



**HAL**  
open science

# Contribution de H2AK119Ub1 dérivée du spermatozoïde au développement embryonnaire précoce du Xénope

Valentin François-Campion

## ► To cite this version:

Valentin François-Campion. Contribution de H2AK119Ub1 dérivée du spermatozoïde au développement embryonnaire précoce du Xénope. Médecine humaine et pathologie. Nantes Université, 2022. Français. NNT : 2022NANU1048 . tel-04041472

**HAL Id: tel-04041472**

**<https://theses.hal.science/tel-04041472>**

Submitted on 22 Mar 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.

# THESE DE DOCTORAT DE

NANTES UNIVERSITE

ECOLE DOCTORALE N° 605

*Biologie Santé*

Spécialité : BioInformatique

Par

**Valentin FRANCOIS--CAMPION**

**Contribution de H2AK119Ub1 dérivée du spermatozoïde au développement embryonnaire précoce du Xénope**

Thèse présentée et soutenue à Nantes, le 27 Octobre 2022 à 14h00

Unité de recherche : UMR 1064 CR2TI Center for Research in Transplantation and Translational Immunology

## Rapporteurs avant soutenance :

Jean-Pierre QUIVY  
Marta RADMAN-LIVAJA

DR CNRS, DYNAMIQUE DU NOYAU (UMR3664), Institut Curie, Paris  
DR2 CNRS, Institut de Génétique moléculaire de Montpellier, Montpellier

## Composition du Jury :

Président : Stéphane MINVIELLE  
Examineurs : Nathalie BEAUJEAN  
Marta RADMAN-LIVAJA  
Yann AUDIC  
Jean-Pierre QUIVY

Directeur de recherche, Institut de Recherche Thérapeutique (IRT1), Nantes  
DR2 – INRAE, Stem cell and Brain Research Institute (SBRI), Lyon  
DR2 CNRS, Institut de Génétique moléculaire de Montpellier, Montpellier  
Chargé de recherche, Institut de génétique et Développement, Rennes  
DR CNRS, DYNAMIQUE DU NOYAU (UMR3664), Institut Curie, Paris

Dir. de thèse : Jérôme JULLIEN

CRCN, Centre de recherche en transplantation et immunologie translationnelle (CR2TI)

# Remerciements

Avant tout, je souhaite remercier Jérôme Jullien de m'avoir fait confiance pour réaliser cette thèse. Travailler ensemble a été à la fois un plaisir, mais surtout une possibilité d'apprendre et découvrir de nouveaux domaines à la fois en biologie qu'en bio-informatique. Merci à toute l'équipe plus globalement : Vanessa pour son expérience apportée, Florian qui a réalisé une grande partie des expériences avec Jérôme, et Maïssa pour toutes nos discussions sur les analyses bio-informatique. Ces trois années et demie ont été superbes à vos côtés, même si mon plus grand regret sera très certainement de ne pas avoir naïtre l'animalerie xénope, que je vous souhaite d'avoir au plus vite.

Je remercie Romain Gibeaux (Université de Rennes) pour la collaboration qui nous a permis de réaliser certaines expériences à partir de l'extrait d'œuf notamment, ainsi que Marta Teperek et Mami Oikawa et Angela Simeone pour le partage d'informations de diverses expériences réalisées par Jérôme à Cambridge.

Merci à tout le laboratoire (CR2TI) pour les diverses réunions et présentations permettant le partage d'informations et d'idées. Je tenais notamment à remercier Laurent David, qui m'a accueilli en Master sur la thématique du développement embryonnaire précoce humain et la différenciation cellulaire. C'est un plaisir d'avoir pu échanger avec son équipe pendant ma thèse (Gaël, Alexandre et Constance). Je remercie aussi tous mes collègues de bureau pendant ces trois années pour les moments de convivialité et la bonne ambiance et humeur générale présente en continu.

Je remercie les membres de mon CSI, Deborah Bourc'his, Géraldine Jean et Laurent David pour leurs conseils et indications pendant ma thèse, ainsi que les membres du jury et rapporteurs pour avoir accepté de participer à la défense de ma thèse.

De même, je tiens à remercier mes amis, Manon, Eole, Adrien, Maxime, Eloïse, Johanna et Venceslas pour nos différentes activités qui m'auront permis de souffler et de penser à autre chose que la thèse pendant ces années, ainsi que votre soutien.

Enfin, je tiens à remercier toute ma famille, en premier mes grands-parents, notamment Annick avec qui je passe toujours de très bons moments et à qui je promets d'essayer de passer plus de temps dans les mois à venir. Ensuite, mes cousins, cousines, oncles et tantes pour leur soutien, et tous ces moments conviviaux qui m'ont tant manqué à l'époque des confinements et du COVID. Enfin, je tiens à remercier mon frère Goulwen, ma sœur Anaïs et mes parents Cécile et Christophe pour tout ce que vous m'avez apporté, c'est tout d'abord grâce à vous si j'ai pu arriver jusqu'ici aujourd'hui. Ces dernières années ont été parsemées de beaucoup de problèmes pour tous, je serai toujours là pour vous comme vous l'avez été pour moi. Enfin, je remercie mes chats que j'adore plus que tout, Hermione et Obi-Wan, pour leur moment de détente et de compagnie très reposante et apaisante.

Merci à tous ceux qui auront contribué à ces trois années.

# TABLE DES MATIERES

<b>REMERCIEMENTS</b> .....	<b>1</b>
<b>LISTE DES FIGURES</b> .....	<b>5</b>
<b>INTRODUCTION</b> .....	<b>8</b>
L'ÉPIGÉNÉTIQUE : DE L'HERÉDITÉ À LA VARIABILITÉ INDIVIDUELLE.....	9
<i>Les fondations du concept de la régulation épigénétique</i> .....	9
<i>Les différents supports de l'épigénétique</i> .....	10
<i>Le nucléosome, une organisation spécifique de la chromatine pour la régulation génique</i> .....	13
DU SPERMATOZOÏDE À L'EMBRYON : TRANSMISSION DE L'INFORMATION GÉNÉTIQUE ET ÉPIGÉNÉTIQUE .....	15
<i>Xenopus laevis, un modèle biologique pertinent</i> .....	15
<i>De la spermatogénèse au développement embryonnaire précoce</i> .....	16
La spermatogénèse, un processus de maturation des spermatozoïdes .....	16
Le développement embryonnaire chez <i>Xenopus laevis</i> .....	18
La contribution épigénétique du spermatozoïde au développement embryonnaire .....	18
<i>H2AK119Ub1 a un rôle central dans l'établissement du complexe polycomb dans les cellules souches embryonnaires.</i> .....	21
 <b>CHAPITRE I- LE PRETRAITEMENT DU SPERMATOZOÏDE PAR LA DEUBIQUITYLASE USP21 ALTERE LE DEVELOPPEMENT EMBRYONNAIRE</b> .....	 <b>24</b>
 <b>CHAPITRE II - LE TRAITEMENT DES SPERMATOZOÏDES PAR LA DEUBIQUITYLASE USP21 ENTRAINE DES DEREGULATIONS DE L'EXPRESSION DES GENES DANS LA GASTRULA</b> .....	 <b>31</b>
OUTILS UTILISÉS POUR L'ANALYSE DU TRANSCRIPTOME PAR RNA-SEQUENCING.....	33
<i>Du séquenceur à la matrice d'expression des gènes : formatage des données</i> .....	33
Fastq, le fichier brut issu du séquenceur.....	33
BAM, l'alignement des reads sur un génome.....	34
Du BAM à la matrice de comptage.....	35
<i>De la matrice d'expression des gènes à l'identification des gènes différentiellement exprimés</i> .....	37
Pré-traitement des données .....	37
Filtration.....	37
La gestion des échantillons aberrants .....	37
La normalisation.....	38
L'effet batch .....	39
L'analyse de gènes différentiellement exprimés .....	40
L'enrichissement fonctionnel .....	41
LE PRE-TRAITEMENT DU SPERMATOZOÏDE PAR USP21 AVANT L'INJECTION DANS L'ŒUF ALTERE LE TRANSCRIPTOME DE L'EMBRYON .....	43
DEFAUT D'ÉLIMINATION DES ARNS MATERNELS OU INDUCTION DE LA TRANSCRIPTION ZYGOTIQUE DES GENES SENSIBLES AU TRAITEMENT USP21 ?50	
 <b>CHAPITRE III- CARACTERISATION DE LA DISTRIBUTION DE H2AK119UB1 DANS LE SPERMATOZOÏDE</b> .....	 <b>53</b>
ANALYSE DE L'ÉPIGÉNOME PAR CHIP-SEQUENCING .....	54

DETAILS DES ETAPES ET OUTILS UTILISES DANS LES ANALYSES CHIP-SEQ .....	56
<i>Élimination des séquences adaptatrices et des reads de faible qualité</i> .....	56
<i>Alignement des reads sur un génome de référence</i> .....	56
<i>Identification des régions du génome présentant un enrichissement pour une modification d'histone</i> .....	58
<i>Visualisation du signal sur le génome</i> .....	62
<i>La gestion des réplicats</i> .....	64
H2AK119Ub1 EST ENRICHIE SUR LES GENES SENSIBLES AU TRAITEMENT USP21 .....	66
<i>H2AK119Ub1 est associé aux régions régulatrices du génome dans le spermatozoïde</i> .....	67
<i>Les gènes dérégulés par le traitement USP21 montrent un enrichissement pour H2AK119Ub1 dans le spermatozoïde</i> .....	70
LES COMBINAISONS DE MARQUES EPIGENETIQUES ASSOCIEES A H2AK119Ub1 DANS LA CHROMATINE DU SPERMATOZOÏDE .....	71
<i>Distribution de H3K4me3 et/ou H3K27me3 dans les régions du génome du spermatozoïde enrichies pour H2AK119Ub1</i> .....	71
<i>H2AK119Ub1 associé avec H3K4me3 est enrichi sur les gènes sensibles au traitement USP21</i> .....	76
 <b>CHAPITRE IV – APRES REPLICATION DE LA CHROMATINE DU SPERMATOZOÏDE, UNE CONFIGURATION BIVALENTE</b>	
<b>H2AK119UB1/H3K4ME3 EST MAINTENUE SUR LES GENES SENSIBLES AU TRAITEMENT USP21 .....</b>	<b>78</b>
REPLICATION DE LA CHROMATINE DU SPERMATOZOÏDE PAR INCUBATION DANS UN EXTRAIT D'ŒUF .....	79
LE DEVENIR DE H2AK119Ub1 ET H3K4ME3 APRES LA PREMIERE REPLICATION .....	82
<i>La réplication de la chromatine du spermatozoïde est associée à un remodelage extensif de la distribution de H2AK119Ub1 et à une stabilité de celle de H3K4me3</i> .....	82
<i>H2AK119Ub1 après réplication est préférentiellement associé avec les gènes sensibles au traitement USP21</i> .....	86
DANS UN CONTEXTE GLOBAL DE DILUTION REPLICATIVE DE H2AK119Ub1, LES GENES SENSIBLES AU TRAITEMENT USP21 MAINTIENNENT UN NIVEAU PRE-REPLICATIF DE CETTE MARQUE .....	90
<i>Pipeline spikChIP appliqué aux échantillons Xenopus laevis couplés à la drosophile</i> .....	93
<b>DISCUSSION .....</b>	<b>98</b>
H2AK119Ub1 DU SPERMATOZOÏDE A L'EMBRYON .....	98
LE TRAITEMENT USP21 ALTERE LE DEVELOPPEMENT DES EMBRYONS ET LE TRANSCRIPTOME.....	99
LA DISTRIBUTION DE H2AK119Ub1 DANS LE SPERMATOZOÏDE ET AU COURS DU DEVELOPPEMENT EMBRYONNAIRE .....	100
L'EFFET SUR LE DEVELOPPEMENT DE L'EMBRYON PAR USP21 EST-IL SPECIFIQUE DE SON ACTIVITE DEUBIQUITYLASE ?.....	101
LA TRANSMISSION DE H2AK119Ub1 DU SPERMATOZOÏDE A L'EMBRYON.....	102
H2AK119Ub1 DANS UN CONTEXTE DE TRANSMISSION INTERGENERATIONNELLE .....	103
<b>PUBLICATION ET EVEIL SCIENTIFIQUE .....</b>	<b>104</b>
<b>REFERENCES .....</b>	<b>105</b>
<b>ABREVIATIONS .....</b>	<b>117</b>
<b>ARTICLE .....</b>	<b>117</b>

# Liste des figures

Figure 1 : Illustration du paysage épigénétique de Waddington .....	9
Figure 2 : Le modèle de régulation de Waddington .....	10
Figure 3 : Modélisation de la méthylation des îlots CpG .....	12
Figure 4 : Organisation de la chromatine par empaiquetage de l'ADN autour d'histones .....	14
Figure 5 : Schéma de la spermatogénèse .....	17
Figure 6 : Schéma du développement embryonnaire de <i>Xenopus laevis</i> .....	19
Figure 7 : Représentation des complexes de répression Polycomb .....	23
Figure 8 : Les stratégies d'interférence avec l'épigénome du spermatozoïde .....	25
Figure 9 : Rétention de H2AK119Ub1 lors de la spermiogénèse .....	26
Figure 10 : H2AK119Ub1 est présent aux stades embryonnaires précédant la ZGA .....	26
Figure 11 : Protocole expérimental du traitement USP21 du spermatozoïde .....	27
Figure 12 : Le traitement avec un extrait d'ovocyte contenant la deubiquitylase USP21 efface spécifiquement H2AK119Ub1 du spermatozoïde .....	28
Figure 13 : Le traitement avec un extrait d'ovocyte ne perturbe pas la structure globale de la chromatine du spermatozoïde .....	28
Figure 14 : Modèle expérimental pour la génération d'embryons haploïdes à partir de spermatozoïdes épigénétiquement édités .....	29
Figure 15 : Les embryons issus de spermatozoïdes traités par l'extrait USP21 présentent des défauts de développement à partir du stade gastrula. ....	30
Figure 16 : Schéma expérimental du RNA-Sequencing dans les gastrulas .....	32
Figure 17 : Structure d'un read dans un fichier fastq .....	34
Figure 18 : Contrôle-qualité des échantillons RNA-Seq .....	35
Figure 19 : Pipeline d'alignement des données RNA-Seq .....	36
Figure 20 : Analyse de la variabilité du jeu de données RNA-Seq par ACP .....	38
Figure 21 : La suppression de l'effet batch .....	39
Figure 22 : Pipeline d'analyse de la table de comptage en RNA-Seq .....	42
Figure 23 : Le traitement à l'extrait d'ovocyte n'altère pas le transcriptome de l'embryon .....	44
Figure 24 : Le traitement USP21 du spermatozoïde mène à des défauts d'expression géniques. ....	45
Figure 25 : Les gènes du cycle cellulaire et du métabolisme de l'embryon sont dérégulés par le traitement USP21 du spermatozoïde .....	46
Figure 26 : La réponse au traitement USP21 dans le jeu de données d'embryons uniques est hétérogène .....	46
Figure 27 : Le profil transcriptomique des gènes dérégulés dans les embryons regroupés est hétérogène dans les embryons uniques .....	47
Figure 28 : Le profil d'expression génique est similaire entre mon analyse et l'analyse Xenbase .....	48

<b>Figure 29 : Le traitement USP21 perturbe l'expression des gènes à la ZGA .....</b>	<b>49</b>
<b>Figure 30 : Le traitement USP21 perturbe la transcription des nouveaux ARN à la ZGA.....</b>	<b>51</b>
<b>Figure 31 : Schéma récapitulatif des échantillons de ChIP-Seq .....</b>	<b>53</b>
<b>Figure 32 : Schéma du protocole de préparation des échantillons de ChIP-Seq pour le séquençage .....</b>	<b>55</b>
<b>Figure 33 : Les réplicats de ChIP-Sequencing sont conformes aux recommandations ENCODE 3 .....</b>	<b>57</b>
<b>Figure 34 : Analyse fingerprintplot de la couverture des reads des fractions ChIP et Input.....</b>	<b>59</b>
<b>Figure 35 : Représentation schématique du principe de détection d'enrichissement par l'outil RECOGNICER .....</b>	<b>60</b>
<b>Figure 36 : Comparaison des pics H2AK119Ub1 déterminés par RECOGNICER et par MACS2 .....</b>	<b>61</b>
<b>Figure 37 : Exemple de pics H2AK119Ub1 détectés par RECOGNICER versus MACS2 .....</b>	<b>63</b>
<b>Figure 38: Heatmap de corrélation de Spearman des réplicats de ChIP-Seq.....</b>	<b>65</b>
<b>Figure 39 : Pipeline d'analyse des données de ChIP-Sequencing .....</b>	<b>65</b>
<b>Figure 40 : Le signal H2AK119Ub1 sur le cluster <i>hoxa.L</i> est semblable entre les trois réplicats</b>	<b>66</b>
<b>Figure 41 : Exemple de deux gènes sensibles au traitement USP21, <i>trib2.S</i> et <i>pacs2.L</i>, présentant un enrichissement pour H2AK119Ub1 dans le spermatozoïde .....</b>	<b>67</b>
<b>Figure 42 : Localisation de H2AK119Ub1 sur le génome du spermatozoïde.....</b>	<b>68</b>
<b>Figure 43 : H2AK119Ub1 enrichi sur les régions régulatrices dans le spermatozoïde .....</b>	<b>69</b>
<b>Figure 44 : Les gènes du développement sont enrichis en H2AK119Ub1 dans le spermatozoïde</b>	<b>69</b>
<b>Figure 45 : Test de permutation de l'enrichissement du signal H2AK119Ub1 sur les gènes sensibles au traitement à l'extrait USP21.....</b>	<b>70</b>
<b>Figure 46 : H2AK119Ub1 et H3K4/K27me3 sont globalement retenus sur différentes particules nucléoprotéiques dans le spermatozoïde .....</b>	<b>73</b>
<b>Figure 47 : H2AK119Ub1, H3K4me3 et H3K27me3 enrichis dans le spermatozoïde sur <i>pacs2.L</i> .....</b>	<b>74</b>
<b>Figure 48 : Proportion des différentes associations possibles entre les pics H2AK119Ub1 et les pics H3K4me3 et/ou H3K27me3 dans le spermatozoïde .....</b>	<b>75</b>
<b>Figure 49 : Distribution génomique des différentes combinaisons épigénétiques associées à H2AK119Ub1 dans le spermatozoïde .....</b>	<b>75</b>
<b>Figure 50 : H2AK119Ub1 couplé à H3K4me3 est enrichi sur les TSSs des gènes sensibles au traitement USP21 .....</b>	<b>77</b>
<b>Figure 51 : Réplication de L'ADN du spermatozoïde après incubation dans un extrait d'œuf...</b>	<b>79</b>
<b>Figure 52 : Présentation des échantillons de H2AK119Ub1, H3K4me3 et H3K27me3 .....</b>	<b>80</b>
<b>Figure 53 : L'organisation canonique de la chromatine est rétablie après la réplication.....</b>	<b>81</b>
<b>Figure 54 : H2AK119Ub1 est activement remodelé pendant le développement embryonnaire...</b>	<b>82</b>

<b>Figure 55 : Les pics H2AK119Ub1 maintenus pendant la réplication sont localisés sur les TSSs</b>	<b>83</b>
<b>Figure 56 : H2AK119Ub1 est déposé sur les TSSs après la première réplication</b> .....	<b>84</b>
<b>Figure 57 : Le signal H2AK119Ub1 est stabilisé dans l’embryon lorsqu’il est couplé à H3K4me3</b> .....	<b>85</b>
<b>Figure 58 : Les gènes sensibles au traitement USP21 maintiennent H2AK119Ub1 après réplication de la chromatine du spermatozoïde</b> .....	<b>87</b>
<b>Figure 59 : H2AK119Ub1 est enrichi sur les gènes sensibles au traitement USP21 dans l’embryon</b> .....	<b>89</b>
<b>Figure 60 : Le devenir de H2AK119Ub1 dans l’embryon</b> .....	<b>90</b>
<b>Figure 61 : Modélisation du pipeline d’analyse ChIP-seq pour la normalisation Drosophile</b> .....	<b>92</b>
<b>Figure 62 : Heatmap de corrélation de Spearman des réplicats de ChIP-Seq quantitatif</b> .....	<b>94</b>
<b>Figure 63 : Contrôle de la réplication de l’ADN du spermatozoïde</b> .....	<b>94</b>
<b>Figure 64 : Boxplot du signal H2AK119Ub1 lors de la normalisation SpikChIP</b> .....	<b>96</b>
<b>Figure 65 : H2AK119Ub1 est préférentiellement retenu sur les gènes sensibles au traitement USP21</b> .....	<b>97</b>
<b>Figure 66 : Le modèle H2AK119Ub1 dans le développement de l’embryon</b> .....	<b>98</b>

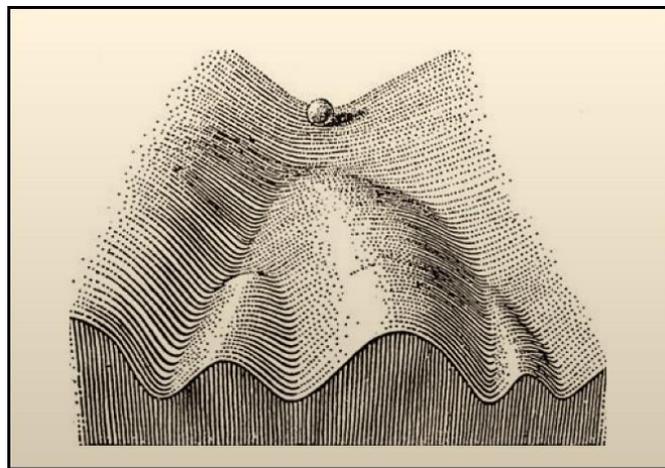
# Introduction

Dans ce manuscrit, je m'intéresse à l'impact des mécanismes de l'épigénétique sur le développement embryonnaire. L'épigénétique représente toute l'information supplémentaire présente sur le génome en dehors de la séquence d'ADN. Cette information permet la régulation des mécanismes cellulaires impliquant l'ADN tels que la réplication, la transcription ou la réparation. Elle est transmise aux cellules filles au cours des divisions cellulaires (Probst, Dunleavy, et Almouzni 2009), mais elle est également maintenue dans des cellules ne se divisant pas, comme dans le système immunitaire inné (Netea et al. 2020) ou les neurones (Kannan et Ravi 2013). Ici, j'ai étudié la transmission de l'information épigénétique d'une génération à l'autre en me focalisant sur la contribution du spermatozoïde à l'embryon (Lacal et Ventura 2018). Je me suis plus particulièrement intéressé à évaluer si une composante de l'épigénome paternel, les histones modifiées, constituent un support épigénétique nécessaire au développement de l'embryon. Dans cette introduction, je décris d'abord brièvement les différents supports de l'information épigénétique et leurs possibles implications dans la programmation épigénétique du spermatozoïde pour le développement embryonnaire. Je présente également le choix et la spécificité de mon modèle d'étude, le xénope, et son intérêt pour l'étude de ces mécanismes épigénétiques. Enfin, je présente la logique qui m'a conduit à cibler mon étude sur la modification post-transcriptionnelle de H2A, H2AK119Ub1.

# L'épigénétique : de l'hérédité à la variabilité individuelle

## *Les fondations du concept de la régulation épigénétique*

Le contexte de paysage épigénétique est introduit en 1957 dans « The strategy of the genes » par Conrad Hal Waddington (Waddington 1957). Dans cet ouvrage, il présente une nouvelle vision du développement cellulaire, et notamment de l'établissement des différents destins cellulaires. Il représente ce phénomène par une figure où une balle est placée au sommet d'un paysage, et à laquelle plusieurs chemins (vallées) s'offrent, symbolisant ainsi différentes trajectoires de différenciation (**Figure 1**, (Goldberg, Allis, et Bernstein 2007)).

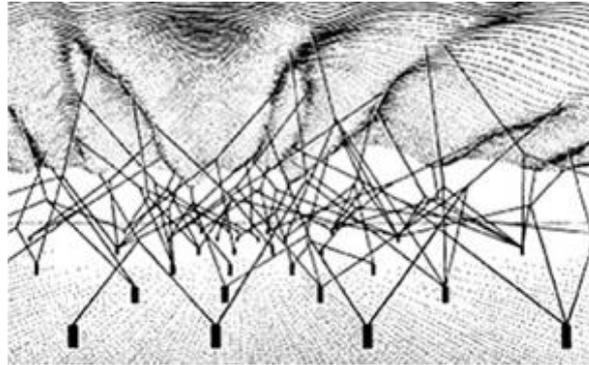


**Figure 1 : Illustration du paysage épigénétique de Waddington**

*Modélisation du modèle du concept de paysage épigénétique décrit par Waddington. Une bille représente la cellule qui peut prendre plusieurs chemins possibles (vallées) pour arriver à différents destins cellulaires, représentant les différentes décisions de choix de devenir cellulaires durant le développement. Figure réimprimée de l'ouvrage de Waddington en 1957 (Waddington 1957).*

Waddington définit ainsi les principes de la régulation du destin cellulaire par un modèle où les gènes et leurs interactions sont des facteurs définissant la forme des montagnes et vallées, qui caractérisent les trajectoires cellulaires de différenciation. Selon ce modèle, ce sont ainsi les combinaisons d'interaction de gènes (**Figure 2**, (Baedke 2013)) qui définissent les branches des trajectoires cellulaires au cours du développement. Ce modèle introduit un concept central selon lequel une cellule ne peut changer de trajectoire de différenciation de façon naturelle une fois engagée dans une branche (vallée). Ce modèle décrit que ce ne sont pas les gènes seuls, mais leurs façons d'interagir et d'être régulés, sans modifier la séquence d'ADN (Watson et Crick

1953), qui influence l'expression des gènes. Ce modèle met également en avant une fondation des mécanismes épigénétique, c'est-à-dire des mécanismes de régulation des gènes qui sont stablement maintenus lors de la progression développementale d'une cellule. Waddington décrit ainsi, en 1968, l'épigénétique comme une branche de la biologie qui étudie le lien entre l'interaction des gènes avec leurs produits et le phénotype en devenir (Waddington 1968).



**Figure 2 : Le modèle de régulation de Waddington**

*Modélisation du système de régulation de Waddington. La forme du paysage que la cellule peut prendre est définie par deux éléments. Le premier élément schématisé par les cylindres à la base représente les gènes, puis les fils montrent les interactions entre les différents gènes spécifiques aux différents chemins, représentant le réseau de régulation. Figure réimprimée de l'ouvrage de Waddington en 1957 (Waddington 1957).*

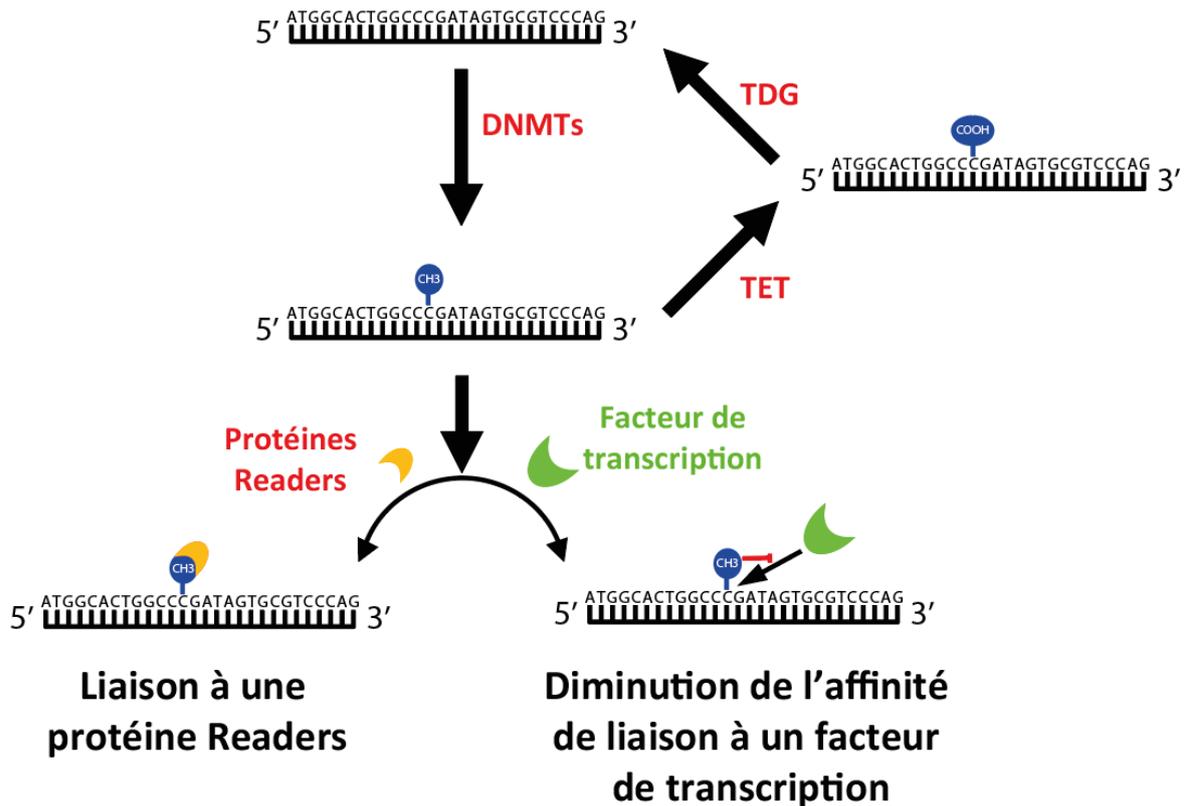
## **Les différents supports de l'épigénétique**

L'épigénome regroupe tous les éléments impliquant des interactions avec l'ADN. L'épigénome est un facteur clé pour réguler la production des protéines dans les cellules, qui seront spécifiques au type cellulaire. Ces marques épigénétiques sont transmissibles pendant les divisions cellulaires (Probst, Dunleavy, et Almouzni 2009) permettant notamment le maintien d'un type cellulaire, mais aussi transmissible à la descendance (Skvortsova, Iovino, et Bogdanović 2018). Contrairement au génome, l'épigénome est plus flexible étant composé de modifications ou d'éléments réversibles qui peuvent être modulés par plusieurs facteurs comme la nutrition, l'environnement, le stress et le mode de vie de l'individu par exemple (Alegría-Torres, Baccarelli, et Bollati 2011). Certains facteurs épigénétiques sont prédictifs d'une sensibilité plus accrue à des pathologies (Moosavi et Motevalizadeh Ardekani 2016). Il existe trois supports principaux d'information épigénétique :

**Modifications chimiques des bases de l'ADN.** Les bases de l'ADN peuvent être modifiées par l'ajout d'un groupement chimique impactant la transcription des gènes (Kumar, Chinnusamy, et Mohapatra 2018). Actuellement, des modifications sur l'adénine et la cytosine seulement ont été découvertes. Par exemple, la méthylation de l'adénine permet de réguler l'expression des gènes, ou de réparer, répliquer et transposer l'ADN chez les procaryotes (Kumar, Chinnusamy, et Mohapatra 2018), mais son abondance est très faible chez les eucaryotes (K.-J. Wu 2020). Les îlots CpG sont des régions du génome avec un fort enrichissement de Cytosine et Guanine dans leur séquence nucléotidique (Bird et al. 1985). Ces séquences sont préférentiellement conservées dans les régions promotrices, en particulier pour les gènes de ménage (Gardiner-Garden et Frommer 1987). Plusieurs enzymes peuvent modifier ces séquences, premièrement les enzymes nommées « Writers » qui méthylent la cytosine au sein des îlots CpG (5-mC). Plusieurs membres de la famille des ADN méthyltransférase réalisent cette fonction, comme DNMT1 qui propage la marque au nouveau brin synthétisé lors de la réplication de l'ADN, ou DNMT3A et DNMT3B qui déposent un groupement méthyle de novo (Moore, Le, et Fan 2013). D'autres protéines, nommées « readers » vont reconnaître le groupement méthyl et l'interpréter, comme les Methyl-CpG Binding Domain (MBD) ou les protéines à doigts de zinc (Zinc Fingers proteins, ZNFs). L'enzyme ten-eleven translocation (TET) permet d'oxyder le groupement méthyl ajouté initialement, pour former un groupement 5-hydroxyméthylcytosine (Tahiliani et al. 2009), 5-formylcytosine (Klungland et Robertson 2017) ou 5-carboxycytosine (He et al. 2011). Enfin, des enzymes nommées « erasers » peuvent enlever les modifications, comme la thymine DNA glycosylase (TDG) qui enlève le groupement carboxyle (He et al. 2011). La méthylation des îlots CpG est généralement associée à la répression des gènes induite par le recrutement de protéines impliquées dans la répression de la transcription (Moore, Le, et Fan 2013), ou par le blocage du recrutement des facteurs de transcription par la présence du groupe méthyl (**Figure 3**).

**Interaction ARN.** Plusieurs types d'ARNs produits dans une cellule sont des ARNs non-codants. Pour autant, ils sont tout aussi important que les ARNs codants pour des protéines. Par exemple, les petits ARNs qui font entre 20 à 25 paires de bases, sont impliqués dans la régulation de l'expression génique (Costa 2008). Parmi eux, on retrouve les Piwi-interacting RNAs (piRNAs) qui sont des petits ARNs capables d'interagir avec la protéine Piwi pour réprimer des gènes, notamment lors du développement de l'embryon avec les gènes du domaine homéobox (Halbach et al. 2020). Les small-interfering RNAs (siRNAs) sont associés à la méthylation de l'ADN pour réguler la transcription des gènes (Morris et al. 2004), tout comme

les micro-ARNs (miRNAs) qui jouent un rôle dans la régulation de la transcription (O'Brien et al. 2018). Enfin, des ARN plus long nommés long-non coding RNAs (lncRNAs) sont des ARNs souvent non-codants de plus de 200pb (Mercer, Dinger, et Mattick 2009). L'ARN Xist impliqué lors de l'inactivation du chromosome X fait partie de ce type d'ARNs notamment (Wei et al. 2017).



**Figure 3 : Modélisation de la méthylation des îlots CpG**

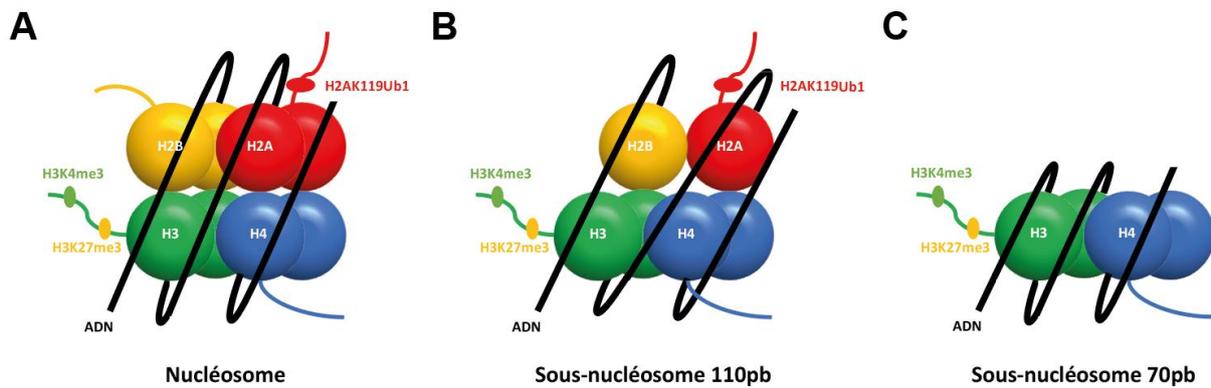
*Une cytosine est méthylée par une DNMT, permettant soit sa lecture par une protéine readers, soit bloquant le recrutement d'un facteur de transcription. Le groupement méthyl est oxydé par l'enzyme TET avant d'être enlevé par l'enzyme TDG pour permettre à l'ADN de retrouver sa séquence originelle non modifiée.*

Mon travail de thèse s'est concentré sur le troisième support de l'information épigénétique représenté par les modifications biochimiques des queues d'histones, que je détaille ci-dessous.

## *Le nucléosome, une organisation spécifique de la chromatine pour la régulation génique*

Au sein du noyau, l'ADN est enroulé autour de protéines basiques nommées histones. L'organisation canonique des histones dans une cellule somatique est sous la forme d'un nucléosome (**Figure 4**). Celui-ci est composé de 2 sous-unités de chacune des quatre histones de base : H2A, H2B, H3 et H4 (Zhou, Gaullier, et Luger 2019), couplé à une histone H1 (Cutter et Hayes 2015). Cet empilement de l'ADN autour des histones permet de compacter l'ADN d'une cellule (Macadangdang et al. 2014)-(Felsenfeld et Groudine 2003). Ces histones ont des variants, comme H3.3 qui a un rôle dans l'infertilité mâle lorsqu'elle est absente dans le spermatozoïde chez la souris (Yuen et al. 2014). Les histones sont composées d'une partie N-terminale nommée queue d'histone, qui est composée principalement d'acide aminé chargés positivement, comme l'arginine et la lysine (Bendandi et al. 2020). Ces résidus peuvent contenir des modifications post-traductionnelles, comme la tri-méthylation de la lysine 4 de l'histone 3 (H3K4me3) ou l'acétylation de la lysine 27 de l'histone 3 (H3K27Ac). Ces marques sont associées à différents états de la chromatine. Par exemple, la marque répressive H3K27me3 est associée à l'hétérochromatine facultative (Bannister et Kouzarides 2011), qui est localisée sur des régions chromosomiques variables contenant usuellement des gènes qui doivent être réprimés, où la chromatine est condensée (Saksouk, Simboeck, et Déjardin 2015). H3K9me3 est aussi une marque répressive, mais elle est plutôt associée avec l'hétérochromatine constitutive, c'est-à-dire sur des régions chromosomiques où la chromatine sera toujours condensée, comme les régions répétées des centromères des chromosomes chez l'Homme (Decombe et al. 2021). Au contraire, des marques sont associées à une chromatine moins condensée, l'euchromatine, comme H3K4me3 ou H3K36me3 (Bannister et Kouzarides 2011). Les variants d'histones peuvent également faire l'objet de modifications post-transcriptionnelles : H3.3 a par exemple un rôle essentiel lors de l'embryogenèse par l'intermédiaire de la phosphorylation de la sérine 31, spécifique de ce variant, nécessaire à la gastrulation chez le xénope (Sitbon et al. 2020).

Lors de ma thèse, je me suis intéressé à plusieurs de ces marques (**Figure 4**). H3K4me3, est une marque localisée à proximité des sites d'initiation de la transcription, et associée à l'activation de la transcription dans des régions euchromatique. Cette modification est déposée par une lysine méthyltransférase (KMT2F/KMT2G), qui se fixe sur les îlots CpG non méthylés (Beacon et al. 2021). Elle est effacée ultérieurement par l'action d'histone déméthylase (i.e. KDM5) (Beacon et al. 2021). Enfin, j'ai investigué le rôle de H3K27me3, une marque répressive de l'initiation de la transcription, et l'ubiquitination de la lysine 119 de l'histone H2A nommée H2AK119Ub1, elle aussi impliquée dans la répression de la transcription, que je détaillerai ultérieurement.



**Figure 4 : Organisation de la chromatine par empilement de l'ADN autour d'histones**

**(A)** Organisation canonique de la chromatine en nucléosome par un empilement avec un octamère d'histones. La séquence d'ADN en noir est enroulée autour du nucléosome, composé de huit histones : deux H2A, deux H2B, deux H3 et deux H4. Ces histones contiennent des queues composées d'acide aminé et peuvent subir des modifications post-traductionnelles. Trois marques d'intérêt sont représentées : H3K4me3 (en vert), un facteur activateur de la transcription, contre H3K27me3 (en jaune) et H2AK119Ub1 (en rouge) qui sont des marques répressives de la transcription. **(B)** Représentation de la structure d'un sous-nucléosome avec la perte d'une histone H2A et une histone H2B. **(C)** Représentation de la structure d'un sous-nucléosome après la perte des deux histones H2A et H2B.

