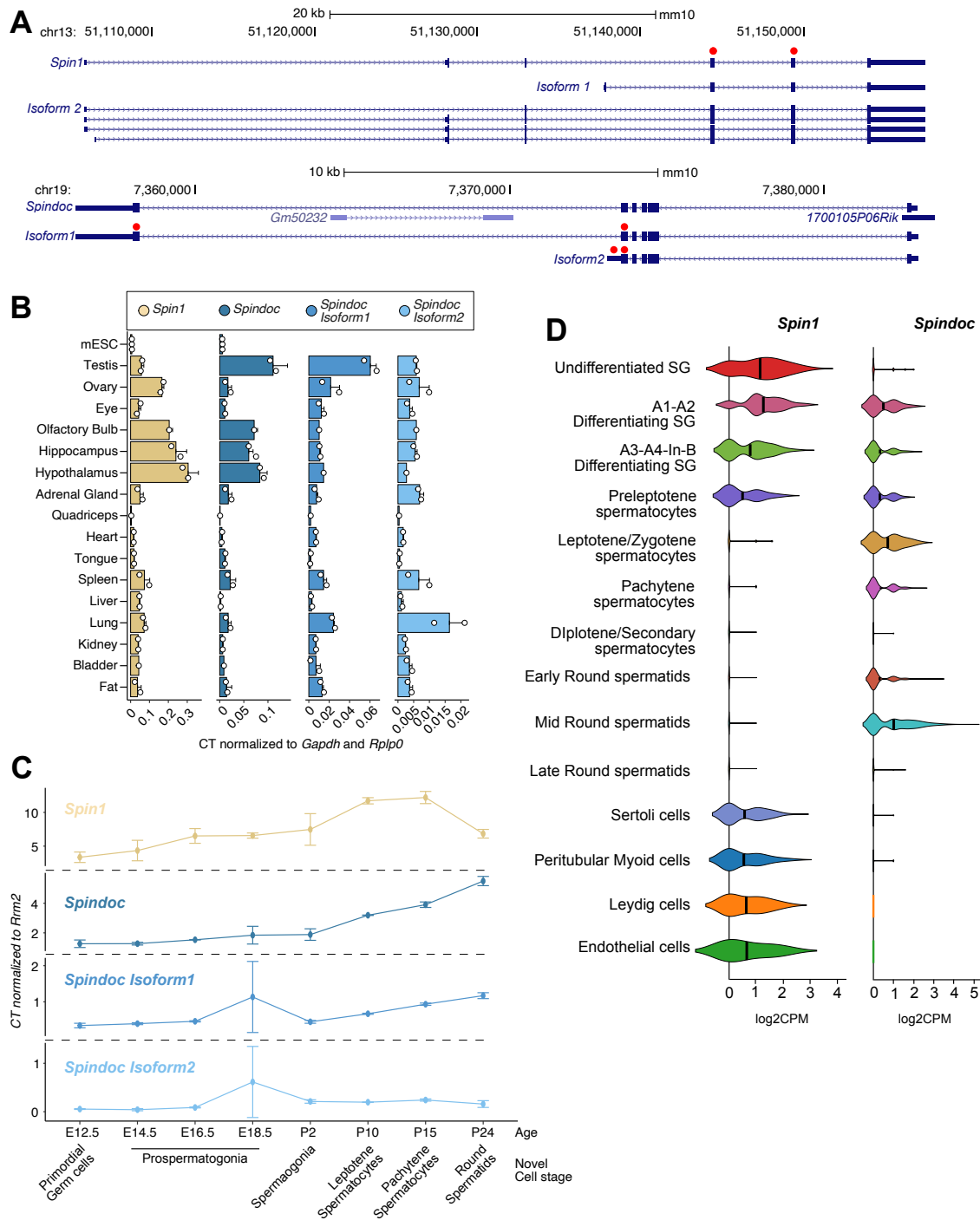
As a first insight into the biological role of SPINDOC *in vivo*, I went on investigating more closely its expression pattern. As the only reported function of this protein is to block SPIN1-mediated transcriptional activation, I profiled *Spin1* and *Spindoc* expression together across tissues. As mentioned in the Introduction, two *Spin1* transcript isoforms exist: a long isoform (isoform 2) with a rather ubiquitous expression and a short isoform (isoform 1) specifically expressed in oocytes (K.-M. Zhang et al., 2008). Two transcripts for *Spindoc* are also reported in the mouse reference Genome browser (C57Bl6/J background) (Figure 14A). To characterize as precisely as possible *Spindoc* and *Spin1* expression pattern, I performed an assay with several qPCR primer pairs (targeting exon junctions) specific for each isoform and each gene target. However, after several attempts, I did not manage to obtain primer pairs amplifying *Spin1* isoform1 specifically. Therefore, I used primers amplifying both isoforms.

I made use of a cDNA tissue bank already available in the lab to measure total *Spin1*, total *Spindoc*, *Spindoc* isoform 1 and *Spindoc* isoform 2 expression (Figure 14B). These results globally confirmed available mouse and human expression datasets (BioGPS and GTex): *Spin1* was highly expressed in ovaries and I also reported high expression in brain tissues. *Spindoc* expression was overall less important than *Spin1*, with the highest expression in the testis (where *Spin1* is also expressed) followed by several brain tissues (Figure 1B). Because of the importance of TE regulation in male germline development (Aravin et al., 2007; Barau et al., 2016; Bourc'his & Bestor, 2004; Zoch et al., 2020), I performed a timeline of *Spin1* and *Spindoc* expression in genital ridges from embryonic day 12.5 (E12.5) to testis at 24 days post-partum (P24) with intermediate timepoints coinciding with key differentiation steps of male germline development (Figure 14C). *Spin1* expression increased gradually in fetal and post-natal life with the highest level at P10 and P15 before decreasing in P24. The P10 to P15 points correspond to the onset and end point of the prophase of the first meiosis during the first spermatogenic wave after birth. *Spindoc* was globally less expressed than *Spin1*, and started increasing only after birth but until P24, with the same trend for both isoforms.

To better characterize the cellular expression of *Spin1* and *Spindoc* in the male germ line, I relied on single cell RNA-seq data from [Hermann et al., 2018](#), which are accessible through a software developed by 10X Genomics, without the need to download or compute data ([Figure 14D](#)). In this dataset, *Spin1* was the most expressed at the beginning of spermatogenesis, in undifferentiated spermatogonial stem cells, differentiating spermatogonia and just prior to meiosis in pre-leptotene cells, and also, but to a lower extent, in several somatic cells of the testis. The lack expression in meiotic cells indicates that my quantification on whole testes at P10-P15 rather measured expression from somatic testicular cells at this stage. Interestingly, *Spindoc* had a distinct cellular specificity, being expressed globally later in spermatogenesis than *Spin1*, in early meiotic cells and then again, in spermatids, and not in somatic testicular cells. This non-overlapping pattern indicates that SPINDOC may exert some functions independent from its interaction with SPIN1, at least in the context of spermatogenesis.

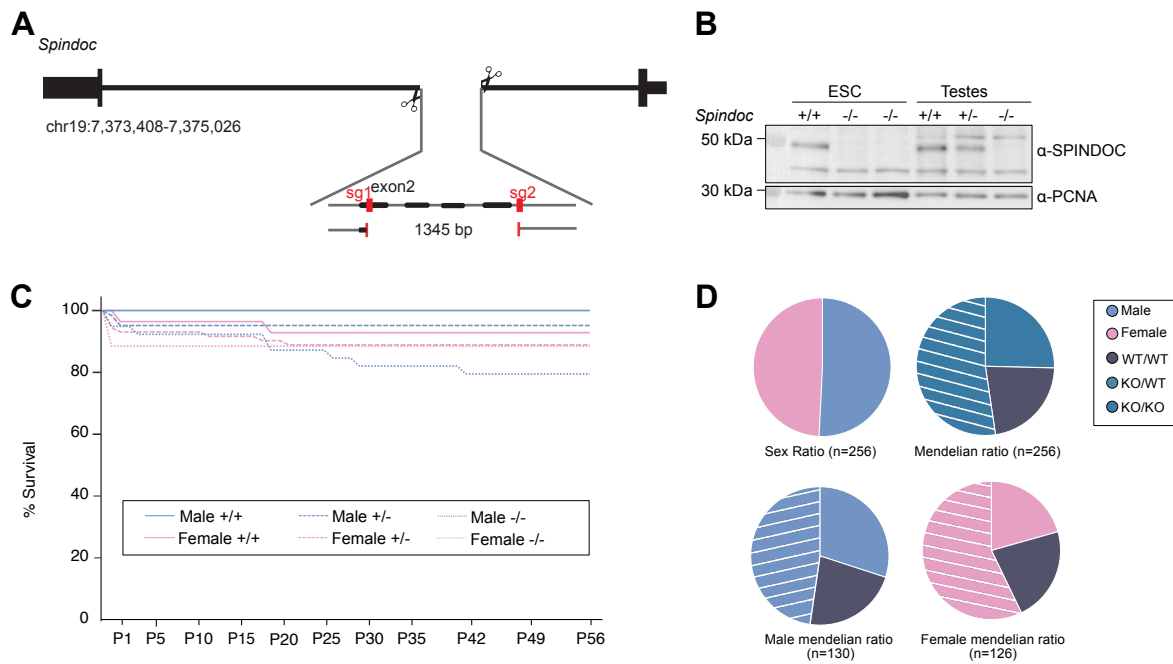
## 2. Generation of a *Spindoc*-KO mouse line

The constitutive deletion of *Spin1* leads to early neonatal (P2) lethality in the mouse, a phenotype that prevents studying the whole spermatogenesis differentiation process after birth in this mutant, and in particular meiosis ([Greschik et al., 2017](#); [Oh et al., 1997](#)). Moreover, from my previous expression analysis, it was clear that only *Spindoc*, not *Spin1*, is expressed in meiotic cells. I went on generating a constitutive *Spindoc*-KO mouse line. The same sgRNAs were used as the ones in ESCs, leading to a 4-exon deletion ([Figure 15A, 15B](#)). As opposed to constitutive *Spin1*-KO, *Spindoc*<sup>-/-</sup> animals were viable. However, *Spindoc*<sup>-/-</sup> had a higher chance of lethality compared to *Spindoc*<sup>+/+</sup> littermates, especially males after weaning (P20) ([Figure 15C](#)). Apart from this slightly reduced survivability, Mendelian and sex ratios at birth were respected, suggesting an absence of defect during embryonic development ([Figure 15D](#)). Overall, from these observations, it seemed that loss of SPINDOC did not affect embryonic development, morphology and viability with no observed damageable phenotype.



**Figure 14 | *Spin1* and *Spindoc* expression pattern *in vivo*** **A**, Genomic tracks extracted from UCSC Genome browser of *Spin1* and *Spindoc* loci, showing their different transcript isoforms. Red dots indicate exon targeted by primer pairs used for RT-qPCR. **B**, Barplot showing expression of *Spin1* (Yellow), *Spindoc* (dark blue) and the two *Spindoc* isoforms (blue shades) in tissues. Values expressed in CT normalized to *Gapdh* and *Rplp0*  $\pm$  standard deviation. Each dot represents distinct animals. **C**, Expression curves of *Spin1* (Yellow), *Spindoc* (dark blue) and *Spindoc* two isoforms (blue shades) in whole fetal gonads and post-natal testes from ales. Values are expressed in CT normalized to *Rrm2*  $\pm$  standard deviation. Two biological replicates were used with a standard deviation shown in vertical bars. Colors used are similar to B. **D**, Violin plot showing expression of *Spin1* and *Spindoc* in the different cell types found in adult testis, expressed in log2CPM. Data are coming from [Hermann et al., 2018](#).



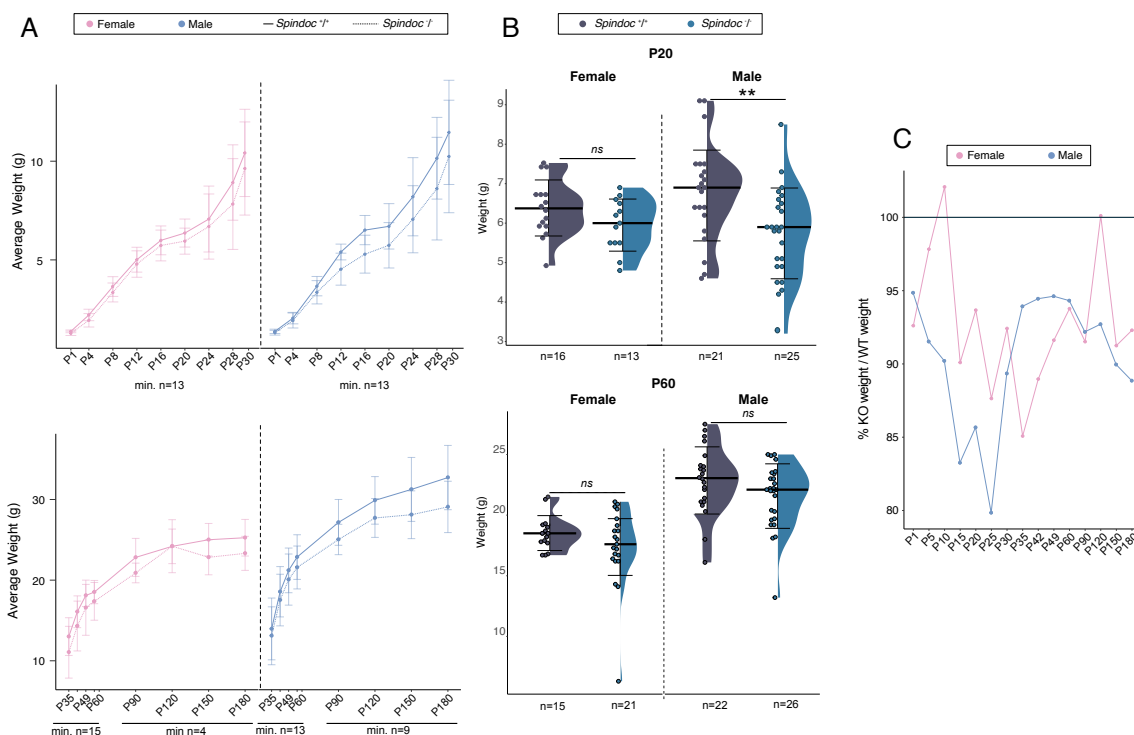


**Figure 15 | Generation of *Spindoc*-KO mouse line** **A**, Scheme of the *Spindoc* locus displaying the sgRNAs used to generate the *Spindoc*-KO mouse line and obtained deletion size. **B**, Immunoblot showing SPINDOC expression in WT and *Spindoc*-KO ESC as well as *Spindoc*<sup>+/+</sup>, *Spindoc*<sup>+/-</sup>, *Spindoc*<sup>-/-</sup> testes. PCNA is used as loading control. **C**, Survival rate in days post-partum of *Spindoc*<sup>+/+</sup>, *Spindoc*<sup>+/-</sup>, *Spindoc*<sup>-/-</sup> males and females born from *Spindoc*<sup>+/+</sup> x *Spindoc*<sup>+/-</sup> intercrosses. **D**, Pie chart showing sex ratio, Mendelian ratio and sex-specific Mendelian ratio of animals at birth, from *Spindoc*<sup>+/+</sup> x *Spindoc*<sup>+/-</sup> intercrosses. Number of animals are displayed under each pie chart.

### 3. SPINDOC is necessary for physiological male growth

Loss of survivability after weaning with no obvious phenotype observed on a daily basis (breathing or mobility) could be caused by weight loss or growth deficiency. To determine if *Spindoc*<sup>-/-</sup> animals had any body weight defect compared to their *Spindoc*<sup>+/+</sup> littermates, I weighed pups from *Spindoc*<sup>+/+</sup> intercrosses from birth to P30 every day, followed by every week up to 8 weeks and once per month until they reached 6 months (P180). Although no difference was observed at birth, *Spindoc*<sup>-/-</sup> males showed a defect in weight gain from P12 to P20 compared to *Spindoc*<sup>+/+</sup> (Figure 16A). This weight defect was not present in females, and seems to be maintained after weaning, while not further increasing. The weight defect was observed only when taking in consideration the average weight of several mice. Indeed, not all *Spindoc*<sup>-/-</sup> males were lighter than age-matched *Spindoc*<sup>+/+</sup> males, as depicted on violin plots (Figure 16B). This might be due to high inter-individual variability, litter size effect and/or variable penetrance of the phenotype. The highest difference was observed at P25, with *Spindoc*<sup>-/-</sup> males weighing on average 80% of *Spindoc*<sup>+/+</sup> weight (Figure

16C). However mild, this phenotype suggests a link between SPINDOC and growth control, feeding behavior or digestion regulation, providing insights into SPINDOC biological relevance. The fact that this phenotype is mostly male-specific also implies a potential role for sex hormones. Another *Spindoc*-KO mouse line (*Spindoc*-KO #2) generated by a short deletion in exon 2 (using the exact same sgRNA than for one of ours) was also described, with a weight defect of 10 to 20% at P21 (F. Yang et al., 2021). However, no distinction was made between males and females. A third independent *Spindoc*-KO mouse line (*Spindoc*-KO #3) generated by a deletion of the whole exon 2, focused on male fertility but failed to mention defective weight gain (Jiang et al., 2021).



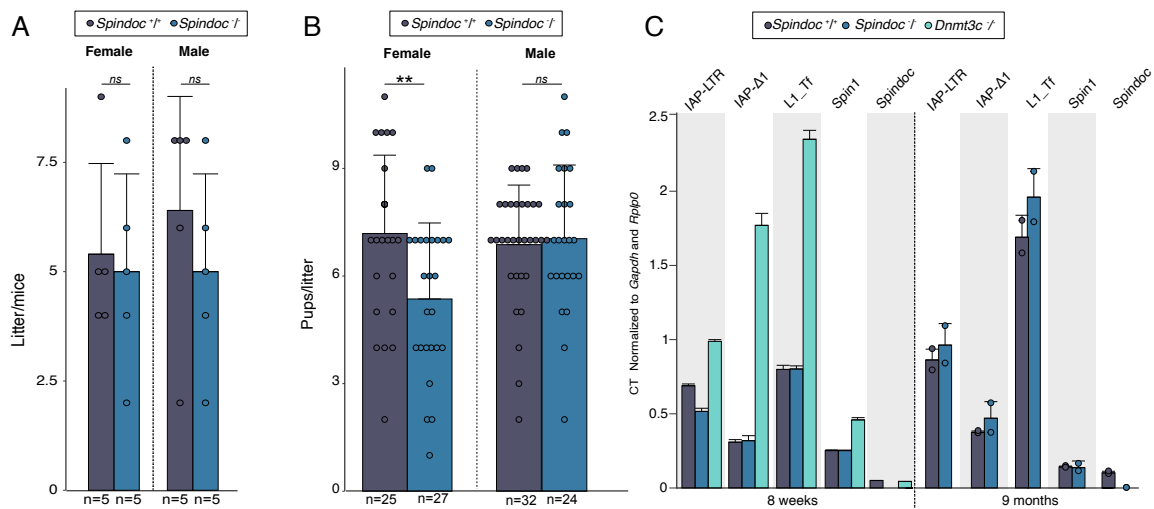
**Figure 16 | Characterization of a defect in weight gain associated to SPINDOC loss** **A**, Growth curves of males (blue) and females (pink) with *Spindoc*<sup>+/+</sup> (plain lines) and *Spindoc*<sup>-/-</sup> (dotted lines) genotypes. Average body weight is expressed in grams from P1 to P30 (top) and from P35 to P180 (bottom) ± standard deviation. Minimal number of animals weighed for each time period is given. Exact number of animals for each timepoint is provided in Table 1 **B**, Violin plots of *Spindoc*<sup>+/+</sup> (grey) and *Spindoc*<sup>-/-</sup> (blue) weights for females (left side) and males (right side) at P20 (top) and P60 (bottom). The number of animals weighed in each category is displayed. All animals came from *Spindoc*<sup>+/+</sup> x *Spindoc*<sup>-/-</sup> intercrosses. P-values obtained from one sided Student's t-test are as follow P20 females = 0.06327; P20 male = 0.00368; P60 females = 0.1244; P60 males = 0.05316 **C**, Ratio in percentage of the body weight of *Spindoc*<sup>-/-</sup> males (blue) and females (pink) compared to *Spindoc*<sup>+/+</sup> every 5 days from P1 to P180 (6 months). All animals came from *Spindoc*<sup>+/+</sup> x *Spindoc*<sup>-/-</sup> intercrosses.

#### 4. SPINDOC loss leads to a minor sub-fertility

Regarding the expression of *Spindoc* and *Spin1* in male and female germ cells, I took interest in the fertility of this line. Before any quantification, I first observed that *Spindoc*<sup>-/-</sup> males and females were able to breed with both WT and *Spindoc*<sup>-/-</sup> individuals. I then quantified the fertility of five pairs of *Spindoc*<sup>+/+</sup> and *Spindoc*<sup>-/-</sup> male and female littermates crossed with WT animals, as described more precisely in the Methods section. Pups obtained from these crosses had globally normal sex ratio ranges, with the highest discrepancy observed for *Spindoc*<sup>-/-</sup> males fostering 58% of females (all data provided in [Table 2](#)). However, *Spindoc*<sup>-/-</sup> males fostered less litters (4.8 litter/male) compared to their WT littermates (6.4 litter/male) in the same time period. Moreover, *Spindoc*<sup>-/-</sup> females had normal litter numbers but fostered less pups per litter with 5.4 pups/litter for *Spindoc*<sup>-/-</sup> against 7.2 pups/litter for *Spindoc*<sup>+/+</sup> over 25 and 27 litters, respectively ([Figure 17A, 17B](#)). This could be interpreted as an effect of the *Spindoc* deletion on mating behavior in males, while it may rather impair oocyte maturation, or have a maternal effect on the embryo in females.

A phenotype of sub-fertility of *Spindoc*<sup>-/-</sup> males was reported for the *Spindoc*-KO #3 line, while female fertility was not apparently tested in that study ([Jiang et al., 2021](#)). However, in contrast to us, the authors reported a reduced litter size of *Spindoc*<sup>-/-</sup> males, which they further linked to lower sperm count in the epididymis and abnormal spermatozoa morphology, including head defects, while meiosis was normal. While using a mouse line with the same genetic background than us (C57BL6/J), the number of pups per litter in WT males they report is higher than what we observe (approx. 9 pups/litter). During our fertility test, we compared pairs of littermates, which implies our WT males are born from *Spindoc*<sup>+/-</sup> x *Spindoc*<sup>+/-</sup> and raised by *Spindoc*<sup>+/-</sup> mothers which could have an impact on their fitness or post-natal development. Therefore, comparing littermates seemed more accurate to remove any of these possible environmental effects. Moreover [Jiang et al., 2021](#) reported a difference in testis size without normalization to body size, which is possibly affected in their *Spindoc*-KO #3 as in our *Spindoc*-KO and the *Spindoc*-KO #2 lines. Overall, the study proposed by [Jiang et al., 2021](#) lacks properly defined controls.

Although our data did not indicate a strong fertility phenotype in *Spindoc*<sup>-/-</sup> males that could be linked to impaired germ cells, I still went on looking at the expression level of IAP and LINE-1 (Tf and A families) on whole testes of *Spindoc*<sup>-/-</sup> males at 8 weeks and 9 months of age, in comparison with *Spindoc*<sup>+/+</sup> males and *Dnmt3c*<sup>-/-</sup> males as a positive control of TE up-regulation (Barau et al., 2016). As expected for the lack of meiotic defect, we did not find increased expression of TEs in *Spindoc*<sup>-/-</sup> males (Figure 17C).



**Figure 17 | Sub-fertility is observed in *Spindoc*<sup>-/-</sup> males and females** **A**, Number of litters fostered by *Spindoc*<sup>+/+</sup> (grey) and *Spindoc*<sup>-/-</sup> (blue) littermate females (left) and males (right) during fertility tests upon breeding with WT animals. Each dot represents a tested animal and the number of tested animals is indicated at the bottom. Non-parametric statistical tests (Wilcoxon) result in p-values for female = 0.511 and male = 0.3274. **B**, Number of pups per litter fostered by *Spindoc*<sup>+/+</sup> (grey) and *Spindoc*<sup>-/-</sup> (blue) littermate females (left) and males (right) during fertility tests. Each dot represents a litter and numbers of litters are indicated at the bottom. P-values obtained from one sided Student's t-test are as follow females = 0.001791 and males = 0.629. **C**, Steady-state levels of IAP transcripts determined by RT-qPCR with two different primer pairs, L1-A, L1-Tf, *Spin1* and *Spindoc* in *Spindoc*<sup>+/+</sup> (grey), *Spindoc*<sup>-/-</sup> (blue) and *Dnmt3C*<sup>-/-</sup> (green, as a positive control) males at 8 weeks ( $n=1$  per genotype, left) and 9 months ( $n=2$  per genotype, right). Values are expressed in CT normalized to *Gapdh* and *Rplp0*  $\pm$  standard deviation on three technical replicates for right panel and biological replicates for left panel.

## METHODS

### Mice

Mice were hosted in a 12 hours light / 12 hours dark cycle with free access to water and food in the pathogen free Animal Care facility of the Institut Curie (agreement C75-05-18). All experimentations were approved by the Institut Curie Animal Care and Use Committee and abide to European and national regulation for the protection of vertebrate animals used for experimental and other scientific purposes (directives 86/609 and 2010/63). For tissue and embryo collection, mice were euthanized by cervical dislocation. Prenatal and post-natal timepoints were obtained by counting following day after pregnancy as E0.5.

### RNA extraction, reverse transcription and qPCR

For genital ridges and early forming gonads (E12.5 and E14.5), tissue collected from two embryos were pooled and extracted with Arcturus Pico Pure RNA isolation kit following manufacturer's instructions. For all other time points, tissues were dissociated using magnetic beads in a Tissue Lyser instrument and RNA was extracted with Trizol (Life Technologies) with manufacturer instruction followed by DNase treatment using Qiagen RNase-free DNase set. RNA concentration was measured with Nanodrop 2000 and reverse transcription was carried out with equal amount of RNA per sample (200ng or 500ng, depending on RNA availability). RNA annealed with 20ng/ $\mu$ L random primers was reverse transcribed by SuperScript III (Invitrogen) in presence of RNase inhibitor (Promega), 500nM dNTPs and 0.1M DTT at 50°C for 1h. 1 $\mu$ L of 5 times diluted reaction mix was used for qPCR with SYBR Green Master Mix (Invitrogen) on a Vii7 instrument (Applied Biosystems). All primers used in qPCR experiment are listed in [Table 1](#). Normalization of RT-qPCR analysis was done as such:  $2^{-CT\_Sample} / \sqrt{[(2^{-CT\_Gapdh}) \times (2^{-CT\_Rplp0})]}$  or  $2^{-CT\_Sample} / (2^{-CT\_Rrm2})$  for timeline expression in male gonads.

Primer name	Forward	Reverse	Application
<i>sgSpindoc1</i>	GCCACCTCGGCCTAACCTGC	GCAGTTAGGCCGAGGTGGC	CRISPR
<i>sgSpindoc2</i>	GTTGCTAGGCCTTCCAACGCT	AGCGTTGGAAGGCCTAGCAAC	CRISPR
<i>Spindoc</i>	CCCTGAGAATGAACCTGCTA	GGTGCTGCGAATGTCTCT	RT-qPCR
<i>Spindoc_iso1</i>	TGCCACTGTAAAGGGACAAC	CGCCACCTTTCTTCTTC	RT-qPCR
<i>Spindoc_iso2</i>	GCCACAAAACCCCTCTTCAG	CACCTCTATCTGGGCTCTGG	RT-qPCR
<i>Spin1</i>	GTACGATGGATTTGACTGTGT	CGATCATTGTGTCCGCTAAG	RT-qPCR
<i>IAP-LTR</i>	CGGGTCGCGGTAATAAAG GT	ACTCTCGTTCCCCAGCTG AA	RT-qPCR
<i>IAP-Δ1</i>	CGCATGTGCCAAGGGTATCT	GAGCATTCTCCTTCGCCTCG	RT-qPCR
<i>L1-Tf</i>	CAGCGGTGCGCCATCTTG	CACCCTCTCACCTGTTTCAGACTAA	RT-qPCR
<i>Actin</i>	AAGTGACGTTGACATCCG	GATCCACATCTGCTGGAAGG	RT-qPCR
<i>Gapdh</i>	TCCATGACAACCTTTGGCATTG	CAGTCTTCTGGGTGGCAGTGA	RT-qPCR
<i>Rrm2</i>	CCGAGCTGGAAAGTAAAGCG	ATGGGAAAGACAACGAAGCG	RT-qPCR
<i>Rplp0</i>	TCCAGAGGCACCATTGAAATT	TCGCTGGTCCACCTT	RT-qPCR
<i>Spindoc</i>	GGGGTGGGACCATGACTAGA	CACCAGAGTGACGGAAGCTC AGTCACCCACAGAGACATTCCG	Genotyping PCR

**Table 1** | List of primers used for this study and their associated application

### Generation of a *Spindoc*-KO mouse line

sgRNAs listed in [Table 1](#) were ordered from *Synthego*, ready to use. 200 ng/uL of sgRNAs were injected by electroporation in 94 B6D2F1/J zygotes together with 400ng/uL CAS9 protein (IDT V3: 1081058). For electroporation, the poring pulse was set to 40V with 4 pulses of 3.5msec every 50msec at a decay rate of 10% and a positive polarity while transfer pulse was set to 5V with 5 pulses of 50msec every 50msec at a decay rate of 40% and an alternative polarity. 92 zygotes were re-implanted in 4 pseudo-pregnant NMRI females. 19 pups were born these zygotes among which 42% were *Spindoc*<sup>+/-</sup> and 16% *Spindoc*<sup>-/-</sup>. Genetic deletions were sequenced (Sanger sequencing service provided by Eurofins) and two founders were kept and crossed for 4 generations with C57BL6/J before starting experiments. Results presented here come from the deletion present in one founder and are under confirmation for a second line emerging from the second founder (data not shown).

### Weighing

Animals born from *Spindoc*<sup>+/-</sup> x *Spindoc*<sup>+/-</sup> intercrosses were weighed every day from their birth (P1) to P30, once a week from P30 to P60 and one a month onward to P180 (6 months). The number of pups weighed for each time point, genotype and sex is given in the following [Table 2](#).

Genotype and sex	Female <sup>+/+</sup>	Female <sup>-/-</sup>	Male <sup>+/+</sup>	Male <sup>-/-</sup>
P1	13	17	18	29
P4	16	20	22	27
P8	20	19	22	31
P12	18	17	22	27
P16	22	20	24	30
P20	16	13	21	25
P24	20	14	13	24
P28	19	15	19	28
P30	18	15	22	28
P35	16	20	24	29
P42	18	18	24	28
P49	16	18	24	28
P60	15	21	22	26
P90	13	8	17	13
P120	14	8	17	12
P150	8	4	12	9
P180	8	4	13	9

**Table 2** | Number of individuals fostered by *Spindoc*<sup>+/-</sup> intercrosses, weighed for growth curves displayed in Figure 3A according to their sex, genotype and age.

### Fertility test

Five pairs of *Spindoc*<sup>+/+</sup> and *Spindoc*<sup>-/-</sup> male and female littermates were used. For the male fertility test, 8 week-old males were given access to two WT C57Bl6/J females (6 to 8 week-old, purchased from Charles River) for a month, after which females were kept separated for 20 days. Pups born during this period were sexed and counted up to 2 days after birth. After a month of break, males were subjected to a second round of assay with two WT C57Bl6/J females again. For the female fertility test, 8 week-old females were in contact for 6 months with one WT C57Bl6/J male (8 to 11 week-old, purchased from Charles River) at the beginning of the assay. All fostered litters during this period were sexed and counted up to 2 days after birth ([Table 3](#)).

<b>Genotype and sex tested parent</b>	<b>Fostered female</b>	<b>Fostered male</b>
<b>Female<sup>+/+</sup></b>	46,7% (n=50)	53,3% (n=57)
<b>Female<sup>-/-</sup></b>	49,2% (n=30)	50,8% (n=31)
<b>Male<sup>+/+</sup></b>	43,5% (n=20)	56,5% (n=26)
<b>Male<sup>-/-</sup></b>	58,5% (n=24)	41,5% (n=17)

**Table 3** | Number of pups fostered during fertility test and sex ratio displayed by sex and genotype of tested parent



# DISCUSSION



## I. General summary

At the start of my PhD project, SPIN1 and SPINDOC had never been studied in ESCs, their role in TE repression was not known and the biological role of SPINDOC *in vivo* was unexplored. Still nowadays, SPINDOC is mentioned in only a handful of publications. This lack of knowledge gave this project both its novelty and challenging aspect, as there was really little evidence to build hypotheses upon. Nonetheless, my project took place in the competitive area of chromatin-based regulation of TEs. I therefore worked on understudied proteins in a dynamic area of research, and had many overarching questions to answer, with the caveat to choose (or cherry pick) among them the most interesting/promising/obvious one with educated guess. Regarding the exact mechanism by which SPIN1 and SPINDOC repress TEs in ESCs, or the exact biological function SPINDOC *in vivo*, my work brings several pieces of highlighting information.

I started this project with the idea to study the role of SPIN1 and SPINDOC independently, by deleting each of them separately in ESCs. The destabilization of SPINDOC that I observed in absence of SPIN1 is in accordance with its reported role as a SPINDOC partner (Bae, Gao, et al., 2017; Devi et al., 2019). However, it prevented me from studying the role of SPIN1 alone in ESCs and TE control. Intriguingly, genetic deletion of both factors led to a close but yet distinct molecular phenotype compared to *Spin1* deletion accompanied with secondary SPINDOC loss. Transcriptional up-regulation linked to SPINDOC deletion fits with the current hypothesis placing SPIN1 as a transcriptional activator and SPINDOC as antagonizing SPIN1 function, and therefore, as a transcriptional repressor. While my work in ESCs tended to describe SPINDOC for its role as a SPIN1 repressor, the distinct cellular specificity of *Spin1* and *Spindoc* expression *in vivo* and the strikingly different phenotypes observed in *Spin1*-KO and *Spindoc*-KO mice open the possibility that SPINDOC may also function independently of SPIN1, at least in some contexts. This idea was confirmed during the course of my PhD by a publication that reported that SPINDOC can form a complex with PARP1, without SPIN1 (F. Yang et al., 2021). Therefore, I am currently performing additional work to decipher the independent and combined roles of SPIN1 and SPINDOC in ESCs and TE control, which will be further discussed in this chapter.

TE control by local heterochromatin formation is a major described and studied pathway, involving H3K9me3 deposition and DNA methylation. Both of these marks are not direct effectors of TE control, but might be needed for effector binding or action associated with transcriptional repression. The presence or absence of these marks is not sufficient for transcriptional repression or activation, respectively, which is well exemplified in my work where DNA methylation loss is not always linked to transcriptional activation, nor is H3K9me3 presence with transcriptional repression. This resonates well with a recent study of DNA methylation and L1-HS expression in a variety of human cell lines, which showed that the lack of DNA methylation at L1-HS loci did not necessarily correlate with their expression ([Lanciano et al., 2023](#)). Although I found that deleting SPIN1 or SPINDOC induced a loss of DNA methylation linked to UHRF1 destabilization in ESCs, several pieces of evidence indirectly suggest that SPIN1 and SPINDOC also rely DNA methylation-independent mechanisms to control TEs, which will be detailed further in this chapter.

Finally, although I could not relate SPINDOC loss to TE reactivation in male germ cells, my work revealed some novelties regarding the role of SPINDOC *in vivo*, notably its importance in physiological growth control. Moreover, my study of SPIN1 and SPINDOC in ESCs brings some new evidence on their role in Wnt signaling and rRNA transcription in this cell type. In a context where SPIN1 is a putative therapeutic target in cancer research, every information concerning its role has to be taken into consideration and I will discuss some perspective regarding SPIN1 and SPINDOC biological relevance in oncology hereafter.