# Du spermatozoïde à l'embryon : transmission de l'information génétique et épigénétique

## *Xenopus laevis*, un modèle biologique pertinent

Le modèle animal utilisé pour les analyses est le xénope (*Xenopus laevis*). Il s'agit d'un amphibien allotétraploïde, dont le génome a été aligné et référencé pour la première fois en 2016 (Session et al. 2016). La spécification entre les espèces *Xenopus laevis* et *Xenopus tropicalis* est estimée à 48 millions d'années, l'apparition des copies S et L de *Xenopus laevis* est estimée à 34 millions d'années, et la divergence entre *Xenopus laevis* et *Xenopus borealis* à 17 millions d'années. Le génome est composé de 18 paires de chromosomes, réparties équitablement entre deux copies du génome nommées L (Long) et S (Short), composées de 9 paires de chromosomes chacune. Les deux copies ont évolué de façon différente dans le temps, avec la copie S qui a subi plus de réarrangements génomiques, et une diminution de son expression génique. Ainsi, les copies L et S d'un même chromosome sont proches en termes de séquences, mais les différences apparues au cours du temps sont désormais distinguables avec les techniques modernes de séquençage. En 2017, un premier assemblage a été réalisé avec la version 9.2 du génome. J'ai utilisé cette version du génome durant ma thèse, elle contient 51 041 gènes, dont 45 093 situés sur des chromosomes. Environ 12% des gènes sont situés sur des scaffolds, qui ne sont pas encore attribués à des localisations chromosomiques. Enfin, depuis 2021, la version 10.1 totalement assemblée du génome est sortie, contenant 44 457 gènes (Fortriede et al. 2019) sans scaffolds.

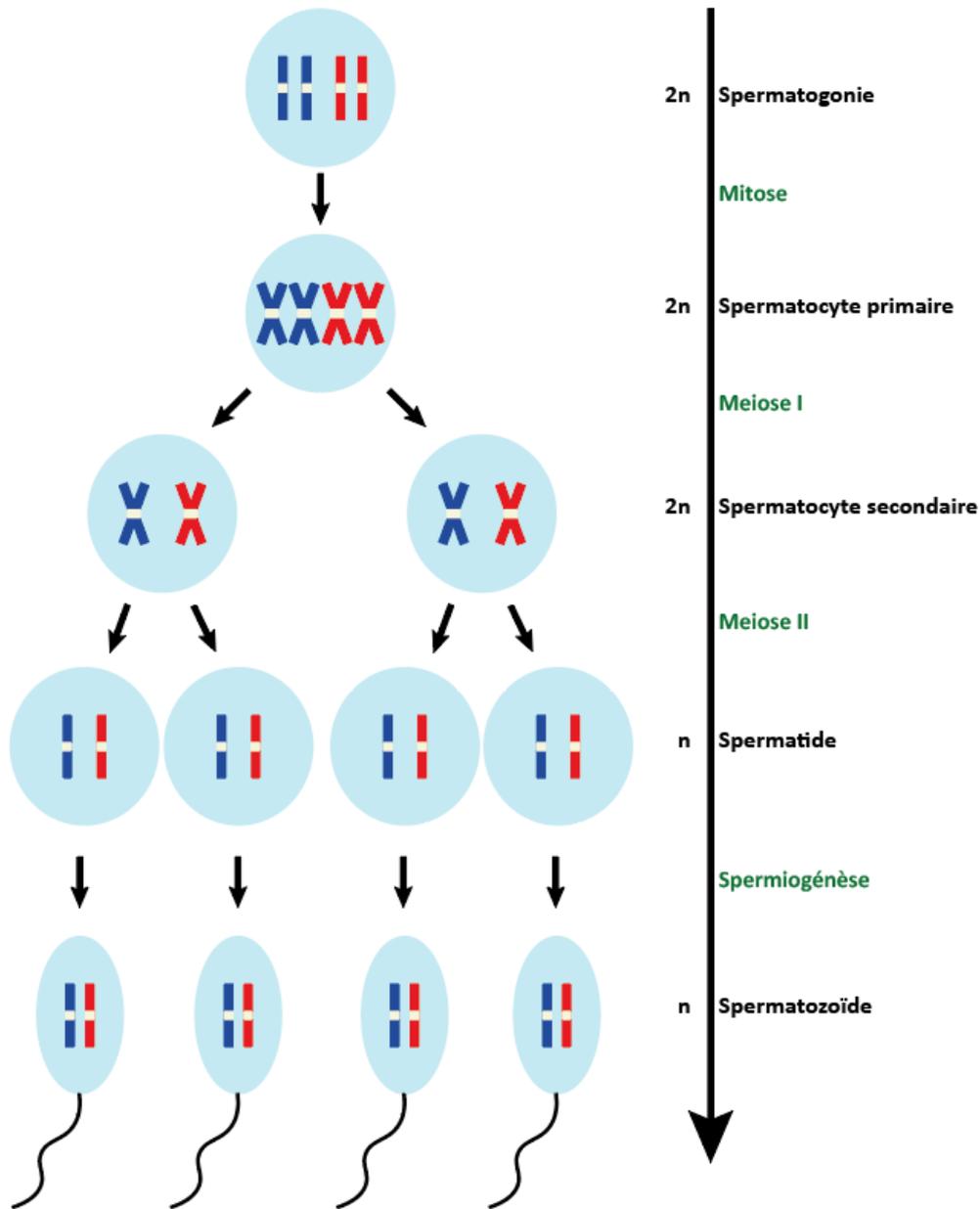
Il y a plusieurs avantages à ce modèle d'études lié à la thématique de recherche pour l'étude du développement embryonnaire :

- L'obtention d'œufs en grande quantité et de grandes taille (Nenni et al. 2019).
- Un développement embryonnaire externe rapide (Fortriede et al. 2019)-(Nenni et al. 2019).
- Un modèle compatible avec plusieurs types de manipulations génomiques (Injection intracytoplasmique de spermatozoïdes, transfert nucléaire, extrait d'œuf) (Nenni et al. 2019)-(Oikawa et al. 2020)-(Teperek et al. 2016).
- La conservation des mécanismes cellulaires régulant le développement embryonnaire et une homologie avec environ 80 à 90% des pathologies chez l'Homme (Nenni et al. 2019)-(Blum et Ott 2018).

## *De la spermatogénèse au développement embryonnaire précoce*

### La spermatogénèse, un processus de maturation des spermatozoïdes

Les cellules d'un organisme peuvent être séparées en deux catégories principales : les cellules somatiques qui représentent la majorité des cellules du corps, et les cellules germinales impliquées dans la transmission du patrimoine génétique à la génération suivante. La production des cellules germinales mâles et femelles nécessitent des processus de maturation spécifiques. L'oogenèse est le processus permettant l'obtention d'un ovocyte mature, et la spermatogénèse décrit le chemin de différenciation vers un spermatozoïde mature (Nishimura et L'Hernault 2017). Cette maturation du spermatozoïde se déroule en plusieurs étapes permettant le passage d'une cellule diploïde vers une cellule haploïde (**Figure 5**). La spermatogénèse représente la transition d'une cellule diploïde initiale à quatre cellules haploïdes. Elle débute par une division mitotique d'une spermatogonie pour donner un spermatocyte primaire. Le spermatocyte primaire réalise un premier cycle de méiose pour former deux spermatocytes secondaires, puis un deuxième cycle de méiose générant quatre spermatides haploïdes. Enfin, la spermiogénèse correspond à la maturation des spermatides en spermatozoïdes, notamment par l'acquisition de composants cellulaires très spécifiques comme le flagelle. Lors de cette transition, certains nucléosomes sont remplacés par de la protamine, ou certaines histones sont perdues ou remplacées par des variantes d'histones comme le variant H3.3 (Yuen et al. 2014) pour condenser très fortement l'ADN (Wang et al. 2019). Environ 1% des histones sont maintenues dans le spermatozoïde chez la souris (Carone et al. 2014)-(Jung et al. 2017) contre 10% chez l'Homme (Brykczynska et al. 2010)-(Hammoud et al. 2009). Au contraire, certaines espèces comme le zébrafish n'ont pas de perte d'histones (S.-F. Wu, Zhang, et Cairns 2011). Dans le modèle *Xenopus laevis*, le remplacement des histones par des protamines est associé à une conservation des histones H3 et H4 mais d'une perte des histones H2A et H2B (Oikawa et al. 2020)-(Katagiri et Ohsumi 1994). Ceci se manifeste par la coexistence, au sein de la chromatine du spermatozoïde, de nucléosome canonique, mais également de différentes formes de sous-nucléosome caractérisées par des niveaux variables de perte de H2A/H2B (**Figure 4**). Cette caractéristique rend ce modèle d'étude intéressant puisque nous étudions une marque spécifique de l'histone H2A.



**Figure 5 : Schéma de la spermatogénèse**

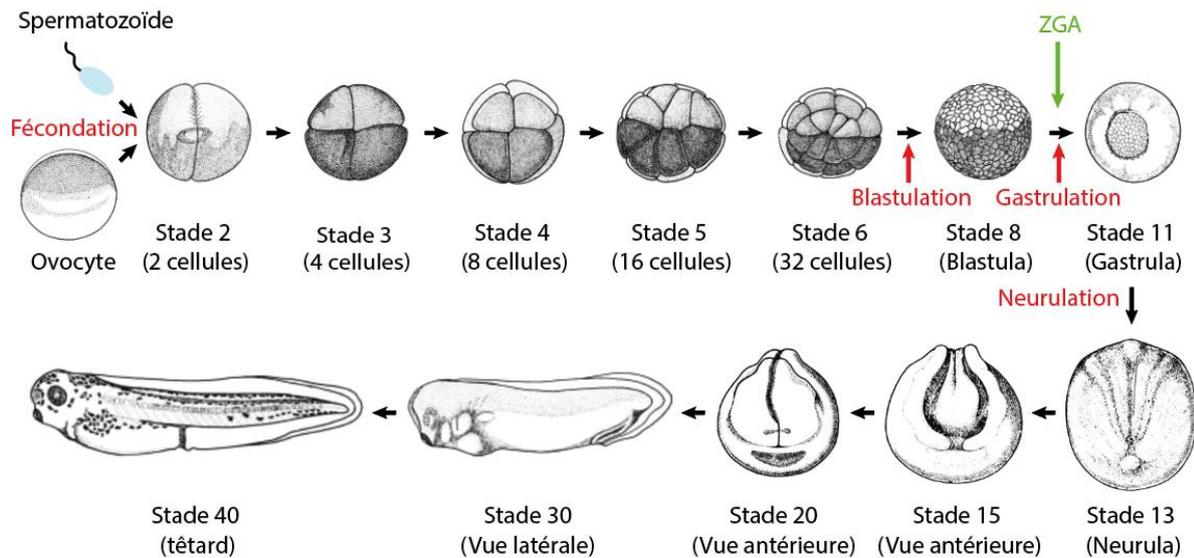
*La spermatogonie donne le spermatocyte primaire après un cycle de mitose. Deux cycles de méiose consécutifs mènent à la formation de quatre spermatides, à partir d'un spermatocyte primaire, par l'intermédiaire de deux spermatocytes secondaires. Enfin, lors de la spermiogénèse, la spermatide acquiert plusieurs modifications structurelles comme le flagelle. Les chromosomes sont représentés en bleu et en rouge, ainsi que la ploïdie à chaque phase.*

## Le développement embryonnaire chez *Xenopus laevis*

L'embryogenèse débute par la formation d'un zygote, cellule unique issue de la fécondation d'un ovocyte par un spermatozoïde. Le zygote subit ensuite une série de divisions cellulaires, passant ainsi aux stades de développement 2 à 6, correspondant aux stades 2, 4, 8, 16 et 32 cellules (Fortriede et al. 2019) (*Copyright © 1994 Pieter D. Nieuwkoop and J. Faber. Digital images created by Xenbase*). Plusieurs phénomènes interviennent ensuite durant le développement de cet embryon. La blastulation est le phénomène qui correspond à la création d'une sphère composée de cellules nommées blastomères entourant une cavité nommée le blastocœle. Ce stade blastula apparaît en moyenne 4 à 5 heures après la fécondation (incubation à 22-23°C). Ultérieurement, le génome du zygote initialement inactif va s'activer vers le stade 8-9 (cycle de clivage 10-12), correspondant à l'activation du génome du zygote (ZGA). S'ensuit la gastrulation qui correspond à la phase de développement où les feuilletts embryonnaires sont mis en place, établissant les premières lignées cellulaires du stade 10 jusqu'au stade 12 (Heasman 2006). La neurula se met en place à partir du stade 12, avec l'apparition du système nerveux central suite à la neurulation et l'apparition du tube neural. L'embryon se développe ensuite avec la mise en place des différents organes jusqu'au stade 40 correspondant au têtard, pour un développement total d'environ 5 jours (**Figure 6**).

## La contribution épigénétique du spermatozoïde au développement embryonnaire

Il a été estimé que 8 à 12% des couples ont des problèmes d'infertilité, avec une cause paternelle dans environ 50% des cas (Agarwal et al. 2021). Lors du développement embryonnaire humain, la copie paternelle du génome est rapidement déméthylée par l'enzyme TET3 après fécondation, tandis que la copie maternelle est déméthylée plus progressivement par l'absence de nouvelle méthylation par dilution du signal avant d'être reméthylée plus tardivement au stade de Blastula après la ZGA (Breton-Larrivée, Elder, et McGraw 2019). Chez le zébrafish, des analyses de la méthylation de l'ADN montrent un niveau de méthylation plus faible sur la copie maternelle, qui est reméthylée après la fécondation pour ressembler au niveau de méthylation du spermatozoïde (Silva 2018)-(Potok et al. 2013).



**Figure 6 : Schéma du développement embryonnaire de *Xenopus laevis***

Un ovocyte est fécondé par un spermatozoïde, donnant l'embryon. Après 1h30 environ, l'ovocyte fécondé devient un embryon de stade 2 constitué de 2 cellules. Des cycles de divisions cellulaires ont lieu ensuite à un rythme de 30 minutes par cycle en moyenne, donnant le stade 3 à 4 cellules, stade 4 à 8 cellules, stade 5 à 16 cellules et le stade 6 à 32 cellules. Après le stade 32 cellules, la blastulation a lieu pour donner naissance à la Blastula. La blastula devient la gastrula suite à la gastrulation. Durant cette transition, le transcriptome de l'embryon s'active, on nomme ce phénomène la ZGA. Enfin, la Gastrulation donnant la gastrula consiste à la génération des premiers organes, avec l'apparition du système nerveux. Après 3 jours, l'embryon devient un têtard, désigné stade 40. Les images sont issues de la base de données Xenbase (Copyright © 1994 Pieter D. Nieuwkoop and J. Faber. Digital images created by Xenbase).

En plus de l'analyse de la méthylation de l'ADN, des études de cartographie des modifications d'histones de plusieurs espèces comme chez la souris (Erkek et al. 2013), chez l'Homme (Brykczynska et al. 2010)-(Hammoud et al. 2009), ou encore le zébrafish (S.-F. Wu, Zhang, et Cairns 2011) documentent l'association des modifications des queues d'histones dans les spermatozoïdes à des gènes associés au développement. Dans un contexte où seulement 1% des histones sont retenues dans le spermatozoïde chez la souris, il a été démontré un enrichissement de ces histones retenues sur les îlots CpG non méthylés par des expériences de ChIP-Sequencing sur les histones H3 (Erkek et al. 2013). Ce mécanisme est conservé chez l'Homme, et suggère un rôle possible de transmission d'informations épigénétique paternelle via des nucléosomes retenus sur les îlots CpG (Erkek et al. 2013).

Il a été démontré chez l'Homme par des expériences d'histones ChIP-Seq contre les marques H3K4me2 et H3K27me3 que la marque H3K4me2 était conservée sur les gènes, comme le cluster Hox, impliqués dans la spermatogénèse et dans l'homéostasie cellulaire, permettant ainsi leur activation (Brykczynska et al. 2010). Au contraire, la marque H3K27me3 est présente sur les gènes du développement, que ce soit dans le spermatozoïde ou dans les cellules somatiques de l'embryon. Ces gènes ne sont pas exprimés dans les gamètes ou l'embryon précoce et présentent en outre une configuration bivalente caractérisée par la présence sur leur promoteur de H3K4me3 et H3K27me3. Ces études corrélatives suggèrent la possible existence d'une programmation épigénétique du spermatozoïde, via les histones modifiées retenues lors de la spermiogénèse, pour supporter le développement embryonnaire (Brykczynska et al. 2010).

Dans une précédente étude de notre équipe, des expériences de cartographie par ChIP-seq ont été menées pour déterminer la distribution de H3K4me3, H3K4me2, H3K27me3 et H3K9me3 dans le spermatozoïde et la spermatide (Teperek et al. 2016). Ces études ont été combinées à des études de RNA-seq sur des embryons dérivés de ces deux types de cellules germinales paternelles. Ceci a permis de démontrer une corrélation entre les différences dans l'épigénome du spermatozoïde et de la spermatide et les différences d'expression des gènes dans les embryons dérivés de ces deux types de cellules. Ainsi, il a été proposé que la programmation de l'épigénome du spermatozoïde par remodelage des modifications d'histones héritées de la spermatide est vitale pour un potentiel développement optimal de l'embryon. Notamment, lors de la spermiogénèse, les gènes du développement perdent H3K4me2 et H3K4me3, mais ils conservent H3K27me3 (Teperek et al. 2016). Compte tenu du fait que la plupart des spermatozoïdes sont capables de supporter de développement embryonnaire, toute information épigénétique nécessaire au développement devrait donc être présente dans tous les spermatozoïdes. Pour vérifier si les histones modifiées retenues dans le spermatozoïde ont cet attribut, des études de ChIP-Seq quantitative pour les marques H3K4me3 et H3K27me3 ont été réalisées (Grzybowski, Chen, et Ruthenburg 2015). Ce type d'analyse a permis de déduire la proportion de spermatozoïde contenant une modification d'histone à un locus donné (Oikawa et al. 2020). Cette approche a permis de déterminer qu'à certains locus, H3K4me3 et H3K27me3 sont présents dans la plupart des spermatozoïdes. De plus, les auteurs ont exploré le devenir de ces marques dérivées du spermatozoïde dans l'embryon après réplication. Il a été ainsi démontré que H3K4me3 et H3K27me3 sont présents dans le spermatozoïde sur des gènes impliqués dans le développement. Après réplication, H3K4me3 est très fortement conservé alors que H3K27me3 subit de forts réarrangements (Oikawa et al. 2020).

Ainsi, il semble que, chez *Xenopus laevis* au moins, les histones modifiées ont les caractéristiques requises pour participer à la programmation épigénétique du spermatozoïde.

### *H2AK119Ub1 a un rôle central dans l'établissement du complexe polycomb dans les cellules souches embryonnaires.*

Le complexe Polycomb est un complexe protéique impliqué dans la modification de l'épigénome essentiel pour la répression des gènes impliqués dans le développement (Boyer et al. 2006) comme le cluster Hox, impliqué dans la survie de l'embryon (Kassis, Kennison, et Tamkun 2017). Ce complexe est important dans le développement embryonnaire (Faust et al. 1995)-(O'Carroll et al. 2001)-(Pasini et al. 2004). Deux groupes de protéines composent ce complexe, les complexes de répression Polycomb 1 et 2, nommés PRC1 et PRC2 (Blackledge et Klose 2021). Le complexe PRC2 tri-méthyle la lysine 27 de l'histone H3 (H3K27me3), tandis que le complexe PRC1 mono-ubiquitine la lysine 119 de l'histone H2A (H2AK119ub1). Il a été mis en évidence deux mécanismes de recrutement des complexes de répression polycomb qui varient selon l'ordre de recrutement des complexes PRC1 et PRC2 (Blackledge et al. 2020).

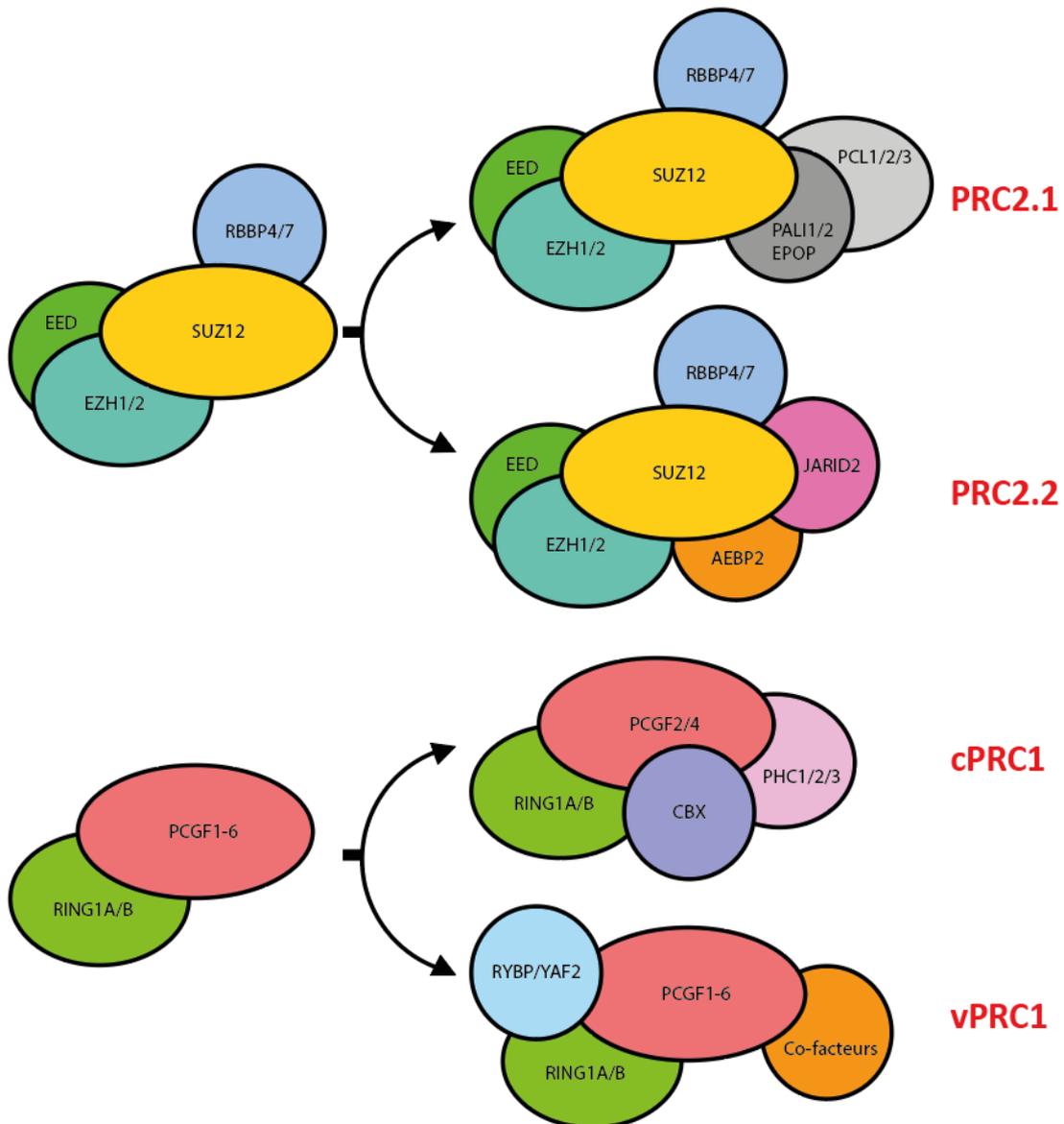
Le complexe PRC2 est composé de la protéine EZH2 (Enhancer of Zeste 2) ou EZH1, couplée à EED (Embryonic Ectoderm Development) permettant la déposition d'un groupement méthyle sur H3K27. Ces deux protéines sont liées à la protéine SUZ12 (Suppressor of Zeste 12), qui se lie elle-même à RBBP4 (Retinoblastoma-binding protein 4) ou RBBP7, et plusieurs d'autres cofacteurs de régulation spécifique à différents variants de PRC2. Le variant PRC2.1 est caractérisé par la liaison de SUZ12 aux cofacteurs PCL1, PCL2 ou PCL3 (Polycomb-like protein), qui reconnaissent des motifs CpG non méthylés, sur des séquences précises de l'ADN (Haojie Li et al. 2017). D'autres co-facteurs comme PALI1 (PRC2-Associated LCOR Isoform 1), PALI2 ou EPOP (Elongin BC and PRC2-associated Protein) sont recrutés pour l'établissement du complexe de répression, mais leur fonction est encore à préciser (Conway et al. 2018). Le deuxième variant est PRC2.2, qui reconnaît la marque H2AK119Ub1 et triméthyle H3K27. Ce variant correspond à la liaison de SUZ12 avec les cofacteurs JARID2 (Jumonji and AT-Rich Interaction Domain-containing protein 2) et AEBP2 (Adipocyte Enhancer-Binding Protein 2).

Le complexe PRC1 est composé de l'enzyme RING1B ou RING1A, et une des six protéines PCGF (Polycomb Group RING Finger). Il existe deux variants PRC1 selon les protéines

associées à celles-ci, canonical PRC1 (cPRC1) et variant PRC1 (vPRC1). Le variant cPRC1, le premier découvert (Shao et al. 1999), est associé aux protéines PCGF2 ou PCGF4, et une des cinq protéines CBX (Chromobox) reconnaissant H3K27me3, ainsi qu'à une sous-unité des trois protéines PHC (polyhomeotic) comme cofacteur à PCGF. La déposition de H2AK119Ub1 de ce variant est totalement dépendante de la déposition de H3K27me3 par PRC2. Au contraire, le variant vPRC1 est associé à une des six protéines PCGF, qui est couplé à différents cofacteurs selon la protéine PCGF associée. Par exemple, PCGF1-vPRC1 est associé à KDM2B qui contient un domaine CXXC de liaison à l'ADN qui reconnaît les îlots CpG par un domaine à doigt de Zinc. PCGF6-vPRC1 est associé des protéines reconnaissant des motifs ADN comme le motif E-Box ou T-Box. Le site catalytique RING1 est lui couplé à RYBP (YY1-binding protein) ou YAP2 (YY1 associated Factor 2) permettant la déposition de H2AK119Ub1 (**Figure 7**).

L'interaction entre la marque H2AK119Ub1 déposée par vPRC1 et les protéines JARID2 et AEBP2 de PRC2.2 a été récemment étudiée par cryo-microscopie électronique (Kasinath et al. 2021) démontrant que ces deux protéines régulent la déposition de H3K27me3 par PRC2.2 en présence de H2AK119Ub1. Des études ont montré que la perturbation de la déposition de H2AK119Ub1 par vPRC1 menait à des défauts de fixations de PRC2 et de déposition de H3K27me3 (Blackledge et al. 2020)-(Tamburri et al. 2020). Ils ont aussi mis en évidence que vPRC1 a la capacité de se lier directement à H2AK119Ub1 nouvellement déposé pour s'épandre, en plus de reconnaître un motif ADN par le domaine CXXC. Ce mécanisme de recrutement du complexe de répression Polycomb conserverait certains gènes du développement inactivés avant la ZGA : chez le zébrafish par exemple la déposition de H2AK119Ub1 par vPRC1 pré-ZGA conduit au recrutement post-ZGA de PRC2.2 via la liaison de AEBP2 (Hickey et al. 2022). Ainsi, H2AK119ub1 maintient la répression des gènes avant la ZGA indépendamment de H3K27me3. Chez la souris également, d'autres expériences d'histone ChIP-Seq pour différentes marques comme H2AK119Ub1, H3K4me3 ou encore H3K27me3 ont montré une déposition initiale de H2AK119Ub1 nécessaire pour la mise en place de H3K27me3 plus tardivement (Mei et al. 2021). D'autre part, des défauts de développement embryonnaire et d'activation des gènes à la ZGA sont observés suite à l'effacement de H2AK119Ub1 dans l'embryon par surexpression de la deubiquitylase PR-DUB (Z. Chen, Djekidel, et Zhang 2021). Tous ces résultats indiquent que la perte de H2AK119Ub1 entre la fécondation et la ZGA provoquerait des arrêts de développement, contrairement à l'effacement de H3K27me3. En effet, la délétion de la sous-unité EED maternelle cause une

perte de H3K27me3 et une perturbation de l'inactivation du chromosome X (Inoue et al. 2018), mais elle provoque un effet limité sur la déposition de H2AK119Ub1 (Z. Chen, Djekidel, et Zhang 2021).



**Figure 7 : Représentation des complexes de répression Polycomb**

Le complexe de répression Polycomb 2 (PRC2) contient deux variants nommés PRC2.1 et PRC2.2. Ces variants contiennent des protéines communes de liaisons SUZ12, RBBP4 ou RBBP7, l'enzyme EZH1 ou EZH2 et EED impliqués dans la déposition de H3K27me3. Le variant PRC2.1 se lie aux îlots CpG de l'ADN en les reconnaissant avec les protéines PCL et PALI/EPOP, au contraire du variant PRC2.2 qui se lie à H2AK119Ub1 avec AEBP2 et JARID2. Le complexe de répression Polycomb 1 (PRC1) contient aussi deux variants, cPRC1 et vPRC1, composé à minimum de l'enzyme RING1A/RING1B et PCGF. Le variant cPRC1 reconnaît H3K27me3 à l'aide du chromodomaine de CBX, contrairement aux cofacteurs de vPRC1 qui reconnaissent des motifs d'ADN ou H2AK119Ub1 nouvellement déposée comme site de fixation.

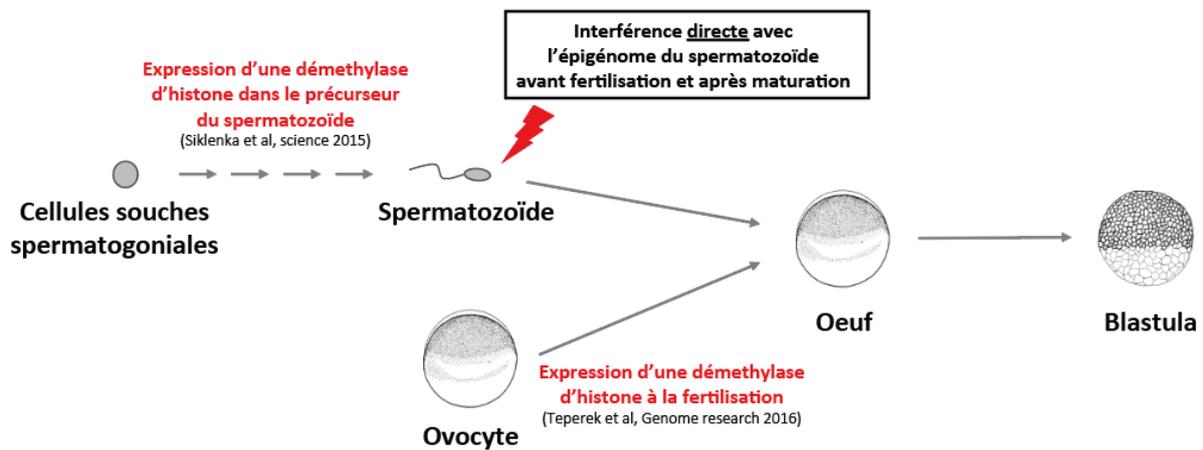
# CHAPITRE I- Le prétraitement du spermatozoïde par la deubiquitylase USP21 altère le développement embryonnaire

Lors des précédents travaux, la contribution paternelle du spermatozoïde n'a pas été étudiée dans la transmission de H2AK119Ub1 à l'embryon, où seule la contribution maternelle était prise en considération (Blackledge et al. 2020)-(Tamburri et al. 2020)-(Mei et al. 2021)-(Z. Chen, Djekidel, et Zhang 2021). La contribution épigénétique paternelle au développement de l'embryon a été prouvée comme importante précédemment, notamment avec la présence de H3K27me3 sur les gènes du développement, mais perdu très rapidement après la fécondation chez la souris (Xia et al. 2019). Pendant ma thèse, je me suis ainsi intéressé à la marque H2AK119Ub1 dans le spermatozoïde, et plus généralement à sa transmission à l'embryon chez *Xenopus laevis* en concomitance avec H3K4me3 et H3K27me3.

L'objectif de ma thèse est d'étudier la participation possible de H2AK119Ub1 dérivé du spermatozoïde dans la régulation du développement embryonnaire précoce. Cette étude repose sur des expériences en wet-lab (prise en charge par Florian Berger, Doctorant & Mami Oikawa, Postdoctorante) et des analyses bio-informatiques réalisées par moi-même. Dans cette section, je décris la stratégie expérimentale utilisée ainsi que les expériences, réalisées par mes collègues, qui constituent la base de cette étude et qui sont à l'origine des données de séquençage sur lesquelles j'ai travaillé.

**Stratégie expérimentale.** Des travaux récents suggèrent un lien entre l'épigénome du spermatozoïde et le développement embryonnaire précoce. Par exemple, chez la souris la déméthylation de H3K4 par expression de l'histone déméthylase KDM1 au cours de la spermiogénèse montre un effet sur la descendance, avec des défauts de développement et de survie (Siklenka et al. 2015). Une autre étude, chez le xénope, a utilisé l'expression des déméthylases KDM5B (ciblant H3K4me2/me3) ou KDM6B (ciblant H3K27me3) par injection de mRNA codant pour ces enzymes dans l'ovocyte, le précurseur immédiat de l'œuf. Il a été observé une dérégulation de l'expression des gènes dans les embryons issus des œufs exprimant ces déméthylases. De façon notable, ces dérégulations de l'expression des gènes embryonnaires par les déméthylases varient en fonction du type de cellule paternelle utilisé (dans ce cas

spermatozoïde ou spermatide) suggérant que l'effet des déméthylases dépend en partie de leurs actions sur l'épigénome paternel (Teperek et al. 2016). Dans ces deux études, cependant, l'interférence avec l'épigénome du spermatozoïde n'est pas directe (**Figure 8**). Par exemple dans l'étude murine, KDM1 est exprimé assez tôt au cours de la spermatogenèse et n'est donc pas strictement limité au spermatozoïde (Siklenka et al. 2015). De la même façon, l'étude chez le xénope repose sur une expression de la déméthylase qui affecte également l'épigénome maternelle (Teperek et al. 2016).

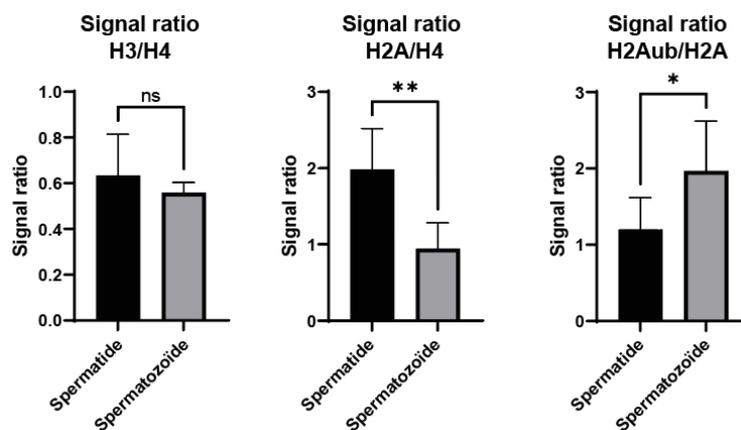


**Figure 8 : Les stratégies d'interférence avec l'épigénome du spermatozoïde**

Les travaux précédents reposent sur une interférence de l'épigénome du spermatozoïde avec des enzymes modifiant la chromatine exprimée soit pendant la spermatogénèse (Siklenka et al 2015, chez la souris), soit dans l'œuf avant fécondation (Teperek et al 2016, chez le xénope). L'approche utilisée pour ce travail de thèse vise à modifier directement l'épigénome du spermatozoïde après complétion de la spermatogenèse et avant l'introduction dans l'œuf.

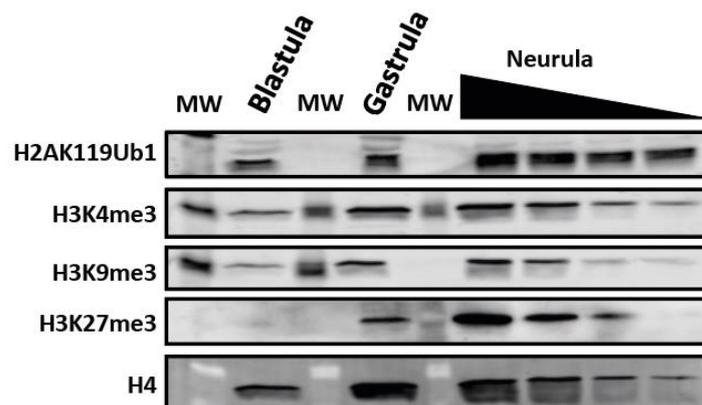
**Choix de H2AK119Ub1 comme marque épigénétique d'étude.** Pour tester l'hypothèse selon laquelle la présence de ces modifications d'histones dans le spermatozoïde serait importante pour supporter le développement de l'embryon, nous nous sommes concentrés sur H2AK119Ub1 et avons établi un protocole pour effacer cette marque directement de la chromatine du spermatozoïde, juste avant son introduction dans l'œuf. Le choix s'est porté sur cette modification, car elle occupe une place centrale dans l'établissement des processus de répression par les complexes PRC (voir introduction). Additionnellement, mes collègues ont pu mettre en évidence que chez le xénope, cette marque portée par H2A est retenue dans le spermatozoïde au cours de la phase de remodelage de la chromatine associée à la transition spermatide-spermatozoïde. En effet, lors de la différenciation de la spermatide en spermatozoïde, il y a une perte globale de H2A. Nous observons ainsi une diminution du ratio

H2A/H4 dans le spermatozoïde comparé à la spermatide (**Figure 9**) (Oikawa et al. 2020), mais dans ce contexte H2AK119Ub1 est retenue avec un signal H2AK119Ub1/H2A dans le spermatozoïde comparable à celui détecté dans la spermatide (**Figure 9**). Enfin, cette marque est détectée dans l'embryon précoce de xénope, avant la ZGA (**Figure 10**) alors qu'une autre marque associée au complexe polycomb, H3K27me3, ne l'est pas. Dans l'ensemble, ceci confère à H2AK119Ub1 toutes les caractéristiques d'une marque épigénétique avec le potentiel d'influencer le développement embryonnaire : rétention dans le spermatozoïde, présence dans l'embryon avant la ZGA, rôle clef dans l'établissement du complexe de répression polycomb.



**Figure 9 : Rétention de H2AK119Ub1 lors de la spermiogénèse**

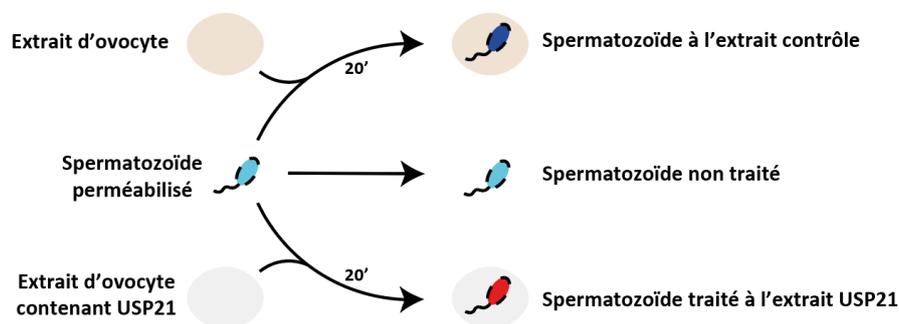
Ratio du signal de l'histone H3 et H2A par rapport à l'histone H4, et du signal H2AK119Ub1 par rapport au signal H2A dans les spermatides et les spermatozoïdes. \*, p-value < 0.05; \*\*, p-value < 0.01 Pearson's Chi-squared test.