## II. The unexpected triangle: SPIN1-SPINDOC, DNA methylation and TE control in ESCs

### A. SPIN1-SPINDOC and UHRF1 stability

#### 1. Hypotheses regarding the role of SPIN1 and SPINDOC in UHRF1 stability

The genome of mESCs grown in 2i medium is lowly methylated, which participates to their close resemblance to cells from the ICM compared to serum-grown ESCs (Ficz et al., 2013; Habibi et al., 2013). In relevance to this low DNA methylation, a destabilization of UHRF1 is known to occur at the protein level in ESCs transitioning from serum to 2i conditions (von Meyenn et al., 2016). While UHRF1 stability and cellular localization has been tangibly linked to the Developmental Pluripotency Associated 3 (DPPA3) protein in the oocyte (W. Du et al., 2019; Y. Li et al., 2018), the origin of UHRF1 down-regulation in 2i-grown ESCs is not fully resolved, although it seems to co-occur with decreased H3K9me2, a mark that is responsible for UHRF1 recruitment to replication foci through its SRA domain, but causal relationships were not inferred (von Meyenn et al., 2016). SPIN1 and SPINDOC seem therefore necessary to avoid degradation of all UHRF1 molecules and to maintain a certain level of DNA methylation in 2i-grown ESCs. UHRF1 is actively targeted for proteasome degradation in 2i-grown ESCs, notably by PRAMEL7 (Graf et al., 2017). SPIN1 and SPINDOC could act either by ensuring transcriptional balance of this proteasome targeting factor or through binding competition with said factor or with UHRF1. However, UHRF1 or PRAMEL7 were not found as interaction partners of SPIN1 in ESCs by IP followed by LC/MS-MS. We could envision that SPINDOC interacts with UHRF1 or PRAMEL7. In immunoblots showing UHRF1 degradation in *Spin1*-KO and *Spindoc*-KO ESCs, I repeatedly observed an even greater loss in *Spindoc*-KO compared to *Spin1*-KO. Moreover, SPINDOC is destabilized but not completely absent in *Spin1*-KO and I observed an up-regulation of *Dppa3* (log<sub>2</sub>FC>3) only in *Spindoc*-KO ESCs. Due to these observations, I propose two hypothetical relationships between SPIN1-SPINDOC and UHRF1 degradation: (1) SPINDOC could interact with UHRF1 in the nucleus, preventing binding to DPPA3 or PRAMEL7, avoiding translocation in the cytoplasm and/or further targeting to the proteasome; (2) SPIN1 could be involved in transcriptional activation of *Dppa3*, and this function would be

antagonized by SPINDOC, explaining *Dppa3* up-regulation in *Spindoc*-KO ESCs. This second interaction may be supported by the down-regulation of *Dppa3* observed in *Spin1*-KO ESCs ( $\log_2FC \approx -2$ ), while *Pramel7* expression is not modified in any mutant cell lines. However, despite down-regulation of *Dppa3* in *Spin1*-KO, I still observe UHRF1 destabilization in these cells. Ongoing complementation studies of SPIN1 and SPINDOC in their respective KO background, and of SPINDOC in *Spin*-KO should bring more information on this matter. Nevertheless, the importance of DNA methylation maintenance in naive mESCs is related to their continuous divisions. *In vivo*, low DNA methylation in the ICM is only temporary. Therefore, the possible increased degradation of UHRF1 in the blastocyst of *Spin1*-KO or *Spindoc*-KO background should not have a considerable phenotypic impact.

## 2. Loss of UHRF1 and TE control

The loss of proteins responsible for DNA methylation maintenance, DNMT1 and UHRF1, have distinct impact on TE up-regulation. While TEs, including IAP elements, are up-regulated in both *Uhrf1*-KO or *Dnmt1*-KO ESCs, this occurs very transiently in *Uhrf1*-KO cells (Sharif et al., 2016). DNA methylation decreases in both mutants, however H3K9me3 is reduced in *Dnmt1*-KO only, likely due to continuous binding of UHRF1 to chromatin and eviction of SETDB1. Consistent with this hypothesis, *Uhrf1*-KO, IAPs are first de-repressed as a consequence to DNA methylation loss but SETDB1-dependent deposition operates to re-silence them. The phenotype I observed in *Spin1*-KO and *Spindoc*-KO ESCs is divergent from the one reported in this publication, despite similar loss of UHRF1 and decreased DNA methylation. More specifically, UHRF1 destabilization in *Spindoc*-KO and *Spin1*-KO ESCs induces long-term TE up-regulation.

First, these discrepancies might merely reflect that DNA methylation loss due to UHRF1 destabilization is not the only factor playing a role in *Spin1*-KO and *Spindoc*-KO ESCs. This adds up to another evidence that DNA methylation loss is neither necessary nor sufficient to explain the level and specificity of TE up-regulation observed in absence of SPIN1 or SPINDOC. Moreover, these studies investigating the role of UHRF1 in ESCs, including the one led by Sharif et al, were performed in serum-based medium, which results in high genomic methylation content. We relied on 2i

medium, and this difference might also explain why we observed long-term TE up-regulation linked to UHRF1 loss as opposed to transient up-regulation observed in *Uhrf1*-KO ESCs in serum. That being said, I observed a slight increase of H3K9me3 in *Spin1*-KO and *Spindoc*-KO compared to WT cells at up-regulated IAPEz elements, which could indicate a similar H3K9me3 accumulation as the one proposed by [Sharif et al., 2016](#) in absence of UHRF1, and also reported to occur upon acute loss of DNA methylation in ESCs ([Walter et al., 2016](#)). However, it seems that in absence of SPIN1 or SPINDOC, this H3K9me3-driven compensatory mechanism is not enough to counteract IAPEz up-regulation, which is maintained upon prolonged mutant ESCs culture in 2i-based medium.

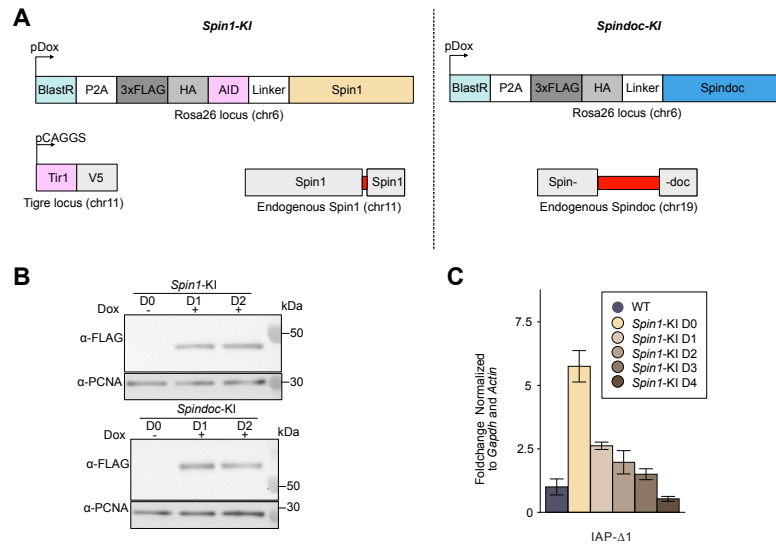
### 3. Long-term culture and passive loss of DNA methylation

UHRF1 destabilization in *Spin1*-KO and *Spindoc*-KO ESCs is a probable cause for the observed DNA methylation loss. Therefore, as DNA methylation maintenance is impaired, its loss is expected to be progressive over cell divisions. Theoretically, after a certain time—that has to be determined—all DNA methylation could be lost. All the experiments I carried out to measure DNA methylation (LUMA, PCR-pyrosequencing and EM-seq) were performed on *Spin1*-KO and *Spindoc*-KO ESCs between passages 10 and 20, which we can roughly estimate to 30 to 60 cell divisions since the genetic ablation of *Spin1* or *Spindoc*. After such time, as DNA methylation should be diluted by half at each cell replication cycle, if UHRF1 was completely absent and DNA methylation maintenance completely abrogated, an almost complete loss of DNA methylation would be expected. Hence the question of how low can DNA methylation decrease in *Spin1*-KO and *Spindoc*-KO ESCs. In the same line of questioning, some transcriptional differences observed between the two mutants could be explained by different levels of DNA methylation due to progressive loss as well as RNA-sequencing performed after different times in culture, which could influence notably TE expression. To resolve this issue, two distinct strategies can be considered: 1) reach the lowest DNA methylation content in *Spin1*-KO and *Spindoc*-KO ESCs and use this lowest point to perform experiments, or 2) make a time-course analysis of progressive loss of DNA methylation and increased TE expression, starting right after SPIN1 or SPINDOC withdrawal. As this last strategy brings more information regarding the dynamic of DNA

methylation loss, TE expression and possibly UHRF1 destabilization, I will present hereby the design of this ongoing experiment.

As mentioned briefly in the previous paragraph, complementation experiments were performed in *Spin1*-KO and *Spindoc*-KO ESCs by inserting at the *Rosa26* locus a coding sequence augmented with 3XFLAG and HA tags of the deleted protein under the dependence of a doxycycline inducible promoter, in their respective KO background (Figure 18A). These cell lines, *Spin1*-KI and *Spindoc*-KI, were obtained thanks to the work of Alexis Cornec under my supervision. After testing correct induction of both proteins (Figure 18B), I performed a prolonged induction of 8 days in order to reset *Spin1*-KO and *Spindoc*-KO cell lines to their original WT state in regards to DNA methylation and TE expression. During the induction and for 7 days after, cell pellets for DNA, RNA and protein extraction were kept every day and will allow to understand the dynamic of UHRF1 destabilization, DNA methylation loss and TE expression right after SPIN1 and SPINDOC loss. In addition, a final cell line of SPINDOC complementation in *Spin1*-KO background may help us to characterize the solo impact of SPIN1 loss in ESCs. For the time being, through preliminary tests, I found that 4 days of induction of the *Spin1*-KI transgene in the *Spin1*-KO background are necessary for IAP- $\Delta$ 1 expression to reach WT levels, with a progressive decrease of expression (Figure 18C). These promising results indicate that loss of SPIN1 is indeed directly responsible for the observed phenotype and that TE up-regulation in *Spin1*-KO is reversible.





**Figure 18 | SPIN1 complementation promote TE repression.** **A**, Schemes of transgenic insertions and deletions in the genome of *Spin1*-KI and *Spindoc*-KI cell lines. For *Spin1*-KI, as an AID tag is added to the construct at the Rosa26 locus, the Tir1 cDNA is inserted at the Tigre locus, which was not needed for *Spindoc*-KI. Transgenes at the Rosa26 locus are inducible through doxycycline treatment. **B**, Immunoblot showing expression of the transgene after 0, 1 and 2 days of doxycycline addition (1µg/mL), renewed every 12 hours in *Spin1*-KI and *Spindoc*-KI, using FLAG antibody for transgenes, and PCNA as loading control. **C**, Barplot showing IAP-Δ1 expression as measured by RT-qPCR, in foldchange normalized to *Gapdh* and *Actin* in WT, *Spin1*-KI without induction and *Spin1*-KI after 1, 2, 3 and 4 days of continuous induction with doxycycline (1µg/mL). One biological replicate was used and shown is mean + standard deviation of three technical replicates.

## B. SPIN1-SPINDOC role in TE control is independent of DNA methylation

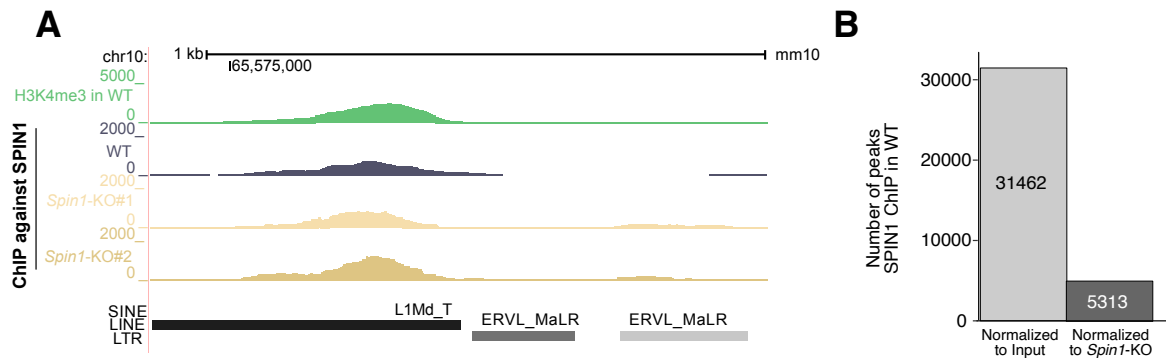
### 1. SPIN1 binding to chromatin in ESCs

SPIN1 is a chromatin reader with abilities to bind H3K4me3. Binding of SPIN1 in several tested cell types is highly correlated to H3K4me3 regions (Franz et al., 2015; Greschik et al., 2017). On peptides, interaction between H3K4me3 and SPIN1 was strengthened by presence of H3K9me3 (Du et al., 2021; Du & Qian, 2022; Zhao et al., 2020). Both chromatin marks are reported to be mutually exclusive and linked to transcriptional activation and repression, respectively. However, the existence of the H3K4me3-dependent H3K9 demethylase KDM4B (Pedersen et al., 2016) indicates that H3K4me3 and H3K9me3 may co-exist in close proximity, in at least a transient manner.

Repeated elements, including TEs, are mainly found in H3K9me3-enriched chromatin. Therefore, it seems unlikely that SPIN1 would bind TE sequences that carry this mark only. A direct effect of SPIN1 on TE expression, through occupancy of these loci, is therefore not really expected. However, SPIN1 binding to H3K4me3-enriched gene promoters could explain the pattern of gene mis-regulation in *Spin1*-KO ESCs. Moreover, SPINDOC could impact or modulate the ability of SPIN1 to bind to chromatin, a matter that was not resolved by previous publications due to contradictory results concerning the ability of SPINDOC-SPIN1 complex to bind chromatin (Du & Qian, 2022; Liu et al., 2020). To resolve this matter, I attempted to perform ChIP-seq against SPIN1 using a previously demonstrated ChIP-proof SPIN1 antibody (provided by R. Schüle, Freiburg University, Franz et al., 2015) in WT, *Spindoc*-KO and *Spin1*-KO ESCs. ChIP-seq in *Spin1*-KO cells was used as a control, with the expectation that we would lose peak enrichment compared to WT conditions. However, a high number of peaks were detected in the *Spin1*-KO ESC clones, at regions overlapping WT peaks (Figure 19A). Two types of normalization can be used in these ChIP-seq analyses: normalize reads from SPIN1-ChIP in WT ESCs against their Input, or normalize reads from SPIN1-ChIP in WT ESCs to SPIN1-ChIP in *Spin1*-KO ESCs. With the latter normalization, regions bound by SPIN1 might be over-estimated, while in the former, they might be under-estimated (Figure 19B). In both scenarios, we cannot warrant that the regions determined as being bound by SPIN1 are indeed, bound. The presence of SPIN1 peaks in *Spin1*-KO could be linked to two technical caveats: (1) SPIN1 binds to H3K4me3-enriched regions, which are in a relaxed chromatin state, easily fragmented by sonication and more easily caught up during IP. ChIP experiments are known to be biased toward open chromatin (D. Jain et al., 2015) and this bias might lead to the presence of open chromatin in *Spin1*-KO ChIP, regions also bound by SPIN1; (2) ChIP, as every antibody-based method, relies on antibody specificity and we could imagine that this antibody binds to other proteins than SPIN1.

To circumvent both potential technical artefacts (sonication and antibody specificity), I decided to use Cleavage Under Targets & Release Using Nuclease (CUT&RUN) in *Spin1*-KI cells before and after induction with doxycycline using an anti-FLAG antibody. Indeed, CUT&RUN relies on binding of antibodies before enzymatic chromatin fragmentation, removing bias due to sonication in ChIP, while the use of the FLAG antibody removes possible SPIN1 antibody unspecific binding. This experiment has

now been done, but sequencing and downstream analyses still have to be performed to uncover the binding pattern of SPIN1 in ESCs. In addition, CUT&RUN against FLAG in both *Spindoc*-KI and *Spindoc*-KI in *Spin1*-KO ESCs would answer the question as to whether SPINDOC binds to chromatin on its own or through SPIN1 only.



**Figure 19 | ChIP against SPIN1 results in enrichment at LINE-1 and in Spin1-KO ESC.** **A**, Genomic tracks extracted from UCSC genome browser showing one L1\_T element bound by SPIN1 in WT ESC (black), *Spin1*-KO #1 and *Spin1*-KO #2 (yellows) and by H3K4me3 in WT ESC (green). Peaks are normalized to inputs. **B**, Number of SPIN1 peaks determined in SPIN1 ChIP in WT ESC determined by normalization to Input and to SPIN1 ChIP in *Spin1*-KO (merge of *Spin1*-KO #1 and *Spin1*-KO #2).

## 2. Hypotheses concerning the role of SPIN1 and SPINDOC in TE control

As formerly mentioned in several instances, I propose that SPIN1 and SPINDOC control TEs not solely through DNA methylation regulation. This other function exerted by SPIN1 and SPINDOC is uncovered in serum-grown *Spin1*-KO and *Spindoc*-KO ESCs, where despite no change in DNA methylation levels, up-regulation of IAP elements still reaches a 5-fold increase compared to WT. Moreover, the initial CRISPR-Cas9 screen that revealed SPIN1 and SPINDOC as putative factors involved in IAP repression was performed in serum-grown ESCs (Chelmicki et al., 2021). The nature of this DNA methylation-independent role of SPIN1 and SPINDOC in TE restriction is, however, not known and can only be hypothesized.

One intriguing observation I made is the *Spindoc*-KO-specific up-regulation of LINE-1. If SPINDOC loss was the only factor responsible for LINE-1 up-regulation, it should also be observed in *Spin1*-KO (where SPINDOC is destabilized) and in dKO ESCs (where SPINDOC is genetically removed). Therefore, our data fit more with a model where SPIN1 acts as transcriptional activator of LINE-1, and SPINDOC antagonizing SPIN1, therefore as a transcriptional repressor of LINE-1. This

hypothesis could be clarified with the identification of SPIN1 binding targets. Notably, in our SPIN1 ChIP-seq data, upon normalization of WT over *Spin1*-KO signal, we could detect SPIN1 binding to LINE-1 elements, onto which H3K4me3 was also detected (Figure 19A). Quantification of such cases and correlation with up-regulated LINE-1 elements in *Spindoc*-KO will be possible thanks to the new SPIN1 CUT&RUN data. TE families and classes that are up-regulated in *Spindoc*-KO and *Spin1*-KO ESCs do not completely overlap, with more diversity in *Spindoc*-KO, which fits with SPIN1 acting as transcriptional activator at a diversity of targets. Complementary to this hypothesis, SINE elements originating from RNA-pol III transcribed elements constitute a major part of up-regulated TE copies in *Spindoc*-KO and several proteins involved in expression of RNA-pol III dependent-RNAs are part of SPIN1 interaction partners.

In contrast, TE families that are up-regulated in *Spin1*-KO and dKO ESCs are more restricted to LTR-retrotransposons and especially, to the ERVK class. In both mutants, the whole SPINDOC-SPIN1 complex is absent and what we witness here might be the direct effect of the complex on TE control. The poor enrichment of KAP1 at TE copies up-regulated in *Spin1*-KO, combined with the lack of KAP1 in our SPIN1-centered proteomics, make it unlikely that SPIN1-SPINDOC bind TEs through KAP1. Moreover, H3K9me3 enrichment is not lost at up-regulated TE copies in *Spin1*-KO ESCs, further dismissing that SPIN1-SPINDOC is linked to KAP1-dependent H3K9me deposition at TEs. Intriguingly, SPIN1 interaction with several chromatin remodelers fosters the hypothesis along which SPIN1-SPINDOC could be involved in chromatin accessibility rather than influence chromatin modifications *per se*. Moreover, SPIN1 interacts with PML, involved in the distribution of H3.3-bearing nucleosomes between heterochromatin and open chromatin regions (Delbarre et al., 2017), knowing that the presence of H3.3 at TEs is related to their complete repression (Elsässer et al., 2015). To test whether SPIN1-SPINDOC action on TEs may involve chromatin remodeling, Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq) will be performed in WT and mutant ESCs.

### III. Studies of TE control and TE expression analyses

#### A. TE annotation

All TE-centered transcriptomic analyses performed for this manuscript are based on genomic annotations of the mouse reference genome. A first caveat in TE annotation is the ability to define a genomic region as a TE, which needs a certain percentage of homology between the considered genomic region and the consensus TE sequence and pre-supposes a 100% accurate genomic assembly. In addition, a variable amount of TEs are polymorphic, meaning present in only a subset of the species population. In regards to the genome of inbred mouse strains, the level of inter-individual variability is expected to be low. However, some heterogeneity can appear between different mouse line backgrounds (Ferraj et al., 2023), as TEs are usually not under selective pressure. Importantly, the reference genome is of C57Bl6/J origin, while the E14 ESCs that we used are from the 129/Ola background, and a certain amount of polymorphic TEs might exist between the two strains. This means that sequencing reads emanating from 129/Ola-specific polymorphic insertions are lost during unique alignment (or even random alignment) and that the estimation of the number of uniquely mappable up-regulated TE copies may be imprecise, if they are not present in the C57Bl6/J genome. More striking is the difficulty to link the different varieties of IAP elements described in the early literature (Kuff & Lueders, 1988)—associated to different levels of transposition autonomy—with the annotation provided today either by Dfam, RepeatMasker or RepBase. For RT-qPCR experiments, I used primers targeting the internally deleted IAP- $\Delta$ 1 elements, which are not specifically annotated as such in these databases and do not appear in my genome-wide RNA-seq analyses. Therefore, when studying of TE transcription, RT-qPCR and RNA-seq are not redundant methods and provide different levels of information. Moreover, between the different TE databases, different names or labels can be used for the same TE copy. In addition to this annotation in genomic databases, lies the question of annotation of our dataset. As TE sequences overlap with other possible annotations like introns or exons, it leads to reads being discarded from annotation because they could belong to several features, leading again to a loss of information.

## B. Unique mapping and control groups

To be able to link chromatin profiling and transcriptional changes, we relied on unique mapping of TE-derived reads. The principal caveat of this mapping strategy is the loss of reads that are not mapped to a specific locus, which mainly occurs for newly TE copies that are lowly divergent from their TE copy of origin, therefore recorded with low mappability. Another reason for this read loss could be linked to newly inserted copies that are unannotated in the reference genome, an event that is more likely to happen when studying cell lines with high TE expression of young and still mobile TEs, such as the case of *Spin1*-KO and *Spindoc*-KO ESCs. Although increased TE expression does not necessarily lead to TE transposition, there is higher chance of transposition when expression is permissive, especially if linked to a relaxed chromatin. However, we chose not to quantify transposition events in our cell lines, first because we wanted to focus on expression control, and then because transposition quantification methods present their own limitations. They are usually based on insertions of transgenic reporters (IAP or L1-HS, as an example) transfected in cells in a plasmid from which transcriptional regulation may differ from the endogenous copies of the genome (C. Lu et al., 2011; Muotri et al., 2005). If any transposition is observed, it is more linked to the availability of transposition proteins produced from endogenous elements than from direct regulation of the plasmid or potential for endogenous elements to actually transpose. All things considered, unique mapping tends to provide partial information regarding the quantity of copies affected in *Spin1*-KO or *Spindoc*-KO ESCs, but certainty towards the uniquely mapped ones. Due to the unexhaustive mapping of sequenced reads, whether a copy is not found up-regulated due to poor mappability or to real absence of up-regulation cannot be untangled. This question was always in our considerations during our analyses, especially when trying to find control groups to compare up-regulated TEs to.

As an example, TE copies that are up-regulated in *Spindoc*-KO ESCs (*Spindoc*-KO-specific TEs) had more various origins than the ones up-regulated in *Spin1*-KO ESCs (*Spin1*-KO-specific TES), which were more intensely up-regulated. We wondered whether this discrepancy was linked to DNA methylation loss having more effects on one or the other group of TEs. Comparing DNA methylation levels in WT ESCs at *Spin1*-KO-specific or *Spindoc*-KO-specific TES was biased by the fact that IAP

elements which are up-regulated in both groups but with more copies affected in *Spin1*-KO retain high level of DNA methylation in 2i-grown ESCs compared to other types of elements like LINE-1 (Ficz et al., 2013; Leitch et al., 2013). Therefore, we narrowed down this analysis to IAPEz-elements, which consist in only few copies among *Spindoc*-KO-specific TEs. Hence the other questions that was present in mind: is an effect observed on few copies a real effect and can it be extended to all copies? Several ideas also relying on comparing categories were finally abandoned due to the difficulty of finding a properly defined control group with enough individuals. While it has certainly become easier and easier over the last decade to study repeated sequences in a genome-wide manner, all limitations are not completely lifted and need at least to be taken in considerations during the study of TE expression.

## IV. SPIN1 and SPINDOC biological relevance

### A. SPIN1 as a therapeutic target in cancer development

The discovery of SPINDOC was part of a global effort to find SPIN1 inhibitors in cancer therapy. This was motivated by the fact that SPIN1 demonstrates oncogenic properties: over-expression of *Spin1* leads to striking phenotypes in human cell lines such as poly-nucleation or increased proliferation (Fang et al., 2018; Gao et al., 2005; Yuan et al., 2007), while high *Spin1* expression correlates with poor cancer prognosis (X. Chen et al., 2016, 2018; Janecki et al., 2018; Lv et al., 2020). In parallel, several small drug inhibitors have been developed to chemically counteract SPIN1 abundance or binding activity (Bae, Viviano, et al., 2017; Fagan et al., 2019; Wagner et al., 2016; Y. Xiong et al., 2019). However, my work in ESCs shows that a complete absence of SPIN1 leads to TE up-regulation, while having no impact on cell proliferation. Therefore, the suppression of SPIN1 might not bring awaited therapeutic effects and might even bring more genomic instability linked to TE reactivation, and to increased chromatin accessibility if our hypothesis reveals to be true. However, we attributed LINE-1 expression in *Spindoc*-KO ESCs to a potential effect of SPIN1 in the transcriptional activation of these elements. This putative role for SPIN1 strengthens its oncogenic impact in humans, where L1-HS are the only elements still able to transpose and their up-regulation is a hallmark of cancerous cells. Regarding my own

findings and previous publications on the role of SPIN1 using over-expression in human cell lines, it indeed seems that the presence of SPINDOC is an important parameter to take into account when studying SPIN1 effects. In cases where *Spin1* is up-regulated and *Spindoc* follows in a similar level, *Spin1* up-regulation might be less detrimental, because balanced by SPINDOC repressive effects. Rather than completely suppressing or inhibiting SPIN1 actions, I would rather suggest to first assess SPINDOC expression levels and raise it to stoichiometric levels of SPIN1. While the first technique uses an already available SPIN1 repressor, enhancing SPINDOC expression level might be technically challenging considering drug development and treatment options.

Another possibility to avoid the negative impact of SPIN1 over-expression is to block its interaction with H3K4me3 and therefore to chromatin. Several molecules have been developed to this aim (Bae, Viviano, et al., 2017; Fagan et al., 2019; Wagner et al., 2016; Y. Xiong et al., 2019). Even though their effect on SPIN1 binding to chromatin has been properly assessed *in vitro*, as well as their dose-effect on cell survival, up to today no data exist on the interplay between these inhibitors and SPINDOC binding and further TE expression. One of this SPIN1 inhibitor, MS31 (Wagner et al., 2016) that specifically targets the second Tudor domain involved in H3K4me3 binding, was recently acquired in our laboratory. Several experiments are ongoing to assess the efficiency of the drug in ESCs, and its SPINDOC-SPIN1 interaction, UHRF1 stability, DNA methylation and impact on TE expression. This works will be done in collaboration with Marie Le Page (now Master student) and could answer several questions regarding the potential effect of blocking SPIN1 in cancer treatment.

## B. Going further with SPINDOC role *in vivo*

Our early work on SPINDOC biological function in mice together with publications leading to similar conclusions open a new research avenue. In fact, SPINDOC expression in male meiotic cells—where *Spin1* is not well expressed—would be an interesting place to study SPINDOC independently from its role as a SPIN1 repressor. On a side note, during my study of *Spin1* and *Spindoc* expression across tissues, I noticed that both are relatively lowly expressed in ESCs, despite my work showing that they have an important role in this cell type. While it seems logical that proteins can



only play a role when they are present, the level of expression is not necessary an appropriate proxy for functional importance. Due to a low impact on fertility, our plan in studying SPINDOC biological relevance will be focused on the male-specific weight defect observed in juvenile animals. First of all, we wish to understand if the weight defect is restricted to some specific organs and tissues or homogeneously distributed over the whole body. To this aim, weighing of organs of *Spindoc*<sup>-/-</sup> and *Spindoc*<sup>+/+</sup> male littermates at P20 (when the weight defect is installed) is ongoing. If we observe a phenotype of organ-specific weight decrease, bulk RNA-seq of this organ will be performed.

Moreover, we now know that homozygous *Spindoc*<sup>-/-</sup> x *Spindoc*<sup>-/-</sup> crosses give fertile and foster viable litters. We (together with Julian Iranzo, technical engineer) currently weigh these litters and compare them to *Spindoc*<sup>+/+</sup> weight metrics obtained previously from heterozygous crosses, to test whether a complete loss of SPINDOC over several generations might have a greater phenotypical impact. As SPINDOC loss influences DNA methylation in ESCs, it might play a role in correct expression of imprinted genes for example, some of which are involved in growth control. Link between weight control and SPINDOC could also be behavioral or linked to less efficient milk digestion. In fact, the weight defect stops increasing after weaning so when the feeding behavior changes drastically, from maternal milk supply for which littermates have to compete, to freely accessible solid food.

*Spindoc*<sup>-/-</sup> x *Spindoc*<sup>-/-</sup> crosses are also used to collect blastocysts and assay by immunofluorescence if loss of SPINDOC in the ICM leads to TE up-regulation, as observed in ESCs. Finally, to provide more precise hypothesis on the biological role of SPINDOC, an IP followed by LC/MS-MS as performed on SPIN1 would be useful. While performant antibodies against SPINDOC are rare, our *Spindoc*-KI cell line could help overcome this limitation.

### C. Going further with SPIN1 role *in vivo*

Constitutive *Spin1*-KO mutation is lethal at the homozygous state shortly after birth, in relation to neuromuscular defects (Greschik et al., 2017). However, the embryonic development of *Spin1*<sup>-/-</sup> animals has never been studied and could bring useful information regarding the reasons for this lethality and the importance of SPIN1 during development. This is especially intriguing for the earliest part of development, after fertilization, when *Spin1* has been described as an abundant transcript inherited from the oocyte (Oh et al., 1997; Rothstein et al., 1992). In the early steps of pre-implantation development, several classes of TEs, including L1s, MERVL and maybe IAPs, are more expressed than in any other cell types. Similarly to what we observe in ESCs, SPIN1 could be a regulator of TE expression in early development.

To study more in detail the role of SPIN1 *in vivo*, we recently acquired a conditional *Spin1*<sup>2lox</sup> mouse model (Greschik et al., 2017), which carries LoxP sites inserted upstream and downstream of *Spin1*-exon4. With this set up, crosses with Cre-driver mouse lines would allow recombination of LoxP sites and exon 4 deletion in specific tissues or cell types of interest, avoiding the early post-natal lethality of the constitutive mutation. For the time being, this mouse line is crossed to a PGK-Cre expressing mouse line, which leads to a constitutive expression in the progeny and used to study embryonic development in absence of SPIN1. This study of SPIN1 role in embryonic development is led by Marie Le Page (now Master and soon to be PhD student).

My host laboratory interests focus on epigenetic decisions that influences development, germline specification and fertility. Previous work in male germ cells has led to the discovery of DNMT3C, a DNA methyltransferase that selectively targets young TEs, among which LINE-1 and IAP are found (Barau et al., 2016). *Dnmt3c* is expressed only in the male developing germline and with a specific window that could link its activity to the piRNA pathway. However, as of now, there is no direct link between the piRNA pathway and DNMT3C in these fetal male germ cells (prospermatogonia) (Zoch et al., 2020). Unpublished work from my host lab revealed that in prospermatogonia, DNMT3C-targeted TEs show a double H3K9me3- H3K4me3 chromatin signature. Co-occurrence of both modifications is not proven yet because

determined in a cell population. However, the ability of SPIN1 to bind both marks concomitantly, together with its expression pattern in the early male germ line, makes it a good candidate to potentially recruit DNMT3C to these dually marked domains. Recent purification of SPIN1 as a partner of MIWI2 and SPOCD1 in fetal male gonads reinforces this intuitive guess ([Zoch et al., 2020](#)), which was even reported in a recent review ([Shirane & Lorincz, 2023](#)).

My work brings light on SPIN1 action in ESCs and the results obtained in this model fuels several hypotheses and projects that are now carried out in my host lab. While I present here the first clues on the implication of SPIN1 in the regulation of TEs, several publications may be expected to come from the lab to strengthen the importance of this new actor in the germ line and early embryo.



# **BIBLIOGRAPHY**



- Ade, C., Roy-Engel, A. M., & Deininger, P. L. (2013). Alu elements: an intrinsic source of human genome instability. *Current Opinion in Virology*, 3, 639–645. <https://doi.org/10.1016/j.coviro.2013.09.002>
- Adey, N. B., Schichman, S. A., Graham, D. K., Peterson, S. N., Edgell, M. H., & Hutchison, C. A. (1994). Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Molecular Biology and Evolution*, 11(5), 778–789. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040158>
- Allouch, A., Di Primio, C., Alpi, E., Lusic, M., Arosio, D., Giacca, M., & Cereseto, A. (2011). The TRIM family protein KAP1 inhibits HIV-1 integration. *Cell Host and Microbe*, 9(6), 484–495. <https://doi.org/10.1016/j.chom.2011.05.004>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. In *Genome Biology* (Vol. 21, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13059-020-1935-5>
- Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports 2019 9:1*, 9(1), 1–5. <https://doi.org/10.1038/s41598-019-45839-z>
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., & Hannon, G. J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, 31(6), 785–799. <https://doi.org/10.1016/J.MOLCEL.2008.09.003>
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316(5825), 744–747. [https://doi.org/10.1126/SCIENCE.1142612/SUPPL\\_FILE/ARAVIN.SOM.PDF](https://doi.org/10.1126/SCIENCE.1142612/SUPPL_FILE/ARAVIN.SOM.PDF)
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., & Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature 2006 442:7099*, 442(7099), 203–207. <https://doi.org/10.1038/NATURE04916>
- Arias, J. F., Koyama, T., Kinomoto, M., & Tokunaga, K. (2012). Retroelements versus APOBEC3 family members: No great escape from the magnificent seven. *Frontiers in Microbiology*, 3(AUG), 275. <https://doi.org/10.3389/FMICB.2012.00275/BIBTEX>
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., & Shirakawa, M. (2008). Recognition of hemimethylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature 2008 455:7214*, 455(7214), 818–821. <https://doi.org/10.1038/NATURE07249>
- Arora, R., Bodak, M., Penouty, L., Hackman, C., & Ciaudo, C. (2022). Sequestration of <sc>LINE</sc> -1 in cytosolic aggregates by <sc>MOV10</sc> restricts retrotransposition. *EMBO Reports*, 23(9). <https://doi.org/10.15252/EMBR.202154458>
- Aziz, R. K., Breitbart, M., & Edwards, R. A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research*, 38(13), 4207–4217. <https://doi.org/10.1093/nar/gkq140>
- 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>
- Badge, R. M., Alisch, R. S., & Moran, J. V. (2003). ATLAS: A System to Selectively Identify Human-Specific L1 Insertions. *The American Journal of Human Genetics*, 72(4), 823–838. <https://doi.org/10.1086/373939>
- Bae, N., Gao, M., Li, X., Premkumar, T., Sbardella, G., Chen, J., & Bedford, M. T. (2017). A transcriptional coregulator, SPIN·DOC, attenuates the coactivator activity of Spindlin1. *Journal of Biological Chemistry*, 292(51), 20808–20817. <https://doi.org/10.1074/jbc.M117.814913>
- Bae, N., Viviano, M., Su, X., Lv, J., Cheng, D., Sagum, C., Castellano, S., Bai, X., Johnson, C., Khalil, M. I., Shen, J., Chen, K., Li, H., Sbardella, G., & Bedford, M. T. (2017). Developing Spindlin1 small-molecule inhibitors by using protein microarrays. *Nature Chemical Biology*, 13(7), 750–756. <https://doi.org/10.1038/nchembio.2377>
- Bailly-Bechet, M., Haudry, A., & Lerat, E. (2014). “One code to find them all”: A perl tool to conveniently parse RepeatMasker output files. *Mobile DNA*, 5(1), 1–15. <https://doi.org/10.1186/1759-8753-5-13/TABLES/3>
- Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., & Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature 2001 410:6824*, 410(6824), 120–124. <https://doi.org/10.1038/35065138>

- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Héroult, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, 354(6314), 909–912. <https://doi.org/10.1126/science.aah5143>
- Bartosovic, M., & Castelo-Branco, G. (2022). Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nature Biotechnology* 2022, 1–12. <https://doi.org/10.1038/S41587-022-01535-4>
- Beyer, S., Pontis, J., Schirwis, E., Battisti, V., Rudolf, A., Le Grand, F., & Ait-Si-Ali, S. (2016). Canonical Wnt signalling regulates nuclear export of Setdb1 during skeletal muscle terminal differentiation. *Cell Discovery* 2016 2:1, 2(1), 1–24. <https://doi.org/10.1038/CELLDISC.2016.37>
- Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A Fraction of the Mouse Genome That Is Derived from Islands of Nonmethylated, CpG-Rich DNA. In *Cell* (Vol. 40).
- Birtle, Z., & Ponting, C. P. (2006). Meisetz and the birth of the KRAB motif. *Bioinformatics*, 22(23), 2841–2845. <https://doi.org/10.1093/BIOINFORMATICS/BTL498>
- Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martínez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., & Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, 500(7461), 222–226. <https://doi.org/10.1038/nature12362>
- Boissinot, S., & Furano, A. V. (2001). Adaptive Evolution in LINE-1 Retrotransposons. *Molecular Biology and Evolution*, 18(12), 2186–2194. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003765>
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., & Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics* 2010 42:12, 42(12), 1093–1100. <https://doi.org/10.1038/NG.708>
- Bosch-Presegué, L., Raurell-Vila, H., Marazuela-Duque, A., Kane-Goldsmith, N., Valle, A., Oliver, J., Serrano, L., & Vaquero, A. (2011). Stabilization of Suv39H1 by SirT1 Is Part of Oxidative Stress Response and Ensures Genome Protection. *Molecular Cell*, 42(2), 210–223. <https://doi.org/10.1016/J.MOLCEL.2011.02.034>
- Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, 317(5845), 1760–1764. <https://doi.org/10.1126/science.1147939>
- Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004), 96–99. <https://doi.org/10.1038/nature02886>
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, 294(5551), 2536–2539. <https://doi.org/10.1126/SCIENCE.1065848>
- Bourque, G., Leong, B., Vega, V. B., Chen, X., Yen, L. L., Srinivasan, K. G., Chew, J. L., Ruan, Y., Wei, C. L., Huck, H. N., & Liu, E. T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research*, 18(11), 1752–1762. <https://doi.org/10.1101/gr.080663.108>
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell*, 128(6), 1089–1103. <https://doi.org/10.1016/J.CELL.2007.01.043>
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Morant, J. V., & Kazazian, H. H. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), 5280–5285. [https://doi.org/10.1073/PNAS.0831042100/SUPPL\\_FILE/1042TABLE5.XLS](https://doi.org/10.1073/PNAS.0831042100/SUPPL_FILE/1042TABLE5.XLS)
- Bu, P., Yagi, S., Shiota, K., Alam, S. M. K., Vivian, J. L., Wolfe, M. W., Rumi, M. A. K., Chakraborty, D., Kubota, K., Dhakal, P., & Soares, M. J. (2017). Origin of a rapidly evolving homeostatic control system programming testis function. *Journal of Endocrinology*, 234(2), 217–232. <https://doi.org/10.1530/JOE-17-0250>
- Bulut-Karslioglu, A., DeLaRosa-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>
- Burns, K. H. (2017). Transposable elements in cancer. In *Nature Reviews Cancer* (Vol. 17, Issue 7, pp. 415–424). <https://doi.org/10.1038/nrc.2017.35>
- Callahan, K. E., Hickman, A. B., Jones, C. E., Ghirlardo, R., & Furano, A. V. (2012). Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. *Nucleic Acids Research*, 40(2), 813–827. <https://doi.org/10.1093/NAR/GKR728>



- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., & Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes and Development*, *28*(13), 1397–1409. <https://doi.org/10.1101/gad.241661.114>
- Chang, Y., Sun, L., Kokura, K., Horton, J. R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J. M., Bedford, M. T., Zhang, X., Shinkai, Y., Fang, J., & Cheng, X. (2011). MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nature Communications* *2011 2:1*, *2*(1), 1–10. <https://doi.org/10.1038/NCOMMS1549>
- 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>
- Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., Landau, N. R., & Weitzman, M. D. (2006). APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. *Current Biology*, *16*(5), 480–485. <https://doi.org/10.1016/J.CUB.2006.01.031>
- Chen, X., Wang, Y. W., & Gao, P. (2018). SPIN1, negatively regulated by miR-148/152, enhances Adriamycin resistance via upregulating drug metabolizing enzymes and transporter in breast cancer. *Journal of Experimental and Clinical Cancer Research*, *37*(1), 1–12. <https://doi.org/10.1186/S13046-018-0748-9/FIGURES/5>
- Chen, X., Wang, Y. W., Xing, A. Y., Xiang, S., Shi, D. B., Liu, L., Li, Y. X., & Gao, P. (2016). Suppression of SPIN1-mediated PI3K–Akt pathway by miR-489 increases chemosensitivity in breast cancer. *Journal of Pathology*, *239*(4), 459–472. <https://doi.org/10.1002/path.4743>
- Chew, T. G., Peaston, A., Lim, A. K., Lorthongpanich, C., Knowles, B. B., & Solter, D. (2013). A Tudor Domain Protein SPINDLIN1 Interacts with the mRNA-Binding Protein SERBP1 and Is Involved in Mouse Oocyte Meiotic Resumption. *PLoS ONE*, *8*(7), 1–10. <https://doi.org/10.1371/journal.pone.0069764>
- Chuang, L. S. H., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., & Li, B. F. L. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21(WAF1). *Science*, *277*(5334), 1996–2000. <https://doi.org/10.1126/SCIENCE.277.5334.1996/ASSET/A9058FDD-11F6-43D7-B58D-347A41B960C8/ASSETS/GRAPHIC/SE3875724004.JPEG>
- Chuma, S., Hosokawa, M., Kitamura, K., Kasai, S., Fujioka, M., Hiyoshi, M., Takamune, K., Noce, T., & Nakatsuji, N. (2006). Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(43), 15894–15899. [https://doi.org/10.1073/PNAS.0601878103/SUPPL\\_FILE/01878FIG8.PDF](https://doi.org/10.1073/PNAS.0601878103/SUPPL_FILE/01878FIG8.PDF)
- Chuong, E. B., Rumi, M. A. K., Soares, M. J., & Baker, J. C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature Genetics* *2013 45:3*, *45*(3), 325–329. <https://doi.org/10.1038/NG.2553>
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Paolo, P., Fiore, D., & Bonapace, I. M. (2004). Np95 Is a Histone-Binding Protein Endowed with Ubiquitin Ligase Activity. *MOLECULAR AND CELLULAR BIOLOGY*, *24*(6), 2526–2535. <https://doi.org/10.1128/MCB.24.6.2526-2535.2004>
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., & Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* *2009 460:7259*, *460*(7259), 1127–1131. <https://doi.org/10.1038/NATURE08248>
- Cusack, M., King, H. W., Spingardi, P., Kessler, B. M., Klose, R. J., & Kriaucionis, S. (2020). Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. *Genome Research*, *30*(10), 1393–1406. <https://doi.org/10.1101/GR.257576.119>
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics* *2017 49:6*, *49*(6), 941–945. <https://doi.org/10.1038/NG.3858>
- Deberardinis, R. J., & Kazazian, H. H. (1999). Analysis of the Promoter from an Expanding Mouse Retrotransposon Subfamily. *Genomics*. <http://www.idealibrary.com>
- Delaney, C. E., Methot, S. P., Guidi, M., Katic, I., Gasser, S. M., & Padeken, J. (2019). Heterochromatic foci and transcriptional repression by an unstructured MET-2/SETDB1 co-factor LIN-65. *Journal of Cell Biology*, *218*(3), 820–838. <https://doi.org/10.1083/JCB.201811038/VIDEO-2>
- Delbarre, E., Ivanauskienė, K., Spirkoski, J., Shah, A., Vekterud, K., Moskaug, J., Bøe, S. O., Wong, L. H., Küntziger, T., & Collas, P. (2017). PML protein organizes heterochromatin domains where it

- regulates histone H3.3 deposition by ATRX/DAXX. *Genome Research*, 27(6), 913–921. <https://doi.org/10.1101/gr.215830.116>
- Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C. N., Diedrich, J. K., Aslanian, A., Ma, J., Moresco, J. J., Moore, L., Hunter, T., Saghatelian, A., & Gage, F. H. (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*, 163(3), 583–593. <https://doi.org/10.1016/J.CELL.2015.09.025>
- Desai, V. P., Chouaref, J., Wu, H., Pastor, W. A., Kan, R. L., Oey, H. M., Li, Z., Ho, J., Vonk, K. K. D., San Leon Granado, D., Christopher, M. A., Clark, A. T., Jacobsen, S. E., & Daxinger, L. (2021). The role of MORC3 in silencing transposable elements in mouse embryonic stem cells. *Epigenetics and Chromatin*, 14(1), 1–14. <https://doi.org/10.1186/S13072-021-00420-9/FIGURES/4>
- Devi, M. S., Meiguilungpou, R., Sharma, A. L., Anjali, C., Devi, K. M., Singh, L. S., & Singh, T. R. (2019). Spindlin docking protein (SPIN.DOC) interaction with SPIN1 (a histone code reader) regulates Wnt signaling. *Biochemical and Biophysical Research Communications*, 511(3), 498–503. <https://doi.org/10.1016/j.bbrc.2019.02.096>
- Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *Journal of Biological Chemistry*, 285(34), 26114–26120. <https://doi.org/10.1074/JBC.M109.089433>
- Ding, D., Liu, J., Dong, K., Midic, U., Hess, R. A., Xie, H., Demireva, E. Y., & Chen, C. (2017). PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. *Nature Communication*, 8(819). <https://doi.org/10.1038/s41467-017-00854-4>
- 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*, 29(1), 15–21. <https://doi.org/10.1093/BIOINFORMATICS/BTS635>
- Dong, K. B., Maksakova, I. A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H. W., Lam, L. L., Mager, D. L., Schübeler, D., Tachibana, M., Shinkai, Y., & Lorincz, M. C. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO Journal*, 27(20), 2691–2701. <https://doi.org/10.1038/EMBOJ.2008.193>
- Doucet, A. J., Hulme, A. E., Sahinovic, E., Kulpa, D. A., Moldovan, J. B., Kopera, H. C., Athanikar, J. N., Hasnaoui, M., Bucheton, A., Moran, J. V., & Gilbert, N. (2010). Characterization of LINE-1 Ribonucleoprotein Particles. *PLOS Genetics*, 6(10), e1001150. <https://doi.org/10.1371/JOURNAL.PGEN.1001150>
- Douse, C. H., Tchasovnikarova, I. A., Timms, R. T., Protasio, A. V., Seczynska, M., Prigozhin, D. M., Albecka, A., Wagstaff, J., Williamson, J. C., Freund, S. M. V., Lehner, P. J., & Modis, Y. (2020). TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-18761-6>
- Du, W., Dong, Q., Zhang, Z., Liu, B., Zhou, T., Xu, R. ming, Wang, H., Zhu, B., & Li, Y. (2019). Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger-type E3 ubiquitin ligase UHRF1. *Journal of Biological Chemistry*, 294(22), 8907–8917. <https://doi.org/10.1074/jbc.RA119.008008>
- Du, Y., & Qian, C. (2022). Non-canonical bivalent H3K4me3K9me3 recognition by Spindlin1/C11orf84 complex. <https://doi.org/10.1002/bies.202100229>
- Du, Y., Yan, Y., Xie, S., Huang, H., Wang, X., Ng, R. K., Zhou, M. M., & Qian, C. (2021). Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-21236-x>
- Ducroux, A., Benhenda, S., Rivière, L., Semmes, O. J., Benkirane, M., & Neuveut, C. (2014). The Tudor Domain Protein Spindlin1 Is Involved in Intrinsic Antiviral Defense against Incoming Hepatitis B Virus and Herpes Simplex Virus Type 1. *PLoS Pathogens*, 10(9), 1004343. <https://doi.org/10.1371/journal.ppat.1004343>
- Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G., & Heidmann, T. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proceedings of the National Academy of Sciences of the United States of America*, 108(46), E1164–E1173. [https://doi.org/10.1073/PNAS.1112304108/SUPPL\\_FILE/PNAS.201112304SI.PDF](https://doi.org/10.1073/PNAS.1112304108/SUPPL_FILE/PNAS.201112304SI.PDF)
- Dura, M., Teissandier, A., Armand, M., Barau, J., Lapoujade, C., Fouchet, P., Bonneville, L., Schulz, M., Weber, M., Baudrin, L. G., Lameiras, S., & Bourc'his, D. (2022). DNMT3A-dependent DNA

- methylation is required for spermatogonial stem cells to commit to spermatogenesis. *Nature Genetics* 2022 54:4, 54(4), 469–480. <https://doi.org/10.1038/S41588-022-01040-Z>
- Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H. M., Turelli, P., & Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Developmental Cell*, 36(6), 611–623. <https://doi.org/10.1016/J.DEVCEL.2016.02.024>
- Ecco, G., Imbeault, M., & Trono, D. (2017). KRAB zinc finger proteins. *Development (Cambridge)*, 144(15), 2719–2729. <https://doi.org/10.1242/dev.132605>
- 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* 2015 522:7555, 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. *BioRxiv*.
- Evsikov, A. V., De Vries, W. N., Peaston, A. E., Radford, E. E., Fancher, K. S., Chen, F. H., Blake, J. A., Bult, C. J., Latham, K. E., Solter, D., & Knowles, B. B. (2004). Systems biology of the 2-cell mouse embryo. *Cytogenetic and Genome Research*, 105(2–4), 240–250. <https://doi.org/10.1159/000078195>
- Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., & Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *20*. <https://doi.org/10.1038/nsmb.2495>
- Fagan, V., Johansson, C., Gileadi, C., Monteiro, O., Dunford, J. E., Nibhani, R., Philpott, M., Malzahn, J., Wells, G., Faram, R., Cribbs, A. P., Halidi, N., Li, F., Chau, I., Greschik, H., Velupillai, S., Allali-Hassani, A., Bennett, J., Christott, T., ... Oppermann, U. (2019). A Chemical Probe for Tudor Domain Protein Spindlin1 to Investigate Chromatin Function. *Journal of Medicinal Chemistry*, 62(20), 9008–9025. <https://doi.org/10.1021/acs.jmedchem.9b00562>
- Fang, Z., Cao, B., Liao, J.-M., Deng, J., Plummer, K. D., Liao, P., Liu, T., Zhang, W., Zhang, K., Li, L., Margolin, D., Zeng, S. X., Xiong, J., & Lu, H. (2018). SPIN1 promotes tumorigenesis by blocking the uL18 (universal large ribosomal subunit protein 18)-MDM2-p53 pathway in human cancer. *ELife*, 7, 1–21. <https://doi.org/10.7554/elife.31275>
- Fanning, T. G. (1983). Size and structure of the highly repetitive BAM HI element in mice. *Nucleic Acids Research*, 11. <https://academic.oup.com/nar/article/11/15/5073/1474425>
- Fehrmann, F., Jung, M., Zimmermann, R., & Kräusslich, H.-G. (2003). Transport of the Intracisternal A-Type Particle Gag Polyprotein to the Endoplasmic Reticulum Is Mediated by the Signal Recognition Particle. *Journal of Virology*, 77(11), 6293–6304. <https://doi.org/10.1128/JVI.77.11.6293-6304.2003/ASSET/238A494C-D960-44BC-B2BC-AF9BB7E52771/ASSETS/GRAPHIC/JV1130027008.JPEG>
- Fehrmann, F., Welker, R., & Kräusslich, H. G. (1997). Intracisternal A-Type Particles Express Their Proteinase in a Separate Reading Frame by Translational Frameshifting, Similar to D-Type Retroviruses. *Virology*, 235(2), 352–359. <https://doi.org/10.1006/VIRO.1997.8708>
- Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, 42(8), e69–e69. <https://doi.org/10.1093/NAR/GKU154>
- Ferraj, A., Audano, P. A., Balachandran, P., Czechanski, A., Flores, J. I., Radecki, A. A., Mosur, V., Gordon, D. S., Walawalkar, I. A., Eichler, E. E., Reinholdt, L. G., & Beck, C. R. (2023). Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements. *Cell Genomics*, 3(5), 100291. <https://doi.org/10.1016/j.xgen.2023.100291>
- Ferry, L., Fournier, A., Arita, K., Shinkai, Y., & Defossez, P.-A. (2017). Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. *Molecular Cell*, 67, 550–565. <https://doi.org/10.1016/j.molcel.2017.07.012>
- Feusier, J., Watkins, W. S., Thomas, J., Farrell, A., Witherspoon, D. J., Baird, L., Ha, H., Xing, J., & Jorde, L. B. (2019). Pedigree-based estimation of human mobile element retrotransposition rates. *Genome Research*, 29(10), 1567–1577. <https://doi.org/10.1101/GR.247965.118/-/DC1>
- Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y. L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., & Reik, W. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*, 13(3), 351–359. <https://doi.org/10.1016/J.STEM.2013.06.004>

- Flasch, D. A., Macia, Á., Sánchez, L., Ljungman, M., Heras, S. R., García-Pérez, J. L., Wilson, T. E., & Moran, J. V. (2019). Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell*, 177(4), 837-851.e28. <https://doi.org/10.1016/J.CELL.2019.02.050>
- Franz, H., Greschik, H., Willmann, D., Ozretić, L., Jilg, C. A., Wardelmann, E., Jung, M., Buettner, R., & Schüle, R. (2015). The histone code reader SPIN1 controls RET signaling in liposarcoma. *Oncotarget*, 6(7). <https://doi.org/10.18632/oncotarget.3000>
- Friedman, J. R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X.-P., Neilson, E. G., & Iii, F. J. R. (1996). *KAP-1, a novel corepressor for the highly conserved KRAB repression domain*.
- Fulka, H., & Langerova, A. (2014). The maternal nucleolus plays a key role in centromere satellite maintenance during the oocyte to embryo transition. *Development*, 141(8), 1694–1704. <https://doi.org/10.1242/DEV.105940>
- Gagnier, L., Belancio, V. P., & Mager, D. L. (2019). Mouse germ line mutations due to retrotransposon insertions. In *Mobile DNA* (Vol. 10, Issue 1, pp. 1–22). Mobile DNA. <https://doi.org/10.1186/s13100-019-0157-4>
- Gao, Y., Yue, W., Zhang, P., Li, L., Xie, X., Yuan, H., Chen, L., Liu, D., Yan, F., & Pei, X. (2005). Spindlin1, a novel nuclear protein with a role in the transformation of NIH3T3 cells. *Biochemical and Biophysical Research Communications*, 335(2), 343–350. <https://doi.org/10.1016/j.bbrc.2005.07.087>
- García-Cao, M., O'Sullivan, R., Peters, A. H. F. M., Jenuwein, T., & Blasco, M. A. (2003). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nature Genetics* 2004 36:1, 36(1), 94–99. <https://doi.org/10.1038/NG1278>
- Gardiner-Garden, M., & Frommer, M. (1987). CpG Islands in Vertebrate Genomes. In *J. Mol. Biol* (Vol. 196).
- Garland, W., Müller, I., Wu, M., Schmid, M., Imamura, K., Rib, L., Sandelin, A., Helin, K., & Jensen, T. H. (2022). Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. *Molecular Cell*, 82(9), 1691-1707.e8. <https://doi.org/10.1016/J.MOLCEL.2022.03.004>
- Gaubatz, J. W., Arcement, B., & Cutler, R. G. (1991). GENE EXPRESSION OF AN ENDOGENOUS RETROVIRUS-LIKE ELEMENT DURING MURINE DEVELOPMENT AND AGING. In *Mechanisms of Ageing and Development* (Vol. 57).
- Gauchier, M., Kan, S., Barral, A., Sauzet, S., Agirre, E., Bonnell, E., Saksouk, N., Barth, T. K., Ide, S., Urbach, S., Wellinger, R. J., Luco, R. F., Imhof, A., & Déjardin, J. (2019). SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Science Advances*, 5(5). [https://doi.org/10.1126/SCIADV.AAV3673/SUPPL\\_FILE/AAV3673\\_TABLE\\_S2.XLSX](https://doi.org/10.1126/SCIADV.AAV3673/SUPPL_FILE/AAV3673_TABLE_S2.XLSX)
- Ge, S. X. (2017). Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. *BMC Genomics* 2017 18:1, 18(1), 1–19. <https://doi.org/10.1186/S12864-017-3566-0>
- Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A. F., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., Ben-Haim, M. S., Eyal, E., Yunger, S., Pinto, Y., Jaitin, D. A., Viukov, S., Rais, Y., Krupalnik, V., Chomsky, E., ... Hanna, J. H. (2015). m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, 347(6225), 1002–1006. <https://doi.org/10.1126/science.1261417>
- Goodier, J. L. (2016). Restricting retrotransposons: a review. *Mobile DNA*, 7(1). <https://doi.org/10.1186/S13100-016-0070-Z>
- Goodier, J. L., Cheung, L. E., & Kazazian, H. H. (2012). MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLOS Genetics*, 8(10), e1002941. <https://doi.org/10.1371/JOURNAL.PGEN.1002941>
- Goodier, J. L., Ostertag, E. M., Du, K., & Kazazian, J. (2001). A novel active L1 retrotransposon subfamily in the mouse. *Genome Research*, 11(10), 1677–1685. <https://doi.org/10.1101/gr.198301>
- Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., & Kazazian, H. H. (2015). The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. *PLOS Genetics*, 11(5), e1005252. <https://doi.org/10.1371/JOURNAL.PGEN.1005252>
- Goodier, J. L., Zhang, L., Vetter, M. R., & Kazazian, H. H. (2007). LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex †. *MOLECULAR AND CELLULAR BIOLOGY*, 27(18), 6469–6483. <https://doi.org/10.1128/MCB.00332-07>
- Graf, U., Casanova, E. A., Wyck, S., Dalcher, D., Gatti, M., Vollenweider, E., Okoniewski, M. J., Weber, F. A., Patel, S. S., Schmid, M. W., Li, J., Sharif, J., Wanner, G. A., Koseki, H., Wong, J., Pelczar, P., Penengo, L., Santoro, R., & Cinelli, P. (2017). Prmel17 mediates ground-state pluripotency