**Figure 10 : H2AK119Ub1 est présent aux stades embryonnaires précédant la ZGA**

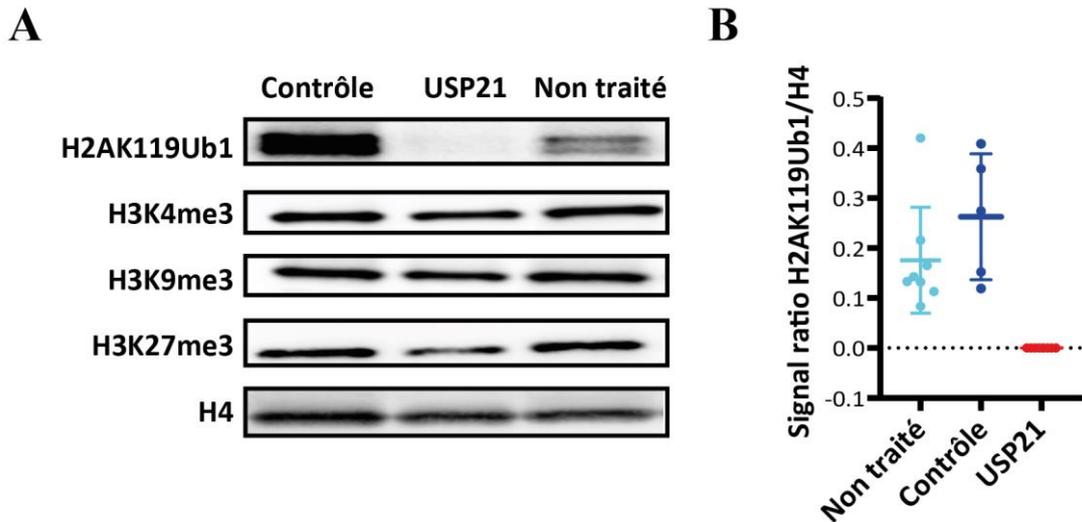
Western blot des marques H2AK119Ub1, H3K4me3, H3K9me3, H3K27me3 sur des noyaux de cellules issues d'embryons au stade Blastula précoce, Gastrula et Neurula. Une révélation de l'histone H4 est utilisée comme témoin de charge des noyaux. La ZGA est au stade de blastula tardive chez le xénope. On observe que l'ensemble des modifications évaluées, à l'exception de H3K27me3, sont détectées à tous les stades testés.

**Effacement de H2AK119Ub1 du spermatozoïde.** Pour obtenir un spermatozoïde dépourvu de H2AK119Ub1, nous avons mis en place un protocole où le spermatozoïde est incubé pendant 20 minutes dans un extrait d'ovocyte contenant une H2A deubiquitylase (**Figure 11**). L'extrait d'ovocyte décompacte partiellement la chromatine compactée du spermatozoïde favorisant ainsi l'accès de l'enzyme USP21 à sa cible au sein de la chromatine. Dans ces conditions, on observe un effacement de la marque H2AK119Ub1 du spermatozoïde par l'extrait USP21 en comparaison à l'extrait « contrôle » ou à l'absence de traitement. Le traitement n'affecte pas d'autres modifications d'histones telles que H3K4me3 ou H3K27me3 (**Figure 12**). Nous avons également vérifié que le traitement à l'extrait d'ovocyte n'affecte pas la structure globale de la chromatine du spermatozoïde. Le remodelage des histones chez *Xenopus laevis* conduit à une structure très caractéristique de la chromatine qui peut être révélée par des analyses de la taille des fragments d'ADNs obtenue après digestion avec une nucléase micrococcale (MNase) (Oikawa et al. 2020)-(Teperek et al. 2016). En effet, la digestion MNase génère une taille de 150pb qui représente l'empreinte caractéristique des nucléosomes qui enveloppent l'ADN dans la chromatine des cellules somatiques. En revanche, la digestion MNase génère des tailles de 150, 110 et 70pb avec la chromatine du spermatozoïde de *Xenopus Laevis*, correspondant à des nucléosomes (150pb) mais également à des sous-nucléosomes (110 et 70pb) issus de la perte de H2A/H2B lors de la spermiogénèse (**Figure 4**). Le traitement à l'extrait d'ovocyte n'a pas d'effet global sur la structure de la chromatine puisqu'un profil similaire à celui du spermatozoïde est maintenu après traitement à l'extrait d'ovocyte (**Figure 13**). Ainsi, le traitement à l'extrait d'ovocyte contenant USP21 permet d'effacer H2AK119Ub1 du spermatozoïde sans affecter la structure globale de la chromatine.



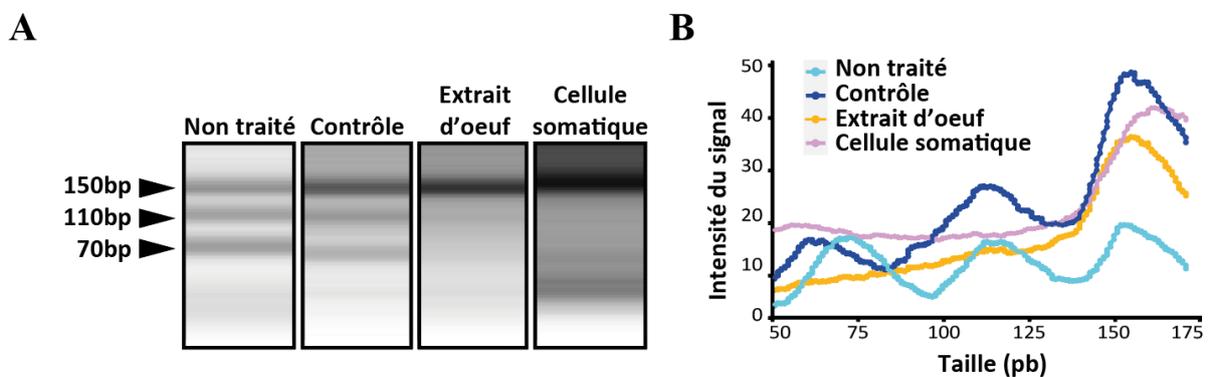
**Figure 11 : Protocole expérimental du traitement USP21 du spermatozoïde**

Des extraits sont produits à partir d'ovocytes 72h après l'injection d'ARNm codant pour la deubiquitylase USP21 ou non. Les spermatozoïdes sont perméabilisés pour permettre à l'extrait d'ovocyte contenant l'enzyme USP21 de pénétrer la membrane. Après 20 minutes d'incubation dans l'extrait contrôle ou couplé à USP21, les spermatozoïdes sont lavés de l'extrait puis utilisés pour les analyses en aval (WB, ChIP, ICSI).



**Figure 12 : Le traitement avec un extrait d'ovocyte contenant la deubiquitylase USP21 efface spécifiquement H2AK119Ub1 du spermatozoïde**

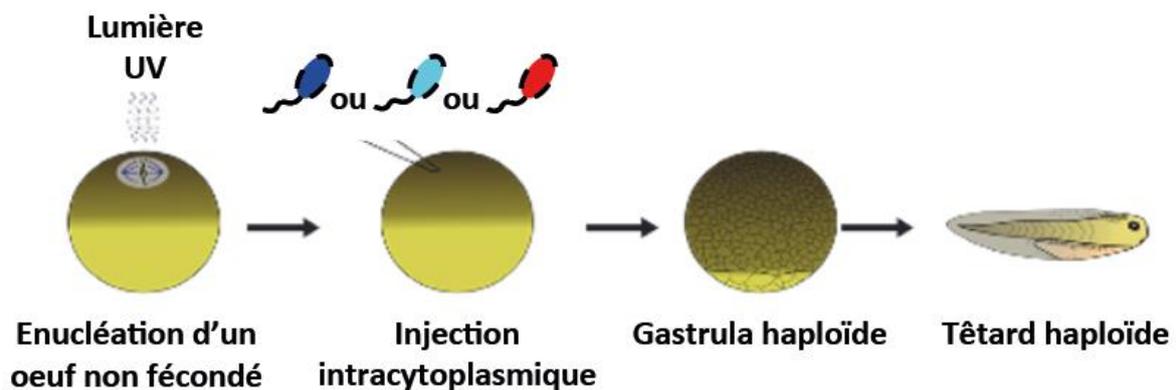
**(A)** Western-blot des modifications d'histones H2AK119Ub1, H3K4me3, H3K9me3, H3K27me3 dans le spermatozoïde traité par l'extrait contrôle, traité par l'extrait couplé à USP21, et non traité. Une révélation de l'histone H4 est utilisée comme témoin de charge. H2AK119Ub1 est efficacement enlevé du spermatozoïde lors du traitement par USP21 alors que les modifications H3K4me3, H3K29me3, et H3K27me3 restent inchangées. **(B)** Quantification du signal H2AK119Ub1/H4 dans les spermatozoïdes non traités, traités par l'extrait contrôle ou traités par l'extrait couplé à UP21.



**Figure 13 : Le traitement avec un extrait d'ovocyte ne perturbe pas la structure globale de la chromatine du spermatozoïde**

**(A)** Analyse par électrophorèse capillaire des fragments d'ADNs produits par digestion MNase du spermatozoïde non traité, traité par l'extrait d'ovocyte contrôle, ou traité par l'extrait œuf. Une digestion MNase de cellule somatique (HEK93) est également incluse comme contrôle. L'extrait d'ovocyte, non répliquant, permet le maintien de la structure de chromatine typique du spermatozoïde (présence de fragments d'ADNs à 150, 110 et 70pb) alors que l'extrait d'œuf, répliquant, entraîne un remodelage de la chromatine vers une structure de cellule somatique (fragment d'ADN d'un nucléosome, 150pb). **(B)** Intensité du signal des différentes tailles du spermatozoïde non traité (cyan), traité avec l'extrait contrôle (bleu), traité avec l'extrait d'œuf répliquant (orange), et d'une cellule somatique (rose).

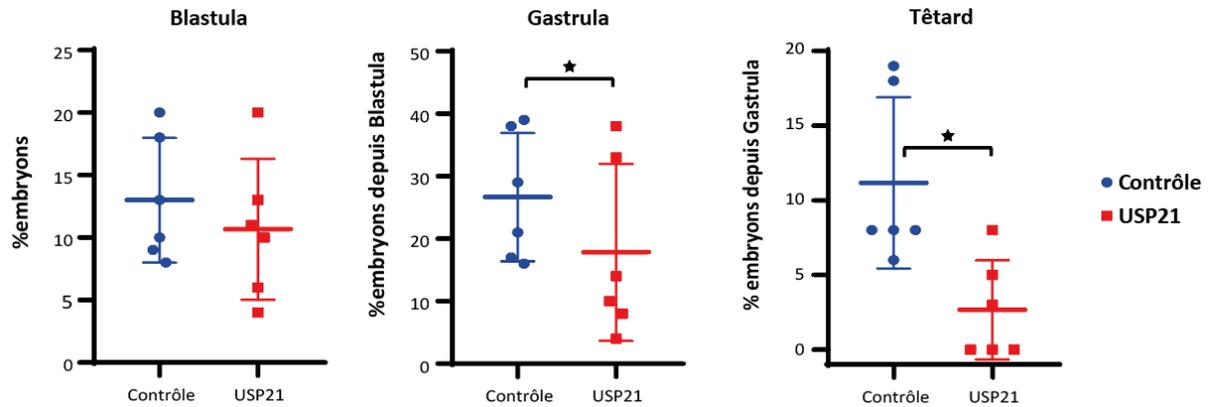
**Les embryons dérivés d'un spermatozoïde traité à USP21 ont un potentiel développemental réduit.** Nous avons ensuite évalué l'impact de la déplétion de H2AK119Ub1 du spermatozoïde sur le développement des embryons. Après énucléation de l'œuf par exposition aux UV pour éliminer la partie épigénétique maternelle, le spermatozoïde traité à USP21 ou non traité est injecté par transfert intracytoplasmique. L'embryon se développe ensuite jusqu'au stade de têtard haploïde avec la copie paternelle comme seule copie du génome pour supporter le développement (**Figure 14**).



**Figure 14 : Modèle expérimental pour la génération d'embryons haploïdes à partir de spermatozoïdes épigénétiquement édités**

*L'œuf est énucléé pour éliminer la contribution maternelle au développement de l'embryon. Le spermatozoïde (non traité = cyan, contrôle = bleu, traité à l'extrait USP21 = rouge) est ensuite injecté dans le cytoplasme de l'œuf énucléé (ICSI) permettant le suivi du développement de l'embryon jusqu'aux stades gastrula puis têtard.*

On observe que les embryons issus d'un spermatozoïde traité par USP21 présentent des défauts de développement à partir du stade gastrula, qui sont encore plus visibles au stade de têtard (**Figure 15**). Nous concluons que le traitement par USP21 altère le potentiel développemental du spermatozoïde. Sur les bases de ces observations, mon travail a consisté à mettre en évidence et à caractériser les défauts d'expressions géniques embryonnaires découlant du traitement USP21 du spermatozoïde (**CHAPITRE II**), de caractériser la distribution de H2AK119Ub1 et son association avec H3K4me3/H3K27me3 dans le spermatozoïde (**CHAPITRE III**) et enfin d'identifier de possibles liens entre la configuration épigénétique du spermatozoïde et la sensibilité des gènes au traitement USP21 au cours de la spermiogénèse et du développement embryonnaire (**CHAPITRE IV**).



**Figure 15 : Les embryons issus de spermatozoïdes traités par l'extrait USP21 présentent des défauts de développement à partir du stade gastrula.**

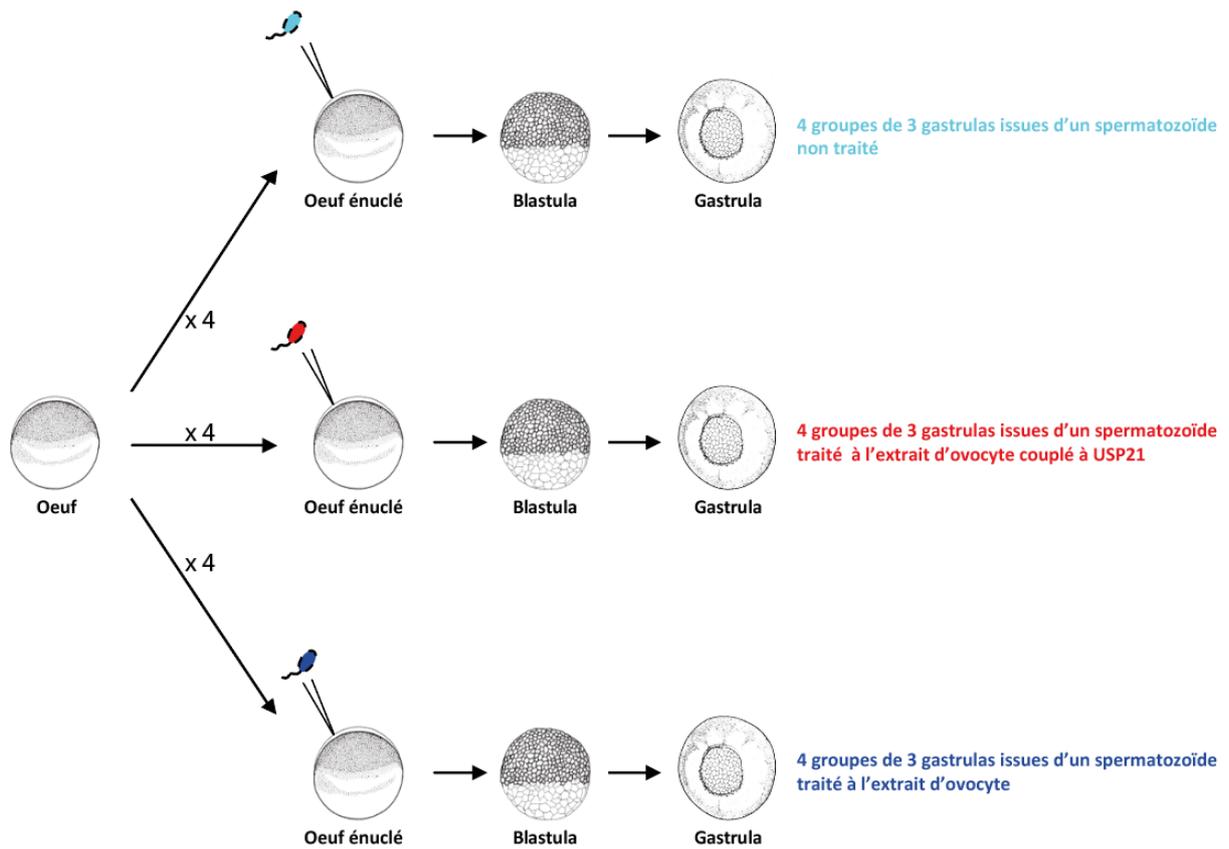
Graphes indiquant le pourcentage d'embryons progressant entre chaque stade de développement (Blastula : de l'ICSI jusqu'à blastula ; Gastrula : de blastula à gastrula ; Têtard : de gastrula à têtard). Les embryons issus d'un spermatozoïde traité à l'extrait « contrôle » ou à l'extrait « USP21 » sont représentés en bleu et en rouge, respectivement. N=6 expériences utilisant 6 sources d'œufs, trois sources de spermatozoïdes, et 3 sources d'extraits. \*, p-value < 0.05, Chi-squared test.

# Chapitre II - Le traitement des spermatozoïdes par la deubiquitylase USP21 entraîne des dérégulations de l'expression des gènes dans la Gastrula

Pour rechercher les causes de ces défauts de développement, nous avons réalisé des expériences de RNA-Sequencing des embryons au stade Gastrula, le premier stade où les défauts de développement apparaissent. L'objectif est d'évaluer si les défauts de développement sont accompagnés par des anomalies d'expression des gènes.

Le protocole repose sur l'utilisation de trois sources différentes de matériels avec une variabilité inhérente : les préparations de spermatozoïdes purifiés, les préparations d'extraits d'ovocyte pour le traitement, et les œufs collectés pour générer les embryons. Nous avons séquencé différents embryons au stade gastrula issues de spermatozoïdes non traités (**non traité**), traité avec l'extrait d'ovocyte seul (**contrôle**), ou traité avec USP21 couplé à l'extrait d'ovocyte (**USP21**). Pour chacune des conditions précédentes, nous avons réalisé plusieurs réplicats biologiques où l'extrait d'ovocyte, la source de spermatozoïde et la source d'œuf étaient différents (au total 3 préparations de spermatozoïdes, 3 préparations d'extraits d'ovocyte, 4 sources d'œuf) (**Figure 16**). Le jeu de données est constitué de pools d'embryons regroupés (3/groupe) par condition en quatre réplicats pour analyse en RNA-Sequencing.

Dans les sections suivantes, je détaillerai d'abord les outils sélectionnés pour l'analyse bio-informatique des données de séquençage d'ARNm, puis je détaillerai par la suite les questions biologiques adressées grâce à ces outils.



**Figure 16 : Schéma expérimental du RNA-Sequencing dans les gastrulas**

Trois conditions sont testées dans le modèle. La première condition (cyan) représente les gastrulas obtenues à partir d'un spermatozoïde non traité, la deuxième condition (bleue) les gastrulas obtenues depuis un spermatozoïde traité par l'extrait USP21, et la dernière condition (rouge) illustre les gastrulas issues d'un spermatozoïde traité avec l'extrait d'ovocyte seul. Chaque échantillon est constitué d'un pool de 3 gastrulas. Quatre réplicats biologiques (préparation indépendante de spermatozoïdes, extrait d'ovocyte, et d'œuf).

## Outils utilisés pour l'analyse du transcriptome par RNA-Sequencing

Après purifications des ARNs totaux (kit RNAeasy, QIAGEN), les bibliothèques d'ADNs complémentaires sont générées en réalisant une étape de reverse transcription qui sélectionne les ARNs 3' polyadénylés (mRNAs). Les séquences ADN complémentaires sont ensuite fragmentées, et les fragments sont marqués avec des adaptateurs qui contiennent un index nucléotidique (8pb) spécifique à chaque échantillon et également une séquence commune. On réalise ensuite une amplification PCR des fragments d'ADNs en utilisant des primers qui cible la partie commune des adaptateurs pour ajouter une séquence supplémentaire à l'extrémité des fragments d'ADNs. Cette dernière séquence est utilisée comme amorce pour former le pont de PCR utilisé lors du séquençage. Les données ont été séquencées sur un NovaSeq 6000 (plateforme bird), en mode 50pb en single-end.

### *Du séquenceur à la matrice d'expression des gènes : formatage des données*

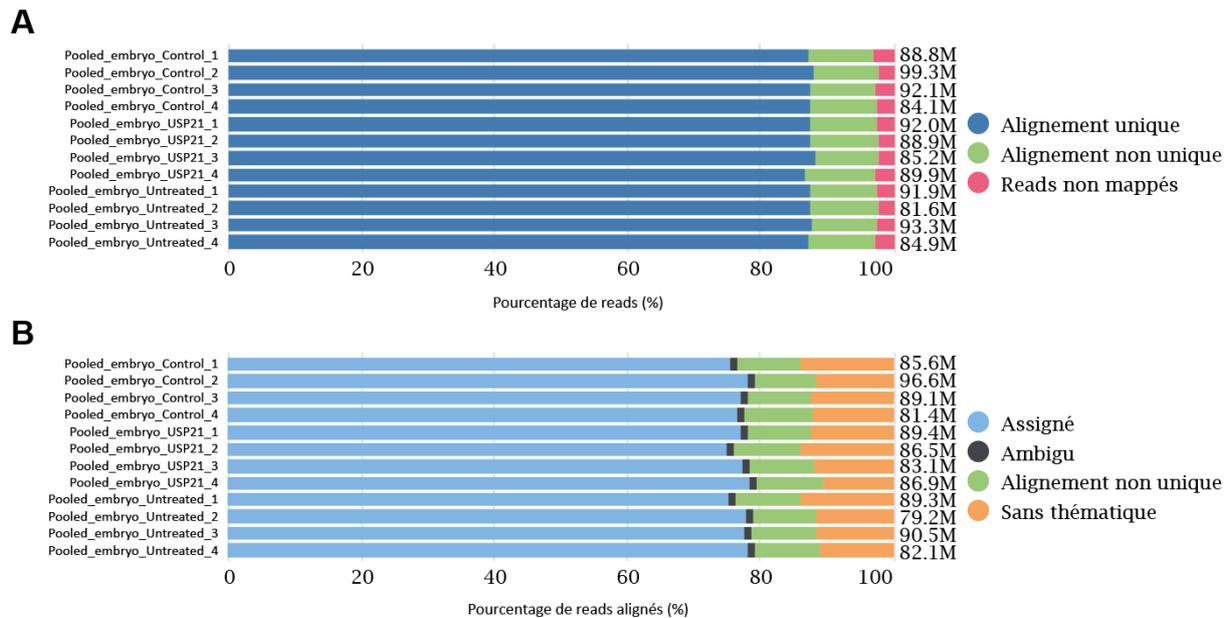
L'analyse bio-informatique de données de RNA-Sequencing débute après le séquençage par l'évaluation des séquences d'ADNs obtenues, appelées reads. L'ensemble des étapes et outils utilisés pour la production de la matrice d'expression des gènes est résumé en figure 19.

#### Fastq, le fichier brut issu du séquenceur

À l'issue du séquençage, nous devons démultiplexer les reads pour identifier l'échantillon duquel il provient à partir des index (étape réalisée par la plateforme de séquençage). Nous obtenons ainsi un fichier d'extension « fastq.gz » par échantillon. Un read est caractérisé dans le fichier « fastq » par un format standard de quatre lignes. La première ligne identifie le read, la deuxième décrit sa séquence nucléotidique, et la dernière ligne indique le score de qualité de chaque nucléotide lors du séquençage (**Figure 17**). L'identifiant utilisé pour décrire le score de qualité d'une base est un codage ASCII en base 33 du score de qualité calculé par le séquenceur. Par exemple, les lettres A, F et J correspondent à une probabilité d'erreur de séquençage de la base de respectivement 0,00063, 0,00020 et 0,00008. La troisième ligne contient le symbole « + » (parfois suivi de l'identifiant de la séquence), pour séparer la séquence nucléotidique de la séquence indiquant le score.



de séquences apparues au cours du temps, permettant ainsi d'associer les reads préférentiellement à la version S ou L du génome. Nous observons un pourcentage d'alignement sur le génome de *Xenopus laevis* entre 97 à 98% selon les échantillons. Environ ~10% des reads totaux ne peuvent être assignés à une seule séquence du génome (alignement non unique), représentant une possible confusion entre les versions S et L, ou pouvant correspondre à un mapping sur les régions répétées du génome telles que les régions satellites (**Figure 18**).



**Figure 18 : Contrôle-qualité des échantillons RNA-Seq**

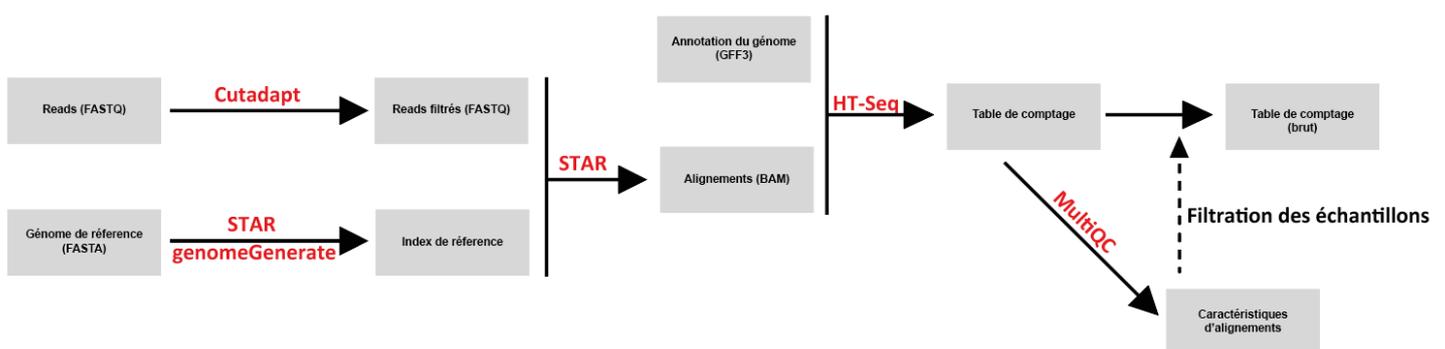
**(A)** Bar graph indiquant le pourcentage de reads à alignement unique (bleu), non unique (vert) ou non alignés (rose) sur le génome 9.2 de *Xenopus laevis*. **(B)** Bar graph indiquant le pourcentage de reads mappés assignés à un exon (assigné, bleu), assignés à plusieurs exons (ambigu, noir), assignés à plusieurs localisations (non unique, vert), assignés à une localisation unique non exonique (sans thématique, orange). Le nombre de reads total après filtration des reads de faible qualité par cutadapt est affiché sur la droite (M, millions).

## Du BAM à la matrice de comptage

La matrice de comptage est produite en déterminant le nombre de reads associés à un élément génomique. Les deux principaux éléments génomiques utilisés en RNA-Seq sont les exons et les gènes (région comprise entre les 5' UTR et 3' UTR). Les coordonnées de chaque exon et de chaque gène sont décrites pour le génome de référence dans un fichier d'extension GTF ou GFF3. J'ai utilisé l'outil HTSeq-Count pour compter pour chaque exon le nombre de reads associés en excluant ceux comportant un overlap intronique (Anders, Pyl, et Huber 2015). Si un read est associé à plusieurs exons, l'exon avec le score le plus élevé est conservé. Si le score

le plus élevé est associé à plusieurs exons, le read est décrit comme ambigu et n'est pas conservé dans la matrice de comptage. Au final, les reads non alignés au génome, les reads avec plusieurs alignements possibles sur le génome de référence à un score d'alignement identique (alignements non uniques), les reads chevauchant entre plusieurs exons (alignements ambigus) et les reads de faible qualité ne sont pas utilisés pour la matrice de comptage. Les valeurs d'expression exoniques sont ensuite additionnées pour un même gène afin d'obtenir une matrice d'expression génique des exons pour chaque échantillon.

L'outil MultiQC permet de contrôler la qualité des reads depuis leur filtration jusqu'à la table de comptage pour chaque échantillon (**Figure 18**) (Ewels et al. 2016). Pour chaque étape, nous avons le nombre de reads filtré, puis le nombre de reads alignés au génome, et enfin le nombre de reads que nous avons pu assigner à un exon : tout ce processus est appelé l'analyse primaire (**Figure 19**). Nous observons ~75% des reads initiaux assignés à un exon pour tous nos échantillons, avec environ 10% d'alignements non unique, ~1% d'alignements ambigus, 2% de reads non assignés au génome et 10 à 12% d'alignements associés à aucun exon. Ces reads associés à aucun exon peuvent être localisés sur des gènes qui ne sont pas encore référencés dans la version 9.2 du génome. En effet l'assemblage en chromosome ainsi que les annotations des gènes évoluent encore significativement (version 10.1 actuellement), indiquant que la version utilisée dans notre étude est probablement suboptimale pour la représentation des exons. Ces résultats suggèrent que le pipeline d'alignement rencontre peu d'ambiguïté entre les copies L et S, et me permet d'obtenir une table de comptage pour la suite des analyses.



**Figure 19 : Pipeline d'alignement des données RNA-Seq**

*Les reads sont filtrés par cutadapt puis alignés par STAR sur le génome indexé par STAR genomeGenerate. Les reads sont attribués aux exons par HT-Seq, puis la qualité de l'analyse est évaluée par MultiQC.*

## *De la matrice d'expression des gènes à l'identification des gènes différemment exprimés*

L'objectif de cette étape est d'identifier des gènes qui seraient exprimés préférentiellement dans une condition. La matrice d'expression génique nécessite un pré-traitement avant son utilisation pour une analyse de gènes différemment exprimés. Les reads assignés à un élément génomique seront appelés counts.

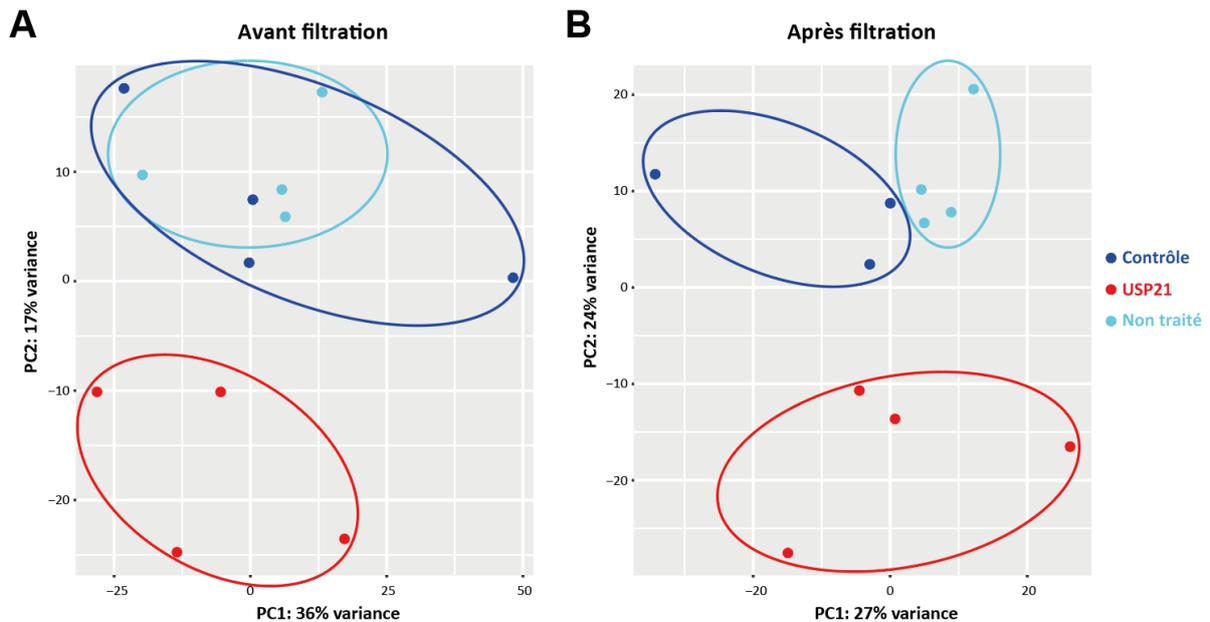
### **Pré-traitement des données**

#### *Filtration*

La filtration consiste à exclure les échantillons qui ne contiendraient pas un certain quota minimum de count. Concernant le RNA-Seq, un seuil de 30M est indiqué par les recommandations ENCODE. Les échantillons suivent largement ces recommandations, avec un total de reads par échantillon compris entre 80M à 90M pour les échantillons provenant de groupe d'embryons. La deuxième étape de filtration consiste à éliminer de l'analyse les gènes qui ne seraient pas détectés dans les échantillons. Ces gènes ne pourront pas être identifiés comme différents entre les conditions, et n'apporteront pas de message biologique, c'est pourquoi je les enlève pour réduire la complexité mathématique du modèle et le temps de calcul. Le seuil utilisé pour l'inclusion d'un gène est la présence d'au moins deux reads dans une condition biologique au sein de l'ensemble des échantillons. Nous conservons ainsi 32 610 gènes sur les 45 093 gènes référencés chez *Xenopus laevis*.

#### *La gestion des échantillons aberrants*

Une représentation en ACP (Analyse en composante principale) des échantillons permet de mettre en évidence des échantillons qui se regroupent selon leur profil transcriptomique. Un échantillon aberrant sera visible sur une représentation à plusieurs dimensions comme l'ACP, où un échantillon peut représenter une très forte variabilité du jeu de données jusqu'à biaiser le message biologique. Lors des préanalyses, un des réplicats contrôle a été identifié comme un échantillon aberrant, contribuant à 17% de la variabilité du jeu de données (axe PC2). L'inertie interclasse entre les différentes conditions est plus grande après déletion de l'échantillon anormal, ce qui signifie que nous différencions mieux les conditions après avoir enlevé cet échantillon (**Figure 20**).



**Figure 20 : Analyse de la variabilité du jeu de données RNA-Seq par ACP**

**(A)** Analyse bidimensionnelle en PCA du jeu de données RNA-Seq avant filtration d'un échantillon contrôle. Les axes PC1 et PC2 représentent les deux axes de variables contribuant le plus à la variabilité du jeu de données. L'inertie intraclasse représentée pour chaque condition (Contrôle = bleu, USP21 = rouge, non traité = cyan) permet de différencier les échantillons traités par USP21 des autres, mais un échantillon traité avec l'extrait « contrôle » représente une partie majeure de la variabilité de l'axe PC1 (36% de la variabilité du jeu de données).

**(B)** Analyse bidimensionnelle en PCA du jeu de données RNA-Seq après filtration d'un échantillon contrôle. Les axes PC1 et PC2 représentent les deux axes de variables contribuant le plus à la variabilité du jeu de données. L'inertie intraclasse représentée pour chaque condition (Contrôle = bleu, USP21 = rouge, non traité = cyan) permet de différencier les échantillons traités par USP21 des autres, notamment par l'axe PC2 (24% de la variabilité du jeu de données).

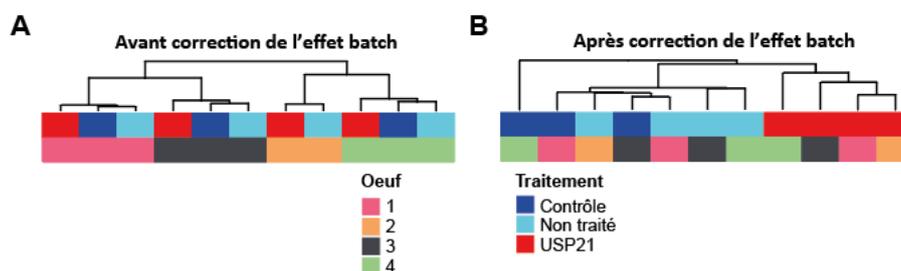
#### La normalisation

La normalisation est un facteur vital pour permettre la comparaison des échantillons. Par exemple, pour un même nombre d'ARNs produits, un gène avec une taille d'exons plus grande aura un nombre absolu de reads plus élevé que celui d'un gène avec des exons de plus petite taille. La différence de profondeur de séquençage entre les échantillons affecte également la valeur absolue de count pour chaque gène. Pour corriger ces événements, une normalisation est appliquée afin de comparer les échantillons à un total de counts équivalents, et de prendre en compte les différentes tailles de gènes. Il existe plusieurs méthodes de normalisation, comme les CPM (Count Per Million mapped reads) qui sont une normalisation des counts en ramenant le nombre de reads totaux à 1 million pour tous les échantillons. Ceci permet par exemple de

visualiser la différence d'expression pour un même gène entre deux conditions, mais ne permet pas une analyse de gènes différentiellement exprimés entre ces conditions. Les RPKM (Reads Per Kilobase of transcript per Million mapped reads) représentent une normalisation CPM en prenant en compte la taille des transcrits. Enfin, les packages comme DESeq2 (Love, Huber, et Anders 2014) ou edgeR (Robinson, McCarthy, et Smyth 2010) permettent de réaliser une normalisation supplémentaire à partir de la matrice de comptage brute en intégrant le fait que l'expression génique suit une loi négative binomiale. Cette normalisation est spécifique et obligatoire (Evans, Hardin, et Stoebel 2018) à l'analyse RNA-Seq permettant de mettre en avant des gènes différentiellement exprimés.

### *L'effet batch*

Lors de l'expérience, certains paramètres expérimentaux peuvent inclure une variabilité indépendante de l'effet biologique observé. Par exemple, dans notre cas, les résultats biologiques peuvent varier selon la source d'œuf, de spermatozoïde ou d'extrait d'ovocyte utilisée. La correction de l'effet batch a donc pour objectif de prendre en compte ces variations pour les corriger. Nous pouvons ainsi compenser ces variations biologiques pour permettre d'évaluer les variations associées aux conditions expérimentales étudiées, dans notre cas l'effet du traitement USP21. Ceci est réalisé à l'aide d'outils comme la fonction `removeBatchEffect` du package `limma` (Ritchie et al. 2015). Après correction de l'effet batch, on observe que la corrélation de Pearson ne regroupe plus les échantillons par l'œuf utilisé, mais par le traitement (Figure 21).



**Figure 21 : La suppression de l'effet batch**

**(A)** Dendrogramme du clustering hiérarchique non supervisé des données RNA-Seq sans la correction de l'effet batch (l'œuf). Les échantillons sont regroupés en fonction de la source d'œuf utilisée (rose = source 1, orange = source 2, noir = source 3, vert clair = source 4).

**(B)** Dendrogramme du clustering hiérarchique non supervisé des données RNA-Seq après la correction de l'effet batch (l'œuf). Les échantillons sont regroupés en fonction du traitement du spermatozoïde (contrôle = bleu, non traité = cyan, USP21 = rouge).

## L'analyse de gènes différentiellement exprimés

Une fois ces étapes complètes, le package DESeq2 permet de comparer l'expression de chaque gène entre deux conditions avec les fonctions DESeq(), counts() et results() sur R en prenant en compte le modèle expérimental (batch). Cette analyse permet de générer deux valeurs : le foldchange et la FDR. Le foldchange est calculé par la formule suivante pour chaque gène :

$$\text{FoldChange} = \frac{\text{moyenne(Expression gene X)}_{\text{condition A}}}{\text{moyenne(Expression gene X)}_{\text{condition B}}}$$

Un foldchange de 1 signifie que le gène a la même expression moyenne entre les deux conditions. Le logarithme du foldchange est utilisé pour la représentation. Un logarithme du foldchange négatif signifie que le gène est plus exprimé dans la condition B. Un foldchange positif indique que le gène est plus exprimé dans la condition A. Néanmoins, le foldchange n'est pas une valeur statistique, mesurée par une p-value. La p-value représente la probabilité que ce qu'on observe soit le fait du hasard. C'est ce qu'on appelle l'hypothèse nulle H0 qui représente le cas où on ne détecte pas de différences statistiques significatives entre les deux groupes, au contraire de l'hypothèse alternative notée H1. Chaque gène est ainsi testé pour évaluer si les différences observées entre les conditions sont statistiquement significatives. Additionnellement, la p-value nécessite une correction. Lors d'un test statistique, nous sélectionnons un seuil  $\alpha$  qui détermine la p-value maximale à atteindre pour rejeter H0. Si ce seuil est de 1%, sachant qu'à chaque test, il y aura 1% de risque d'avoir un faux positif, nous aurons un faux positif tous les cents tests soit cents gènes. Pour résoudre ce souci de répétition des tests, il existe plusieurs méthodes.

La méthode de Bonferroni est la plus simple, mais aussi la plus conservatrice, elle peut donc contenir encore plusieurs faux négatifs. Elle consiste à diviser le seuil  $\alpha$  par le nombre N de tests à réaliser.

$$\alpha_{\text{Bonferroni}} = \frac{\alpha}{N}$$

La FDR est la correction de Benjamini-Hochberg. Elle consiste à prendre en compte le nombre de tests N réalisés comme Bonferroni, mais tout en tenant compte du rang  $k$  de p-value du gène.