- through proteasomal-epigenetic combined pathways. *Nature Cell Biology*, 19(7), 763–773. <https://doi.org/10.1038/ncb3554>
- Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, 20(October). <https://doi.org/10.1038/s41580-019-0159-6>
- Gregersen, L. H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., & Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs. *Molecular Cell*, 54, 573–585. <https://doi.org/10.1016/j.molcel.2014.03.017>
- Greschik, H., Duteil, D., Messaddeq, N., Willmann, D., Arrigoni, L., Sum, M., Jung, M., Metzger, D., Manke, T., Günther, T., & Schüle, R. (2017). The histone code reader Spin1 controls skeletal muscle development. *Cell Death & Disease*, 8(1), 1–15. <https://doi.org/10.1038/cddis.2017.468>
- Groh, S., Milton, A. V., Marinelli, L. K., Sickinger, C. V., Russo, A., Bollig, H., de Almeida, G. P., Schmidt, A., Forné, I., Imhof, A., & Schotta, G. (2021). Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nature Communications* 2021 12:1, 12(1), 1–18. <https://doi.org/10.1038/S41467-021-26288-7>
- Groh, S., & Schotta, G. (2017). Silencing of endogenous retroviruses by heterochromatin. *Cellular and Molecular Life Sciences*, 74(11), 2055–2065. <https://doi.org/10.1007/s00018-017-2454-8>
- Grummt, I., & Längst, G. (2013). Epigenetic control of RNA polymerase I transcription in mammalian cells. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(3–4), 393–404. <https://doi.org/10.1016/j.bbagr.2012.10.004>
- Guo, H., Chitiprolu, M., Gagnon, D., Meng, L., Perez-Iratxeta, C., Lagace, D., & Gibbins, D. (2014). Autophagy supports genomic stability by degrading retrotransposon RNA. *Nature Communications* 2014 5:1, 5(1), 1–11. <https://doi.org/10.1038/ncomms6276>
- Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., Benedetti, R., Altucci, L., Jansen, J. H., Walter, J., Gut, I. G., Marks, H., & Stunnenberg, H. G. (2013). Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell*, 13(3), 360–369. <https://doi.org/10.1016/J.STEM.2013.06.002>
- Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A. S., Mattei, A. L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., Giehr, P., Buschow, R., Galonska, C., von Meyenn, F., Pappalardi, M. B., McCabe, M. T., Wittler, L., Giesecke-Thiel, C., Mielke, T., ... Meissner, A. (2021). Dnmt1 has de novo activity targeted to transposable elements. *Nature Structural and Molecular Biology*, 28(7), 594–603. <https://doi.org/10.1038/s41594-021-00603-8>
- Han, M., Li, J., Cao, Y., Huang, Y., Li, W., Zhu, H., Zhao, Q., Han, J. D. J., Wu, Q., Li, J., Feng, J., & Wong, J. (2020). A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. *Nucleic Acids Research*, 48(21), 12116–12134. <https://doi.org/10.1093/NAR/GKAA1003>
- Hancks, D. C., & Kazazian, H. H. (2012). Active human retrotransposons: variation and disease. *Current Opinion in Genetics and Development*, 22, 191–203. <https://doi.org/10.1016/j.gde.2012.02.006>
- Harten, S. K., Bruxner, T. J., Bharti, V., Blewitt, M., Nguyen, T. M. T., Whitelaw, E., & Epp, T. (2014). The first mouse mutants of D14Abb1e (Fam208a) show that it is critical for early development. *Mammalian Genome*, 25(7–8), 293–303. <https://doi.org/10.1007/s00335-014-9516-0>
- Hashimoto, H., Horton, J. R., Zhang, X., Bostick, M., Jacobsen, S. E., & Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 2008 455:7214, 455(7214), 826–829. <https://doi.org/10.1038/NATURE07280>
- Hermann, B. P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K. N., Chen, I. C., Gildersleeve, H., Lehle, J. D., Mayo, M., Westernströer, B., Law, N. C., Oatley, M. J., Velte, E. K., Niedenberger, B. A., Fritze, D., Silber, S., Geyer, C. B., Oatley, J. M., & McCarrey, J. R. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Reports*, 25(6), 1650–1667. <https://doi.org/10.1016/j.celrep.2018.10.026>
- Hu, S., Li, J., Xu, F., Mei, S., Le Duff, Y., Yin, L., Pang, X., Cen, S., Jin, Q., Liang, C., & Guo, F. (2015). SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. *PLOS Genetics*, 11(7), e1005367. <https://doi.org/10.1371/JOURNAL.PGEN.1005367>
- Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylitz, J. J., Marioni, J. C., Hackett, J. A., & Surani, M. A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *ELife*, 6. <https://doi.org/10.7554/ELIFE.22345>

- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genetics*, *6*(9), e1001134. <https://doi.org/10.1371/JOURNAL.PGEN.1001134>
- Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, *543*(7646), 550–554. <https://doi.org/10.1038/nature21683>
- Inoue, A., Matoba, S., & Zhang, Y. (2012). Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. *Cell Research* *2012* *22*:12, *22*(12), 1640–1649. <https://doi.org/10.1038/CR.2012.160>
- Ivanov, A. V., Peng, H., Yurchenko, V., Yap, K. L., Negorev, D. G., Schultz, D. C., Psulkowski, E., Fredericks, W. J., White, D. E., Maul, G. G., Sadofsky, M. J., Zhou, M. M., & Rauscher, F. J. (2007). PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. *Molecular Cell*, *28*(5), 823–837. <https://doi.org/10.1016/J.MOLCEL.2007.11.012>
- Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., & Torres-Padilla, M. E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics*, *49*(10), 1502–1510. <https://doi.org/10.1038/ng.3945>
- Jain, D., Baldi, S., Zabel, A., Straub, T., & Becker, P. B. (2015). Active promoters give rise to false positive ‘Phantom Peaks’ in ChIP-seq experiments. *Nucleic Acids Research*, *43*(14), 6959–6968. <https://doi.org/10.1093/NAR/GKV637>
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O’Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., ... Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. <https://doi.org/10.1038/nbt.4060>
- Janecki, D. M., Sajek, M., Smialek, M. J., Kotecki, M., Ginter-Matuszewska, B., Kuczynska, B., Spik, A., Kolanowski, T., Kitazawa, R., Kurpisz, M., Jaruzelska, J., Janecki, D. M., Sajek, M., Smialek, M. J., Kotecki, M., Ginter-Matuszewska, B., Kuczynska, B., Spik, A., Kolanowski, T., ... Jaruzelska, J. (2018). SPIN1 is a proto-oncogene and SPIN3 is a tumor suppressor in human seminoma. *Oncotarget*, *9*(65), 32466–32477. <https://doi.org/10.18632/ONCOTARGET.25977>
- Jiang, X., Zhu, X., Cheng, Y., Azhar, M., Xing, X., Li, W., Cao, Y., Shi, Q., & Bao, J. (2021). The Spin1 interactor, Spindoc, is dispensable for meiotic division, but essential for haploid spermatid development in mice. *Reproductive Biology and Endocrinology* *2021* *19*:1, *19*(1), 1–11. <https://doi.org/10.1186/S12958-021-00828-8>
- Kamaletdinova, T., Fanaei-Kahrani, Z., & Wang, Z. Q. (2019). The Enigmatic Function of PARP1: From PARylation Activity to PAR Readers. *Cells* *2019*, *Vol. 8*, *Page 1625*, *8*(12), 1625. <https://doi.org/10.3390/CELLS8121625>
- Kaneda, M., Hirasawa, R., Chiba, H., Okano, M., Li, E., & Sasaki, H. (2010). Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes to Cells*, *15*(3), 169–179. <https://doi.org/10.1111/J.1365-2443.2009.01374.X>
- Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., & Lorincz, M. C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs. *Cell Stem Cell*, *8*(6), 676–687. <https://doi.org/10.1016/j.stem.2011.04.004>
- Kauzlaric, A., Ecco, G., Cassano, M., Duc, J., Imbeault, M., & Trono, D. (2017). The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PLoS ONE*, *12*(3). <https://doi.org/10.1371/journal.pone.0173746>
- Kazazian, H. H., & Moran, J. V. (2017). Mobile DNA in Health and Disease. *New England Journal of Medicine*, *377*(4), 361–370. [https://doi.org/10.1056/NEJMRA1510092/SUPPL\\_FILE/NEJMRA1510092\\_DISCLOSURES.PDF](https://doi.org/10.1056/NEJMRA1510092/SUPPL_FILE/NEJMRA1510092_DISCLOSURES.PDF)
- Kigami, D., Minami, N., Takayama, H., & Imai, H. (2003). MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biology of Reproduction*, *68*(2), 651–654. <https://doi.org/10.1095/BIOLREPROD.102.007906>
- Kimball, A. S., Davis, F. M., denDekker, A., Joshi, A. D., Schaller, M. A., Bermick, J., Xing, X., Burant, C. F., Obi, A. T., Nysz, D., Robinson, S., Allen, R., Lukacs, N. W., Henke, P. K., Gudjonsson, J. E., Moore, B. B., Kunkel, S. L., & Gallagher, K. A. (2019). The Histone Methyltransferase Setdb2 Modulates Macrophage Phenotype and Uric Acid Production in Diabetic Wound Repair. *Immunity*, *51*(2), 258–271.e5. <https://doi.org/10.1016/J.IMMUNI.2019.06.015>

- Kirino, Y., & Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature Structural & Molecular Biology* 2007 14:4, 14(4), 347–348. <https://doi.org/10.1038/NSMB1218>
- Kissinger, J. C., & DeBarry, J. (2011). Genome cartography: charting the apicomplexan genome. *Trends in Parasitology*, 27(8), 345–354. <https://doi.org/10.1016/J.PT.2011.03.006>
- Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, 20(1), 1–13. <https://doi.org/10.1186/S13059-019-1910-1/FIGURES/6>
- Kowald, A. (1997). *Possible Mechanisms for the Regulation of Telomere Length*.
- Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571–1572. <https://doi.org/10.1093/BIOINFORMATICS/BTR167>
- Kuff, E. L., & Lueders, K. K. (1988). The Intracisternal A-Particle Gene Family: Structure and Functional Aspects. *Advances in Cancer Research*, 51(C), 183–276. [https://doi.org/10.1016/S0065-230X\(08\)60223-7](https://doi.org/10.1016/S0065-230X(08)60223-7)
- 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* 2001 410:6824, 410(6824), 116–120. <https://doi.org/10.1038/35065132>
- Lahaye, X., Gentili, M., Silvin, A., Conrad, C., Picard, L., Jouve, M., Zueva, E., Maurin, M., Nadalin, F., Knott, G. J., Zhao, B., Du, F., Rio, M., Amiel, J., Fox, A. H., Li, P., Etienne, L., Bond, C. S., Colleaux, L., & Manel, N. (2018). NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell*, 175(2), 488-501.e22. <https://doi.org/10.1016/J.CELL.2018.08.062>
- Lanciano, S., Philippe, C., Sarkar, A., Pratella, D., Domrane, C., Doucet, A. J., Essen, D. van, Sacconi, S., Ferry, L., Defossez, P.-A., & Cristofari, G. (2023). Comprehensive locus-specific L1 DNA methylation profiling reveals the epigenetic and transcriptional interplay between L1s and their integration sites. *BioRxiv*, 2023.01.03.522582. <https://doi.org/10.1101/2023.01.03.522582>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., & Peters, A. H. F. M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, 13(14), 1192–1200. [https://doi.org/10.1016/S0960-9822\(03\)00432-9](https://doi.org/10.1016/S0960-9822(03)00432-9)
- Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grable, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A., Surani, M. A., & Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural and Molecular Biology*, 20(3), 311–316. <https://doi.org/10.1038/nsmb.2510>
- Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., & Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), 14075–14080. <https://doi.org/10.1073/PNAS.1008850107/-/DCSUPPLEMENTAL>
- Li, T., Wang, L., Du, Y., Xie, S., Yang, X., Lian, F., Zhou, Z., & Qian, C. (2018). Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Research*, 46(6), 3218–3231. <https://doi.org/10.1093/nar/gky104>
- Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., Wang, M., Duan, X., Tan, J., Zheng, Y., Zhang, P., Fan, G., Wong, J., Xu, G. L., Wang, Z., ... Zhu, B. (2018). Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* 2018 564:7734, 564(7734), 136–140. <https://doi.org/10.1038/S41586-018-0751-5>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. <https://doi.org/10.1093/BIOINFORMATICS/BTT656>
- Lim, S. L., Qu, Z. P., Kortschak, R. D., Lawrence, D. M., Geoghegan, J., Hempfling, A. L., Bergmann, M., Goodnow, C. C., Ormandy, C. J., Wong, L., Mann, J., Scott, H. S., Jamsai, D., Adelson, D. L., & O'Bryan, M. K. (2015). HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. *PLoS Genetics*, 11(10), e1005620. <https://doi.org/10.1371/JOURNAL.PGEN.1005620>
- Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghghi, F. G., ... Ecker, J. R. (2013). Global epigenomic reconfiguration

- during mammalian brain development. *Science*, 341(6146).  
<https://doi.org/10.1126/SCIENCE.1237905>
- 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* 2021 591:7849, 591(7849), 322–326. <https://doi.org/10.1038/S41586-021-03313-9>
- Liu, L., Oliveira, N. M. M., Cheney, K. M., Pade, C., Dreja, H., Bergin, A. M. H., Borgdorff, V., Beach, D. H., Bishop, C. L., Dittmar, M. T., & McKnight, A. (2011). A whole genome screen for HIV restriction factors. *Retrovirology*, 8(1), 94. <https://doi.org/10.1186/1742-4690-8-94/FIGURES/6>
- Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, 553(7687), 228–232. <https://doi.org/10.1038/nature25179>
- Liu, S., Brind'Amour, J., Karimi, M. M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., & Lorincz, M. C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development*, 28(18), 2041–2055. <https://doi.org/10.1101/GAD.244848.114>
- Liu, X., Zhang, Y., Wen, Z., Hao, Y., Banks, C. A. S., Lange, J. J., Slaughter, B. D., Unruh, J. R., Florens, L., Abmayr, S. M., Workman, J. L., & Washburn, M. P. (2020). Driving integrative structural modeling with serial capture affinity purification. *PNAS*, 117(50), 31861–31870. <https://doi.org/10.1073/pnas.2007931117/-/DCSupplemental>
- Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., Imhof, A., Nakatani, Y., Dent, S. Y. R., & Almouzni, G. (2009). The HP1 $\alpha$ -CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Reports*, 10(7), 769–775. <https://doi.org/10.1038/EMBOR.2009.90>
- Lu, C., Contreras, X., & Peterlin, B. M. (2011). P Bodies Inhibit Retrotransposition of Endogenous Intracisternal A Particles. *Journal of Virology*, 85(13), 6244–6251. <https://doi.org/10.1128/jvi.02517-10>
- Lu, R., & Wang, G. G. (2014). Tudor: a versatile family of histone methylation “readers” *Histone modification and its “reader” proteins in gene regulation*. 38(11), 1–18. <https://doi.org/10.1016/j.tibs.2013.08.002>
- Lun, A. T. L., & Smyth, G. K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Research*, 44(5), e45–e45. <https://doi.org/10.1093/NAR/GKV1191>
- Lv, B. bei, Ma, R. ran, Chen, X., Zhang, G. hao, Song, L., Wang, S. xia, Wang, Y. wen, Liu, H. ting, & Gao, P. (2020). E2F1-activated SPIN1 promotes tumor growth via a MDM2-p21-E2F1 feedback loop in gastric cancer. *Molecular Oncology*, 14(10), 2629–2645. <https://doi.org/10.1002/1878-0261.12778>
- Maclennan, M., García-Cañadas, M., Reichmann, J., Khazina, E., Wagner, G., Playfoot, C. J., Salvador-Palomeque, C., Mann, A. R., Peressini, P., Sanchez, L., Dobie, K., Read, D., Hung, C. C., Eskeland, R., Meehan, R. R., Weichenrieder, O., García-Pérez, J. L., & Adams, I. R. (2017). Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *ELife*, 6. <https://doi.org/10.7554/ELIFE.26152>
- Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J., & Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. *Proceedings of the National Academy of Sciences of the United States of America*, 109(19), 7385–7390. [https://doi.org/10.1073/PNAS.1200913109/SUPPL\\_FILE/PNAS.201200913SI.PDF](https://doi.org/10.1073/PNAS.1200913109/SUPPL_FILE/PNAS.201200913SI.PDF)
- Malik, H. S., Burke, W. D., & Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*, 16(6), 793–805. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026164>
- Mankan, A. K., Schmidt, T., Chauhan, D., Goldeck, M., Höning, K., Gaidt, M., Kubarenko, A. V., Andreeva, L., Hopfner, K.-P., & Hornung, V. (2014). Cytosolic RNA:DNA hybrids activate the cGAS–STING axis. *The EMBO Journal*, 33(24), 2937–2946. <https://doi.org/10.15252/EMBJ.201488726>
- Martens, J. H. A., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., & Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO Journal*, 24(4), 800–812. <https://doi.org/10.1038/SJ.EMBOJ.7600545>
- Martin, S. L., & Bushman, F. D. (2001). Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1 Retrotransposon. *MOLECULAR AND CELLULAR BIOLOGY*, 21(2), 467–475. <https://doi.org/10.1128/MCB.21.2.467-475.2001>



- 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), 927–931. <https://doi.org/10.1038/nature08858>
- Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., Fukami, K., Osborne, T. F., Kodama, T., Aburatani, H., & Sakai, J. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Molecular Cell*, *60*(4), 584–596. <https://doi.org/10.1016/j.molcel.2015.10.025>
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, *36*, 345–355.
- Michaud, E. J., Van Vugt, M. J., Bultman, S. J., Sweet, H. O., Davisson, M. T., & Woychik, R. P. (1994). *Differential expression of a new dominant agouti allele (A is correlated with methylation state and is influenced by parental lineage)*.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., ... Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. <https://doi.org/10.1038/nature06008>
- Mita, P., Wudzinska, A., Sun, X., Andrade, J., Nayak, S., Kahler, D. J., Badri, S., LaCava, J., Ueberheide, B., Yun, C. Y., Fenyö, D., & Boeke, J. D. (2018). LINE-1 protein localization and functional dynamics during the cell cycle. *ELife*, *7*. <https://doi.org/10.7554/ELIFE.30058>
- Mitra, R., Li, X., Kapusta, A., Mayhew, D., Mitra, R. D., Feschotte, C., & Craig, N. L. (2013). Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 234–239. [https://doi.org/10.1073/PNAS.1217548110/SUPPL\\_FILE/ST04.DOCX](https://doi.org/10.1073/PNAS.1217548110/SUPPL_FILE/ST04.DOCX)
- Modzelewski, A. J., Gan Chong, J., Wang, T., & He, L. (2022). Mammalian genome innovation through transposon domestication. *Nature Cell Biology* *2022* *24*:9, 24(9), 1332–1340. <https://doi.org/10.1038/S41556-022-00970-4>
- Modzelewski, A. J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., Tjokro, K., Sales, G., Biton, A., Anand, A., Speed, T. P., Xuan, Z., Wang, T., Risso, D., & He, L. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*, *184*(22), 5541–5558.e22. <https://doi.org/10.1016/J.CELL.2021.09.021>
- Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, *99*(3), 371–382. <https://doi.org/10.1242/DEV.99.3.371>
- Montavon, T., Shukeir, N., Erikson, G., Engist, B., Onishi-Seebacher, M., Ryan, D., Musa, Y., Mittler, G., Meyer, A. G., Genoud, C., & Jenuwein, T. (2021). Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nature Communications*, *12*(1). <https://doi.org/10.1038/S41467-021-24532-8>
- Morazzani, E. M., Wiley, M. R., Murreddu, M. G., Adelman, Z. N., & Myles, K. M. (2012). Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. *PLOS Pathogens*, *8*(1), e1002470. <https://doi.org/10.1371/JOURNAL.PPAT.1002470>
- Morgani, S., Nichols, J., & Hadjantonakis, A. K. (2017). The many faces of Pluripotency: In vitro adaptations of a continuum of in vivo states. In *BMC Developmental Biology* (Vol. 17, Issue 1). <https://doi.org/10.1186/s12861-017-0150-4>
- Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., & Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Molecular Cell*, *53*(2), 277–289. <https://doi.org/10.1016/J.MOLCEL.2013.12.005>
- Müller, I., Moroni, A. S., Shlyueva, D., Sahadevan, S., Schoof, E. M., Radzishchanskaya, A., Højfeldt, J. W., Tatar, T., Koche, R. P., Huang, C., & Helin, K. (2021). MPP8 is essential for sustaining self-renewal of ground-state pluripotent stem cells. *Nature Communications*, *12*(1). <https://doi.org/10.1038/s41467-021-23308-4>
- Muotri, A. R. (2023). Interchromosomal translocation in neural progenitor cells exposed to L1 retrotransposition. *Genetics and Molecular Biology*, *46*(1). <https://doi.org/10.1590/1678-4685-GMB-2022-0268>
- Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., & Gage, F. H. (2005). *Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition*. <https://doi.org/10.1038/nature03663>

- Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J. S., Grindheim, J. M., Blahnik, K., Garcia, B. A., Tan, K., Bonasio, R., Jenuwein, T., & Zaret, K. S. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science*, *363*(6424), 294–297. <https://doi.org/10.1126/science.aau0583>
- Nishiyama, A., Mulholland, C. B., Bultmann, S., Kori, S., Endo, A., Saeki, Y., Qin, W., Trummer, C., Chiba, Y., Yokoyama, H., Kumamoto, S., Kawakami, T., Hojo, H., Nagae, G., Aburatani, H., Tanaka, K., Arita, K., Leonhardt, H., & Nakanishi, M. (2020). Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nature Communications* *2020* *11*:1, *11*(1), 1–17. <https://doi.org/10.1038/S41467-020-15006-4>
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* *2012* *485*:7398, *485*(7398), 381–385. <https://doi.org/10.1038/NATURE11049>
- Oh, B., Hwang, S.-Y., Solter, D., & Knowles, B. B. (1997). Spindlin, a major maternal transcript expressed in the mouse during transition from oocyte to embryo. *Development*, *124*, 493–503. <http://dev.biologists.org.ezproxyhost.library.tmc.edu/content/develop/124/2/493.full.pdf>
- Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., & Suzuki, T. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nature Structural & Molecular Biology* *2007* *14*:4, *14*(4), 349–350. <https://doi.org/10.1038/NSMB1220>
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, *99*(3), 247–257. [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6)
- Oliver, K. R., McComb, J. A., & Greene, W. K. (2013). Transposable elements: Powerful contributors to angiosperm evolution and diversity. In *Genome Biology and Evolution* (Vol. 5, Issue 10, pp. 1886–1901). <https://doi.org/10.1093/gbe/evt141>
- Öllinger, R., Childs, A. J., Burgess, H. M., Speed, R. M., Lundegaard, P. R., Reynolds, N., Gray, N. K., Cooke, H. J., & Adams, I. R. (2008). Deletion of the Pluripotency-Associated Tex19.1 Gene Causes Activation of Endogenous Retroviruses and Defective Spermatogenesis in Mice. *PLoS Genetics*, *4*(9), e1000199. <https://doi.org/10.1371/JOURNAL.PGEN.1000199>
- Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* *2007* *448*:7154, *448*(7154), 714–717. <https://doi.org/10.1038/NATURE05987>
- Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., & Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics*, *20*, 89–108. <https://doi.org/10.1038/s41576-018-0073-3>
- Pedersen, M. T., Kooistra, S. M., Radziszewska, A., Laugesen, A., Johansen, J. V., Hayward, D. G., Nilsson, J., Agger, K., & Helin, K. (2016a). Continual removal of H3K9 promoter methylation by Jmjd2 demethylases is vital for ESC self-renewal and early development. *The EMBO Journal*, *35*(14), 1550–1564. <https://doi.org/10.15252/embj.201593317>
- Peters, A. H. F. M., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., & Jenuwein, T. (2001). Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell*, *107*(3), 323–337. [https://doi.org/10.1016/S0092-8674\(01\)00542-6](https://doi.org/10.1016/S0092-8674(01)00542-6)
- Philippe, C., Vargas-Landin, D. B., Doucet, A. J., Van Essen, D., Vera-Otarola, J., Kuciak, M., Corbin, A., Nigumann, P., & Cristofari, G. (2016). Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *ELife*, *5*(MARCH2016). <https://doi.org/10.7554/ELIFE.13926>
- Piko, L., Hammons, M. D., & Taylor, K. D. (1984). Amounts, synthesis, and some properties of intracisternal A particle-related RNA in early mouse embryos. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(2 I), 488–492. <https://doi.org/10.1073/PNAS.81.2.488>
- Popova, K., Rybnikova, E., Chesnokova, E., Beletskiy, A., & Kolosov, P. (2022). *The Role of Transposable Elements of the Human Genome in Neuronal Function and Pathology*. <https://doi.org/10.3390/ijms23105847>
- Porro, A., Feuerhahn, S., Delafontaine, J., Riethman, H., Rougemont, J., & Lingner, J. (2014). *Functional characterization of the TERRA transcriptome at damaged telomeres*. <https://doi.org/10.1038/ncomms6379>

- Poullet, P., Carpentier, S., & Barillot, E. (2007). myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *PROTEOMICS*, 7(15), 2553–2556. <https://doi.org/10.1002/PMIC.200600784>
- Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., & Trono, D. (2012). The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. *Cell Reports*, 2(4), 766. <https://doi.org/10.1016/J.CELREP.2012.08.043>
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A., & Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell*, 44(3), 361–372. <https://doi.org/10.1016/J.MOLCEL.2011.08.032>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/NAR/GKW257>
- Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., & Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Research*, 18(4), 597–609. <https://doi.org/10.1101/gr.073486.107>
- Richardson, S. R., Gerdes, P., Gerhardt, D. J., Sanchez-Luque, F. J., Bodea, G. O., Muñoz-Lopez, M., Jesuadian, J. S., Kempen, M. J. H. C., Carreira, P. E., Jeddelloh, J. A., Garcia-Perez, J. L., Kazazian, H. H., Ewing, A. D., & Faulkner, G. J. (2017). Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Research*, 27(8), 1395–1405. <https://doi.org/10.1101/GR.219022.116/-DC1>
- Riggs, C. L., Kedersha, N., Ivanov, P., & Anderson, P. (2020). Mammalian stress granules and P bodies at a glance. *Journal of Cell Science*, 133(16). <https://doi.org/10.1242/jcs.242487>
- Robbez-Masson, L., Tie, C. H. C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I. A., Timms, R. T., Herrero, J., Lehner, P. J., & Rowe, H. M. (2018). The hush complex cooperates with trim28 to repress young retrotransposons and new genes. *Genome Research*, 28(6), 836–845. <https://doi.org/10.1101/gr.228171.117>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/BIOINFORMATICS/BTP616>
- Rothbart, S. B., Dickson, B. M., Ong, M. S., Krajewski, K., Houliston, S., Kireev, D. B., Arrowsmith, C. H., & Strahl, B. D. (2013). Multivalent histone engagement by the linked tandem tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes and Development*, 27(11), 1288–1298. <https://doi.org/10.1101/gad.220467.113>
- Rothstein, J. L., Johnson, D., Deloia, J. A., Skowronski, J., Solter, D., & Knowles, B. (1992). *Gene expression during preimplantation mouse development*.
- 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), 237–240. <https://doi.org/10.1038/nature08674>
- Rugg-Gunn, P. J., Cox, B. J., Ralston, A., & Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24), 10783–10790. <https://doi.org/10.1073/pnas.0914507107>
- Rushton, M. D., Saunderson, E. A., Patani, H., Green, M. R., & Ficz, G. (2022). An shRNA kinase screen identifies regulators of UHRF1 stability and activity in mouse embryonic stem cells. *Epigenetics*, 17(12), 1590–1607. <https://doi.org/10.1080/15592294.2022.2044126>
- Sachs, P., Ding, D., Bergmaier, P., Lamp, B., Schlagheck, C., Finkernagel, F., Nist, A., Stiewe, T., & Mermoud, J. E. (2019). SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nature Communications*, 10(1335). <https://doi.org/10.1038/s41467-019-09078-0>
- Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F. J., & Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. *EMBO Reports*, 16(7), 836–850. <https://doi.org/10.15252/EMBR.201439937>
- Saito, E. S., Keng, V. W., Takeda, J., & Horie, K. (2008). Translation from nonautonomous type IAP retrotransposon is a critical determinant of transposition activity: Implication for retrotransposon-

- mediated genome evolution. *Genome Research*, 18(6), 859–868.  
<https://doi.org/10.1101/gr.069310.107>
- Sakashita, A., Kitano, T., Ishizu, H., Guo, Y., Masuda, H., Ariura, M., Murano, K., & Siomi, H. (2023). Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nature Genetics*. <https://doi.org/10.1038/s41588-023-01324-y>
- Santos-Barriopedro, I., Bosch-Presegué, L., Marazuela-Duque, A., De La Torre, C., Colomer, C., Vazquez, B. N., Fuhrmann, T., Martínez-Pastor, B., Lu, W., Braun, T., Bober, E., Jenuwein, T., Serrano, L., Esteller, M., Chen, Z., Barceló-Batllori, S., Mostoslavsky, R., Espinosa, L., & Vaquero, A. (2018). SIRT6-dependent cysteine monoubiquitination in the PRE-SET domain of Suv39h1 regulates the NF- $\kappa$ B pathway. *Nature Communications*, 9(1).  
<https://doi.org/10.1038/S41467-017-02586-X>
- Sato, S., Gillette, M., de Santiago, P. R., Kuhn, E., Burgess, M., Doucette, K., Feng, Y., Mendez-Dorantes, C., Ippoliti, P. J., Hobday, S., Mitchell, M. A., Doberstein, K., Gysler, S. M., Hirsch, M. S., Schwartz, L., Birrer, M. J., Skates, S. J., Burns, K. H., Carr, S. A., & Drapkin, R. (2023). LINE-1 ORF1p as a candidate biomarker in high grade serous ovarian carcinoma. *Scientific Reports* 2023 13:1, 13(1), 1–15. <https://doi.org/10.1038/S41598-023-28840-5>
- Scarola, M., Comisso, E., Pascolo, R., Chiaradia, R., Marion, R. M., Schneider, C., Blasco, M. A., Schoeftner, S., & Benetti, R. (2015). *ARTICLE Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA*. <https://doi.org/10.1038/ncomms8631>
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., ... Wilson, R. K. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science*, 326(5956), 1112–1115.  
<https://doi.org/10.1126/SCIENCE.1178534>
- Schneider, W. M., Chevillotte, M. D., & Rice, C. M. (2014). Interferon-Stimulated Genes: A Complex Web of Host Defenses. <https://doi.org/10.1146/Annurev-Immuno-032713-120231>, 32, 513–545. <https://doi.org/10.1146/ANNUREV-IMMUNOL-032713-120231>
- Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, 16(8), 919–932. <https://doi.org/10.1101/GAD.973302>
- Schulz, M., Teissandier, A., De, E., Mata, L. A., Armand, M., Iranzo, J., Marjou, F. El, Gestraud, P., Walter, M., Kinston, S., Göttgens, B., Greenberg, M. V. C., & Bourc'his, D. (2023). DNA methylation restricts coordinated germline and neural fates in embryonic stem cell differentiation. *BioRxiv*. <https://doi.org/10.1101/2022.10.22.513040>
- Scott, E. C., Gardner, E. J., Masood, A., Chuang, N. T., Vertino, P. M., & Devine, S. E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Research*, 26(6), 745–755. <https://doi.org/10.1101/gr.201814.115>
- Seczynska, M., Bloor, S., Cuesta, S. M., & Lehner, P. J. (2022). Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature*, 601(7893), 440–445.  
<https://doi.org/10.1038/s41586-021-04228-1>
- Shanle, E. K., Shinsky, S. A., Bridgers, J. B., Bae, N., Sagum, C., Krajewski, K., Rothbart, S. B., Bedford, M. T., & Strahl, B. D. (2017). Histone peptide microarray screen of chromo and Tudor domains defines new histone lysine methylation interactions. *Epigenetics & Chromatin*, 10(1), 12. <https://doi.org/10.1186/s13072-017-0117-5>
- Sharif, J., Endo, T. A., Nakayama, M., Karimi, M. M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M. A., Sun, Z., Ishikura, T., Mizutani-Koseki, Y., Ohara, O., Shinkai, Y., Nakanishi, M., Xie, H., Lorincz, M. C., & Koseki, H. (2016). Activation of Endogenous Retroviruses in Dnmt1<sup>-/-</sup> ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell*, 19(1), 81–94.  
<https://doi.org/10.1016/j.stem.2016.03.013>
- Shi, H., Strogantsev, R., Takahashi, N., Kazachenka, A., Lorincz, M. C., Hemberger, M., & Ferguson-Smith, A. C. (2019). ZFP57 regulation of transposable elements and gene expression within and beyond imprinted domains. *Epigenetics and Chromatin*, 12(1), 1–13.  
<https://doi.org/10.1186/S13072-019-0295-4/FIGURES/6>
- Shi, H., Wei, J., & He, C. (2019). Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular Cell*, 74(4), 640–650.  
<https://doi.org/10.1016/j.molcel.2019.04.025>

- Shirane, K., & Lorincz, M. (2023). Epigenetic Mechanisms Governing Female and Male Germline Development in Mammals. *Sexual Development*, 16(5–6), 1–23. <https://doi.org/10.1159/000529336>
- Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., Hata, K., Martin, S. L., Noce, T., Kuramochi-Miyagawa, S., Nakano, T., Sasaki, H., Pillai, R. S., Nakatsuji, N., & Chuma, S. (2009). The TDRD9-MIWI2 Complex Is Essential for piRNA-Mediated Retrotransposon Silencing in the Mouse Male Germline. *Developmental Cell*, 17(6), 775–787. <https://doi.org/10.1016/J.DEVCEL.2009.10.012>
- Siguié, P., Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiology Reviews*, 38(5), 865–891. <https://doi.org/10.1111/1574-6976.12067>
- Song, J., Rechkoblit, O., Bestor, T. H., & Patel, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*, 331(6020), 1036–1040. [https://doi.org/10.1126/SCIENCE.1195380/SUPPL\\_FILE/SONG-SOM.PDF](https://doi.org/10.1126/SCIENCE.1195380/SUPPL_FILE/SONG-SOM.PDF)
- Sookdeo, A., Hepp, C. M., McClure, M. A., & Boissinot, S. (2013). Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA*, 4(1). <https://doi.org/10.1186/1759-8753-4-3>
- Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378), 490–495. <https://doi.org/10.1038/nature10716>
- Stetson, D. B., Ko, J. S., Heidmann, T., & Medzhitov, R. (2008). Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell*, 134, 587–598. <https://doi.org/10.1016/j.cell.2008.06.032>
- Su, X., Zhu, G., Ding, X., Lee, S. Y., Dou, Y., Zhu, B., Wu, W., & Li, H. (2014). Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. *Genes & Development*, 28(6), 622–636. <https://doi.org/10.1101/gad.233239.113>
- Sultana, T., van Essen, D., Siol, O., Bailly-Bechet, M., Philippe, C., Zine El Aabidine, A., Pioger, L., Nigumann, P., Sacconi, S., Andrau, J. C., Gilbert, N., & Cristofari, G. (2019). The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular Cell*, 74(3), 555-570.e7. <https://doi.org/10.1016/J.MOLCEL.2019.02.036>
- Sun, X. J., Wei, J., Wu, X. Y., Hu, M., Wang, L., Wang, H. H., Zhang, Q. H., Chen, S. J., Huang, Q. H., & Chen, Z. (2005). Identification and Characterization of a Novel Human Histone H3 Lysine 36-specific Methyltransferase. *Journal of Biological Chemistry*, 280(42), 35261–35271. <https://doi.org/10.1074/JBC.M504012200>
- 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>
- Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., & Shinkai, Y. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes and Development*, 16(14), 1779–1791. <https://doi.org/10.1101/gad.989402>
- Takahashi, N., Coluccio, A., Thorball, C. W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A. C., & Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. *Genes and Development*, 33(1–2), 49–54. <https://doi.org/10.1101/GAD.320069.118/-/DC1>
- Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O. G., Shiekhata, R., Picketts, D. J., & Yang, X. (2004). A Novel Transcription Regulatory Complex Containing Death Domain-associated Protein and the ATR-X Syndrome Protein\*. <https://doi.org/10.1074/jbc.M401321200>
- Taylor, M. S., Wu, C., Fridy, P. C., Zhang, S. J., Senussi, Y., Wolters, J. C., Cheng, W.-C., Heaps, J., Miller, B. D., Mori, K., Cohen, L., Jiang, H., Molloy, K. R., Chait, B. T., Goggins, M., Bhan, I., Franses, J. W., Yang, X., Taplin, M.-E., ... Burns, K. H. (2023). Ultrasensitive detection of circulating LINE-1 ORF1p as a specific multi-cancer biomarker. *BioRxiv*, 2023.01.25.525462. <https://doi.org/10.1101/2023.01.25.525462>
- Tchasovnikarova, I. A., Timms, R. T., Matheson, N. J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M. A., & Lehner, P. J. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*, 348(6242), 1481–1485. <https://doi.org/10.1126/science.aaa7227>
- Teissandier, A., & Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO Journal*, 36(11), 1471–1473. <https://doi.org/10.15252/EMBJ.201796812>