Cette méthode est moins conservatrice que la p-value de Bonferroni, plus un gène aura un rang proche de 1, représentant le gène le plus significatif, plus la p-value  $p$  sera pénalisée.

$$\text{FDR} = \frac{p * N}{k}$$

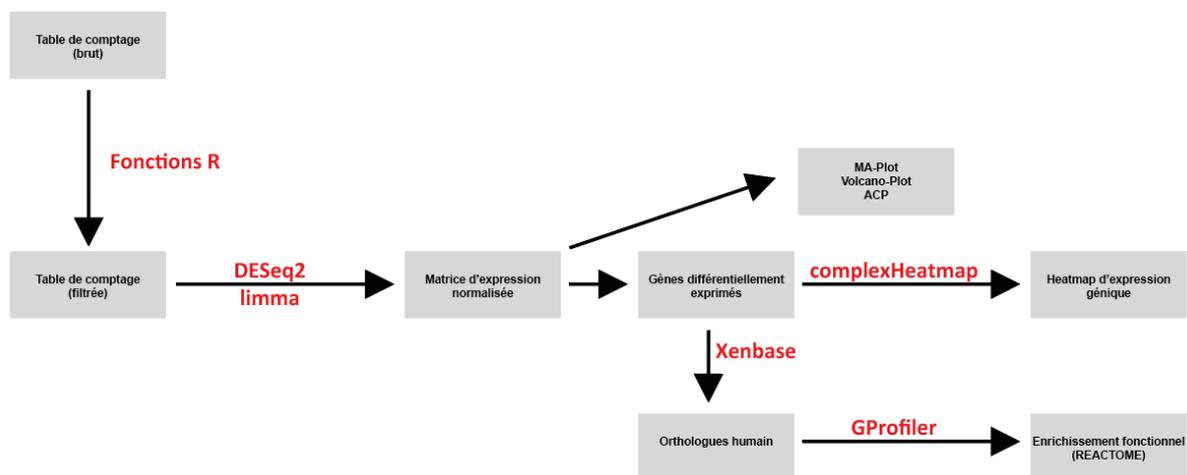
Pour représenter les gènes différentiellement exprimés entre les deux conditions, plusieurs données et graphiques sont utilisés. Le MA-Plot est une représentation du logarithme du foldchange en fonction du nombre de counts dans les échantillons pour chaque gène. Le volcano-plot indique la FDR en fonction du logarithme du foldchange. Enfin, les heatmaps (Gu, Eils, et Schlesner 2016) d'expression de gènes permettent de montrer l'expression génique d'un groupe de gènes prédéfinis dans plusieurs échantillons. Les données d'expression géniques sont centrées et réduites pour la heatmap, nommé aussi z-score. Cette modification mathématique consiste à modifier les données d'expression de façon à ce que pour l'ensemble des conditions évaluées, la moyenne soit égale à 0, et que l'écart-type soit de 1, ce qui permet de représenter les tendances d'expression relative de l'ensemble des gènes au sein d'une échelle identique.

### L'enrichissement fonctionnel

L'enrichissement fonctionnel consiste à déterminer si un groupe de gènes est associé à une activité biologique particulière. Plusieurs gènes d'une même voie biologique peuvent avoir leur expression perturbée dans une condition par rapport à une autre. Les bases de données disponibles ne permettent pas de réaliser directement une analyse d'enrichissement fonctionnel pour le modèle *Xenopus laevis* car il n'existe pas de données de GO directement associées aux gènes du xénope.

La base de données Xenbase (Fortriede et al. 2019) référence les orthologues Humain de *Xenopus laevis*, permettant d'établir une liste de gènes de référence constituée des gènes Humains possédant un orthologue de *Xenopus laevis*. Environ 65% des gènes du xénope non issus de scaffold contiennent un orthologue humain (29262/45093). La deuxième étape consiste à sélectionner l'algorithme d'enrichissement à utiliser. L'algorithme ORA est une méthode d'analyse de surreprésentation des voies biologiques. Le principe est de tester la proportion des gènes différentiellement exprimés présente au sein des différentes voies biologiques répertoriées, afin d'identifier un enrichissement. Plusieurs bases de données comme

REACTOME (Gillespie et al. 2022), Go (Ashburner et al. 2000) ou KEGG (Kanehisa 2000) décrivent des voies biologiques et peuvent être utilisées. Néanmoins, il existe plusieurs limites à cette stratégie en raison de l'absence de hiérarchie de classement entre les gènes selon leur FDR ou leur expression, et en absence de la prise en compte des réseaux de régulation. La méthode GSEA consiste à prendre en compte la hiérarchie des gènes à partir de la FDR ou de l'expression génique. Enfin, la méthode NTA prend en compte le réseau de régulation en plus de la hiérarchie de classement entre les gènes. Néanmoins, cette méthode NTA dépend des bases de données où sont référencées les interactions géniques décrivant les réseaux de régulations. Pour réaliser l'enrichissement fonctionnel, j'ai utilisé la méthode nommée ORA. Cette méthode prend en compte uniquement l'identité gènes, sans prendre en compte de leur niveau d'expression ou des réseaux de régulation auxquels ils peuvent appartenir. Il suffit d'une seule des deux copies S ou L de *Xenopus laevis* identifiées pour sélectionner l'orthologue Humain associé. J'ai testé les outils WebGesalt (Liao et al. 2019) et GProfiler (Raudvere et al. 2019) pour réaliser mes analyses. L'outil REVIGO (Supek et al. 2011) m'a permis ensuite de filtrer la redondance des GO Term depuis la base de données Go. Les analyses d'enrichissement fonctionnel seront réalisées avec GProfiler. Toutes les étapes de l'analyse sont résumées dans la **Figure 22**.

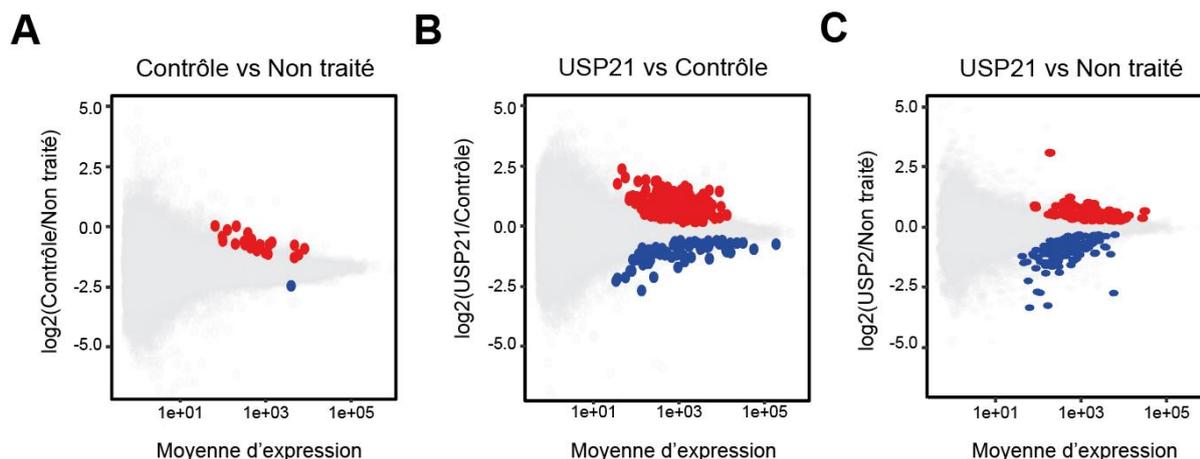


**Figure 22 : Pipeline d'analyse de la table de comptage en RNA-Seq**

La table de comptage est filtrée à l'aide de fonctions sur R, puis normalisée par le package DESeq2. L'effet batch est corrigé par le package limma. Les heatmaps sont générées à l'aide du package complexHeatmap, et l'enrichissement fonctionnel est réalisée sur les orthologues humain à partir de l'outil GProfiler.

## Le pré-traitement du spermatozoïde par USP21 avant l'injection dans l'œuf altère le transcriptome de l'embryon

J'ai déployé ces outils pour explorer dans un premier temps un jeu de donnée constitué de douze réplicats biologiques d'échantillons reparties en 3 conditions expérimentales : des gastrulas issues de spermatozoïde non traité appelé « non traité », de spermatozoïdes traités à l'extrait d'ovocyte nommé « contrôle », et enfin de spermatozoïdes traités par l'extrait d'ovocyte et la désubiquitinase USP21 nommé « USP21 » (**Figure 16**). Les œufs utilisés ont été irradiés par UV pour détruire la copie maternelle, afin d'obtenir des gastrulas haploïdes au sein desquelles la transcription du génome est exclusivement issue de la copie paternelle (**Figure 14**). Seulement 24 gènes sont identifiés comme significativement différents entre les gastrulas issues d'un spermatozoïde traité par l'extrait d'ovocyte seul et celles issues d'un spermatozoïde non traité (**Figure 23A**). Le traitement à l'extrait d'ovocyte contrôle semble donc avoir un effet mineur sur le transcriptome de l'embryon, ce résultat est conforté par le dendrogramme (**Figure 21**) qui regroupe ces échantillons ensemble. En revanche, les comparaisons entre les gastrulas dérivées de spermatozoïdes traités à l'extrait « USP21 » et celles dérivées de spermatozoïdes traités à l'extrait « contrôle » (**Figure 23B**) ou « non traité » (**Figure 23C**) retournent un nombre bien plus élevé de gènes différentiellement exprimés (269 et 325, respectivement) indiquant une dérégulation transcriptionnelle associée au traitement à USP21. 27,5% des gènes dérégulés dans la comparaison « USP21 » vs « contrôle » sont identifiés dans la condition « USP21 » vs « non traité », dont 28% des gènes up-régulés (**Tableau 1**). Pour la suite des analyses, afin de cibler les effets dérivés du traitement à USP21, nous avons comparé les gastrulas issues des conditions où seul l'expression de cette enzyme était différent (extrait « USP21 » vs extrait « contrôle »).



**Figure 23 : Le traitement à l'extrait d'ovocyte n'altère pas le transcriptome de l'embryon**

**(A)** MA-Plot montrant le  $\log_2$  du foldchange de l'expression des gènes entre les conditions contrôles et non traités en fonction de la moyenne d'expression des gènes dans les échantillons. (Points rouges : gènes surexprimés dans la condition contrôle  $N=3$ . Points bleus : gènes surexprimés dans la condition non traitée  $N=4$ ,  $FDR < 0.1$ )

**(B)** MA-Plot montrant le  $\log_2$  du foldchange de l'expression des gènes entre les conditions contrôles et non traités en fonction de la moyenne d'expression des gènes dans les échantillons. (Points rouges : gènes surexprimés dans la condition USP21  $N=4$ . Points bleus : gènes surexprimés dans la condition contrôle  $N=3$ ,  $FDR < 0.1$ )

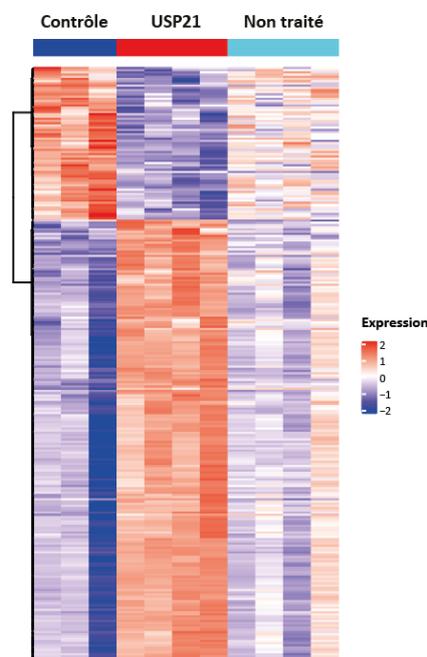
**(C)** MA-Plot montrant le  $\log_2$  du foldchange de l'expression des gènes entre les conditions contrôles et non traités en fonction de la moyenne d'expression des gènes dans les échantillons. (Points rouges : gènes surexprimés dans la condition USP21  $N=4$ . Points bleus : gènes surexprimés dans la condition non traitée  $N=4$ ,  $FDR < 0.1$ )

<b>FDR</b>	<b>Nombre de gènes UP</b>	<b>Nombre de gènes DOWN</b>
<b>0.1</b>	<b>200</b>	<b>69</b>
<b>0.05</b>	<b>106</b>	<b>44</b>

**Tableau 1 : Effet du seuil de FDR sur le nombre de gènes différentiellement exprimés identifiés**

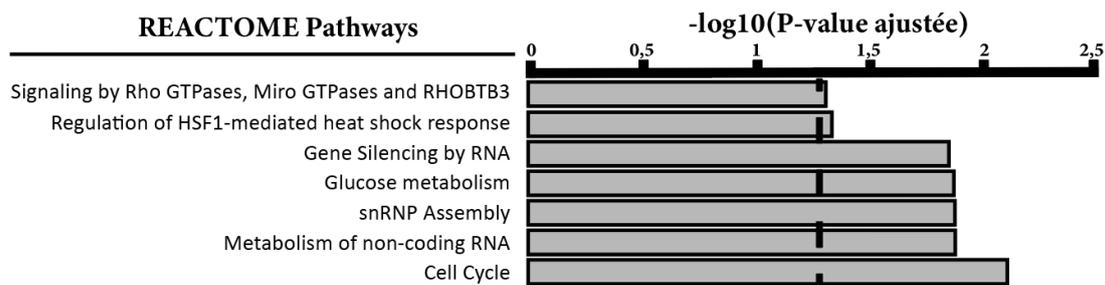
Les gènes UP sont les gènes sur-exprimés dans le cas du traitement à l'extrait USP21, et les gènes down ceux qui sont sous-exprimés suite au traitement à l'extrait USP21.

L'analyse DGE-seq2 a mis en évidence 269 gènes dérégulés par le traitement USP21 (**Figure 24**, FDR 0.1). On observe que ~75% des gènes dérégulés (200 gènes sur les 269) sont surexprimés en réponse au traitement USP21 du spermatozoïde. L'enrichissement fonctionnel des orthologues des gènes identifiés comme différentiellement exprimés met en avant une perturbation des gènes associés au cycle cellulaire dans la gastrula lors du traitement du spermatozoïde par USP21, comme RPA1, RPA2, MAPK1 ou HAUS8 (**Figure 25**). En conclusion, il est possible que la dérégulation des gènes par le traitement à l'extrait USP21, qui sont associés au cycle cellulaire, perturbe le développement embryonnaire et soit à l'origine des défauts de développement observés précédemment (**Figure 15**). L'analyse du devenir des embryons USP21 indique un comportement hétérogène avec une mortalité progressive des embryons au cours du développement du stade gastrula au stade de têtard. Pour mieux définir les événements transcriptomiques associés au traitement USP21, nous avons étendu notre analyse d'expression génique à des embryons uniques plutôt qu'à des pools d'embryons comme précédemment. Pour ce faire, j'ai eu à ma disposition des données de RNA-Seq provenant de 8 embryons uniques pour chaque condition (**Figure 26**, spermatozoïdes « non traités », traitement « contrôle » ou traitement « USP21 »).



**Figure 24 : Le traitement USP21 du spermatozoïde mène à des défauts d'expression géniques**

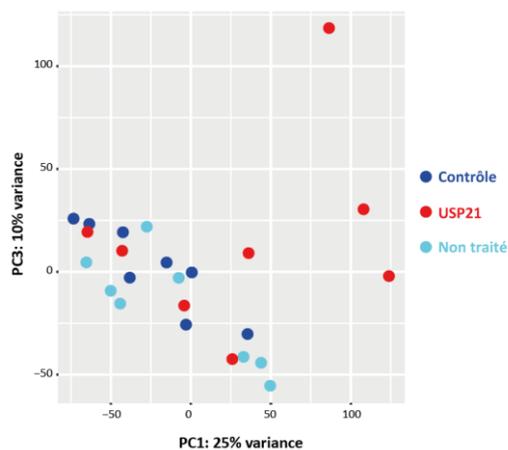
*Profil d'expression z-scoré des 269 gènes différentiellement exprimés dans les gastrulas pour les différentes conditions de traitement (contrôle = bleu, USP21 = rouge, non traité = cyan). 200 gènes ont une augmentation de leur expression après traitement du spermatozoïde par l'extrait USP21.*



**Figure 25 :** Les gènes du cycle cellulaire et du métabolisme de l’embryon sont dérégulés par le traitement USP21 du spermatozoïde

Le set des orthologues Humain des 269 gènes dérégulés dans la gastrula par le traitement du spermatozoïde à l’extrait USP21 est confronté à la base de données REACTOME. La p-value ajustée représente la significativité de l’enrichissement des gènes pour différentes voies biologiques.

Ces embryons passent les recommandations ENCODE de 30M de reads par échantillon. La réponse au traitement est similaire avec les embryons haploïdes uniques, mais on observe une variation plus grande dans la réponse au traitement à l’extrait USP21 (**Figure 26**).

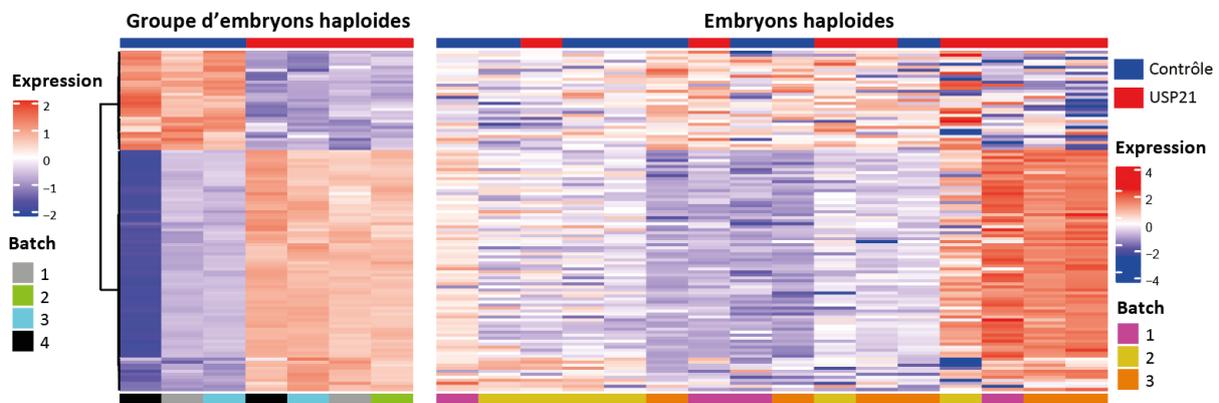


**Figure 26 :** La réponse au traitement USP21 dans le jeu de données d’embryons uniques est hétérogène

ACP des gastrulas d’embryons haploïdes issues d’un spermatozoïde traité par l’extrait USP1 (USP21 = rouge), d’un spermatozoïde traité par l’extrait d’ovocyte seul (contrôle = bleu) ou d’un spermatozoïde non traité (non traité = cyan).

Nous observons que sur 8 échantillons de gastrulas issues d’un spermatozoïde traité par l’extrait USP21, seulement 4 soit 50% montrent la même dynamique en embryon haploïde unique qu’en embryons regroupés, indépendamment du batch (**Figure 27**). L’autre moitié a un profil transcriptomique qui est plus proche des embryons « contrôle » et issus de spermatozoïdes non traités par USP21. Ce résultat est en adéquation avec les observations précédentes, où on

observe qu'environ 1/4 des gastrulas dérivées de spermatozoïdes traités par l'extrait USP21 survivent jusqu'au stade de têtard (**Figure 15**). On observe ainsi que la variabilité du traitement mise en avant par des observations morphologiques est également mise en avant avec des observations transcriptomiques après analyse de l'expression génique en embryons uniques.

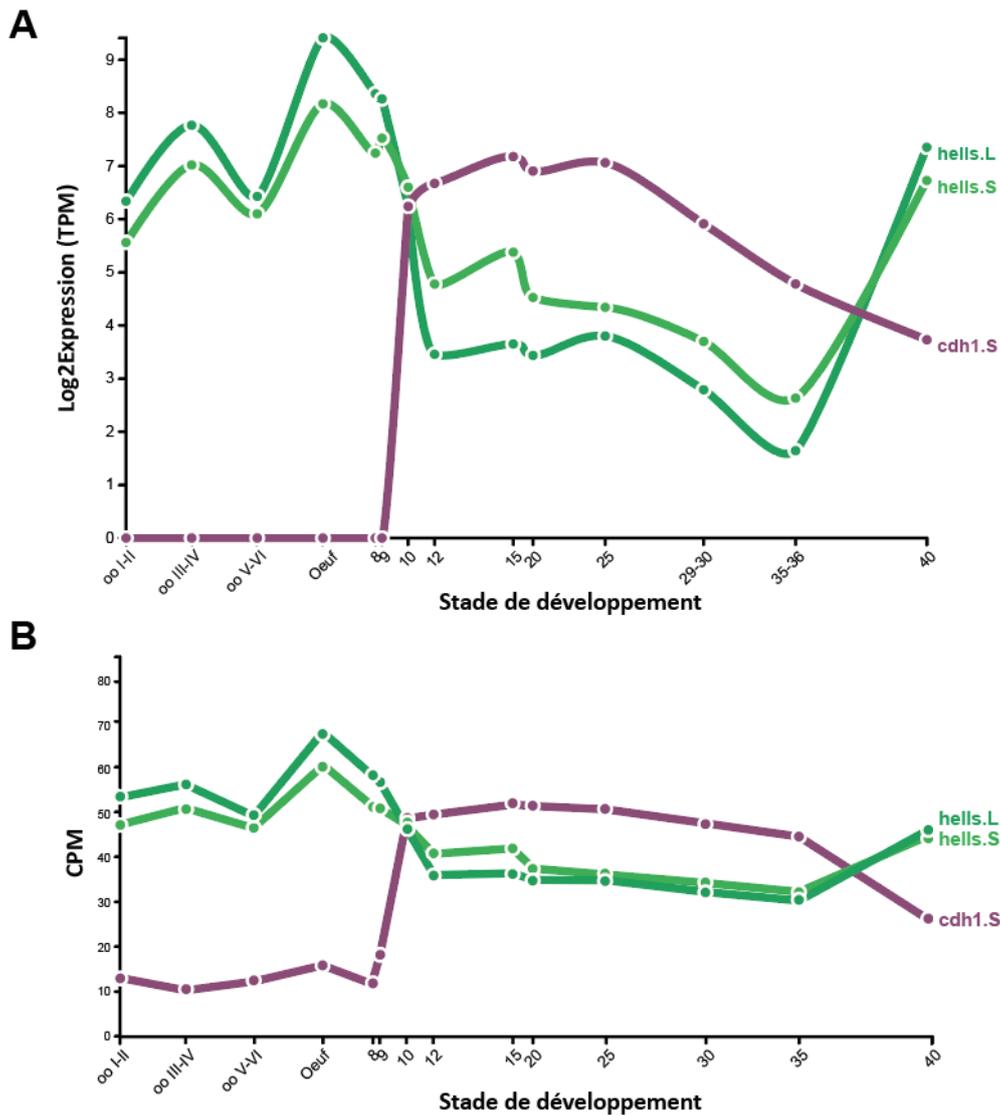


**Figure 27 : Le profil transcriptomique des gènes dérégulés dans les embryons regroupés est hétérogène dans les embryons uniques**

Profil d'expression des gènes différenciellement exprimés identifiés précédemment dans les analyses de groupes de gastrulas (à gauche) et l'expression de ces mêmes gènes dans des embryons uniques (droite). L'expression des gènes est représenté par une transformation z-score, indépendamment pour les deux heatmaps. Le batch indique des expériences indépendantes (différentes sources d'œufs pour générer les embryons).

Nous avons actuellement le profil transcriptomique des gènes sensibles au traitement USP21 uniquement dans la gastrula. Dans l'objectif de caractériser ces gènes, j'ai exploré leur profil transcriptomique durant le développement précoce de l'embryon. J'ai utilisé un jeu de données public (Session et al. 2016) pour déterminer leur dynamique d'expression au cours du développement embryonnaire de *Xenopus laevis*. Ce jeu de données contient des échantillons RNA-Seq qui incluent les étapes de développement depuis l'œuf jusqu'au stade têtard, données disponibles en ligne à la visualisation sur Xenbase (Fortriede et al. 2019).

Ces échantillons passent les recommandations ENCODE. J'ai calculé les Counts par Millions issus de mon pipeline d'analyse pour les comparer avec les valeurs de transcrits par million obtenu par Session et Al (Session et al. 2016) disponible sur Xenbase. On observe que les tendances d'expressions géniques illustrées pour les gènes *hells.L* et *hells.S* (up-régulés par USP21) et *cdh1.S* (down-régulé par USP21) sont similaires entre les deux analyses (**Figure 28**). Ceci valide indirectement l'ensemble de mon pipeline d'analyse RNA-Seq.



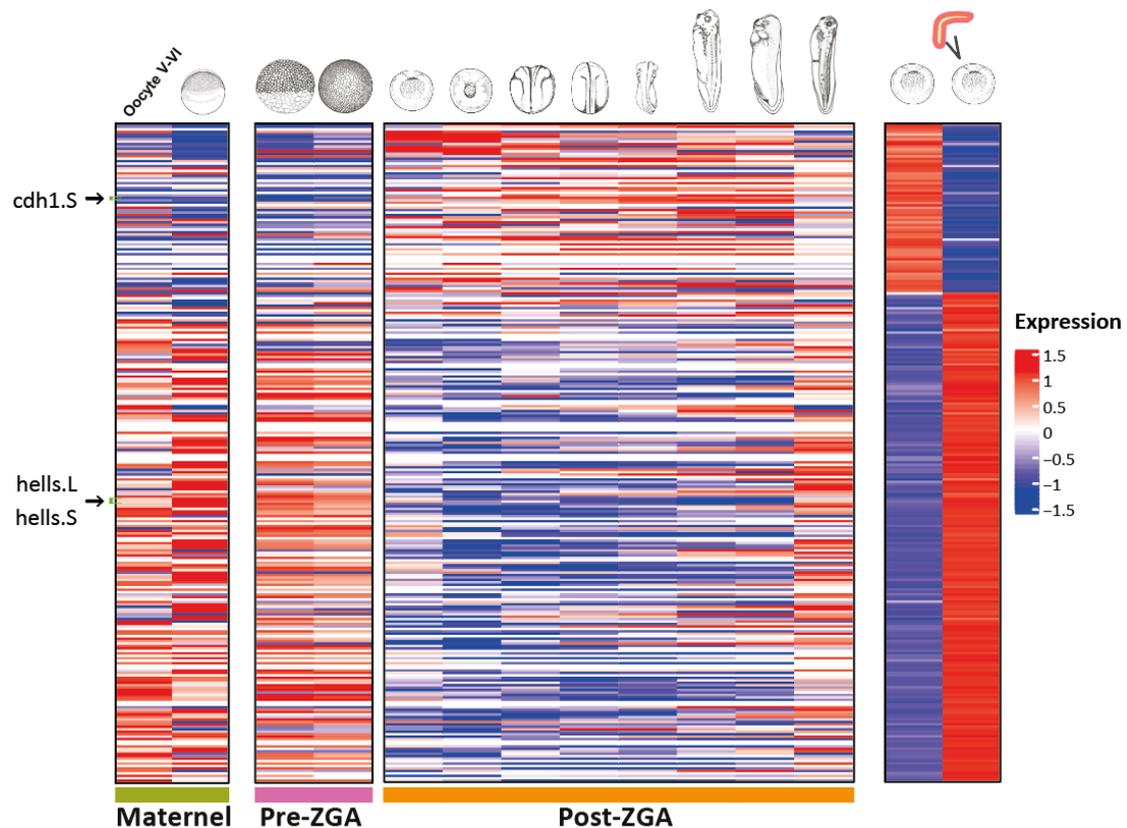
**Figure 28 : Le profil d'expression génique est similaire entre mon analyse et l'analyse Xenbase**

**(A)** Profil d'expression des gènes *hells.L* (vert foncé), *hells.S* (vert clair) et *cdh1.S* (violet) dans l'analyse de Session disponible sur Xenbase en  $\log_2(\text{TPM})$ . *cdh1.S* est exprimé après le stade 9, soit après la ZGA, puis son expression diminue dans le temps. Au contraire, les gènes *hells* sont exprimés avant la ZGA, puis le nombre de transcrit diminue avant une réactivation de leur transcription au stade 40, caractéristique du têtard.

**(B)** Profil d'expression des gènes *hells.L* (vert foncé), *hells.S* (vert clair) et *cdh1.S* (violet) dans mon analyse en CPM. Les tendances sont similaires à l'analyse de Session disponible sur Xenbase.

J'ai ensuite réanalysé ces données avec mon pipeline d'analyse. On observe que les gènes *hells.S* et *hells.L*, représentant des gènes surexprimés dans la gastrula après le traitement du spermatozoïde par l'extrait USP21, ont une composante maternelle élevée. Ils sont très exprimés avant l'activation du génome du zygote (ZGA), soit au stade maternel, œuf, stade 8/9. L'expression de ces gènes diminue fortement dans l'embryon après la ZGA à partir du stade 10, soit au stade gastrula, jusqu'au stade 40 représentant le têtard, où ce gène est à nouveau

déecté. Cette tendance est caractéristique des gènes up-régulés par USP21. A contrario, le gène *cdh1.S*, dans le groupe des gènes sous-exprimés lors du traitement par l'extrait USP21, est inactif avant la ZGA et est activé après la ZGA lors du développement de l'embryon dans une dynamique inverse du gène *hells*. Dans les gastrulas issues de spermatozoïdes traités à l'extrait USP21, ces gènes ne sont pas exprimés après la ZGA au stade gastrula, pouvant impliquer un blocage de leur transcription (**Figure 29**).



**Figure 29 : Le traitement USP21 perturbe l'expression des gènes à la ZGA**

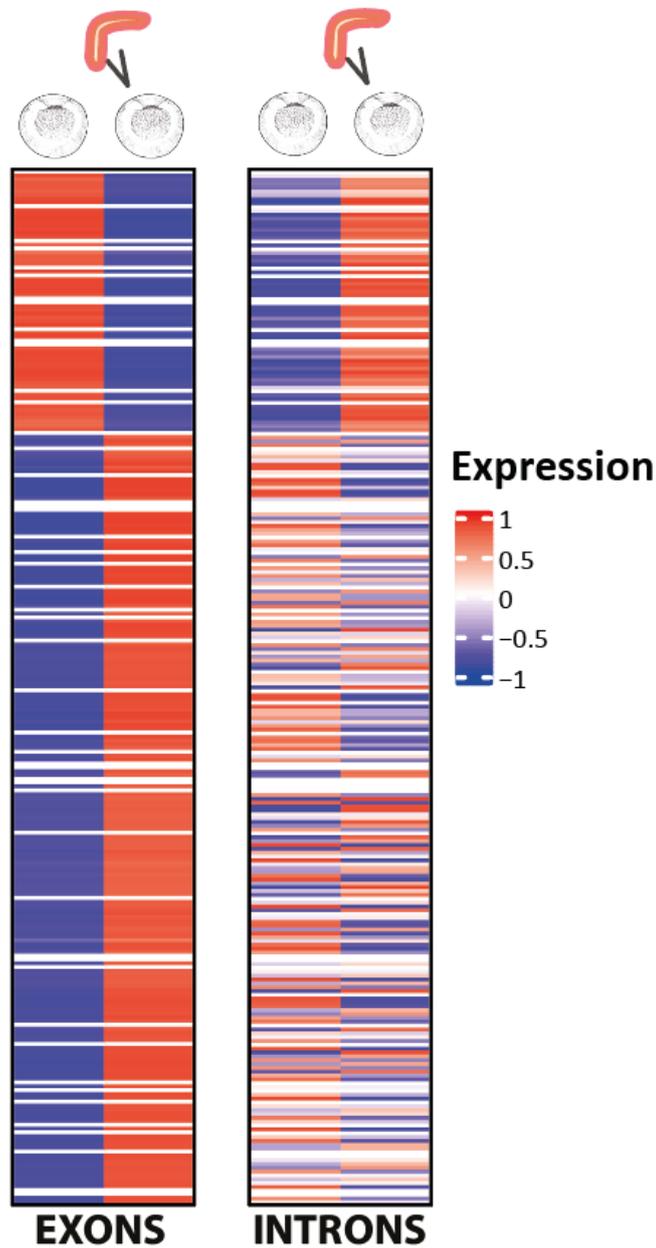
*Profil d'expression z-scoré, durant le développement embryonnaire, des 269 gènes différentiellement exprimés suite au traitement du spermatozoïde par l'extrait USP21. Les deux copies du gènes *hells* et le gène *cdh1.S* sont mis en évidence. Le groupe de gènes *cdh1.S* contiennent des gènes avec peu de transcrits pré-ZGA, et plus de transcrits post-ZGA, contrairement aux gènes *hells* qui caractérise le groupe de gènes avec des transcrits pré-ZGA (maternels), puis une activation de leur transcription beaucoup plus tardive (stade de têtard). Les deux jeux de données ont été analysés indépendamment de la matrice de comptage jusqu'au z-score, malgré que l'échelle de z-score soit unique aux deux heatmaps.*

## Défaut d'élimination des ARNs maternels ou induction de la transcription zygotique des gènes sensibles au traitement USP21 ?

Le profil d'expression des gènes dérégulés en réponse au traitement USP21 est dominé par un effet de surexpression qui s'opère sur des gènes caractérisés, au cours du développement, par une forte contribution maternelle associée à une élimination à la ZGA. En conséquence, l'effet du traitement USP21 pourrait être expliqué soit par une induction de la transcription zygotique de ces gènes soit par une perturbation de l'élimination des ARNs maternels, ou une combinaison de ces processus.

Pour tenter de distinguer entre ces scénarios, j'ai exploré la possibilité d'utiliser la présence de reads introniques dans les ARNs comme indicateur de leur transcription récente. En effet, une partie des ARNs nouvellement transcrits contient encore des introns non épissés, une caractéristique utilisée en analyse scRNA-seq pour déterminer la récente activation d'un gène (La Manno et al. 2018). J'ai donc calculé pour chaque gène le nombre de reads introniques et de reads exoniques (Gaidatzis et al. 2015) dans les échantillons « contrôle » versus USP21. La projection des données exoniques et introniques normalisées et transformées est réalisée après normalisation avec DESeq2. Comme attendu, on observe la même tendance avec cette approche (**Figure 30**, panel de gauche), pour les exons, qu'avec l'analyse classique utilisant STAR (**Figure 29**, panel de droite). Concernant les introns, on observe des tendances différentes entre les gènes up-régulés et down-régulés par USP21. Pour les gènes up-régulés, on n'observe pas de différences prononcées pour les reads introniques entre les conditions « contrôle » et « USP21 ». Ceci pourrait indiquer que les ARNs accumulés de façon aberrante dans les embryons « USP21 » ne sont pas issus d'une transcription récente. Dans le cas des gènes down-régulés, on constate plus de reads introniques dans les gastrulas dérivées de spermatozoïdes traités à l'extrait USP21 que dans les gastrulas contrôle. En considérant le modèle selon lequel la présence de reads introniques dans un ARN est une indication de sa transcription récente, ceci suggère qu'au moment de la collecte des échantillons, les ARNs des gènes down-régulés par le traitement USP21 émanent d'un événement de transcription plus récent, dans les gastrulas USP21, que dans les gastrulas contrôles. Ceci pourrait suggérer que ces ARNs sont transcrits avant le stade gastrula dans les échantillons contrôle et que le traitement USP21 retarde l'induction de ces gènes avec un début de transcription à ce stade. Alternativement, cette

accumulation d'introns pourrait découler d'une inhibition de la transcription couplée à une perturbation des phénomènes d'épissage. En conclusion, ces analyses ne permettent pas de conclusions définitives et demandent des contrôles additionnels pour être en mesure de conclure.



**Figure 30 : Le traitement USP21 perturbe la transcription des nouveaux ARN à la ZGA**

*Profil d'expression z-scoré de la matrice exonique (à gauche) et intronique (à droite) des gastrulas issues des spermatozoïdes traités par l'extrait USP21 ou traités par l'extrait d'ovocyte. Les gènes sensibles au traitement USP21 non identifiés dans le jeu de données intronique ou exoniques sont affichés en blanc. Les deux jeux de données ont été analysés indépendamment de la matrice de comptage jusqu'au z-score, malgré que l'échelle de z-score soit unique aux deux heatmaps.*

J'ai finalement caractérisé plus en détail les gènes sensibles à USP21 en explorant plus en détail l'expression zygotique au cours du développement. Pour cela, j'ai exploré les données d'une étude récente chez *Xenopus laevis* qui a mis en évidence non seulement les gènes avec une transcription zygotique (Z) mais également des gènes qui présentent une transcription zygotique dans le contexte d'une contribution maternelle importante (MZ) (H. Chen et Good 2022). Cette distinction n'est pas possible en RNA-Seq classique et requiert une étape supplémentaire de marquage des ARNs nouvellement transcrits (H. Chen et Good 2022).

J'ai vérifié si les gènes sensibles au traitement USP21 sont enrichis dans une de ces deux catégories de gènes. J'ai calculé le nombre de gènes up-régulés ou down-régulés en commun avec les sets de gènes MZ (contribution maternelle-zygotique) ou Z (expression zygotique uniquement). J'ai ensuite réalisé 10000 permutations de gènes pour déterminer si cet overlap était significatif par rapport à une distribution aléatoire. On observe ainsi que les gènes affectés par USP21 présentent un enrichissement pour les gènes à expression zygotiques dans le cadre d'une contribution maternelle forte (**Tableau 2**). Ces dernières observations indiquent donc que les événements de dérégulation observés en réponse au traitement USP21 affectent des gènes qui sont normalement activés lors de la ZGA.

Gènes	Gènes up-régulés (200)				Gènes down-régulés (69)			
	Z-score	P-Value	nObs	nTh	Z-score	P-Value	nObs	nTh
<b>M Genes</b>	13.616	0.0001	99	30	-0.938	0.8682	8	11
<b>Z Genes</b>	-1,68	1	0	3	2,101	0,0721	3	1
<b>MZ Genes</b>	3,894	0,0005	21	9	2,665	0,0173	8	3

**Tableau 2 : Les gènes up-régulés par le traitement USP21 ont une contribution maternelle et zygotique**

Tableau récapitulatif du nombre de gènes up-régulés (à gauche, en rouge) et down-régulés (à droite, en bleu) par le traitement USP21 qui ont une expression maternelle (M), zygotique (Z) ou maternelle-zygotique (MZ). Le nombre observé (nObs), le nombre théorique (nTh) obtenue avec 10 000 permutations permettent d'obtenir une p-value ainsi qu'un Z-score pour décrire l'enrichissement du groupe de gènes issu du jeu de données Z et MZ. Les listes de gènes Z et MZ sont issues de Chen et Good, 2022.

# Chapitre III- Caractérisation de la distribution de H2AK119Ub1 dans le spermatozoïde

Jusqu'à présent, notre analyse indique que le traitement USP21 du spermatozoïde entraîne des défauts d'expression génique au stade gastrula. Cependant, ces premières observations n'indiquent pas si l'effet observe découle directement de l'activité H2A deubiquitylase de USP21. Pour évaluer cette possibilité, j'ai tout d'abord caractérisé la localisation de la marque cible de USP21, H2AK119Ub1, dans le spermatozoïde. J'ai ensuite recherché si une corrélation existait entre la présence de H2AK119Ub1 et la sensibilité des gènes au traitement à l'extrait USP21. Nous avons utilisé la méthode de ChIP-Sequencing pour répondre à cette question de la localisation de H2AK119Ub1 sur le génome du spermatozoïde. Dans cette partie, j'ai utilisé trois réplicats pour la marque H2AK119Ub1 dans le spermatozoïde, trois réplicats pour la marque H3K4me3 et deux réplicats pour la marque H3K27me3 (issus de (Oikawa et al. 2020)), tous séquencés en paired-end. (**Figure 31**). Enfin, nous avons plusieurs échantillons de spermatide pour les trois marques H3K27me3, H3K4me3 et H2AK119Ub1.