- The, M., MacCoss, M. J., Noble, W. S., & Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *Journal of the American Society for Mass Spectrometry*, 27(11), 1719. <https://doi.org/10.1007/S13361-016-1460-7>
- Timms, R. T., Tchasovnikarova, I. A., Antrobus, R., Dougan, G., & Lehner, P. J. (2016). ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. *Cell Reports*, 17(3), 653–659. <https://doi.org/10.1016/j.celrep.2016.09.050>
- Touré, A., Szot, M., Mahadevaiah, S. K., Ine Rattigan, A. , Ojarikre, O. A., & Burgoyne, P. S. (2004). A New Deletion of the Mouse Y Chromosome Long Arm Associated With the Loss of Ssty Expression, Abnormal Sperm Development and Sterility. *Genetics*, 166, 901–912.
- Tsusaka, T., Shimura, C., & Shinkai, Y. (2019). ATF7IP regulates SETDB1 nuclear localization and increases its ubiquitination. *EMBO Reports*, 20(12), e48297. <https://doi.org/10.15252/EMBR.201948297>
- Ukai, H., Ishii-Oba, H., Ukai-Tadenuma, M., Ogiu, T., & Tsuji, H. (2003). Formation of an active form of the interleukin-2/ 15 receptor  $\beta$ -chain by insertion of the intracisternal A particle in a radiation-induced mouse thymic lymphoma and its role in tumorigenesis. *Molecular Carcinogenesis*, 37(2), 110–119. <https://doi.org/10.1002/mc.10128>
- Urrutia, R. (2003). Protein family review KRAB-containing zinc-finger repressor proteins. *Genome Biology*, 4, 231. <http://genomebiology.com/2003/4/10/231>
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785), 320–324. <https://doi.org/10.1126/SCIENCE.1129333>
- Valot, B., Langella, O., Nano, E., & Zivy, M. (2011). MassChroQ: A versatile tool for mass spectrometry quantification. *PROTEOMICS*, 11(17), 3572–3577. <https://doi.org/10.1002/PMIC.201100120>
- Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., & Gorbunova, V. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nature Communications* 2014 5:1, 5(1), 1–10. <https://doi.org/10.1038/ncomms6011>
- Vitullo, P., Sciamanna, I., Baiocchi, M., Sinibaldi-Vallebona, P., & Spadafora, C. (2012). LINE-1 retrotransposon copies are amplified during murine early embryo development. *Molecular Reproduction and Development*, 79(2), 118–127. <https://doi.org/10.1002/MRD.22003>
- von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., Kroeze, L. I., Nesterova, T. B., Jansen, J. H., Xie, H., He, C., Reik, W., & Stunnenberg, H. G. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Molecular Cell*, 62(6), 848–861. <https://doi.org/10.1016/j.molcel.2016.04.025>
- Wagner, T., Greschik, H., Burgahn, T., Schmidtkunz, K., Schott, A. K., McMillan, J., Baranauskiene, L., Xiong, Y., Fedorov, O., Jin, J., Oppermann, U., Matulis, D., Schule, R., & Jung, M. (2016). Identification of a small-molecule ligand of the epigenetic reader protein Spindlin1 via a versatile screening platform. *Nucleic Acids Research*, 44(9). <https://doi.org/10.1093/nar/gkw089>
- Walter, M., Teissandier, A., Pérez-Palacios, R., Bourc'His, D., & Pe, R. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *ELife*, 5(JANUARY2016), 1–30. <https://doi.org/10.7554/eLife.11418.001>
- Wang, J.-X., Zeng, Q., Chen, L., Du, J.-C., Yan, X.-L., Yuan, H.-F., Zhai, C., Zhou, J.-N., Jia, Y.-L., Yue, W., & Pei, X.-T. (2012). SPINDLIN1 Promotes Cancer Cell Proliferation through Activation of WNT/TCF-4 Signaling. *Molecular Cancer Research*, 10(3), 326–335. <https://doi.org/10.1158/1541-7786.MCR-11-0440>
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., & Liu, J. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell*, 157(4), 979–991. <https://doi.org/10.1016/J.CELL.2014.04.017>
- Wang, Q., Young Gie Chung, DeVries, W. N., Struwe, M., & Latham, K. E. (2001). Role of Protein Synthesis in the Development of a Transcriptionally Permissive State in One-Cell Stage Mouse Embryos. *Biology of Reproduction*, 65(3), 748–754. <https://doi.org/10.1095/BIOLREPROD65.3.748>
- Wang, W., Chen, Z., Mao, Z., Zhang, H., Ding, X., Chen, S., Zhang, X., Xu, R., & Zhu, B. (2011). Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. *EMBO Reports*, 12(11), 1160–1166. <https://doi.org/10.1038/embor.2011.184>
- Wang, Z., Fan, R., Russo, A., Cernilogar, F. M., Nuber, A., Schirge, S., Shcherbakova, I., Dzhilyanova, I., Ugur, E., Anton, T., Richter, L., Leonhardt, H., Lickert, H., & Schotta, G. (2022). Dominant role

- of DNA methylation over H3K9me3 for IAP silencing in endoderm. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-32978-7>
- Watanabe, T., Cui, X., Yuan, Z., Qi, H., & Lin, H. (2018). MIWI2 targets RNAs transcribed from piRNA-dependent regions to drive DNA methylation in mouse prospermatogonia. *The EMBO Journal*, 37(18), e95329. <https://doi.org/10.15252/EMBJ.201695329>
- Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics* 2007 39:4, 39(4), 457–466. <https://doi.org/10.1038/NG1990>
- Wei, J., Yu, X., Yang, L., Liu, X., Gao, B., Huang, B., Dou, X., Liu, J., Zou, Z., Cui, X. L., Zhang, L. S., Zhao, X., Liu, Q., He, P. C., Sepich-Poore, C., Zhong, N., Liu, W., Li, Y., Kou, X., ... He, C. (2022). FTO mediates LINE1 m6A demethylation and chromatin regulation in mESCs and mouse development. *Science*, 376(6596). [https://doi.org/10.1126/SCIENCE.ABE9582/SUPPL\\_FILE/SCIENCE.ABE9582\\_DATA\\_S1\\_TO\\_S3.ZIP](https://doi.org/10.1126/SCIENCE.ABE9582/SUPPL_FILE/SCIENCE.ABE9582_DATA_S1_TO_S3.ZIP)
- Weiss, R. A. (1969). *The Host Range of BRYAN Strain Rous Sarcoma Virus Synthesized in the Absence of Helper Virus*. 5, 511–528.
- Wincker, P., Jubier-Maurin, V., & Roizes, G. (1987). Unrelated sequences at the 5' end of mouse LINE-1 repeated elements define two distinct sub-families. In *Nucleic Acids Research* (Vol. 15). <https://academic.oup.com/nar/article/15/21/8593/2378006>
- Wolf, D., & Goff, S. P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242), 1201–1204. <https://doi.org/10.1038/nature07844>
- 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 and Development*, 29(5), 538–554. <https://doi.org/10.1101/gad.252767.114>
- Wu, H., Chen, X., Xiong, J., Li, Y., Li, H., Ding, X., Liu, S., Chen, S., Gao, S., & Zhu, B. (2010). Histone methyltransferase G9a contributes to H3K27 methylation in vivo. *Cell Research* 2011 21:2, 21(2), 365–367. <https://doi.org/10.1038/CR.2010.157>
- Xie, S., & Qian, C. (2018). The growing complexity of UHRF1-mediated maintenance DNA methylation. *Genes*, 9(12), 1–12. <https://doi.org/10.3390/genes9120600>
- Xiong, F., Wang, R., Lee, J. H., Li, S., Chen, S. F., Liao, Z., Hasani, L. Al, Nguyen, P. T., Zhu, X., Krakowiak, J., Lee, D. F., Han, L., Tsai, K. L., Liu, Y., & Li, W. (2021). RNA m6A modification orchestrates a LINE-1–host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Research* 2021 31:8, 31(8), 861–885. <https://doi.org/10.1038/S41422-021-00515-8>
- Xiong, Y., Greschik, H., Johansson, C., Seifert, L., Bacher, J., Park, K. S., Babault, N., Martini, M., Fagan, V., Li, F., Chau, I., Christott, T., Dilworth, D., Barsyte-Lovejoy, D., Vedadi, M., Arrowsmith, C. H., Brennan, P., Fedorov, O., Jung, M., ... Jin, J. (2019). Discovery of a Potent and Selective Fragment-like Inhibitor of Methyllysine Reader Protein Spindlin 1 (SPIN1). *Journal of Medicinal Chemistry*, 62(20), 8996–9007. <https://doi.org/10.1021/acs.jmedchem.9b00522>
- 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* 2021 591:7849, 591(7849), 317–321. <https://doi.org/10.1038/S41586-021-03210-1>
- Yang, B. X., El Farran, C. A., Guo, H. C., Yu, T., Fang, H. T., Wang, H. F., Schlesinger, S., Seah, Y. F. S., Goh, G. Y. L., Neo, S. P., Li, Y., Lorincz, M. C., Tergaonkar, V., Lim, T. M., Chen, L., Gunaratne, J., Collins, J. J., Goff, S. P., Daley, G. Q., ... Loh, Y. H. (2015). Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell*, 163(1), 230–245. <https://doi.org/10.1016/J.CELL.2015.08.037>
- Yang, F., Chen, J., Liu, B., Gao, G., Sebastian, M., Jeter, C., Shen, J., Person, M. D., & Bedford, M. T. (2021). SPINDOC binds PARP1 to facilitate PARylation. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-26588-y>
- Yang, F., Eckardt, S., Leu, N. A., McLaughlin, K. J., & Wang, P. J. (2008). Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis. *Journal of Cell Biology*, 180(4), 673–679. <https://doi.org/10.1083/JCB.200709057>
- Yang, F., Huang, X., Zang, R., Chen, J., Fidalgo, M., Sanchez-Priego, C., Yang, J., Caichen, A., Ma, F., Macfarlan, T., Wang, H., Gao, S., Zhou, H., & Wang, J. (2020). DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell*, 26(2), 234–250.e7. <https://doi.org/10.1016/J.STEM.2020.01.004>

- Yang, L., Xia, L., Wu, D. Y., Wang, H., Chansky, H. A., Schubach, W. H., Hickstein, D. D., & Zhang, Y. (2002). Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. *Oncogene* 2002 21:1, 21(1), 148–152. <https://doi.org/10.1038/SJ.ONC.1204998>
- Yang, N., & Kazazian, H. H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Structural & Molecular Biology* 2006 13:9, 13(9), 763–771. <https://doi.org/10.1038/NSMB1141>
- Yang, N., Wang, W., Wang, Y., Wang, M., Zhao, Q., Rao, Z., Zhu, B., & Xu, R. M. (2012). Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44). <https://doi.org/10.1073/pnas.1208517109>
- Yang, P., Wang, Y., & Macfarlan, T. S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. In *Trends in Genetics* (Vol. 33, Issue 11, pp. 871–881). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2017.08.006>
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356(6337). [https://doi.org/10.1126/SCIENCE.AAJ2239/SUPPL\\_FILE/AAJ2239\\_YIN\\_SM\\_TABLES\\_S1-S6.XLSX](https://doi.org/10.1126/SCIENCE.AAJ2239/SUPPL_FILE/AAJ2239_YIN_SM_TABLES_S1-S6.XLSX)
- Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., & Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 2008 453:7194, 453(7194), 519–523. <https://doi.org/10.1038/NATURE06968>
- Yuan, H., Zhang, P., Qin, L., Chen, L., Shi, S., Lu, Y., Yan, F., Bai, C., Nan, X., Liu, D., Li, Y., Yue, W., & Pei, X. (2007). Overexpression of SPINDLIN1 induces cellular senescence, multinucleation and apoptosis. <https://doi.org/10.1016/j.gene.2007.11.019>
- Zamudio, N., & Bourc'His, D. (2010). Transposable elements in the mammalian germline: A comfortable niche or a deadly trap. *Heredity*, 105(1), 92–104. <https://doi.org/10.1038/hdy.2010.53>
- Zhang, A., Dong, B., Doucet, J., Moldovan, J. B., Moran, J. V., & Silverman, R. H. (2014). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Research*, 42(6), 3803–3820. <https://doi.org/10.1093/nar/gkt1308>
- Zhang, H., Zhang, F., Chen, Q., Li, M., Lv, X., Xiao, Y., Zhang, Z., Hou, L., Lai, Y., Zhang, Y., Zhang, A., Gao, S., Fu, H., Xiao, W., Zhou, J., Diao, F., Shi, A., Su, Y., Zeng, W., ... Li, J. (2021). The piRNA pathway is essential for generating functional oocytes in golden hamsters. *Nature Cell Biology*. <https://doi.org/10.1038/s41556-021-00750-6>
- Zhang, K.-M., Wang, Y.-F., Huo, R., Bi, Y., Lin, M., Sha, J.-H., & Zhou, Z.-M. (2008). Characterization of Spindlin1 isoform2 in mouse testis. *Asian J Androl*, 10(5), 741–748. <https://doi.org/10.1111/j.1745-7262.2008.00424.x>
- Zhang, X., Zhu, G., Su, X., Li, H., & Wu, W. (2018). Nucleolar localization signal and histone methylation reader function is required for SPIN1 to promote rRNA gene expression. *Biochemical and Biophysical Research Communications*, 505(1), 325–332. <https://doi.org/10.1016/j.bbrc.2018.09.098>
- Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., & Jeltsch, A. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Research*, 38(13), 4246–4253. <https://doi.org/10.1093/NAR/GKQ147>
- Zhao, F., Liu, Y., Su, X., Lee, J.-E., Song, Y., Wang, D., Ge, K., Gao, J., Zhang, M. Q., & Li, H. (2020). Molecular basis for histone H3 “K4me3-K9me3/2” methylation pattern readout by Spindlin1. *Journal of Biological Chemistry*, 295, 16877–16887. <https://doi.org/10.1074/jbc.ra120.013649>
- Zoch, A., Auchynnikava, T., Berrens, R. V., Kabayama, Y., Schöpp, T., Heep, M., Vasiliauskaitė, L., Pérez-Rico, Y. A., Cook, A. G., Shkumatava, A., Rappsilber, J., Allshire, R. C., & O'Carroll, D. (2020). SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature*, 584(7822), 635–639. <https://doi.org/10.1038/s41586-020-2557-5>



## RÉSUMÉ

---

Les éléments transposables (ET) sont des éléments génétiques mobiles présent chez toutes les espèces vivantes. Pour limiter leurs effets néfastes, les organismes hôtes ont développé différents mécanismes de contrôle des ET au cours de l'évolution, ce qui est particulièrement essentiel pour assurer l'intégrité génomique des cellules germinales et pluripotentes. En supprimant leur transcription initiale, la méthylation de l'ADN est un répresseur efficace des ET. Cependant, pendant la phase de reprogrammation épigénétique qui accompagne le développement, la méthylation de l'ADN est largement perdue dans les cellules embryonnaires, à l'exception de quelques régions qui incluent certains ET. Ces observations soulèvent deux questions fondamentales : Comment les ET sont-ils réprimés en absence de méthylation pendant cette période ? Et comment certains ET résistent à la perte de méthylation globale ?

Pendant ma thèse, j'ai étudié le rôle de SPINDLIN1 (SPIN1) et SPINDOC, un activateur transcriptionnel et son co-facteur, comme nouveau complexe de lecture chromatinienne impliqué dans le contrôle des ET dans l'embryon. En utilisant des cultures de cellules souches embryonnaires comme système modèle et des outils d'édition du génome basés sur la technologie CRISPR-Cas9, j'ai généré des lignées mutantes pour l'une ou l'ensemble de ces deux protéines et analysé leur transcriptome (y compris aux ET), leur profil chromatinien et leur méthylation de l'ADN. J'ai pu ainsi démontrer que SPIN1 et SPINDOC sont nécessaires au contrôle des ET, au moins en partie en assurant le bon fonctionnement de la machinerie de maintenance de la méthylation de l'ADN sur l'ensemble du génome. J'ai aussi mis en évidence l'implication de SPINDOC dans la croissance et la fertilité chez la souris. Dans l'ensemble, mes travaux mettent en évidence un lien précédemment inconnu entre SPIN1-SPINDOC et la biologie des transposons et un nouveau rôle pour ce complexe : la régulation de la méthylation de l'ADN.

## MOTS CLÉS

---

Transposons, Chromatine, Stabilité du génome, Épigénétique

## ABSTRACT

---

Transposable elements (TEs) are mobile genetic elements present in every living species. To limit their harmful effects, species have developed over evolutionary time-scales several mechanisms to control TE expression and mobilization, which is of utmost importance to protect genomic integrity in germ cells and pluripotent cells. By suppressing their initial transcription, DNA methylation is a potent TE repressor. However, during the epigenetic reprogramming that accompanies pre-implantation development, embryonic cells lose most of their genomic methylation, except at a few regions including some TEs. These observations raise two overarching questions: How are TEs repressed in absence of DNA methylation at this particular time? How some TEs resist the global DNA methylation erasure?

My PhD work focused on the study of SPINDLIN1 (SPIN1) and SPINDOC—a transcriptional activator and its specific co-factor—as a new chromatin reading complex involved in TE control in embryonic cells. Using cultures of embryonic stem cells as a modeling system and CRISPR-Cas9-based genome editing tools, I generated knock-out and rescue cell lines for either or both of these factors and analyzed their transcriptome (including TE expression), chromatin and DNA methylation profile. I found both factors to be required for TE control, in part through promoting the function of the DNA methylation maintenance machinery genome-wide. I also found SPINDOC to be involved in growth and fertility in mice. Altogether, my work reveals a previously unknown link between SPIN1-SPINDOC and transposon biology, and a new role for this complex: the regulation of DNA methylation.

## KEYWORDS

---

Transposable elements, Chromatin, Genomic stability, Epigenetic