Spermatide	Spermatozoïde
3x H2AK119Ub1 (SE)	3x H2AK119Ub1 (PE)
2x H3K4me3 (SE)	3x H3K4me3 (PE)
3x H3K27me3 (SE)	2x H3K27me3 (PE)

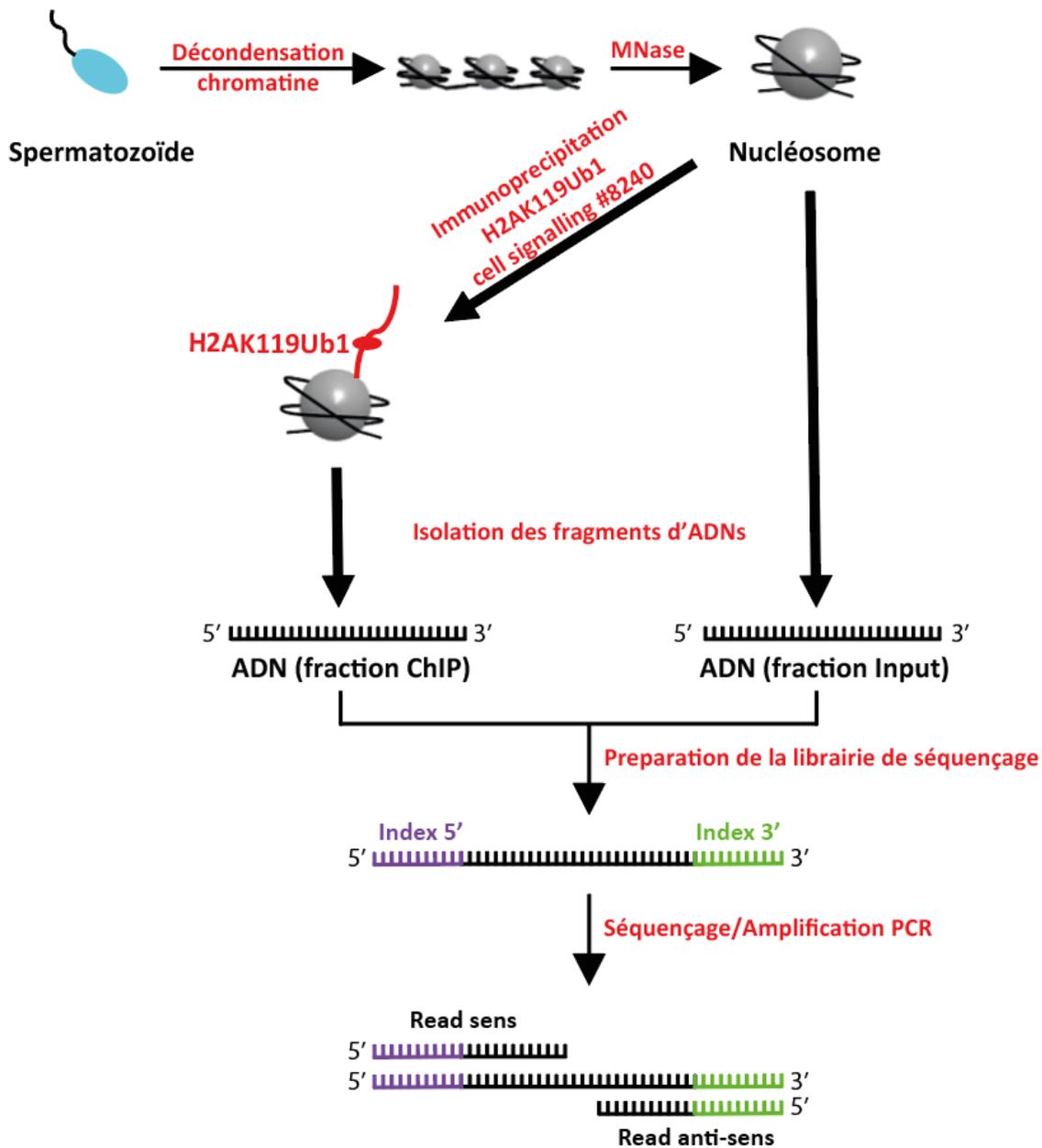
**Figure 31 : Schéma récapitulatif des échantillons de ChIP-Seq**

*Les trois réplicats de H2AK119Ub1 et H3K4me, ainsi que les deux réplicats de H3K27me3 sont séquencés en paired-end (en rouge) dans le spermatozoïde. Trois réplicats de H3K27me3 et H2AK119Ub1, ainsi que 2 réplicats de H3K4me3 ont été séquencé en single-end dans le progéniteur du spermatozoïde : la spermatide. (SE = Single-End, PE= Paired-End)*

## Analyse de l'épigénome par ChIP-Sequencing

Nous avons employé une technique de ChIP Native, basée sur une fragmentation de la chromatine par une nucléase micrococcale (MNase) qui digère les portions d'ADNs accessible entre les particules nucléoprotéiques qui constituent la chromatine, dans le cas du spermatozoïde de *Xenopus laevis* des nucléosomes et/ou sous-nucléosome (**Figure 4, Figure 13**). Un protocole adapté à la chromatine très condensée du spermatozoïde est utilisé (décrit dans (Oikawa et al. 2020)-(Teperek et al. 2016)). Une fraction de la chromatine fragmentée ainsi obtenue est directement mise de cote pour fournir un échantillon dit « Input ». Cet échantillon est très important car il permet d'identifier de potentielle variation dans l'efficacité de digestion de la chromatine au sein de différentes régions du génome. Ces variations proviennent par exemple de différence dans la composition en base (Tolstorukov et al. 2009) ou bien de différence dans l'accessibilité, comme par exemple au sein des régions fortement hétérochromatique difficile d'accès (Henikoff et al. 2009) - (Liang et Keleş 2012). Ainsi, la fraction Input permet de prendre en compte dans notre analyse les disparités dans la digestion de la chromatine par la MNase (Park 2009). Le reste de la chromatine fragmentée est soumise à immunoprécipitation avec des anticorps reconnaissant les marques d'histones d'intérêt (dans notre cas H2AK119Ub1, cell signalling#8240, H3K4me3, abcam#8580, H3K27me3 laboratoire de T Jenuwein). Cette fraction enrichie pour les fragments d'ADNs associés avec une modification d'histone est dénotée « ChIP ». L'ADN associé aux fractions ChIP et Input est ensuite isolé, indexé par ligation d'adaptateurs puis amplifié par PCR avant le séquençage en paired-end 50pb sur un appareil Novaseq 6000. Le démultiplexage a été réalisé par la plateforme de séquençage, fournissant les fichiers fastQ des séquences des fractions ChIP et Input pour chaque échantillon (**Figure 32**).

Dans les sections suivantes, je détaillerai d'abord les outils sélectionnés pour l'analyse bio-informatique des données de séquençage, puis je détaillerai par la suite les questions biologiques adressées grâce à ces outils.



**Figure 32 : Schéma du protocole de préparation des échantillons de ChIP-Seq pour le séquençage**

Après perméabilisation des membranes du spermatozoïde, la chromatine est fragmentée par une MNase permettant de libérer les particules nucléoprotéiques tels que les nucléosomes et sous-nucléosomes. Une fraction est immuno-précipitée par l'anticorps anti-H2AK119Ub1 (la fraction ChIP), et l'autre fraction est utilisée en l'état (fraction Input). Les fragments d'ADNs autour des nucléosomes retenus dans les deux fractions sont isolés, séquencés et amplifiés.

## Détails des étapes et outils utilisés dans les analyses ChIP-Seq

L'étape d'alignement comprend toutes les étapes de manipulation des reads issues du démultiplexage jusqu'à la production d'un fichier BAM qui répertorie la position de chaque read sur le génome de référence.

### *Élimination des séquences adaptatrices et des reads de faible qualité*

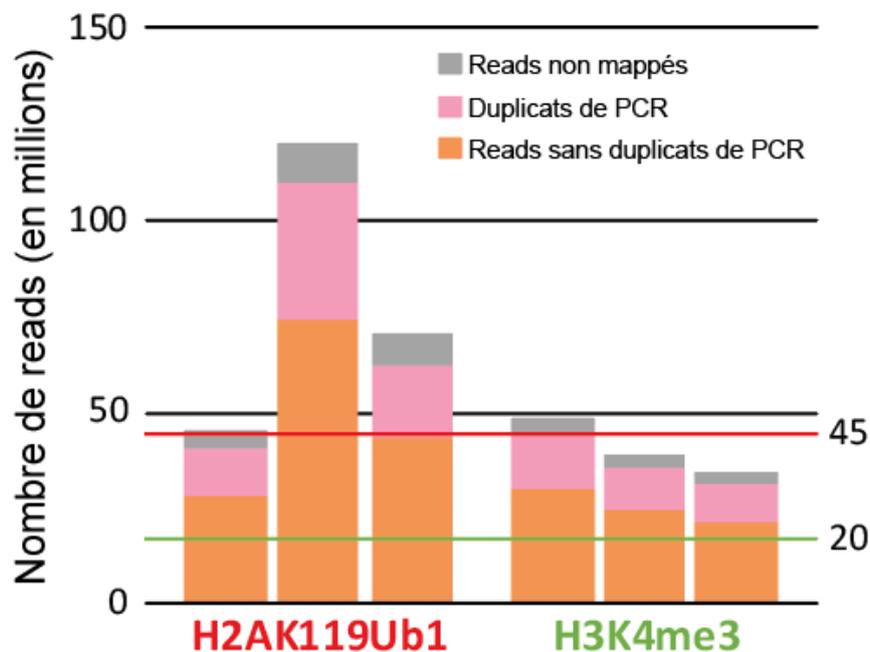
Les reads issus du démultiplexage comportent des séquences adaptatrices qui ne correspondent pas au matériel biologique de départ et qu'il est nécessaire d'éliminer. J'ai utilisé le logiciel cutadapt (Martin 2011) pour à la fois enlever les séquences adaptatrices, mais aussi exclure les reads de faible qualité de séquençage, avec les mêmes paramètres que pour le RNA-Seq (-m 20, -q 20). Contrairement au RNA-Seq séquençé en single-end, nous avons un séquençage en paired-end dans ce cas. Il est important de prendre en compte l'appariement des reads lors de cette étape de filtration, pour modifier les séquences des deux brins conjointement en enlevant le read associé (--pair-filter=any). La qualité de la filtration est analysée avec le logiciel fastqc (Andrews S 2010).

### *Alignement des reads sur un génome de référence*

Les séquences sont alignées au génome de référence à l'aide de l'outil bwa, avec la fonction bwa mem (H. Li et Durbin 2009). Le génome d'alignement utilisé est le génome *Xenopus laevis* version 9.2. Nous devons ensuite isoler les reads issus de duplicats liés à l'amplification PCR (Tian et al. 2019). Ces duplicats induisent un biais du signal non présent dans l'échantillon initial et totalement dépendant de la phase d'amplification. Le package Picard (Broad Institute of MIT and Harvard, s. d.), avec la fonction markduplicates identifie les duplicats de PCR par comparaison des séquences 5' des reads. Ces duplicats de PCR induisent une surreprésentation d'un signal non biologique. Les reads identifiés comme duplicats sont marqués pour les différencier du read originel. Cette différenciation est basée sur le score de qualité du read, le read avec la meilleure somme des qualités de séquençage des bases est conservé. Le read anti-sens associé au read identifié comme duplicat de PCR est lui aussi marqué. Les statistiques d'alignements des reads sur les génomes sont obtenues par samtools flagstat (H. Li et al. 2009), ou encore la suite deepTools (Ramírez et al. 2016) avec la fonction plotFingerprint qui permet d'évaluer la qualité de la fraction ChIP, ainsi que de la fraction Input (Diaz et al. 2012).

Enfin, la fonction bamPEFragmentSize de la suite deepTools nous permet de vérifier la taille des fragments d'ADNs que nous avons, en prenant en compte la taille de la séquence génomique recouverte par les reads paired-end appariés.

Au final, les recommandations ENCODE 3 préconisent un nombre de 20M de reads cartographiés sur le génome pour les marques localisées telle que H3K4me3, et de 45M de reads pour les marques diffuses telle que H2AK119Ub1 (Landt et al. 2012). Un des échantillons pour la marque H2AK119Ub1 contient un nombre légèrement inférieur de reads de *Xenopus laevis* que les recommandations ENCODE 3 de 45M de reads pour la marque H2AK119Ub1 (**Figure 33**).



**Figure 33 : Les réplicats de ChIP-Sequencing sont conformes aux recommandations ENCODE 3**

Les recommandations ENCODE 3 préconisent un minimum de 45M de reads (ligne rouge) pour la marque H2AK119Ub1, et un minimum de 20M de reads (ligne verte) pour H3K4me3. Les trois réplicats utilisés pour chaque marque sont représentés.

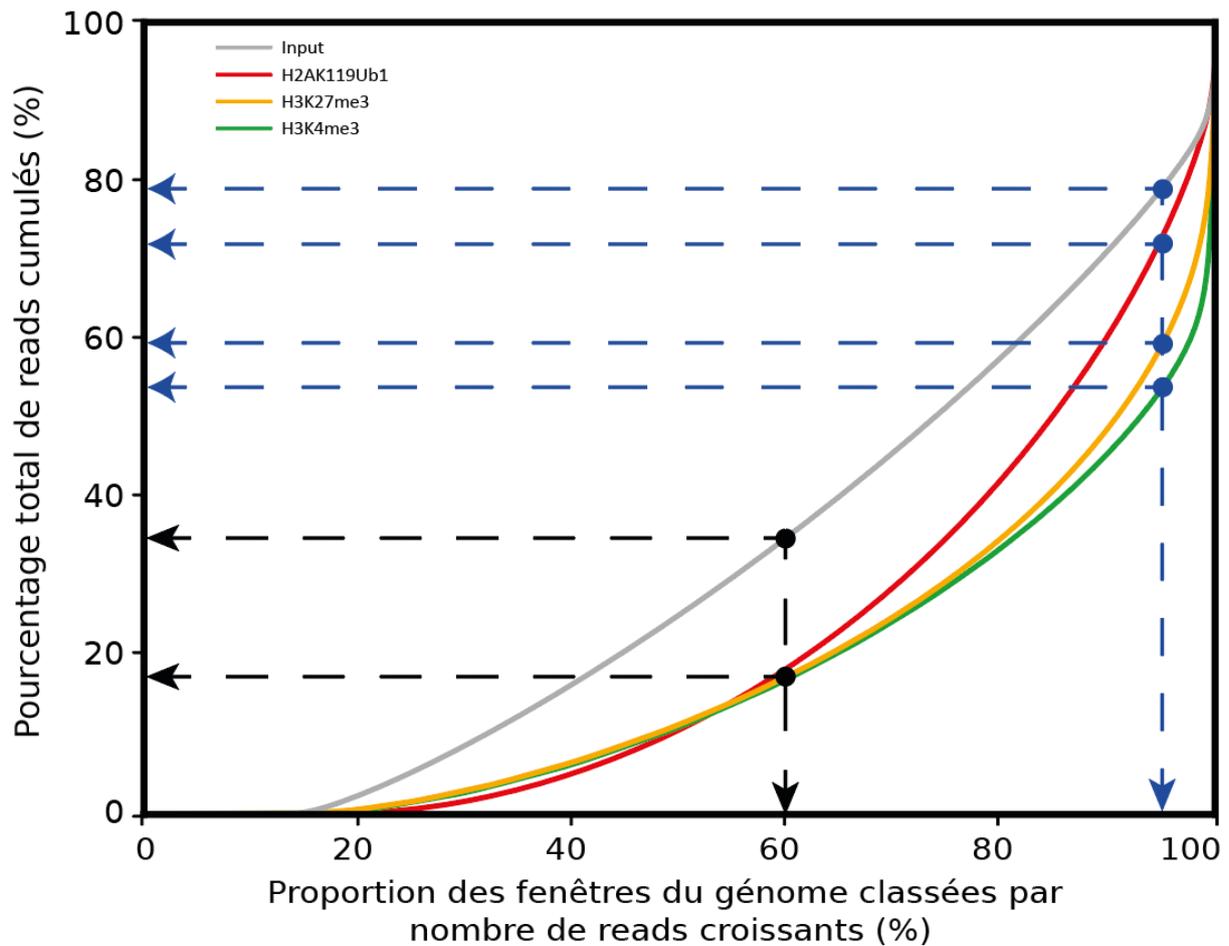
Samtools view permet ensuite de filtrer les reads issus de réplicats de PCR à partir du tag déposé par markduplicates selon la méthodologie décrite précédemment.

## *Identification des régions du génome présentant un enrichissement pour une modification d'histone*

Lors de ma thèse, j'ai utilisé deux outils pour détecter un enrichissement du signal sur des portions du génome. Un signal correspond à un empilement de reads sur une région donnée dans la fraction ChIP, normalisée par rapport au signal basal présent dans la fraction Input.

Ces outils permettent d'identifier statistiquement si le signal est enrichi sur des régions du génome. Le premier outil utilisé est MACS2 (Zhang et al. 2008), dont les fonctionnalités permettent d'identifier des enrichissements de signal, soit très localisé (paramètre par défaut), soit relativement large sur le génome (option `--broad`). Cet outil a été utilisé pour la détection du signal des marques H3K4me3 et H3K27me3, avec l'option complémentaire `--broad` pour H3K27me3. En revanche, il n'y a pas de consensus pour les outils les plus appropriés pour l'analyse d'enrichissement de H2AK119Ub1. Contrairement à la marque H3K4me3 qui est présente sur environ 0.1 à 0.5% des histones H3 des cellules (LeRoy et al. 2013), la marque H2AK119Ub1 est présente sur 10% des histones H2A des cellules (Fursova et al. 2019) (Huseyin et Klose 2021). En outre, l'inspection de la distribution des reads sur un génome browser (IGV) indique un profil très diffus de H2AK119Ub1 sur le génome (Blackledge et al. 2020)-(Tamburri et al. 2020)-(Mei et al. 2021)-(Z. Chen, Djekidel, et Zhang 2021). Enfin, l'analyse fingerprint du profil de la fraction ChIP de H2AK119Ub1 montre une couverture beaucoup plus proche de celle de l'Input que la fraction ChIP de la marque localisée H3K4me3, ou de la marque H3K27me3, renforçant la notion d'une distribution très diffuse de H2AK119ub1 (**Figure 34**).

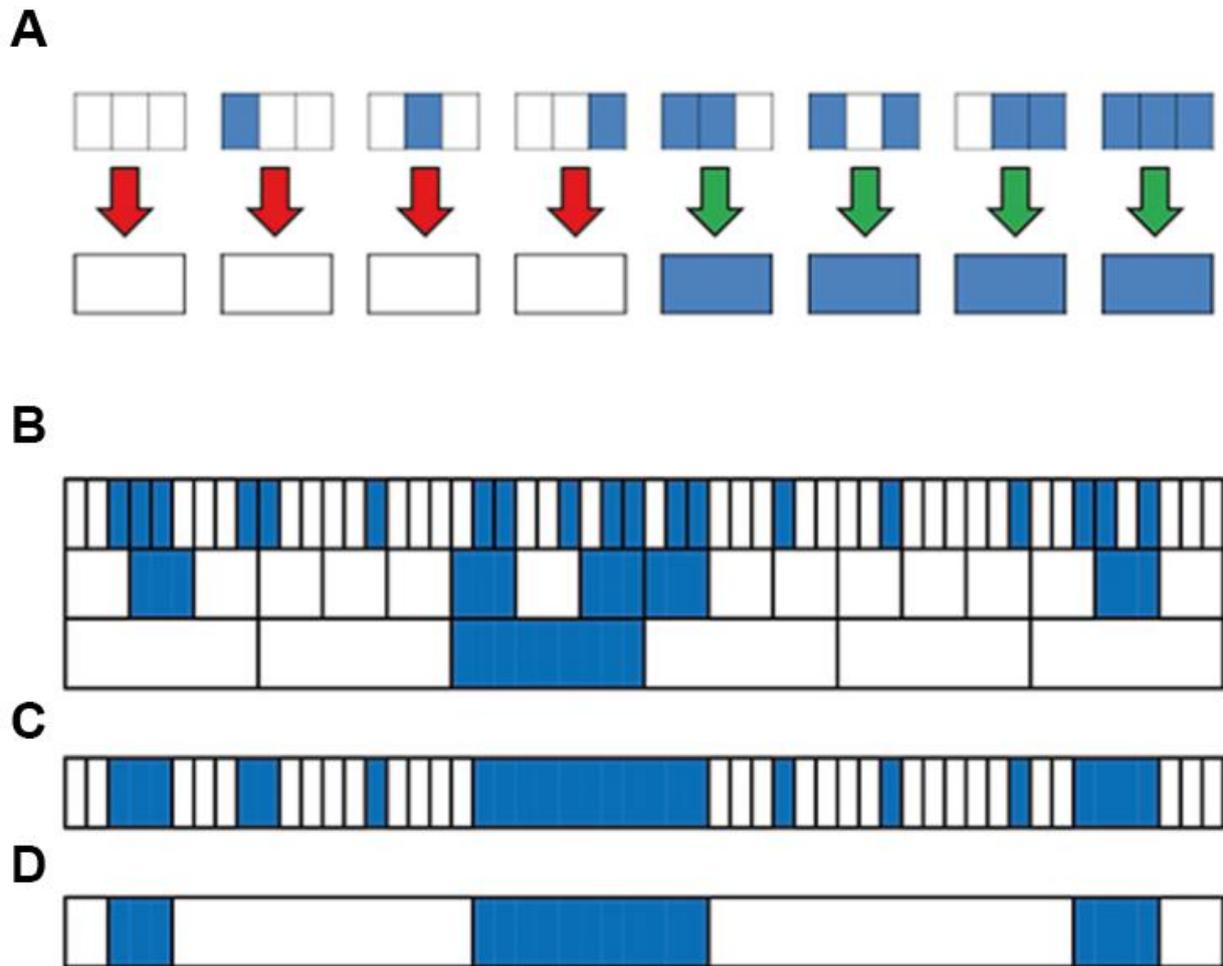
De nombreuses études de H2AK119Ub1 se concentrent sur les sites de fixations de RING1B, l'enzyme déposant H2AK119Ub1 (Blackledge et al. 2020) - (Tamburri et al. 2020) - (Andrea Scelfo et al. 2019) (**Figure 8**) ou caractérisent le signal sur des éléments définis du génome tel que les TSSs (heatmap, metaplot) (Kuznetsov et al. 2019). Pour notre étude, nous voulons définir les caractéristiques générales de la distribution de H2AK119Ub1 et être en mesure de tester l'association de cette marque avec notre jeu de gènes sensibles à USP21. Pour ce faire, nous avons donc besoin d'un outil pour déterminer sans a priori, l'enrichissement de la marque H2AK119Ub1 sur le génome. J'ai réalisé l'analyse avec un autre outil nommé RECOGNICER (Zang, Wang, et Peng 2020). Cet outil est développé par les auteurs du logiciel SICER (Zang et al. 2009), un autre logiciel de détection d'enrichissement du signal similaire à MACS2.



**Figure 34 : Analyse fingerprintplot de la couverture des reads des fractions ChIP et Input.**

La courbe représente la proportion de reads cumulés (ordonnée) qui sont couverts par la proportion de fenêtre du génome issue d'une segmentation du génome en fenêtres de 500pb (abscisse). Les fenêtres sont rangées en fonction du nombre croissant de reads qu'elles contiennent. La proportion du génome non séquencée est représentée par l'intersection de la courbe avec l'axe des abscisses (proportion des fenêtres de 500pb qui contiennent aucun read). La forme de la courbe représente le profil de la marque, plus le profil est proche de l'Input, plus la marque est diffuse. Par exemple, si on prend 60% des fenêtres classées du génome (flèches noires), nous couvrons environ 20% des reads des ChIP H3K4me3 (courbe verte), H2AK119Ub1 (courbe rouge) et H3K27me3 (courbe jaune) contre 40% des reads de l'Input (courbe grise). Si on prend 95% des fenêtres (flèches bleues), nous couvrons ~55% des reads de la ChIP H3K4me3 (courbe verte), ~60% des reads de la ChIP H3K27me3 (courbe jaune), ~70% des reads de la ChIP H2AK119Ub1 (courbe rouge) et ~80% des reads de l'Input (courbe grise). ~45% des reads H3K4me3 sont donc contenus dans les 5% des fenêtres du génome avec le plus de reads, ce qui est caractéristique d'une marque localisée. Au contraire ~30% des reads H2AK119Ub1 sont donc contenus dans les 5% des fenêtres du génome avec le plus de reads, ce qui est caractéristique d'une marque diffuse.

Contrairement à SICER, RECOGNICER utilise un algorithme de recherche du signal utilisant de façon itérative différents niveaux de segmentation du génome (**Figure 35**).

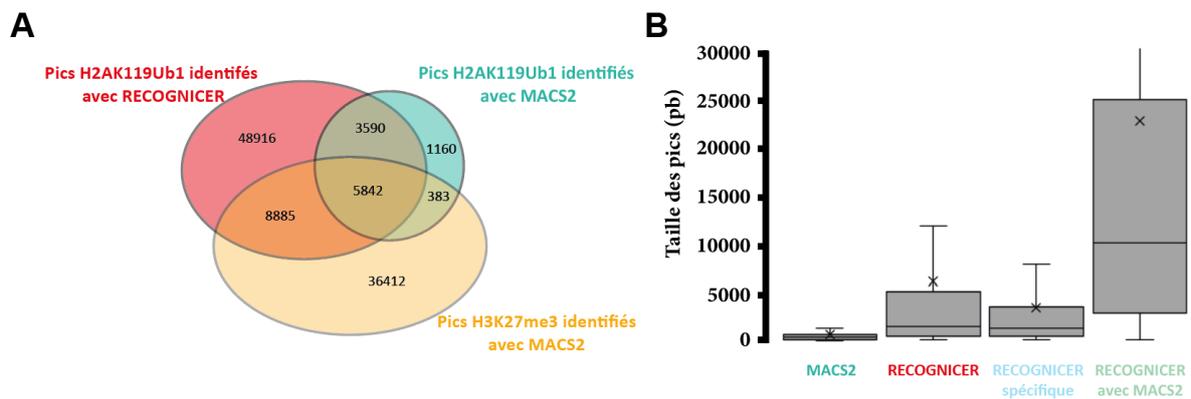


**Figure 35 : Représentation schématique du principe de détection d'enrichissement par l'outil RECOGNICER**

*Figure reproduite de Zang, Wang, et Peng 2020*

**(A)** Le génome est partitionné en fenêtre à l'échelle la plus fine. Pour chaque bin le signal est classé comme positif (1=bleu) ou négatif (0=blanc). On procède ensuite à une transformation en bloc où les fenêtres sont regroupées pour obtenir une partition du génome à une échelle plus étendue. Durant cette étape, le statut positif du bloc est déterminé par le statut des fenêtres le composant à l'échelle inférieure. Pour cela, une règle de la majorité est appliquée. Dans le cas illustré, si moins de 2 fenêtres sur 3 de l'échelle inférieure sont positives, la fenêtre de l'échelle supérieure devient négative (flèche rouge). Dans le cas contraire, (2 ou 3 fenêtres sur 3 positives) la fenêtre de l'échelle supérieure est conservée positive (flèche verte). **(B-D)** Procédure d'analyse : **(B)** Une analyse itérative du processus décrit en **(A)** est réalisée afin d'identifier les domaines d'enrichissement à différentes échelles. **(C)** Après l'analyse itérative, les domaines d'enrichissement des régions candidates à toutes les échelles sont rassemblés **(D)** puis leurs importances déterminées pour produire un set final de zones d'enrichissement.

L'outil MACS2 a été utilisé avec les paramètres de détection d'une marque diffuse, et une FDR de 0,01. RECOGNICER a été utilisé avec un paramètre de FDR à 0,001. Lors de la comparaison de la détection de H2AK119Ub1 par les deux outils, nous avons observé que 86% des régions identifiées par MACS2 sont également identifiées par RECOGNICER, qui identifie en outre six fois plus de régions (**Figure 36A**). Nous capturons un plus grand overlap des pics H3K27me3 avec les pics H2AK119Ub1 avec RECOGNICER (28.5%) qu'avec MACS2 (12%). RECOGNICER reconnaît des régions enrichies en H2AK119Ub1 plus grandes que celles identifiées par MACS2 (**Figure 36B**).



**Figure 36 : Comparaison des pics H2AK119Ub1 déterminés par RECOGNICER et par MACS2**

**(A)** Diagramme de Venn des pics H2AK119Ub1 identifiés par RECOGNICER (67 233 pics, rouge) et MACS2 (10 975 pics, bleu-vert), et des pics H3K27me3 identifiés par MACS2 (51 522 pics, jaune). **(B)** Barplot de la taille des pics H2AK119Ub1 identifiés par MACS2 (bleu-vert), RECOGNICER (rouge), seulement par RECOGNICER (bleu) ou RECOGNICER et MACS2 (vert).

Ces résultats suggèrent que RECOGNICER permet de reconnaître le signal H2AK119Ub1 que MACS2 identifie, tout en identifiant de nouvelles régions candidates que MACS2 ne détecte pas. Après détection du signal, l'outil ChIPseeker (Yu, Wang, et He 2015) recherche la localisation du signal sur les régions du génome (telles que les Introns, Exons, TSSs...), mais ne réalise pas de test statistique pour évaluer d'éventuels enrichissements sur ces éléments. Dans l'objectif de rechercher un enrichissement du signal dans certaines régions du génome, j'ai utilisé le package regioneR (Gel et al. 2015). Ce package permet de réaliser des tests statistiques de permutation de pics selon plusieurs mesures, comme l'overlap qui représentent le nombre de régions qui se chevauchent entre deux groupes, ou la distance entre les pics et des éléments du génome (TSSs, enhancers...). Les régions définies comme enhancers chez *Xenopus laevis* dans les analyses sont issues de Elurbe et al (Elurbe et al. 2017), enhancers qui ont été défini à partir des sites de fixation de P300 dans l'embryon. Ces régions représentent

donc des éléments de régulation actifs dans l’embryon et sont donc des candidats potentiels pour une programmation épigénétique dans le spermatozoïde. Néanmoins, les enhanceurs ne sont pas assignés aux gènes de façon spécifique actuellement. Ainsi, chaque enhanceur a été associé empiriquement au gène le plus proche (Fishilevich et al. 2017).

Pour déterminer si la marque H2AK119Ub1 présente une association préférentielle avec certains éléments du génome, j’ai choisi d’utiliser la mesure d’overlap entre les pics et ces éléments plutôt que la mesure de distance. Pour ce faire, une première étape consiste à générer d’abord une loi de distribution représentant l’overlap théorique entre des régions d’intérêt du génome (par exemple, les TSSs) et un groupe de pics redistribués de façon aléatoire, nommée l’étape de randomisation. L’overlap observé entre le set de pics réels et les régions d’intérêt du génome est ensuite confronté à la loi de distribution d’overlap théorique ainsi créée, permettant de calculer un Z-Score et une P-Value pour déterminer si les pics observés sont effectivement enrichis ou non sur ces régions d’intérêt.

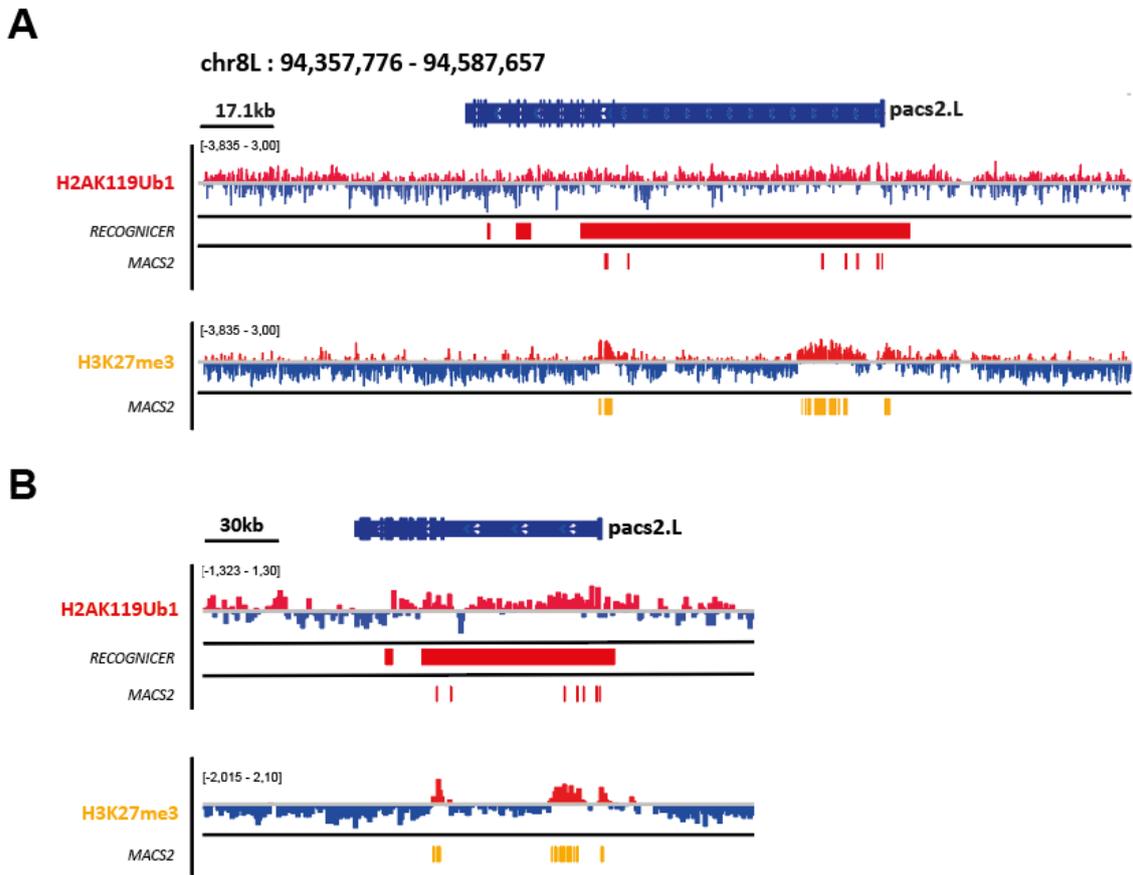
Enfin, j’ai pu attribuer la localisation des régions enrichies à certains gènes selon différents paramètres. Un gène peut être classé positif pour la présence de H2AK119Ub1 si un pic H2AK119Ub1 est localisé autour du TSS de ce gène à plus ou moins (+/-) 5kb (kilobases) ou bien si ce gène est le plus proche d’un enhanceur qui présente un pic H2AK119Ub1. Cette liste de gènes m’a ensuite permis de réaliser de l’enrichissement fonctionnel selon le même protocole que pour le RNA-Sequencing, basé sur l’utilisation d’orthologues humains.

### *Visualisation du signal sur le génome*

La représentation des résultats, pour visualiser les régions enrichies, est possible directement depuis le logiciel IGV (Thorvaldsdottir, Robinson, et Mesirov 2013), qui offre une interface de visualisation du signal sur le génome. deepTools permet de normaliser le nombre de reads entre les fractions ChIP et Input en fonction du nombre total de reads par fraction, suivi du log2 du ratio des deux fractions, à l’aide de la fonction bamcompare.

$$\text{Signal}(\text{bin}) = \log_2 \left( \frac{\text{NbReadsChIP}(\text{bin})/\text{NbReadsChIP}(\text{Echantillon})}{\text{NbReadsInput}(\text{bin})/\text{NbReadsInput}(\text{Echantillon})} \right)$$

Ce signal est ensuite visualisé dans IGV, et les régions enrichies sont mises en évidence avec les pics identifiés précédemment. On observe de cette façon que H2AK119Ub1 est mieux couvert par les régions détectées enrichies avec RECOGNICER plutôt que MACS2 selon différents niveaux de zoom, avec des régions plus grandes identifiées (**Figure 37**).



**Figure 37 : Exemple de pics H2AK119Ub1 détectés par RECOGNICER versus MACS2**

**(A)** Log2(ChIP/Input) de la marque H2AK119Ub1 (en haut) et H3K27me4 (en bas) sur le gène pacs2.L. Les pics de H2AK119Ub1 sont identifiés par RECOGNICER ou MACS2 (en rouge), et les pics de H3K27me3 par MACS2 (en jaune). **(B)** Identique à (A) pour un niveau de zoom différent (x0.57)

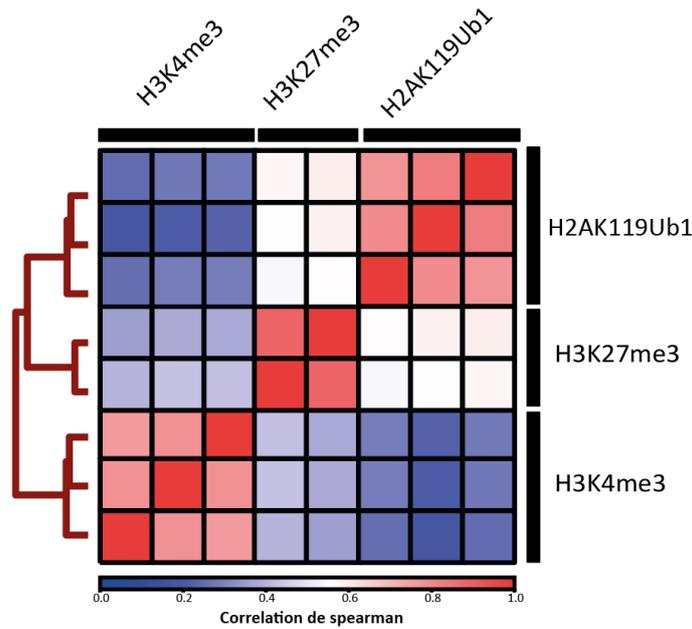
Enfin, le package deepTools contient deux fonctions pour représenter le signal, sous forme de metaplot dans un premier cas avec une visualisation de l'intensité du signal en fonction de la localisation sur le génome par rapport à une référence, comme les pics, les TSSs ou des éléments de régulation du génome. Par exemple, le TSS-plot est un metaplot du signal moyen autour d'une distance prédéfinie d'un subset donné de TSS. La dernière représentation permet de différencier toutes les régions une par une, à l'aide d'une heatmap de visualisation, et d'un algorithme de clustering identifiant des groupes de pics où la dynamique du signal peut suivre différentes tendances.

## *La gestion des réplicats*

Pendant ma thèse, j'ai exploré deux méthodes pour la gestion des réplicats en ChIP-Seq :

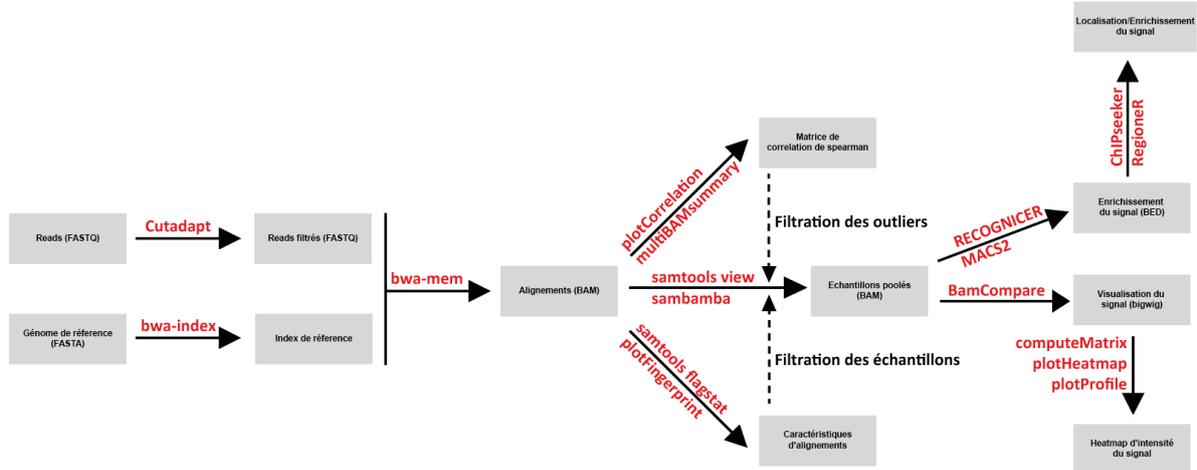
La première méthode consiste à analyser tous les réplicats individuellement, puis de comparer les listes des régions enrichies (pics) pour la marque d'intérêt. La suite bedtools permet de sélectionner un pool de régions communes aux réplicats selon différents paramètres d'overlap prédéfinis (Quinlan et Hall 2010). Ces paramètres divergent selon les analyses, par exemple tous les pics sont parfois conservés (Andrea Scelfo et al. 2019). Néanmoins, cette méthode nécessite d'avoir un pool de pics suffisant pour réaliser l'overlap et devient très rapidement restrictive dans le nombre final de pics conservés. Le nombre final de pics conservés dépendra du seuil de restriction appliqué dans la sélection du nombre d'échantillons minimum où le pic doit être présent. Ce dernier élément est un a priori du résultat dépendant du nombre minimum choisi d'échantillons devant contenir le pic. De plus, cette méthode ne permet pas une représentation facile sur un seul échantillon regroupant les réplicats (heatmap, metaplot, visualisation IGV).

La deuxième méthode consiste ainsi à regrouper les échantillons après avoir vérifié qu'ils corrélaient bien entre eux. Pour cela, on utilise la corrélation de Spearman pour le regroupement des réplicats entre eux. Pour cela, les outils multibamssummary et plotcorrelation de la suite deepTools (Ramírez et al. 2016) permettent de réaliser cette identification en comparant le signal entre échantillons sur des fenêtres de 10kb couvrant l'ensemble du génome (**Figure 38**). Les échantillons sont regroupés selon le type de modifications, et on observe également une plus grande corrélation entre les échantillons H2AK119Ub1 et H3K27me3 qu'avec les échantillons H3K4me3, un résultat en accord avec la colocalisation observée dans les cellules somatiques entre les marques associées aux complexes PRC1 et PRC2. Ensuite, j'ai réalisé un sous-échantillonnage des réplicats. Un nombre de reads identique est sélectionné de façon aléatoire et sans remise (outil samtools) pour chaque réplicat, puis ils sont rassemblés pour générer un échantillon unique (outil : sambamba (Tarasov et al. 2015)). Nous évitons ainsi tout biais qui pourrait provenir de la surreprésentation d'un échantillon. Cette méthodologie est utilisée dans plusieurs études récentes sur H2AK119Ub1 en ChIP-Seq (Blackledge et al. 2020)- (Mei et al. 2021)-(Z. Chen, Djekidel, et Zhang 2021), et est celle que j'ai utilisée ultérieurement pour mon analyse ChIP-Seq (**Figure 39**).



**Figure 38: Heatmap de corrélation de Spearman des réplicats de ChIP-Seq**

Les échantillons sont regroupés selon la corrélation de Spearman calculée sur des fenêtres de 10kb. Les trois réplicats de H3K4me3 et H2AK119Ub1, et les deux réplicats de H3K27me3 corrélient entre eux respectivement, indiquant une bonne corrélation entre les réplicats biologique. Deux échantillons de ChIP-Seq H3K27me3 ont aussi été utilisés lors de l'analyse.

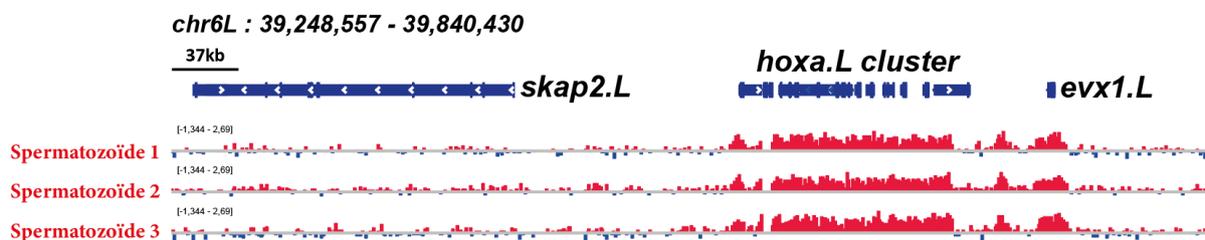


**Figure 39 : Pipeline d'analyse des données de ChIP-Seqencing**

Les reads sont filtrés par cutadapt puis alignés par bwa-mem sur le génome indexé par bwa-index. La qualité de l'alignement est évaluée par samtools flagstat et plotFingerprint, et la qualité des réplicats est évaluée avec plotCorrelation et multiBAMsummary. Après filtration des échantillons de mauvaise qualité, les échantillons sont regroupés par sous-échantillonnage avec samtools view et sambamba. L'enrichissement du signal est analysé avec MACS2 et RECOGNICER puis caractérisé sur R par les packages RegioneR et ChIPseeker. Les données sont représentées à l'aide des fonctions computeMatrix, plotHeatmap et plotProfile de la suite deepTools.

## H2AK119Ub1 est enrichie sur les gènes sensibles au traitement USP21

Récemment, il a été mis en évidence le rôle proéminent de H2AK119Ub1 dans la mise en place du complexe de régulation Polycomb dans les cellules souches embryonnaires (Tamburri et al. 2020). Lors de mon analyse RNA-Seq des gastrulas (**CHAPITRE II**), j'ai mis en évidence une liste de gènes dont l'expression est dérégulée dans la gastrula lors du traitement du spermatozoïde par l'enzyme USP21 ciblant la marque H2AK119Ub1. Dans cette partie, j'ai tout d'abord exploré le profil de la marque H2AK119Ub1 dans le spermatozoïde très peu documenté dans la littérature jusqu'à maintenant, puis j'ai recherché si cette marque était présente préférentiellement sur les gènes sensibles au traitement à l'extrait USP21. Les trois réplicats H2AK119Ub1 ont été analysés selon le protocole bio-informatique décrit précédemment, et fusionné après sous-échantillonnage, une fois la vérification de la corrélation de Spearman supérieure à 0,7 entre les réplicats (**Figure 38**), de leur profondeur de séquençage supérieure aux recommandations ENCODE 3 de 45M (**Figure 33**), et enfin de la visualisation d'un signal à tendance similaire entre les réplicats sur des régions contenant du signal H2AK119Ub1, comme le cluster *hoxa.L* (**Figure 40**).

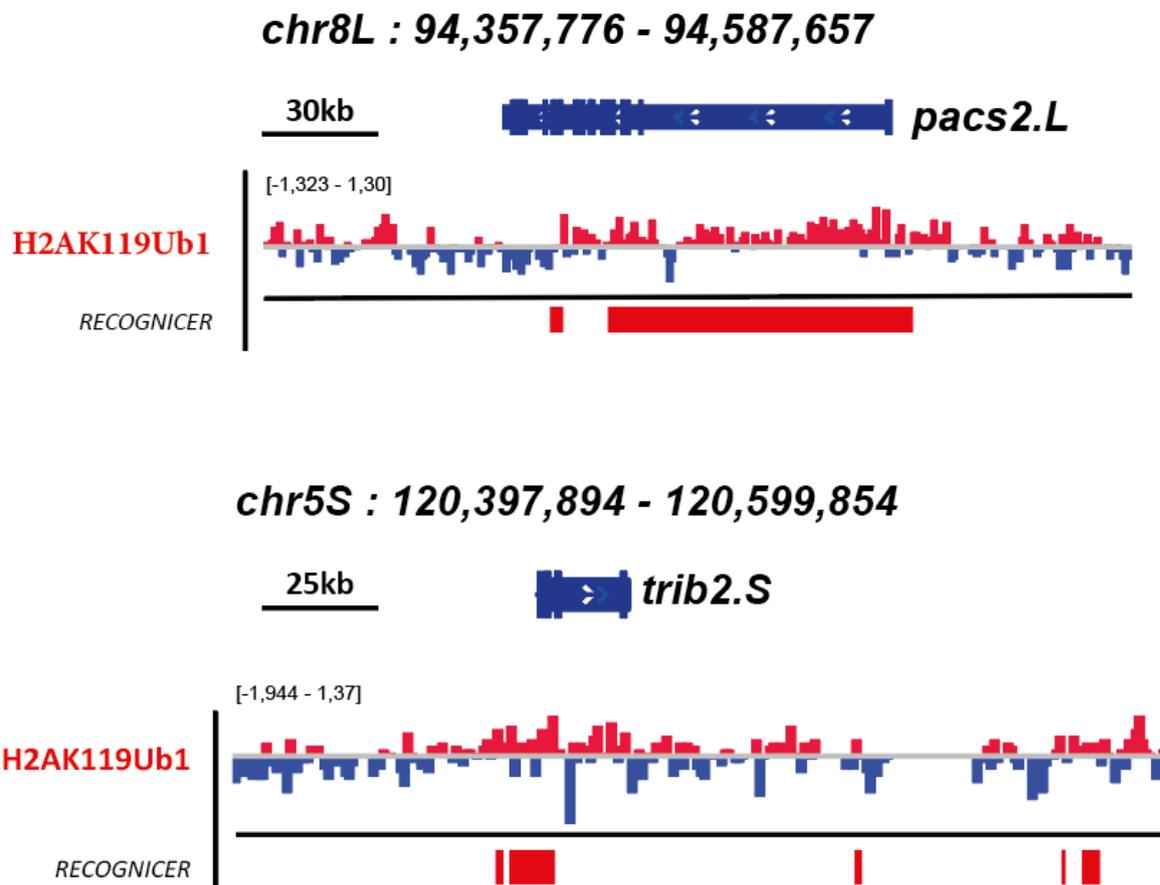


**Figure 40** : Le signal H2AK119Ub1 sur le cluster *hoxa.L* est semblable entre les trois réplicats

*Log2(ChIP/Input)* de la marque H2AK119Ub1 pour les trois réplicats de spermatozoïdes. Le bleu indique une intensité plus importante dans la fraction Input que ChIP, au contraire du rouge indiquant une intensité plus importante dans la ChIP que l'Input. L'échelle de signal *log2(ChIP/Input)* est identique entre les trois réplicats. On observe du signal autour du cluster *hoxa.L* et du gène *evx1.L*, au contraire du gène *skap2.L*. La visualisation est issue du logiciel IGV.

## *H2AK119Ub1 est associé aux régions régulatrices du génome dans le spermatozoïde*

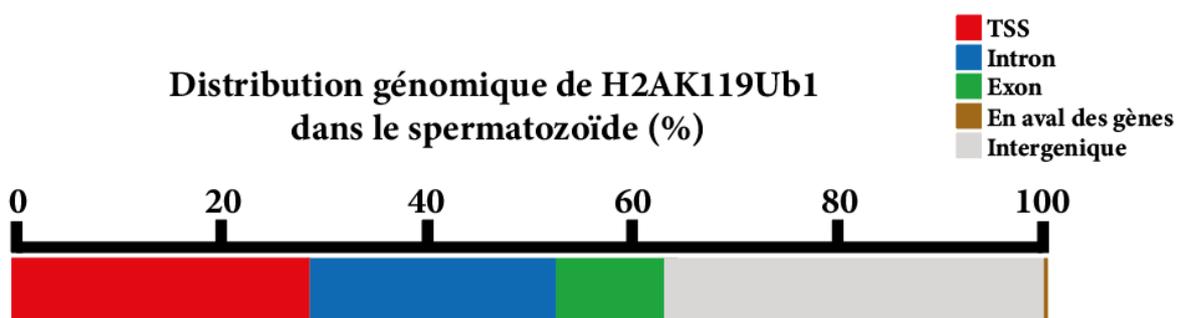
Pour caractériser H2AK119Ub1 dans le spermatozoïde, j'ai tout d'abord évalué sa localisation en identifiant avec l'outil RECOGNICER 67233 régions génomiques, nommées pics, identifiées comme enrichies en H2AK119Ub1. Par exemple, les gènes *pacs2.L* et *trib2.S* contiennent un enrichissement du signal H2AK119Ub1 autour de leurs TSSs détecté par RECOGNICER (Figure 41).



**Figure 41 : Exemple de deux gènes sensibles au traitement USP21, *trib2.S* et *pacs2.L*, présentant un enrichissement pour H2AK119Ub1 dans le spermatozoïde**

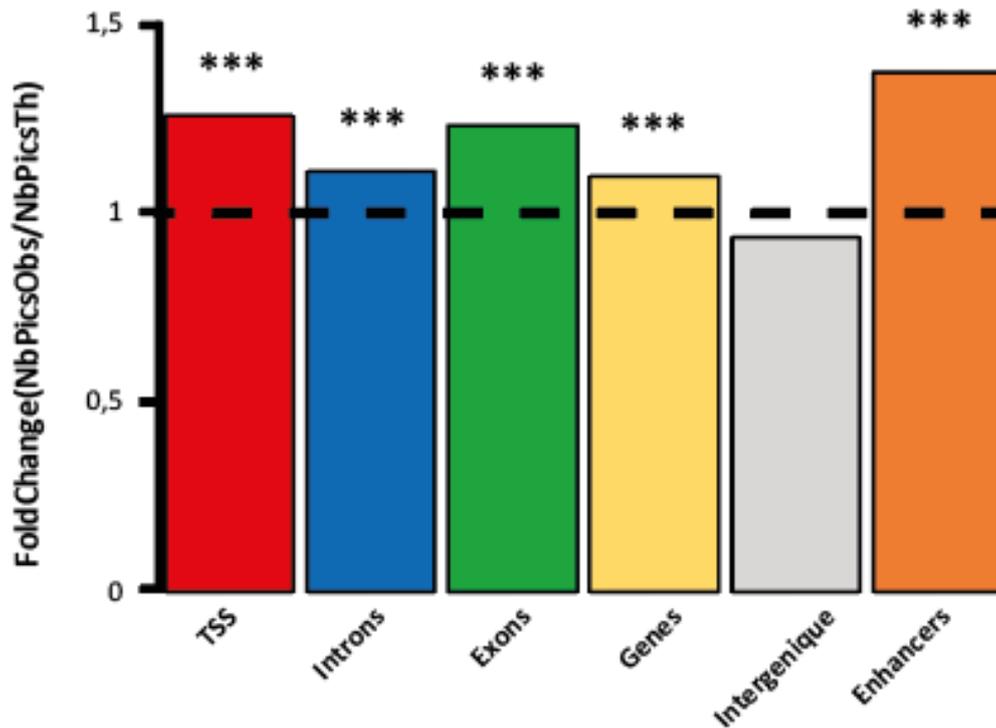
*Log2(ChIP/Input)* de la marque H2AK119Ub1 dans le spermatozoïde est représenté par le  $\log_2$  du ratio entre la fraction ChIP et Input, pour les gènes *trib2.S* et *pacs2.L* identifiés comme sensibles au traitement USP21. L'enrichissement du signal détecté par RECOGNICER est représenté par des blocs rouges.

J'ai mis au point une approche pour identifier la localisation de la marque H2AK119Ub1 sur le génome, à l'aide de ChIPseeker introduit précédemment. On observe qu'approximativement 30% des pics H2AK119Ub1 sont localisés sur des TSSs, et un peu moins de 35% sont localisés sur les introns ou exons des gènes. Le reste des pics est situé entre les gènes (**Figure 42**). J'ai ensuite comparé cette distribution génomique des pics H2AK119Ub1 à celle d'un même groupe de pics aléatoirement distribués pour identifier si cette marque s'associe préférentiellement à certains éléments génomiques dans le spermatozoïde. Ceci est réalisé grâce à des tests de permutation à l'aide du package Regioner. On observe que nous identifions ainsi 1.2 fois plus de pics H2AK119Ub1 localisés sur des TSSs comparé à une distribution théorique décrivant une équiprobabilité d'un enrichissement sur toutes les parties du génome. Cette valeur est représentée par une p-value de permutation significative (N=1000), à une valeur inférieure à 0,001 (**Figure 43**). On observe ainsi que les pics H2AK119Ub1 sont préférentiellement localisés sur les régions régulatrices du génome tel que les TSSs, les enhancers ou les gènes. Enfin, nous avons identifié que la marque H2AK119Ub1 était enrichie sur les TSSs des gènes impliqués dans certains pathways décrit dans la base de données REACTOME, en suivant le même protocole décrit dans la partie RNA-Seq. On observe que H2AK119Ub1 dans le spermatozoïde est présent sur des gènes impliqués dans le développement et des voies de signalisation comme les Tyrosine Kinases, les GTPases, ou encore la voie de signalisation au FGF (**Figure 44**). Ces résultats indiquent qu'une partie des gènes impliqués dans le développement de l'embryon présente un marquage H2AK119Ub1 dans le spermatozoïde.



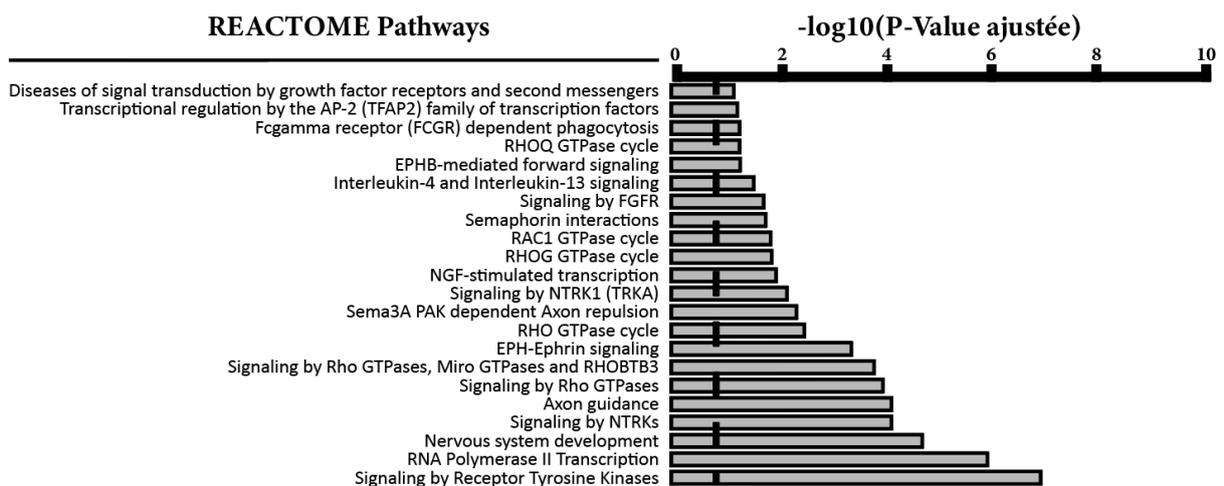
**Figure 42 : Localisation de H2AK119Ub1 sur le génome du spermatozoïde**

Proportion des pics H2AK119Ub1 présents à +/- 5kb des TSSs (en rouge), localisés sur les introns (en bleu), sur les exons (en vert), dans les régions intergénomiques (en gris) et en amont des gènes (en marron).



**Figure 43 : H2AK119Ub1 enrichi sur les régions régulatrices dans le spermatozoïde**

Foldchange du nombre de pics observés (NbPicsObs) par rapport au nombre de pics théorique randomisés (NbPicsTh) issus d'une distribution aléatoire de pics chevauchant d'au moins une paire de base les TSSs +/- 5kb (en rouge), les introns (en bleu), les exons (en vert), les gènes (en jaune), les régions intergéniques (en gris) et les enhancers identifiés par Elurbe et al (en orange). \*\*\*P-Value < 0.001, test de randomisation des pics sur le génome (N=1000).

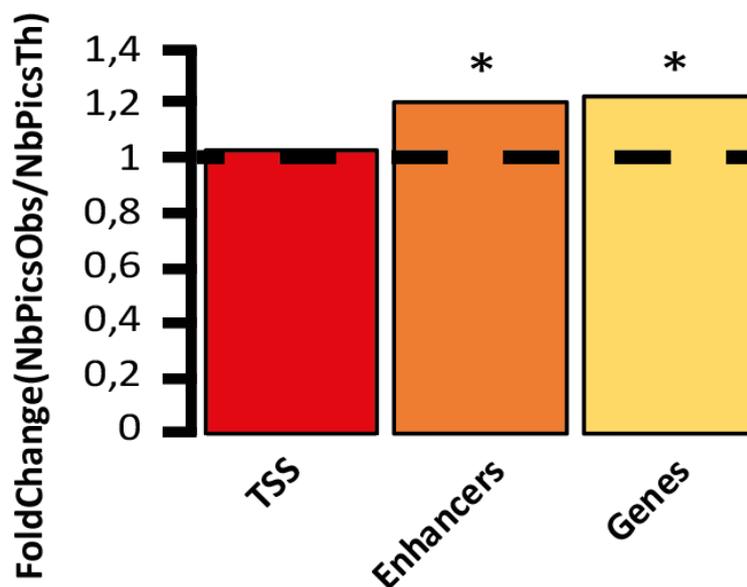


**Figure 44 : Les gènes du développement sont enrichis en H2AK119Ub1 dans le spermatozoïde**

Les orthologues Humain des gènes associés au set des pics H2AK119Ub1 identifiés dans le spermatozoïde sont confrontés à la base de données REACTOME. La p-value ajustée représente la significativité de l'enrichissement des gènes pour les différentes voies biologiques significativement enrichies.

## Les gènes dérégulés par le traitement USP21 montrent un enrichissement pour H2AK119Ub1 dans le spermatozoïde

Nous avons mis en évidence que la marque H2AK119Ub1 était enrichie sur les régions régulatrices dans le spermatozoïde. J'ai ensuite évalué si les gènes sensibles au traitement à l'extrait USP21 présentent une association particulière avec H2AK119Ub1. Pour cela, j'ai adapté la méthodologie de permutation précédente (N=10000) en testant si les pics H2AK119Ub1 ont un overlap statistiquement plus élevé sur les gènes détectés comme sensibles au traitement à l'extrait USP21 par rapport à des gènes distribués de façon aléatoire. J'ai utilisé comme métrique l'overlap d'au moins une paire de base entre un pic et la localisation génomique (TSSs, enhancers, gènes) d'un gène pour considérer un pic associé à un gène, selon la même stratégie qui m'a permis d'identifier que les pics H2AK119Ub1 étaient enrichis sur les régions régulatrices du génome (**Figure 43**). On observe un enrichissement des pics H2AK119Ub1 sur les gènes sensibles au traitement à l'extrait USP21 et leurs enhancers associés avec des p-value d'enrichissement à 0,04 et 0,026 respectivement (**Figure 45**).



**Figure 45 : Test de permutation de l'enrichissement du signal H2AK119Ub1 sur les gènes sensibles au traitement à l'extrait USP21**

Foldchange du nombre de pics (NbPicsObs) chevauchant d'au moins une paire de base sur les gènes sensibles au traitement USP21 par rapport au nombre de pics (NbPicsTh) chevauchant d'au moins une paire de base une distribution aléatoire de gènes issue d'un sous-échantillonnage, autour des TSSs +/- 5kb (en rouge), sur les enhancers identifiés par Elurbe et al (en orange), et sur les gènes (en jaune) sensibles au traitement USP21. \*P-Value < 0.05, test de permutation (N=10000).

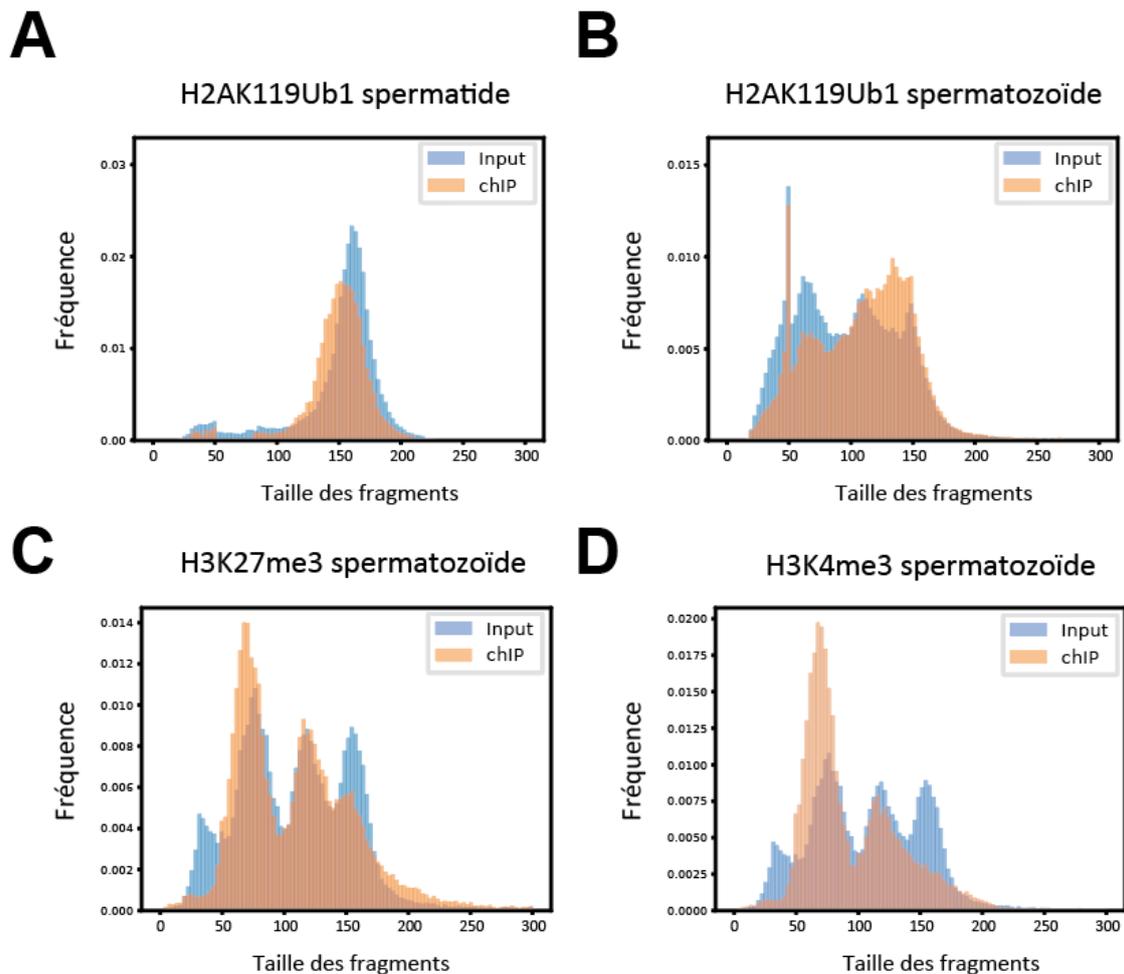
## Les combinaisons de marques épigénétiques associées à H2AK119Ub1 dans la chromatine du spermatozoïde

L'analyse précédente suggère que la débubiquitination de H2AK119Ub1 dans le spermatozoïde est un mécanisme d'action possible pour expliquer l'effet de USP21 sur un groupe de gènes. Cependant, étant donné qu'un nombre important de gènes est également associé à H2AK119Ub1 dans le spermatozoïde sans pour autant être affecté par le traitement USP21, ceci suggère que des facteurs additionnels déterminent la sensibilité à la débubiquitination. Pour explorer cette possibilité, nous avons décidé de déterminer le contexte épigénétique plus large dans lequel H2AK119Ub1 opère en déterminant son association avec d'autres modifications d'histones. Pour cela, nous nous sommes intéressés à H3K27me3, cette marque étant souvent combinée avec H2AK119Ub1 lors de la mise en place du complexe de répression Polycomb (Blackledge et al. 2020). Nous avons aussi étudié la présence de H3K4me3 dans le spermatozoïde. Les gènes du développement présentent un double marquage H3K4me3/H3K27me3 dans les spermatozoïdes (Brykczynska et al. 2010)-(Hammoud et al. 2009)-(S.-F. Wu, Zhang, et Cairns 2011)-(Erkek et al. 2013). H3K4me3 est fortement maintenu durant le développement embryonnaire précoce sur les régions régulatrices du génome, au contraire de H3K27me3 qui est faiblement maintenu, avec l'hypothèse qu'une autre marque comme H2AK119Ub1 pourrait être responsable de la mise en place de mécanismes répressifs (Oikawa et al. 2020).

### *Distribution de H3K4me3 et/ou H3K27me3 dans les régions du génome du spermatozoïde enrichies pour H2AK119Ub1*

Pour cette analyse, j'ai utilisé des données de ChIP-Seq dans le spermatozoïde pour les marques H3K4me3 (trois réplicats), H3K27me3 (deux réplicats), H2AK119Ub1 (trois réplicats), et dans le progéniteur du spermatozoïde pour H2AK119Ub1 (**Figure 31**). Ces réplicats ont à chaque fois été regroupés en un seul réplicat avec la même méthodologie de sous-échantillonnage que celle utilisée précédemment pour H2AK119Ub1 dans le spermatozoïde, après vérification des recommandations ENCODE 3 sur le nombre de reads (**Figure 33**) et de la corrélation entre les réplicats (**Figure 38**). À partir de données en paired-end, nous pouvons déterminer la taille des fragments correspondant à chaque paire de reads.

Ainsi, on observe que les reads des fractions Input présentent des tailles de ~150pb dans le progéniteur (**Figure 46A**, en bleu) correspondant à des nucléosomes alors qu'on observe des fragments à ~150pb, ~110pb et ~70pb dans le spermatozoïde **Figure 46**(**Figure 46B-D**, en bleu) en accord avec la structure en nucléosome/sous-nucléosome de la chromatine du spermatozoïde (**Figure 4**). De façon intéressante, on constate que dans les échantillons de ChIP, cette répartition n'est pas maintenue. En effet, les fragments des tailles spécifiques aux sous-nucléosomes sont surreprésentés dans les ChIP H3K27me3 (**Figure 46C**, en rouge) et H3K4me3 (**Figure 46D**, en rouge). Ceci indique que les marques H3K4me3 et H3K27me3 sont préférentiellement présentes dans le spermatozoïde sur des nucléosomes remodelés avec une perte des sous-unités H2A et H2B. Au contraire, on observe que H2AK119Ub1 est majoritairement localisé sur les fragments à 150pb, soit les nucléosomes non recombines conservant les histones H2A et H2B (**Figure 46B**, en rouge). Ce résultat est logique puisque la présence de H2A est évidemment un prérequis pour la déposition de H2AK119Ub1. Ces observations indiquent donc que globalement lors de la spermiogénèse, les nucléosomes arborant H3K4me3 et H3K27me3 sont préférentiellement remodelés en sous-nucléosome. Par nature, H2AK119Ub1 est au contraire retenue sur des nucléosomes. De façon globale, ceci suggère que H3K4me3 et H3K27me3 sont retenues dans le spermatozoïde sur des nucléoparticules différentes de celles présentant H2AK119Ub1. Au-delà de ces tendances globales, je me suis ensuite intéressé à la combinaison des signaux H3K27me3 et H3K4me3 avec H2AK119Ub1 sur les régions du génome présentant un enrichissement pour ces marques.



**Figure 46 : H2AK119Ub1 et H3K4/K27me3 sont globalement retenus sur différentes particules nucléoprotéiques dans le spermatozoïde**

**(A)** Histogramme de la distribution de la taille des fragments H2AK119Ub1 dans les spermatides.

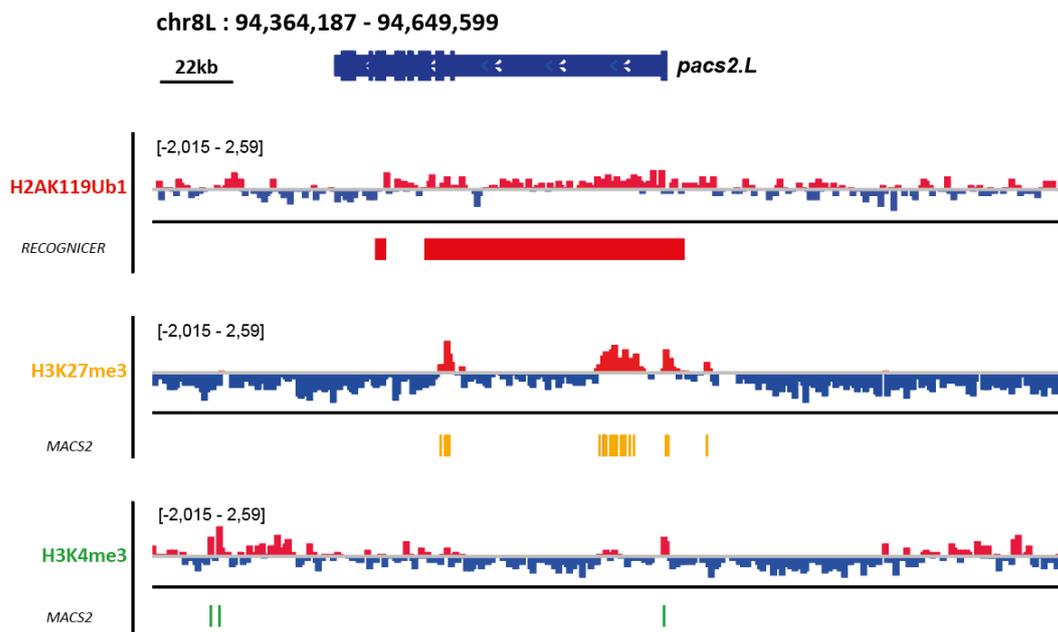
**(B)** Histogramme de la distribution de la taille des fragments H2AK119Ub1 dans les spermatozoïdes.

**(C)** Histogramme de la distribution de la taille des fragments de H3K27me3 dans les spermatozoïdes.

**(D)** Histogramme de la distribution de la taille des fragments de H3K4me3 dans les spermatozoïdes.

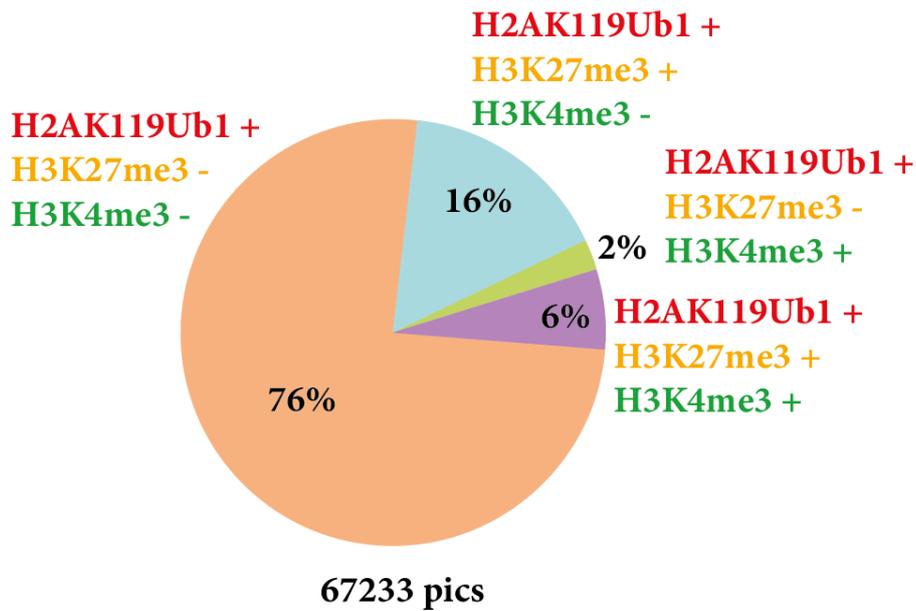
La fraction ChIP (en rouge) est spécifique aux fragments contenant uniquement la marque d'intérêt, au contraire de la fraction Input (en bleu) contenant la répartition des tous les types de nucléosomes dans l'échantillon. Les nucléosomes produisent des fragments de 150bp, alors que les sous-nucléosomes dans le spermatozoïde produisent des fragments de 70 ou 110bp.

On observe par exemple que le gène *pac2.L* sensible au traitement à l'extrait USP21 contient à la fois un enrichissement du signal pour H2AK119Ub1, mais aussi pour H3K4me3 et H3K27me3 (**Figure 47**). Cette observation nous montre que ces trois marques peuvent être combinées. J'ai recherché pour chaque région enrichie en H2AK119Ub1 si elle était aussi identifiée comme enrichie en H3K27me3 et/ou H3K4me3. J'ai ainsi défini l'association de H3K27me3 et/ou H3K4me3 avec H2AK119Ub1 à la présence d'un overlap d'au moins une paire de base entre leurs pics respectifs. On observe que 76% des pics H2AK119Ub1 ne sont pas colocalisés avec H3K4me3 ou H3K27me3, contre 16% des pics H2AK119Ub1 combinés à H3K27me3, 2% combinés à H3K4me3, et 6% qui sont associés à H3K4me3 et H3K27me3 (**Figure 48**). Enfin, les 8% des pics H2AK119Ub1 associés avec H3K4me3 sont localisés majoritairement sur les TSSs à 90%, contrairement à seulement 10% des pics H2AK119Ub1 seuls et 30% des pics H2AK119Ub1 associés à H3K27me3 seul (**Figure 49**).



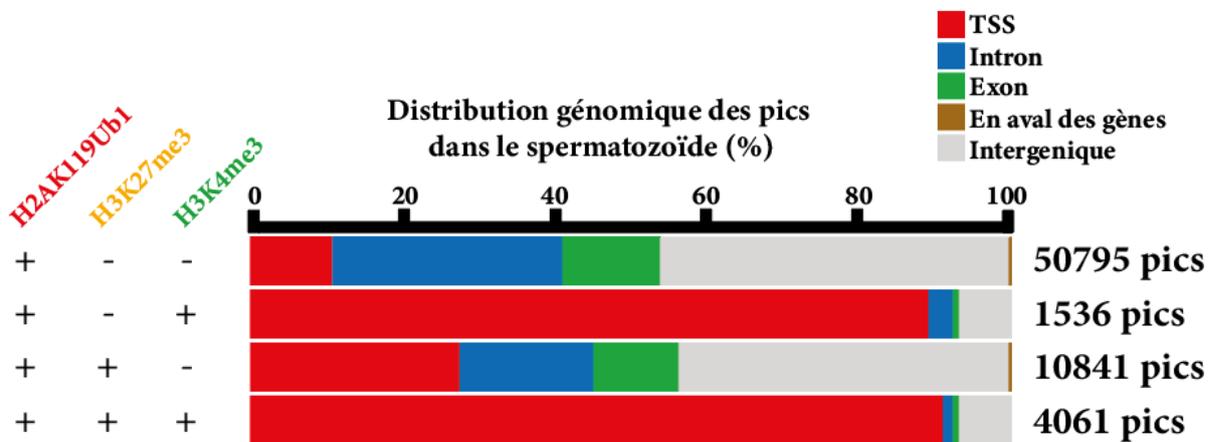
**Figure 47 : H2AK119Ub1, H3K4me3 et H3K27me3 enrichis dans le spermatozoïde sur *pac2.L***

*Log2(ChIP/Input)* des marques H2AK119Ub1, H3K27me3 et H3K4me3 autour du gène *pac2.L* visualisé sur le logiciel IGV. Le signal est mis à la même échelle pour les trois marques. L'enrichissement du signal des différentes marques est représenté par des blocs (H2AK119Ub1 = rouge, H3K27me3 = jaune, H3K4me3 = vert).



**Figure 48 : Proportion des différentes associations possibles entre les pics H2AK119Ub1 et les pics H3K4me3 et/ou H3K27me3 dans le spermatozoïde**

Diagramme circulaire représentant les proportions de pics H2AK119Ub1 seuls (orange), combinés avec H3K4me3 seul (vert), H3K27me3 seul (bleu), ou avec H3K4me3 et H3K27me3 (violet). Le nombre total de pics H2AK119Ub1 est affiché en bas.

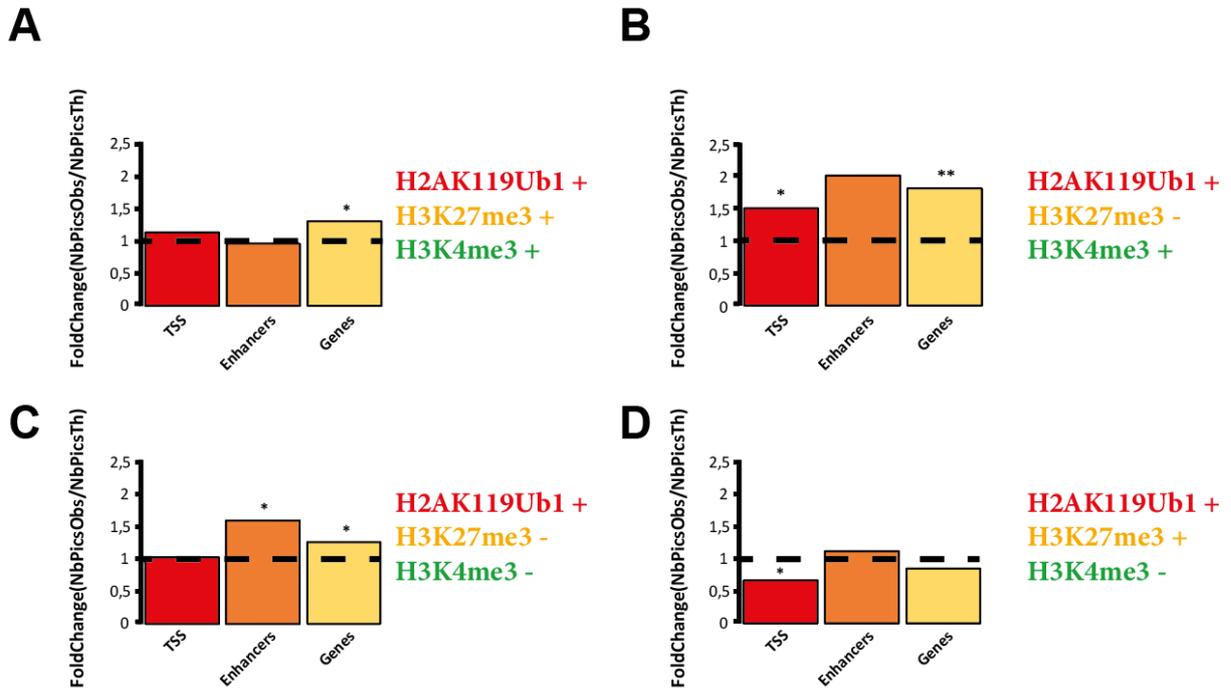


**Figure 49 : Distribution génomique des différentes combinaisons épigénétiques associées à H2AK119Ub1 dans le spermatozoïde**

Proportion des pics H2AK119Ub1 seuls, associés à H3K4me3 seulement, associés à H3K27me3 seulement et associés à H3K4me3/H3K27me3 localisés à +/- 5kb des TSSs (en rouge), ou localisés sur les introns (en bleu), sur les exons (en vert), dans les régions intergéniques (en gris) et en amont des gènes (en marron). Le nombre de pics de chaque combinaison est affiché à droite.

## *H2AK119Ub1 associé avec H3K4me3 est enrichi sur les gènes sensibles au traitement USP21*

L'observation que les pics H2AK119Ub1 sont associés à H3K4me3 sur les TSSs des gènes du spermatozoïde suggère un possible rôle de cette association dans la régulation de l'expression des gènes dans l'embryon. Précédemment, nous avons mis en évidence un enrichissement de H2AK119Ub1 sur les enhancers et le corps des gènes issus des gènes détectés comme sensibles au traitement à l'extrait USP1 (**Figure 45**). Ainsi, j'ai ensuite déterminé si la sensibilité d'un gène au traitement USP21 était associée à une combinaison de H2AK119Ub1 avec les marques H3K4me3 et/ou H3K27me3. À partir de l'extraction des différentes combinaisons de H2AK119Ub1 avec H4K3me3 et/ou H3K27me3 (**Figure 49**), j'ai déterminé pour chaque combinaison s'il existait un enrichissement des pics de ces combinaisons sur les gènes sensibles au traitement à l'extrait USP21 comparé à une distribution aléatoire de gènes avec le test de permutation (N=10000) présenté précédemment (**Figure 50**). On retrouve la même tendance pour les régions enrichies avec seulement H2AK119Ub1 que celle observée pour l'ensemble des pics H2AK119Ub1 (**Figure 45**). Ce résultat est en adéquation avec le fait que ce groupe représente 76% des pics totaux (**Figure 48**). On observe un sous-enrichissement de la combinaison H2AK119Ub1/H3K27me3 sur les TSSs des gènes sensibles au traitement USP21, c'est-à-dire qu'ils sont significativement dépourvus en H2AK119Ub1/H3K27me3 à leurs TSSs dans le spermatozoïde. Au contraire, la combinaison H2AK119Ub1/H3K4me3 est enrichie sur les TSSs des gènes sensibles au traitement USP21, et de façon plus prononcée sur les enhancers et le corps de ces gènes. Enfin, les régions trivalentes montrent un enrichissement sur le corps des gènes sensibles au traitement à l'extrait USP21. Ces résultats suggèrent que la sensibilité des gènes au traitement à USP21 se manifeste dans le contexte d'une association de H2AK119Ub1 à H3K4me3 dans le spermatozoïde. Pour mieux délimiter les caractéristiques des gènes sensibles au traitement USP21, la dernière partie de mon travail a consisté à déterminer le devenir des marques épigénétiques dérivées du spermatozoïde dans l'embryon.



**Figure 50 : H2AK119Ub1 couplé à H3K4me3 est enrichi sur les TSSs des gènes sensibles au traitement USP21**

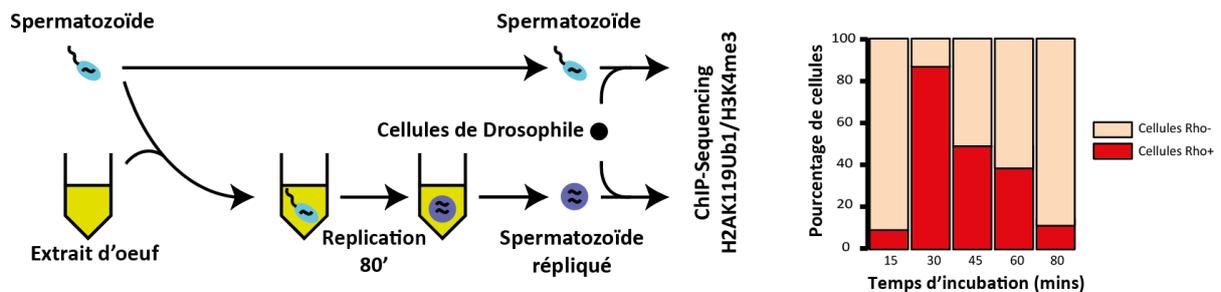
**(A)** Foldchange du nombre de pics H2AK119Ub1/H3K4me3/H3K27me3 (NbPicsObs) chevauchant d'au moins une paire de base les gènes sensibles au traitement USP21 par rapport au nombre de pics (NbPicsTh) chevauchant d'au moins une paire de base une distribution aléatoire de gènes issus par sous-échantillonnage, autour des TSSs +/- 5kb (en rouge), sur les enhancers identifiés par Elurbe et al 2017 (en orange), et sur les gènes (en jaune) sensibles au traitement USP21. **(B)** Identique à (A) pour les pics H2AK119Ub1 associés à H3K4me3 seulement. **(C)** Identique à (A) pour les pics associés à H2AK119Ub1 uniquement **(D)** Identique à (A) pour les pics H2AK119Ub1 associés à H3K27me3 seulement. \*P-Value<0.05, \*\*P-Value<0.01. Test de permutation (N=10000).

# CHAPITRE IV – Après réplication de la chromatine du spermatozoïde, une configuration bivalente H2AK119Ub1/H3K4me3 est maintenue sur les gènes sensibles au traitement USP21

J'observe une corrélation entre la présence d'une configuration H2AK119Ub1/H3K4me3 sur les gènes dans le spermatozoïde et la sensibilité de l'expression embryonnaire de ces gènes au traitement par la H2A deubiquitylase USP21. Je raisonne que si cette information présente sur le spermatozoïde est importante pour la régulation de l'expression des gènes dans l'embryon, cette information doit être transmise lors des 8 divisions cellulaires qui précèdent l'initiation de la ZGA chez le xénope. Dans cette partie, je vais évaluer la possibilité d'une transmission du signal H2AK119Ub1, en l'état, du spermatozoïde à l'embryon en développement. Pour réaliser cette analyse, nous avons utilisé des extraits d'œuf qui permettent de mimer les étapes de réplifications de la chromatine qui débutent à la fécondation (Gaggioli et al. 2013). J'assume que le devenir des marques épigénétiques du spermatozoïde après réplication dans l'extrait d'œuf représente l'état de la chromatine transmise durant la phase pré-ZGA de l'embryogenèse. Ce postulat est valide par une analyse précédente qui a démontré une très forte similitude entre les pics d'histones H3K4me3 après traitement du spermatozoïde par un extrait d'œuf et ceux d'une blastula pré-ZGA (Oikawa et al. 2020). Avec cette approche, j'ai d'abord réalisé une analyse ChIP-Seq classique du comportement de H2AK119Ub1 et de H3K4me3, au cours de la réplication à partir des pics avant et après la première réplication. Pour élargir notre vision du comportement de H2AK119Ub1/H3K4me3 au cours du développement, j'ai également utilisé des données de ChIP-Seq réalisées sur des embryons post-ZGA (gastrula stade 12, obtenus par sonication de la chromatine (Kuznetsov et al. 2019)) et des spermatides afin de déterminer la dynamique de ces marques de la spermatide jusqu'à l'embryogenèse précoce. Dans un second temps, j'ai réalisé une analyse quantitative qui vise à déterminer comment la marque est transmise durant la réplication, c'est-à-dire si elle est transmise sur aucune, une seule ou les deux copies. Ceci n'est pas mesurable par un protocole de ChIP-Seq classique, nous avons ainsi ajouté des cellules de *Drosophila melanogaster* dans nos échantillons pour réaliser une ChIP-seq normalisée.

## Réplication de la chromatine du spermatozoïde par incubation dans un extrait d'œuf.

Ces expériences ont été réalisées par Florian Berger en collaboration avec Romain Gibeaux (IGDR, Rennes, production d'extrait d'œuf). Les spermatozoïdes sont incubés à une concentration de 2000 noyau/ $\mu\text{l}$  d'extrait d'œuf (Kitaoka, Heald, et Gibeaux 2018) et la réplication est évaluée en réalisant à différents temps d'incubation, une impulsion de 5 minutes avec de l'UTP-Rhodamine, un précurseur fluorescent de la synthèse ADN (**Figure 51**). On observe un maximum de noyau en réplication à 30 mins d'incubation puis une baisse progressive jusqu'à 80 mins, suggérant la complétion de la réplication lorsque les noyaux sont prélevés pour analyse ChIP-Seq.



**Figure 51 : Réplication de L'ADN du spermatozoïde après incubation dans un extrait d'œuf**

**(A)** Le spermatozoïde est incubé dans un extrait d'œuf pendant 80 minutes. Une fraction des échantillons est soumise à une impulsion d'UTP-Rhodamine à différents temps d'incubation. **(B)** Analyse en microscopie à fluorescence (bargraph) pour déterminer le pourcentage de noyau ayant incorporé l'UTP-Rhodamine, indiquant un processus de réplication.

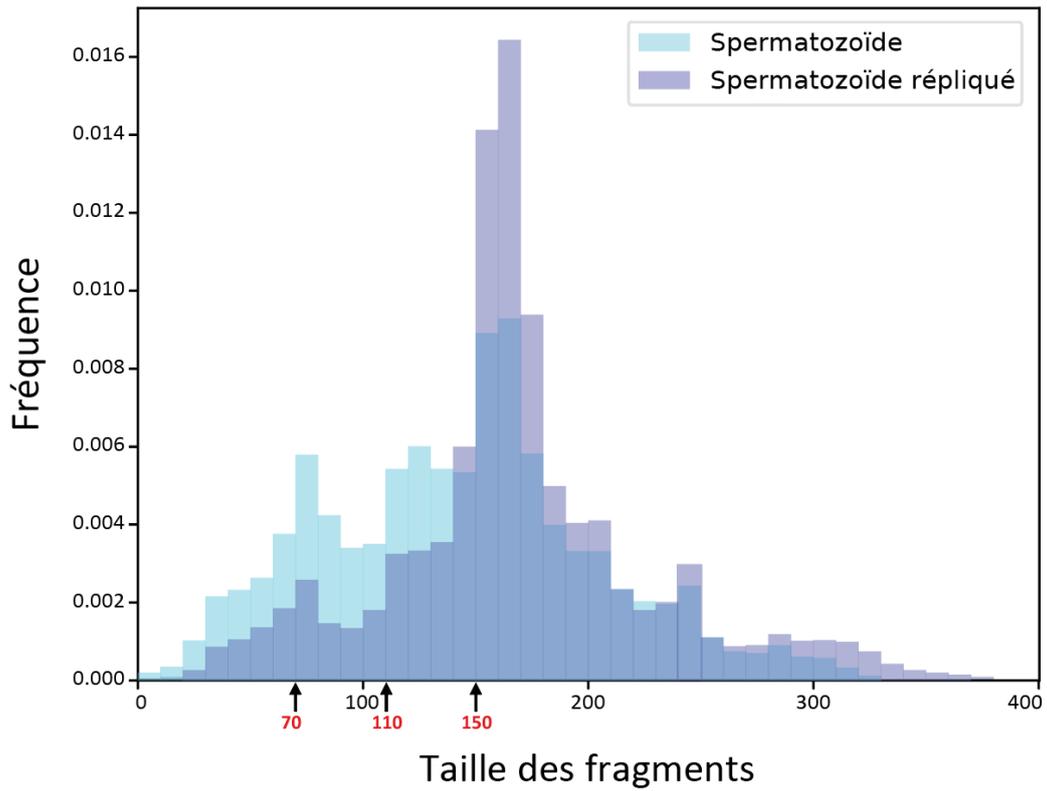
Le jeu de données généré pour l'analyse correspond à trois réplicats H3K4me3 et trois réplicats H2AK119Ub1 pour le spermatozoïde avant et après réplication, auquel est ajouté 200 000 cellules de drosophile pour 1 000 000 de spermatozoïdes. Nous avons aussi des réplicats des trois marques H2AK119Ub1, H3K4me3 et H3K27me3 dans la spermatide et la gastrula après ZGA (**Figure 52**).

Spermatoïde	Spermatozoïde avec cellules Drosophila	Spermatozoïde répliqué dans l'extrait d'oeuf avec cellules Drosophila	Gastrula (Stade 12)
3x H2AK119Ub1 (SE)	3x H2AK119Ub1 (PE)	3x H2AK119Ub1 (PE)	2x H2AK119Ub1 (SE)
2x H3K4me3 (SE)	3x H3K4me3 (PE)	3x H3K4me3 (PE)	2x H3K4me3 (SE)
3x H3K27me3 (SE)			2x H3K27me3 (SE)

**Figure 52 : Présentation des échantillons de H2AK119Ub1, H3K4me3 et H3K27me3**

Le nombre de réplicat et le mode de séquençage de chacun est indiqué (PE = Paired-End, SE=Single-End). En plus des échantillons présentés précédemment (**Figure 31**), nous possédons trois réplicats H2AK119Ub1 et H3K4me3 séquencés en paired-end dans le spermatozoïde répliqué avec des cellules de drosophile. Les échantillons utilisés pour l'analyse des phénomènes associés à la réplication sont en rouge. Nous avons aussi à disposition deux réplicats dans la gastrula pour H3K4me3, H2AK119Ub1 et H3K27me3, tous séquencés en single-end et issus de Kuznetsov et al 2019 après sonication de la chromatine.

L'analyse de la taille des fragments d'ADNs obtenus par digestion MNase confirme également que l'incubation dans l'extrait d'œuf a bien rétabli une organisation canonique de la chromatine avec disparition des sous-nucléosomes caractéristiques du spermatozoïde (70pb et 110pb) au profit des nucléosomes (150pb) après la première réplication (**Figure 53**), comme attendu (**Figure 13**). Ces observations suggèrent que les échantillons de spermatozoïdes incubés dans des extraits d'œufs ont bien subi une phase de réplication et peuvent être utilisés pour l'analyse du comportement des modifications d'histones.



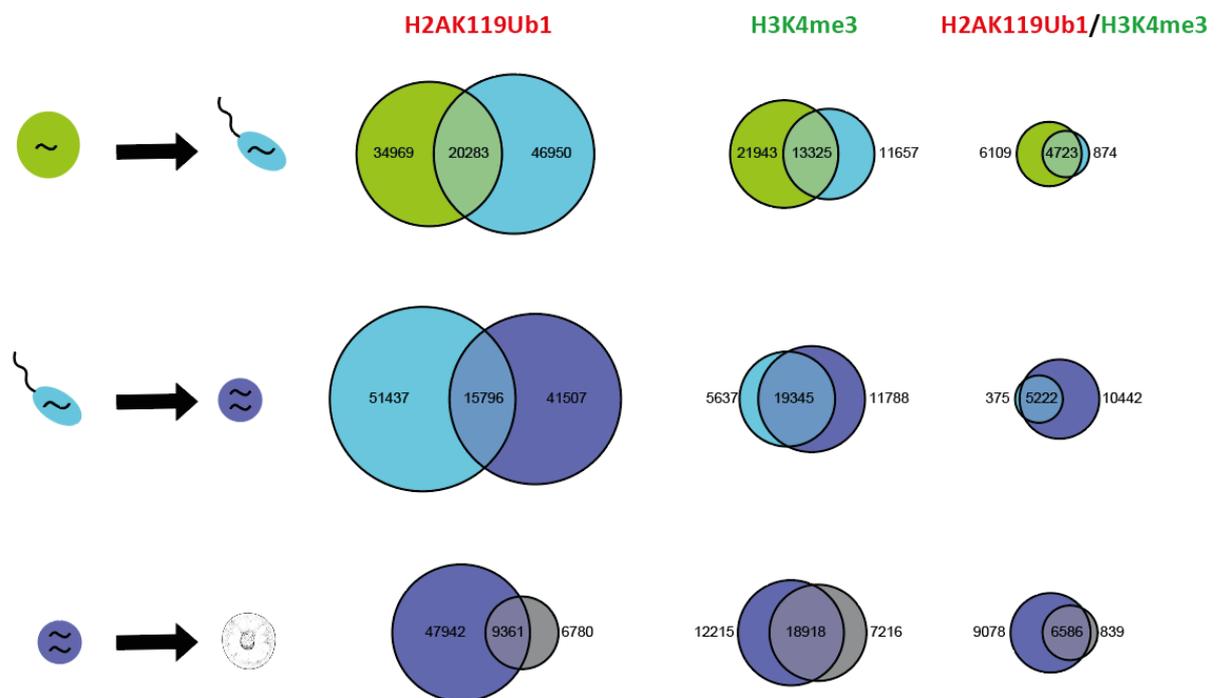
**Figure 53 : L'organisation canonique de la chromatine est rétablie après la réplication**

*Histogramme de la distribution de la taille des fragments dans le spermatozoïde avant réplication (cyan) et après réplication (violet). Les tailles de fragments correspondant aux nucléosomes (150pb) et sous-nucléosomes (70 et 110pb) sont affichées en rouge.*

## Le devenir de H2AK119Ub1 et H3K4me3 après la première réplication

*La réplication de la chromatine du spermatozoïde est associée à un remodelage extensif de la distribution de H2AK119Ub1 et à une stabilité de celle de H3K4me3*

J'ai tout d'abord comparé la distribution de ces marques dans le spermatozoïde avant et après incubation dans l'extrait d'œuf. J'ai identifié les zones d'accumulation de H2AK119Ub1 et H3K4me3 par une analyse de pics en ChIP-Seq classique. Dans un premier temps, je me suis intéressé à l'overlap des régions détectées comme enrichies (pics) avant ou après la première réplication. 23,5% des pics H2AK119Ub1 du spermatozoïde sont conservés après réplication, alors que les pics H3K4me3 sont conservés à 77,4% des pics (**Figure 54**).

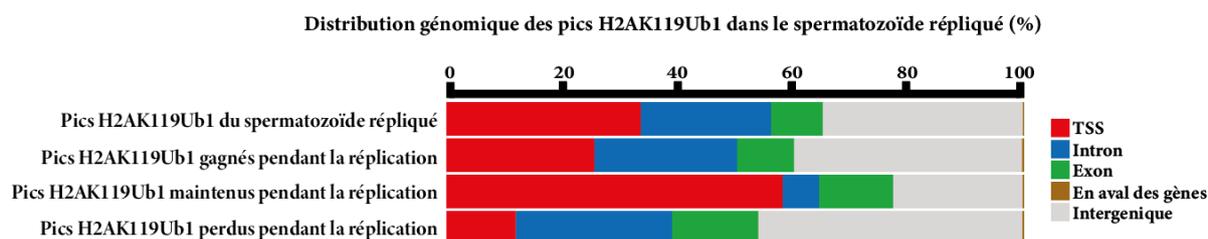


**Figure 54 : H2AK119Ub1 est activement remodelé pendant le développement embryonnaire**

Diagramme de Venn de la dynamique des pics H2AK119Ub1, H3K4me3 et bivalents pour H2AK119Ub1/H3K4me3 détectés lors de la spermiogénèse (spermatide = vert, spermatozoïde = cyan), de la première réplication (spermatozoïde = cyan, spermatozoïde répliqué = violet) et lors de la ZGA (spermatozoïde répliqué = violet, gastrula = gris). L'échelle des diagrammes est commune.

Ce résultat pour H3K4me3 est en adéquation avec une précédente analyse ayant démontré que 75% des pics H3K4me3 étaient conservés après la première réplication (Oikawa et al. 2020). De façon générale, la réplication de la chromatine du spermatozoïde s'accompagne donc d'une stabilité du signal H3K4me3 et d'un remodelage extensif du signal H2AK119Ub1 avec plus de 75% de perte des pics H2AK119Ub1 durant la réplication (51437/67233 pics) et le gain de nombreux pics (41 507 pics présents uniquement dans le spermatozoïde répliqué). Enfin, on observe que 93,3% (5222/5597) des pics bivalents H2AK119Ub1/H3K4me3 du spermatozoïde sont conservés lors de la première réplication (**Figure 54**), indiquant que le statut de bivalence H2AK119Ub1/H3K4me3 présent dans le spermatozoïde est très fortement retenu post-réplication. La spermiogénèse (transition spermatide/spermatozoïde) montre un fort remodelage de H2AK119Ub1 et de H3K4me3. De façon notable, on observe cependant qu'une large majorité des régions bivalentes du spermatozoïde sont héritées de la spermatide. Enfin, l'embryon post-ZGA semble caractérisé par une réduction du nombre de régions marquées pour l'ensemble des marques considérées.

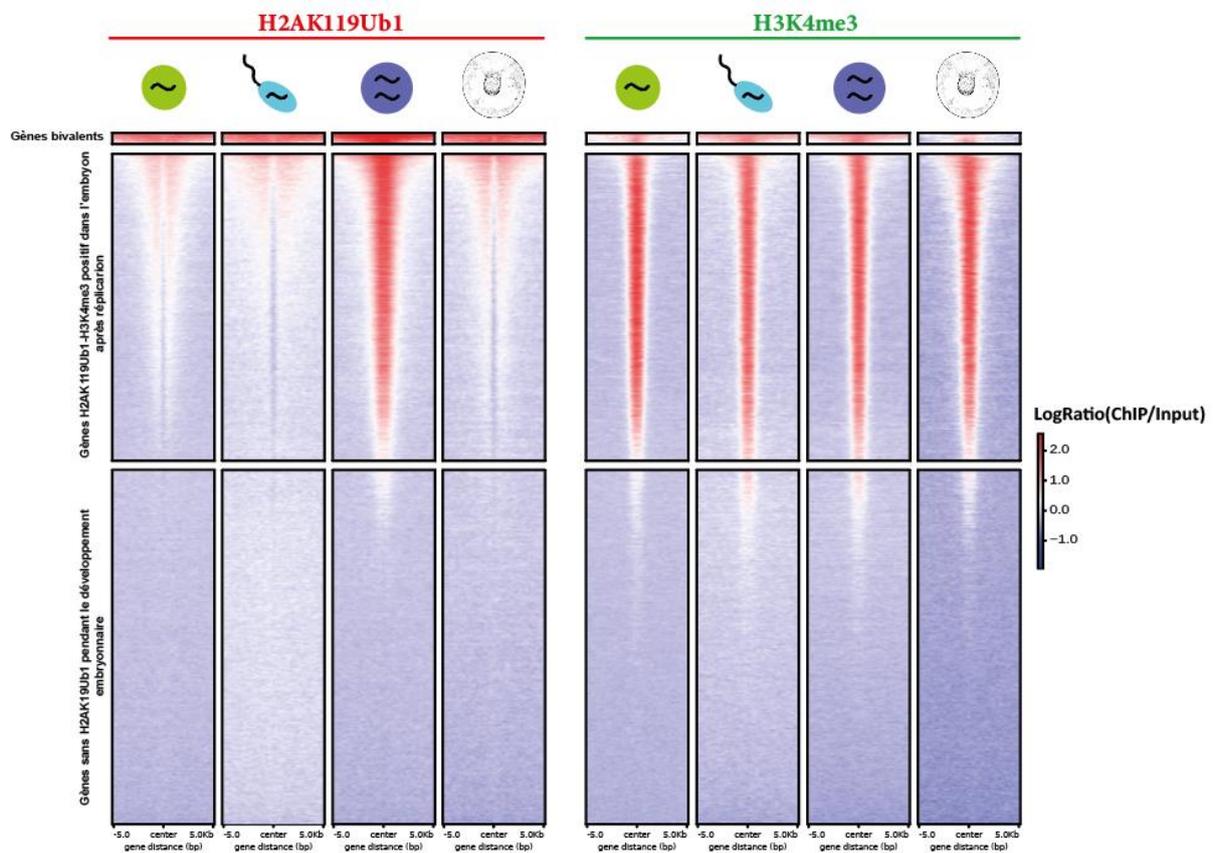
Nous avons mis en évidence précédemment un enrichissement de H2AK119Ub1 seul ou avec H3K4me3 dans le spermatozoïde sur les régions régulatrices des gènes (TSS, enhancers). J'ai donc évalué si le remodelage observé de H2AK119Ub1 durant la réplication s'opère sur ces éléments de régulation. Malgré une proportion minoritaire de pics H2AK119Ub1 retenus lors de la première réplication, on observe que ces pics sont majoritairement situés au niveau des TSSs des gènes (~60%), au contraire des pics perdus (~10%) ou des pics gagnés (~25%) durant la première réplication (**Figure 55**).



**Figure 55 : Les pics H2AK119Ub1 maintenus pendant la réplication sont localisés sur les TSSs**

Le pourcentage de pics sur différentes thématiques génomique (TSS en rouge, intron en bleu, exon en vert, intergénique en gris, en amont du gène en marron) est mis en évidence pour l'ensemble des pics H2AK119Ub1 dans le spermatozoïde après la première réplication, puis pour ces pics séparés selon si H2AK119Ub1 est perdu, conservé ou gagné durant la première réplication. Le nombre de pics dans chaque groupe est indiqué à droite. La distribution génomique du spermatozoïde est disponible dans la **Figure 42**.

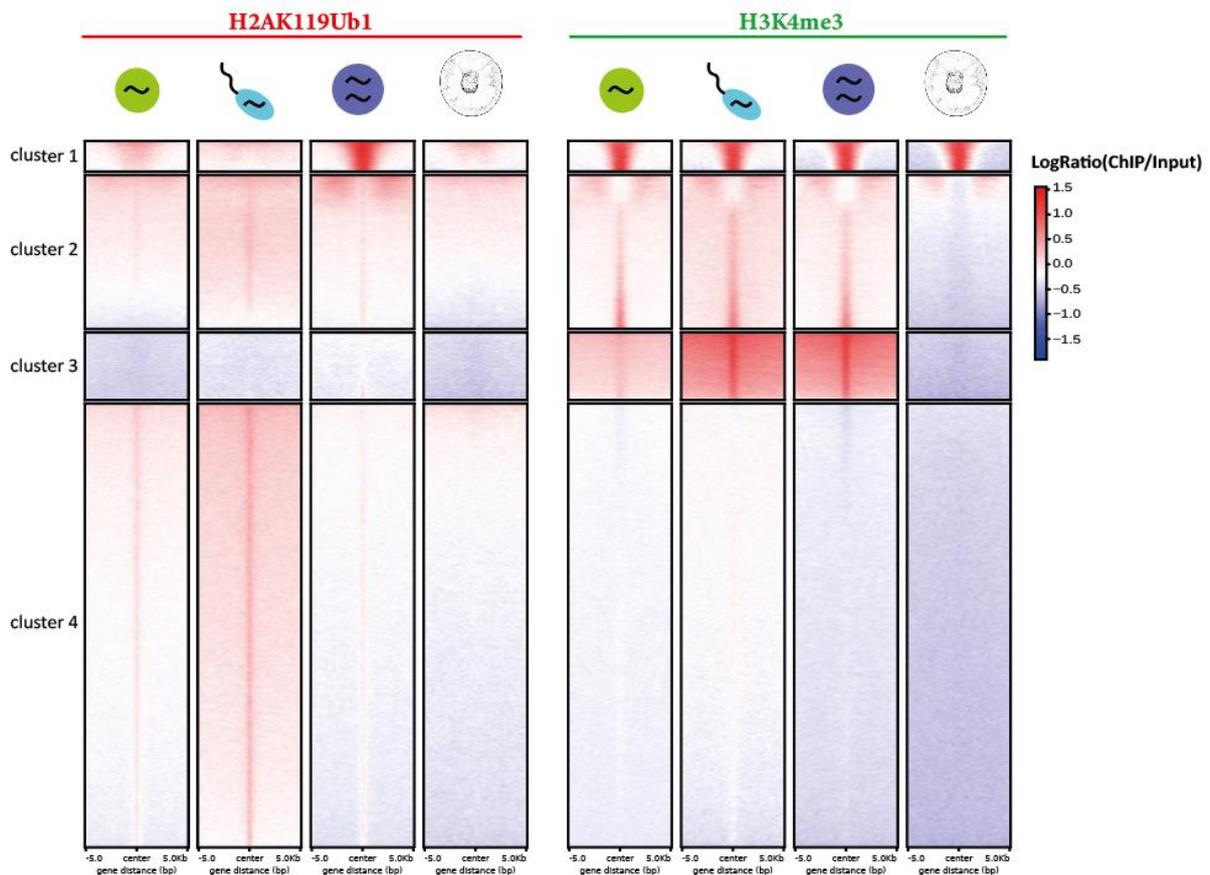
La visualisation du signal H2AK119Ub1 et H3K4me3 sur les TSSs des gènes pendant le développement est présentée en **Figure 56**. Les gènes ont été organisés avec un clustering hiérarchique non supervisé. On observe 760 gènes dans le cluster le plus haut, soit ~1,5% des gènes qui semblent être caractérisés par une bivalence pour H2AK119Ub1/H3K4me3 de la spermatide jusqu'à la ZGA, où le signal H3K4me3 semble plus restreint. Cette bivalence sur les TSSs se caractérise par un domaine d'enrichissement de H2AK119Ub1 sur un large domaine autour du TSS. De plus, on observe dans le cluster central que H2AK119Ub1 serait gagné sur 16 473 gènes post-réplication, soit ~36,5% des gènes, en adéquation avec l'observation de la **Figure 55**. Ces analyses indiquent donc que le remodelage intensif de la distribution de H2AK119Ub1 au cours de la réplication du spermatozoïde se caractérise par un maintien sélectif de cette marque sur les TSSs ainsi que par son association avec H3K4me3.



**Figure 56 : H2AK119Ub1 est déposé sur les TSSs après la première réplication**

*Heatmap du  $\log_2(\text{ChIP}/\text{Input})$  de H2AK119Ub1 et H3K4me3 dans la spermatide, le spermatozoïde, le spermatozoïde après réplication, et la gastrula (colonnes) sur tous les gènes (rangées). Les gènes sont regroupés avec un clustering hiérarchique non supervisé (hclust).*

Enfin, après cette caractérisation des TSSs (**Figure 56**), j'ai caractérisé plus en détail la cinétique des marques H2AK119Ub1 et H3K4me3 en élargissant l'analyse à l'ensemble des pics identifiés sur cette période de développement (**Figure 57**). Dans cet objectif, j'ai représenté le signal H2AK119Ub1 et H3K4me3 (**Figure 57**) au cours du développement, pour chaque set de pics identifiés à une étape donnée du développement. Par exemple, si l'on considère la première rangée de la **Figure 57**, représentant les pics H2AK119Ub1 détectés dans la spermatide, on constate que le signal H2AK119Ub1 est maintenu de la spermatide à la gastrula uniquement sur la fraction des pics qui présentent un marquage relativement étendu de H2AK119Ub1 et sont colocalisés avec H3K4me3. Cette représentation confirme une dynamique forte de H2AK119Ub1 avec de nombreuses régions présentant un signal uniquement à un stade de développement et en absence de H3K4me3. En outre, ceci confirme également la stabilité de H2AK119Ub1 sur les régions où une colocalisation avec H3K4me3 est observée.



**Figure 57 : Le signal H2AK119Ub1 est stabilisé dans l'embryon lorsqu'il est couplé à H3K4me3**

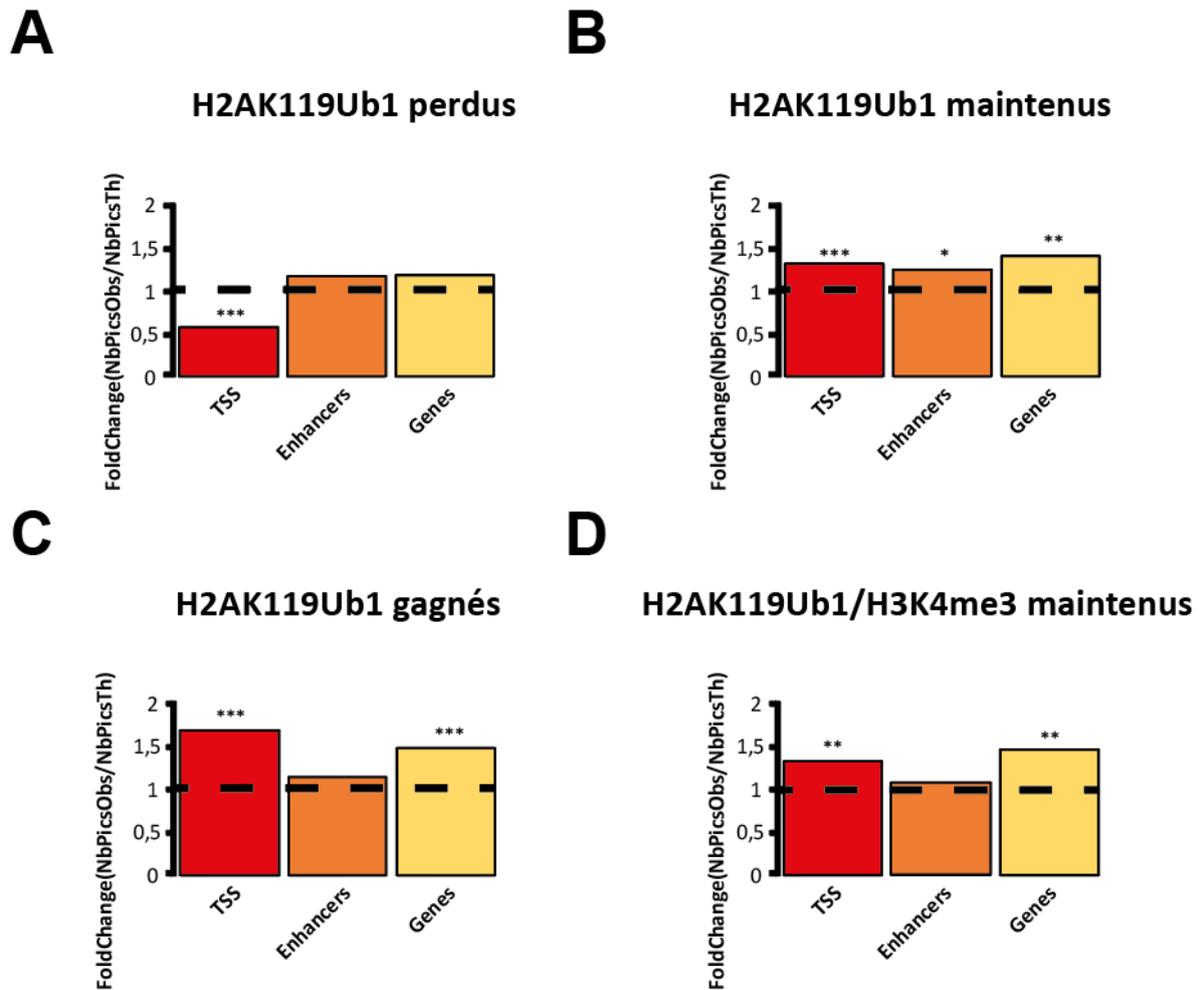
Heatmap du  $\log_2(\text{ChIP}/\text{Input})$  de H2AK119Ub1 et H3K4me3 dans la spermatide, le spermatozoïde, le spermatozoïde après réplication, et la gastrula (colonnes) pour les sets de pics identifiés à chaque stade (rangées). Le nombre de pics est affiché à droite pour chaque set (rangées) caractérisé dans la **Figure 54**.

## *H2AK119Ub1 après réplication est préférentiellement associé avec les gènes sensibles au traitement USP21.*

Nous avons précédemment montré que les gènes sensibles au traitement à l'extrait USP21 étaient enrichis pour la bivalence H2AK119Ub1/H3K4me3 sur leurs TSSs et les gènes en entier. Nous avons aussi mis en évidence que les pics bivalents H2AK119Ub1/H3K4me3 étaient conservés durant la transmission à l'embryon. Ainsi, dans cette partie, nous allons investiguer comment la bivalence H2AK119Ub1/H3K4me3 est transmise à l'embryon du côté paternel, notamment sur les gènes sensibles au traitement à l'extrait USP21.

Dans un premier temps, j'ai testé l'enrichissement des pics H2AK119Ub1, dans le spermatozoïde répliqué, sur les gènes sensibles au traitement à l'extrait USP21. Les pics ont été catégorisés en plusieurs groupes selon leur dynamique d'apparition, disparition ou de maintenance durant la première réplication (**Figure 55**). La mesure utilisée est à nouveau le nombre d'overlap des pics avec les gènes à l'aide d'un test de permutation (N=10000). On observe que les pics H2AK119Ub1 perdus durant la première réplication sont sous-enrichis sur les TSSs des gènes sensibles au traitement à l'extrait USP21 (**Figure 58A**). Au contraire, ce set de gènes est caractérisé par un maintien et/ou un gain de H2AK119Ub1 sur le TSS durant la première réplication (**Figure 58B-C**), ainsi qu'un maintien de H2AK119Ub1 sur leur enhancers.

J'ai ensuite évalué le devenir de la bivalence H2AK119Ub1/H3K4me3 dans le spermatozoïde après la première réplication. Les gènes sensibles au traitement USP21 sont enrichis pour la bivalence H2AK119Ub1/H3K4me3 dans le spermatozoïde, qui est conservée lors de la première réplication (**Figure 58D**). L'ensemble de ces résultats indiquent que l'association H2AK119Ub1/H3K4me3 avec les gènes sensibles au traitement USP21 est maintenue post-réplication.



**Figure 58 : Les gènes sensibles au traitement USP21 maintiennent H2AK119Ub1 après réplication de la chromatine du spermatozoïde**

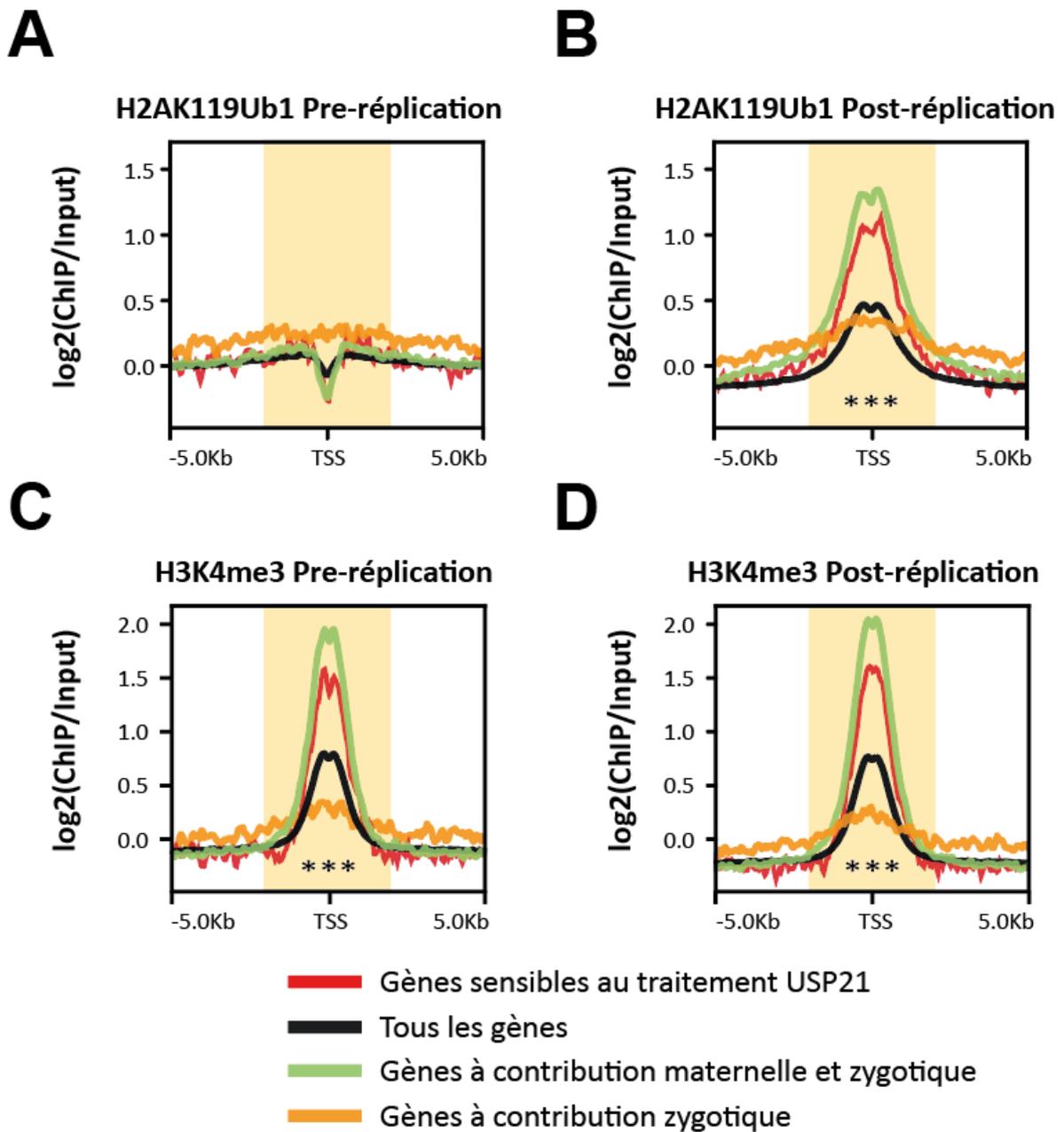
**(A)** Foldchange du nombre de pics H2AK119Ub1 (NbPicsObs) chevauchant d'au moins une paire de base les gènes sensibles au traitement USP21 par rapport au nombre de pics (NbPicsTh) chevauchant d'au moins une paire de base une distribution aléatoire de gènes issus par sous-échantillonnage, autour des TSSs +/- 5kb (en rouge), sur les enhancers identifiés par Elurbe et al (en orange), et sur les gènes (en jaune) sensibles au traitement USP21. **(B)** Identique à (A) pour les pics H2AK119Ub1 maintenus pendant la réplication. **(C)** Identique à (A) pour les pics H2AK119Ub1 gagnés pendant la réplication **(D)** Identique à (A) pour les pics H2AK119Ub1 associés à H3K4me3 et maintenus pendant la réplication. P-Value<0.05, \*\*P-Value<0.01, \*\*\*P-Value<0.001. Test de permutation (N=10000)

Pour compléter l'analyse basée sur la détection de site d'enrichissement du signal H2AK119Ub1, j'ai également calculé directement le signal ChIP/Input sur les TSSs afin de comparer cette mesure du groupe de gènes sensible au traitement USP21 à l'ensemble des gènes. À partir des metaplots réalisés par la fonction plotProfile, j'ai obtenu les valeurs d'intensité pour chaque fenêtre de 50pb sur la région de 2kb autour du TSS. J'ai comparé sur

ces fenêtres les distributions des intensités entre les différents groupes de gènes avec un test non paramétrique de Wilcoxon-Mann-Whitney.

On observe que les TSSs des gènes sensibles au traitement USP21 présentent pré-réplication un signal H2AK119Ub1 similaire à celui de l'ensemble des gènes (**Figure 59A**). En revanche post-réplication, les gènes sensibles au traitement USP21 montrent un signal H2AK119Ub1 plus élevé que la moyenne (courbe bleue et verte). De façon notable, il semble que le profil des gènes sensibles à USP21 présente un signal H2AK119Ub1 similaire aux gènes à contribution maternelle zygotique (H. Chen et Good 2022) (**Figure 59B**, courbe bleue et rouge). H3K4me3 est présent à une intensité significativement supérieure sur les gènes sensibles au traitement USP21 dans le spermatozoïde avant et après réplication, avec un signal H3K4me3 similaire aux gènes à contribution maternelle et zygotique à nouveau (**Figure 59C-D**). Ces résultats concordent avec les résultats précédents indiquant un maintien/gain de pics H2AK119Ub1 sur les gènes sensibles au traitement USP21 et un maintien de la bivalence H2AK119Ub1/H3K4me3.

En conclusion, nous observons que la présence de H2AK119Ub1 sur les régions régulatrices de gènes dans le spermatozoïde est associée à la sensibilité au traitement USP21. Cet effet de USP21 ne se manifeste que sur une fraction des gènes marqués H2AK119Ub1. Ceux-ci sont caractérisés par une co-occurrence H2AK119Ub1/H3K4me3 dans le spermatozoïde, qui semble être maintenue préférentiellement sur ces gènes post-réplication. Cependant, ces données ne permettent pas de comparer l'intensité du signal entre les échantillons. Par exemple, on ne peut pas comparer l'échelle de valeur du signal H2AK119Ub1 pré-réplication et post-réplication. Pour notre compréhension mécanistique de l'effet de H2AK119Ub1, ceci est limitant. En effet, il est possible que post-réplication, les gènes sensibles au traitement USP21 soit plus fortement marqués que la moyenne, mais que ce marquage soit cependant plus faible que celui observé pré-réplication (**Figure 59**). Ceci pourrait avoir lieu dans le cadre d'une dilution globale des marques H2AK119Ub1 au cours de la réplication. Cela suggère que H2AK119Ub1 dérivé du spermatozoïde pourrait être continuellement dilué lors des divisions embryonnaires, ne permettant pas de transmettre directement l'information jusqu'à la ZGA, après ~8 cycles de divisions cellulaires et pourrait par exemple requérir un mécanisme de relai. Pour pouvoir distinguer entre ces possibilités, nous avons appliqué une CHIP normalisée dont l'objectif est précisément de pouvoir réaliser de telles comparaisons.



**Figure 59 : H2AK119Ub1 est enrichi sur les gènes sensibles au traitement USP21 dans l'embryon**

**(A)** Metaplot du log<sub>2</sub>(ChIP/Input) de H2AK119Ub1 dans le spermatozoïde avant réplication sur les TSS des gènes sensibles au traitement USP21 (rouge), des gènes à contribution zygotique (orange), des gènes à contribution maternelle-zygotique (vert) et de tous les gènes (noir). **(B)** Identique à (A) pour H2AK119Ub1 après réplication. **(C)** Identique à (A) pour H3K4me3 avant réplication **(D)** Identique à (A) pour H3K4me3 après réplication.

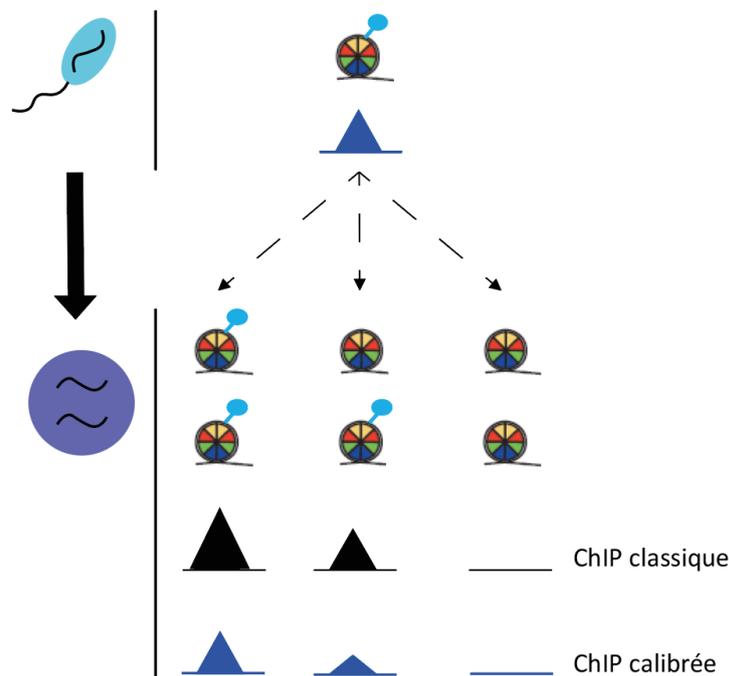
\*\*\*P-Value < 0.001. Test de Wilcoxon-Mann-Whitney +/-2kb du TSS (en jaune)

Dans un contexte global de dilution replicative de H2AK119Ub1, les gènes sensibles au traitement USP21 maintiennent un niveau pré-replicatif de cette marque

Bien que l'analyse précédente permette l'identification de changement de distribution relative le long du génome au sein d'un même échantillon, elle ne permet cependant pas de comparer l'intensité du signal entre les échantillons. Or, nous voulons étudier en particulier si un changement global de l'abondance des marques épigénétiques est observé lors de la réplication.

Lors de la réplication, le matériel génétique est doublé. Durant cette réplication, on peut envisager trois scénarios pour le devenir du signal H2AK119Ub1 et/ou H3K4me3 (**Figure 60**) :

- Effacé, avec une perte du signal sur la première copie.
- Répliqué, avec un signal répliqué sur la deuxième copie.
- Dilué, avec une répartition du signal initial sur les deux copies post-réplication.



**Figure 60 : Le devenir de H2AK119Ub1 dans l'embryon**

Lors de la réplication du spermatozoïde, la marque H2AK119Ub1 peut être totalement effacée (à droite), conservée sur une seule des deux copies (centre) ou répliquée sur la deuxième copie (à gauche). La visualisation de l'intensité du signal attendu est schématisée entre une analyse ChIP-Seq classique (avant correction drosophile) et une ChIP quantitative normalisée avec des cellules de drosophile (après correction drosophile).

La comparaison de l'intensité du signal entre deux échantillons est évaluée par une analyse de ChIP quantitative qui utilise une référence identique. Plusieurs méthodes de calibration des expériences de ChIP-Seq ont été développées pour répondre à ce problème (Orlando et al. 2014)-(Egan et al. 2016). La méthode que nous avons utilisée consiste à ajouter des cellules d'un autre génome de référence nommé spike-in, différent du génome d'intérêt (Blackledge et al. 2020)-(Tamburri et al. 2020)-(Hickey et al. 2022)-(Conway et al. 2021).

Nous avons choisi d'utiliser des cellules de drosophile dans notre cas, que nous avons ajoutées à chaque échantillon de ChIP à l'étape de digestion MNase. L'utilisation de cellules du modèle *Drosophila melanogaster* contient plusieurs avantages comme son génome totalement assemblé, la facilité de culture des cellules, la faible similarité de séquences avec les mammifères, et les similitudes des marques épigénétiques de la drosophile par rapport à celles reportées chez l'Homme ou la souris (Orlando et al. 2014). Le génome de la drosophile est composé de 4 paires de chromosomes, dont une paire sexuelle et trois paires d'autosomes, référencées sur la base de données FlyBase (Gramates et al. 2022), pour une longueur totale du génome de 137.5Mpb contre 2507.8Mpb chez *Xenopus laevis*. La petite taille de ce génome comparé à celui de *Xenopus laevis* (plus de 10 fois plus petit) est également un critère important, car une petite fraction des reads obtenus lors du séquençage est suffisante pour obtenir une bonne couverture du génome de la drosophile. La quantité de matériel spike-in doit être identique entre les échantillons, avec au moins un million de reads (Egan et al. 2016).

Il existe ensuite plusieurs méthodes de calibration du signal. La normalisation ChIP-Rx (Chromatin Immuno Precipitation with Reference exogenous genome) consiste à appliquer un facteur de correction dépendant du nombre total de reads du génome de spike-in (Orlando et al. 2014), au contraire d'une normalisation standard qui prendrait en compte la totalité des reads des deux génomes.

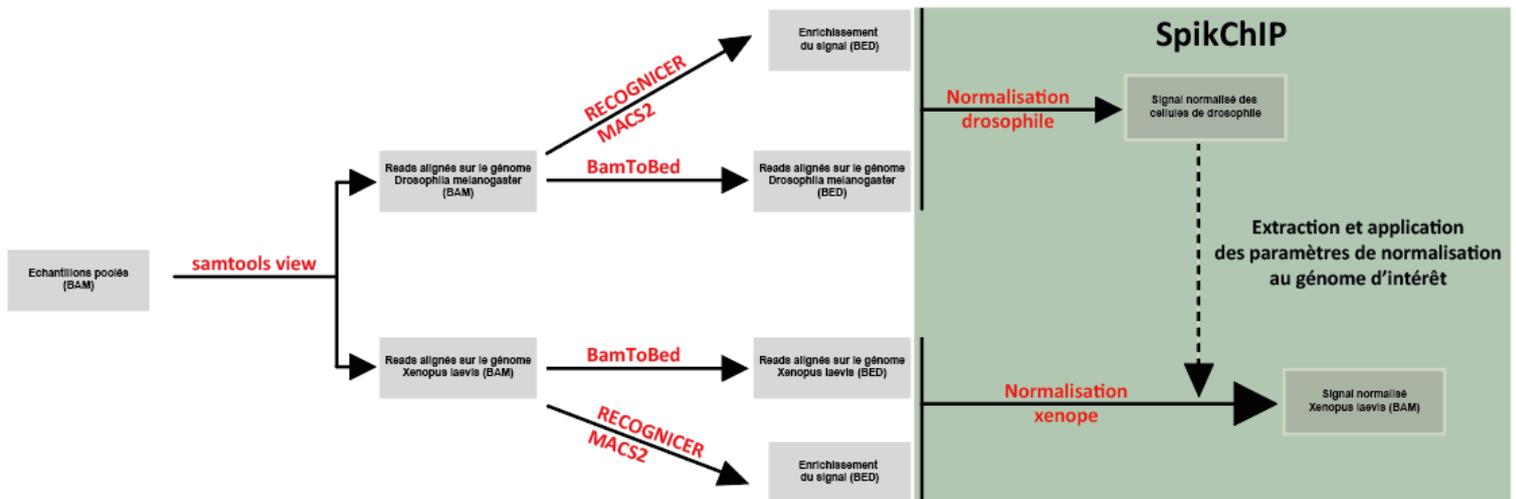
$$\text{ChIP-Rx}(\text{bin}) = \frac{\text{NbReadsXL}(\text{bin})}{\text{NbReadsDM}(\text{Genome})}$$

$$\text{Standard}(\text{bin}) = \frac{\text{NbReadsXL}(\text{bin})}{\text{NbReadsXL}(\text{Genome}) + \text{NbReadsDM}(\text{Genome})}$$

La normalisation Tag removal prend en compte directement la comparaison des conditions dans le modèle de normalisation. Elle calcule un facteur de normalisation entre les deux conditions A et B, à partir du nombre de reads alignés sur le génome spike-in pour ces deux conditions, la condition B étant celle contenant le plus de reads. Ce facteur de normalisation du génome spike-in est ensuite utilisé pour réaliser un sous-échantillonnage des reads du génome d'intérêt de l'échantillon avec le plus de reads (condition B ici), pour conserver la même proportion de reads entre les deux conditions, proportion calculée à partir du génome spike-in (Egan et al. 2016).

$$\text{SubSampleReads (CondB)} = \text{Reads (Cond B)} \times \frac{\text{NbReads Spike-in(Cond A)}}{\text{NbReads Spike-in(Cond B)}}$$

Enfin, la normalisation spikChIP est une méthode utilisant une normalisation des données du génome d'intérêt segmenté par un modèle de régression locale calculé sur le génome spike-in (Blanco, Di Croce, et Aranda 2021). J'ai choisi de réaliser cette normalisation qui réduit l'influence du bruit de séquençage, tout en minimisant l'effet des régions avec peu de signal, en séparant les régions enrichies en signal (pics) des autres (background). Toutes les étapes de la normalisation avec les cellules de drosophile sont présentées dans la **Figure 61**.



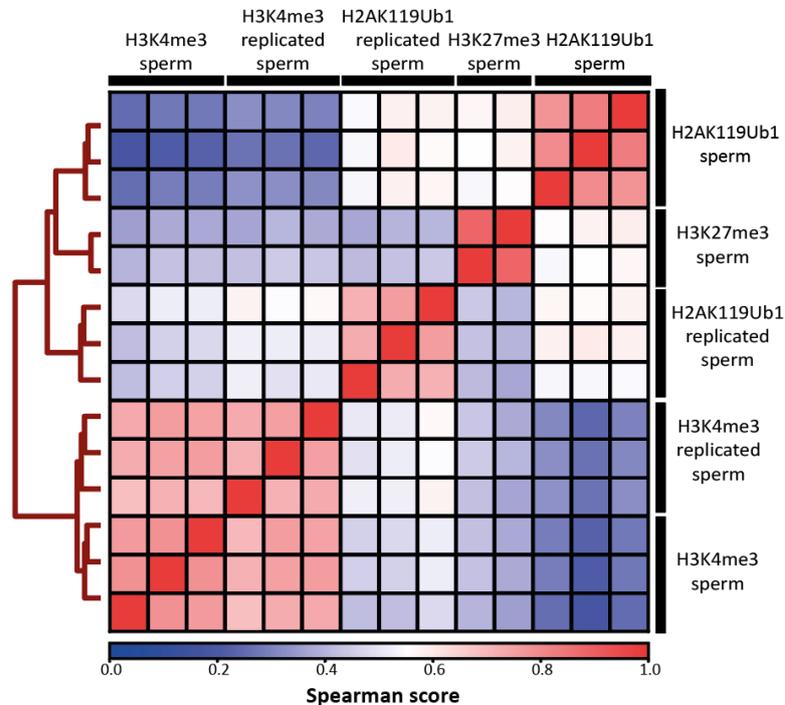
**Figure 61 : Modélisation du pipeline d'analyse ChIP-seq pour la normalisation Drosophile**

Les reads des génomes de *Xenopus laevis* et *Drosophila Melanogaster* sont séparés respectivement dans deux fichiers par samtools view. La détection de l'enrichissement du signal est réalisée avec RECOGNICER pour H2AK119Ub1 et MACS2 pour H3K4me3 sur les deux génomes. La normalisation drosophile est réalisée par l'algorithme SpikChIP, puis les paramètres de normalisation sont appliqués à *Xenopus laevis*, pour normaliser le signal de *Xenopus laevis* des échantillons avant et après la répliation. Les logiciels et méthodologies sont écrits en rouge. Les étapes réalisées par l'algorithme SpikChIP sont mises en évidence dans un fond vert.

## *Pipeline spikChIP appliqué aux échantillons Xenopus laevis couplés à la drosophile*

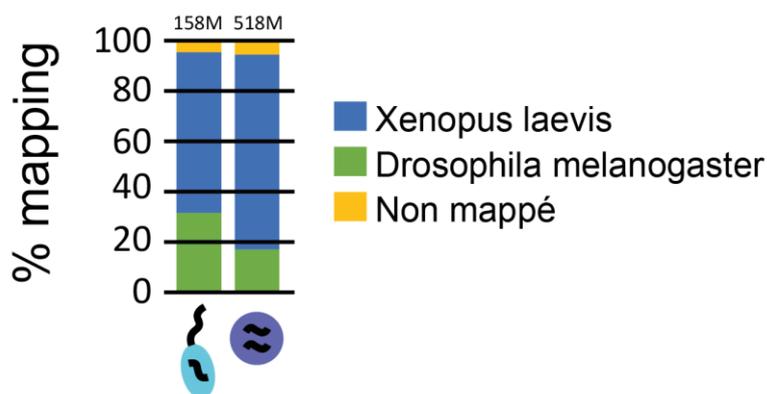
Il est parfois possible de comparer les intensités de signal ChIP-Seq entre différentes conditions si l'on considère que le signal global entre les conditions testées est similaire. Par exemple, lors du traitement de cellules en culture par une drogue, des changements locaux de la distribution d'histones modifiées peuvent s'opérer dans un contexte où la majorité des sites d'accumulation de ces marques est stable. Dans cette situation, il est concevable de comparer le changement d'intensité des pics d'histones entre les échantillons. Dans notre cas, nous ne pouvons pas assumer ce postulat de stabilité globale étant donné qu'il y a eu un cycle de réplication entre nos deux conditions, modifiant potentiellement l'ensemble du signal dans un des échantillons. L'objectif de la normalisation drosophile est de nous permettre de comparer le signal H2AK119Ub1 après un cycle de réplication, notamment pour différencier les trois états d'effacement, dilution ou réplication de la marque décrit précédemment (**Figure 60**). Le protocole est similaire à celui présenté pour le ChIP-Seq précédemment, avec plusieurs étapes modifiées ou ajoutées (**Figure 39**, **Figure 61**). La première différence est à l'étape d'alignement. J'ai créé un génome hybride artificiel couplant la version 9.2 du génome de *Xenopus laevis* à la version 6.40 de *Drosophila melanogaster*. Après vérification de la corrélation des réplicats sur le génome du xénope (**Figure 62**), j'ai réalisé un sous-échantillonnage des reads de la drosophile et du xénope des trois réplicats, comme décrit précédemment. La heatmap de corrélation de Spearman indique une modification du signal H2AK119Ub1 lors de la réplication. En revanche, les échantillons H3K4me3 restent bien corrélés pré/post-réplication en accord avec la stabilité des pics comme observé précédemment (**Figure 54**). On extrait ensuite les alignements sur les chromosomes spécifiques de *Xenopus laevis* et de *Drosophila melanogaster* dans deux fichiers différents à l'aide de samtools view. Dans la fraction Input du spermatozoïde, la proportion de reads drosophile est de 31% pour 64% de reads du xénope, nous avons ainsi la quantité minimale de 1 millions de reads du génome spike-in nécessaire. Après réplication, on observe une modification du ratio des reads au profit de *Xenopus Laevis*. Cette observation est cohérente avec le fait que nous avons 200 000 cellules de drosophile ajoutées à 1 million de spermatozoïdes ayant subi une réplication dans l'extrait d'œuf (soit un équivalent de ~2 millions de spermatozoïdes si la réplication a bien eu lieu). Ainsi, le bon déroulement de la réplication suggéré par les tests précédents (**Figure 51**) est également confirmé par l'augmentation, après incubation dans l'œuf, de la proportion d'ADN issu de *Xenopus laevis*

observée dans les analyses en ChIP quantitative (**Figure 63**). Ce résultat confirme que la réplication des spermatozoïdes de xénope dans les extraits s'est bien opérée, diminuant la part de la drosophile dans les échantillons après réplication.



**Figure 62 : Heatmap de corrélation de Spearman des réplicats de ChIP-Seq quantitatif**

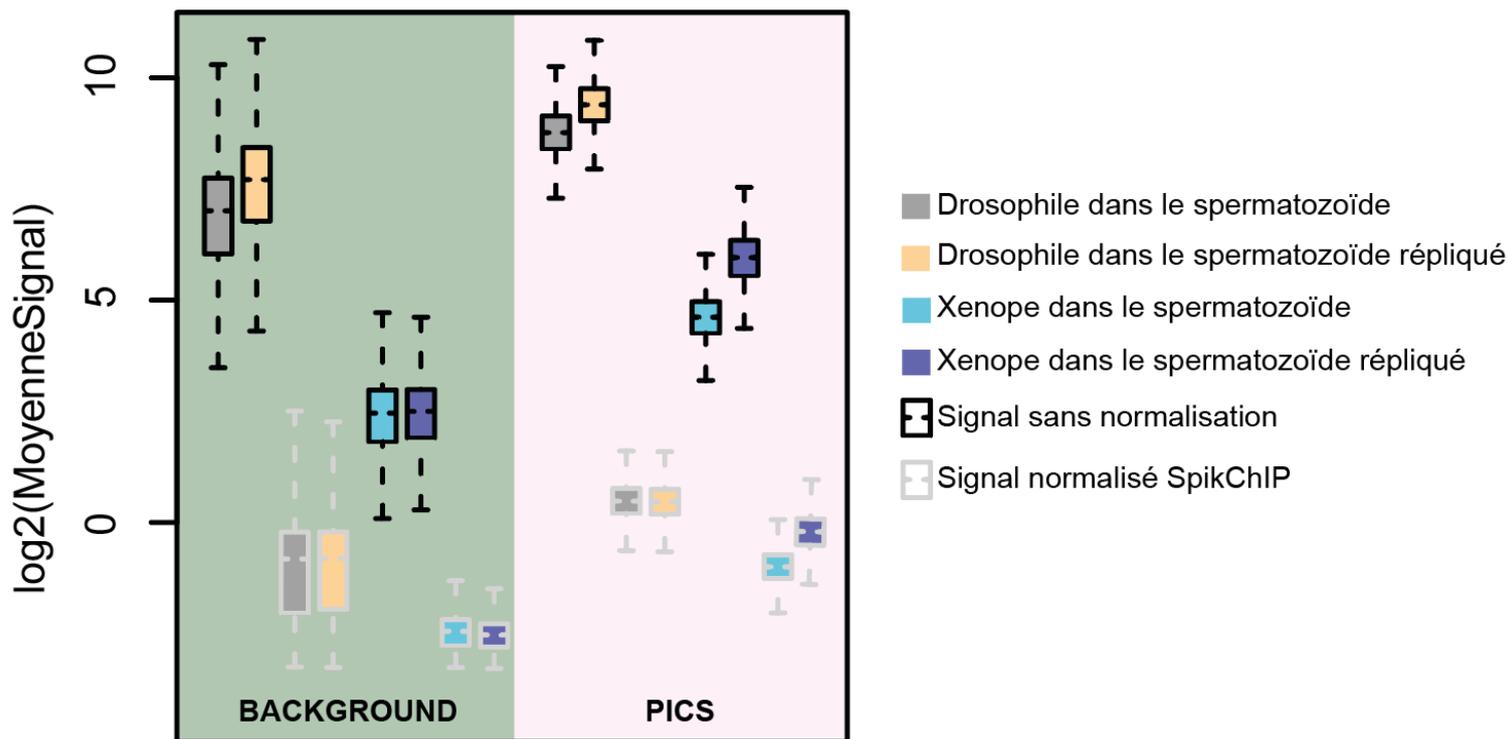
Les échantillons sont regroupés selon la corrélation de Spearman. Les trois réplicats pour H3K4me3 et H2AK119Ub1 avant et après la première réplication sont représentés. Les deux échantillons de la marque H3K27me3 utilisés précédemment sont aussi ajoutés comme référence.



**Figure 63 : Contrôle de la réplication de l'ADN du spermatozoïde**

Le nombre de reads alignés sur le génome de *Xenopus laevis* (bleu), de *Drosophila melanogaster* (vert) sont représenté en barplot pour le spermatozoïde non répliqué (gauche) et le spermatozoïde répliqué (droite). Le nombre total de reads est affiché en haut

Après contrôle de la qualité des échantillons, les pics H2AK119Ub1 et H3K4me3 sont identifiés pour les deux génomes, à l'aide de MACS2 pour H3K4me3 et RECOGNICER pour H2AK119Ub1. L'algorithme SpikChIP découpe les deux génomes en fenêtre de taille égale de 200pb dans ce cas. Les fenêtres sont ultérieurement catégorisées comme « background » ou « pics », l'attribution à la catégorie « pics » dépendant de l'existence d'un overlap d'une base minimum entre un pic et une fenêtre. Cette identification d'overlap est réalisée à l'aide de la fonction matchpeaks de l'outil seqcode (Blanco, González-Ramírez, et Di Croce 2021). Le signal final sur une fenêtre est calculé de deux façons. La première consiste à prendre le nombre de reads moyen tout au long de la fenêtre, la deuxième à prendre le nombre de reads maximum observé sur la fenêtre. Le nombre de reads est dans un premier temps calculé pour toutes les fenêtres des deux génomes, pour obtenir un signal sur chaque fenêtre avec la fonction recoverChIPlevels de seqcode, déterminé sur le nombre de reads. Ensuite, une normalisation par un modèle de régression LOESS est réalisée sur les fenêtres du génome de la drosophile pour générer un modèle permettant de normaliser le signal spike-in entre les spermatozoïdes avant ou après la première réplication. Les paramètres du modèle de la normalisation du signal de la drosophile sont ensuite extraits, et appliqués aux intensités du signal des fenêtres du génome de *Xenopus laevis*, pour corriger le signal. La normalisation par les cellules de drosophile permet ainsi de normaliser indépendamment le signal H2AK119Ub1 dans les fenêtres classées comme « background » de celles classées comme « pics ». On observe une intensité du signal identique dans les cellules de drosophile ajoutées au spermatozoïde avant ou après la première réplication sur les deux types de fenêtre (« background » et « pics ») après la normalisation SpikChIP, validant ainsi l'implémentation du pipeline (**Figure 64**, boxplot gris et jaune avec un contour gris). Le signal H2AK119Ub1 du spermatozoïde sur les fenêtres « background » du spermatozoïde est équivalent avant ou après la réplication, mais il est plus faible que le signal présent sur les fenêtres de « pics » contenant un enrichissement en H2AK119Ub1 détecté par RECOGNICER (**Figure 64**, boxplot bleu et violet avec un contour gris). Toutes ces observations montrent que l'algorithme SpikChIP permet de normaliser le signal de la drosophile sur les régions génomique enrichie en H2AK119Ub1 ou non. Enfin, sur les fenêtres des pics H2AK119Ub1 dans le spermatozoïde avant ou après réplication, on observe un signal d'intensité moyenne plus élevée dans le spermatozoïde après réplication.



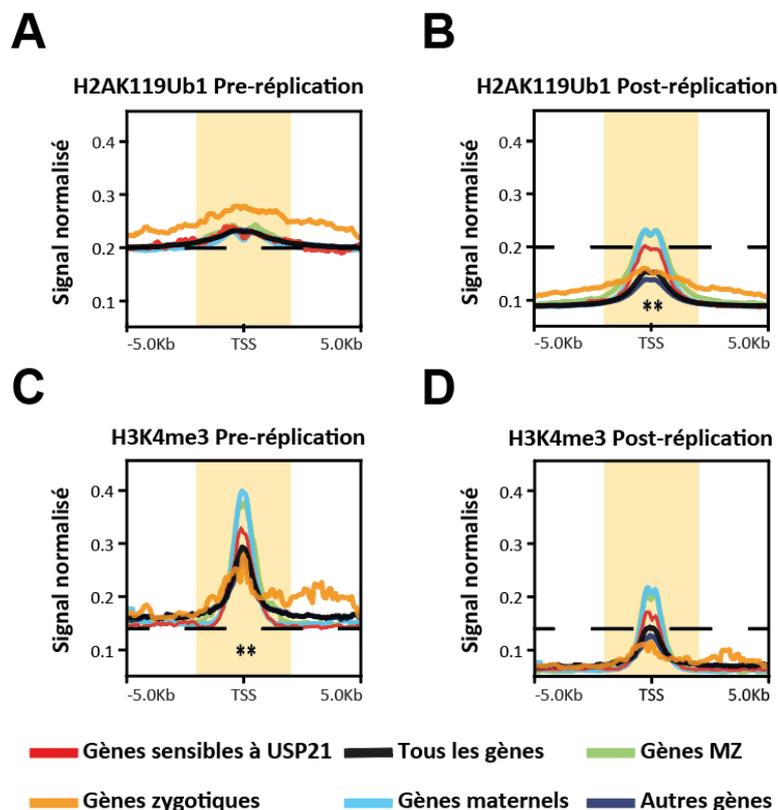
**Figure 64 : Boxplot du signal H2AK119Ub1 lors de la normalisation SpikChIP**

La distribution de l'intensité moyenne du signal H2AK119Ub1 pour chaque fenêtre de 200pb est représentée sur les régions enrichies (Pics = rose) ou les régions non enrichies (Background = vert). Les boxplots d'intensité de signal non normalisés sont encadrés en noir, contrairement au signal normalisé avec SpikChIP encadré en gris clair. Quatre types de cellules sont différenciés, les cellules de drosophile dans les spermatozoïdes avant répllication (jaune) et après répllication (gris), et les cellules de xénope dans le spermatozoïde avant (cyan) et après répllication (violet).

Néanmoins, une correction additionnelle est requise pour compenser le changement de ratio drosophile/xénope induit par la répllication dans les extraits. Ainsi, le signal mesuré dans le spermatozoïde après répllication est corrigé en fonction du ratio de cellule de drosophile et de xénope dans les conditions pré-répllication et post-répllication, qui peut être inféré des proportions de reads alignés sur les deux génomes dans la fraction INPUT de ces échantillons :

$$\text{SignalCHIPCorrigé\_XL (Post-Replication)} = \text{SignalCHIP\_XL (Post-Replication)} \times \frac{\frac{\text{NbReadsINPUT\_XL(Pre-Replication)}}{\text{NbReadsINPUT\_DM(Pre-Replication)}}}{\frac{\text{NbReadsINPUT\_XL(Post-Replication)}}{\text{NbReadsINPUT\_DM(Post-Replication)}}}$$

Lorsque cette normalisation est appliquée, on constate une diminution du signal pour H2AK119Ub1 (**Figure 65A-B**, ligne noire) et H3K4me3 (**Figure 65C-D**, ligne noire) sur l'ensemble des TSSs. Ceci suggère que bien qu'il y ait toujours une accumulation du signal sur les TSSs post-réplication, le signal restant est cependant diminué comparé à l'état pré-réplication. Ceci suggère un mécanisme général de dilution des marques épigénétiques du spermatozoïde post-réplication. Les gènes sensibles à USP21 présentent cependant un profil TSS sensiblement différent lors de la réplication avec un niveau H2AK119Ub1 retenu post-réplication plus élevé sur les gènes sensibles à USP21 que celui observé pour l'ensemble des gènes (**Figure 65B**, ligne rouge et noire). En revanche, le signal H3K4me3 est dilué de la même façon sur les gènes sensibles à USP21 comparés à l'ensemble des gènes (**Figure 65D**, ligne rouge et noire).



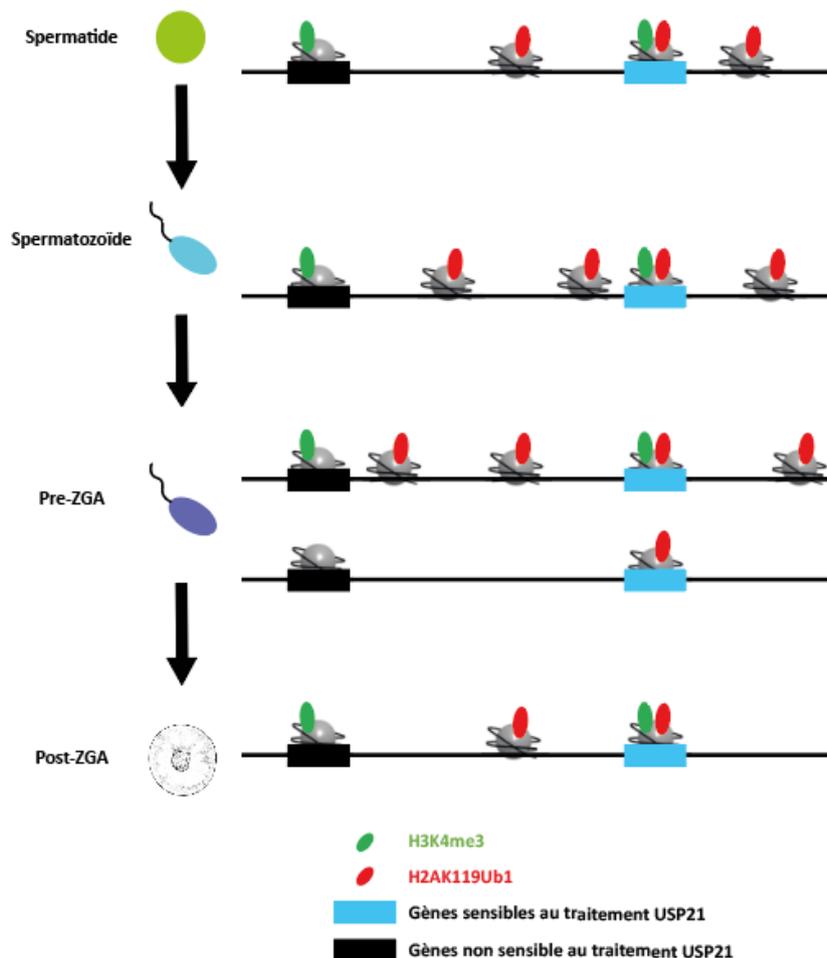
**Figure 65 : H2AK119Ub1 est préférentiellement retenu sur les gènes sensibles au traitement USP21**

**(A)** Metaplot du signal de H2AK119Ub1 dans le spermatozoïde avant réplication sur les TSSs des gènes sensibles au traitement USP21 (rouge), des gènes à contribution maternelle stricte (cyan), maternelle et zygotique (vert), zygotique stricte (orange), des gènes non inclus dans ces groupes (bleu) et de tous les gènes (noir). **(B)** Identique à (A) pour H2AK119Ub1 après réplication. **(C)** Identique à (A) pour H3K4me3 avant réplication. **(D)** Identique à (A) pour H3K4me3 après réplication. Le signal est normalisé avec SpikChIP et corrigé par le ratio d'Input, et une courbe noire en pointillé représente le niveau basal du signal pour l'ensemble des gènes dans le spermatozoïde avant réplication. \*\*P-Value < 0.01. Test de Wilcoxon-Mann-Whitney +/-2kb du TSS (en jaune).

# Discussion

## H2AK119Ub1 du spermatozoïde à l'embryon

Durant ma thèse, j'ai mis en évidence que dans un contexte où H3K4me3 est associé avec H2AK119Ub1, l'effacement de H2AK119Ub1 par USP21 sur le spermatozoïde pourrait mener à des défauts de développement. Les gènes sensibles à USP21 auraient besoin de H2AK119Ub1 sur le spermatozoïde pour transmettre l'information dans l'embryon jusqu'à la ZGA en relai de H3K4me3 dilué pré-ZGA (**Figure 66**). Le signal H3K4me3 et H2AK119Ub1 est détecté post-ZGA, néanmoins nous ne pouvons pas conclure sur son intensité de signal par rapport au stade pré-ZGA dans notre modèle actuel.



**Figure 66 : Le modèle H2AK119Ub1 dans le développement de l'embryon**

Profil du signal en log2 de H2AK119Ub1, H3K4me3 et de la fraction INPUT dans la spermatozoïde (noire), le spermatozoïde avant la première réplication (orange) ou après la première réplication (bleu), et dans la gastrula au stade 12 de développement (rouge).

J'ai aussi observé que H3K27me3 disparaissait après fécondation, avec une perte significative de H3K27me3 sur les gènes sensibles au traitement USP21, un résultat cohérent avec le modèle où H2AK119Ub1 serait primordial à H3K27me3 pré-ZGA, observé aussi chez la souris (Mei et al. 2021)-(Z. Chen, Djekidel, et Zhang 2021) et le zébrafish (Hickey et al. 2022). Notre étude teste ainsi pour la première fois la transmission de H2AK119Ub1 du spermatozoïde à l'embryon.

## Le traitement USP21 altère le développement des embryons et le transcriptome

Les embryons traités par l'extrait USP21 présentent à la fois des défauts de développement, mais aussi des anomalies transcriptomiques. Avec ces données, nous ne pouvons pas définitivement établir un lien de causalité entre dérégulations transcriptionnelles et les défauts de développement. Néanmoins, nous avons mis en évidence plusieurs corrélations entre les analyses morphologiques et transcriptomiques. Par exemple, j'ai mis en évidence que les gènes dérégulés par le traitement USP21 étaient impliqués dans le cycle cellulaire, et pourraient engendrer les défauts de développement de l'embryon. Nous avons aussi observé un parallèle entre l'hétérogénéité du développement des embryons qui meurent à différents stades et l'hétérogénéité des anomalies transcriptomiques avec certains embryons uniques qui avaient un profil transcriptomique similaire à un embryon issu d'un spermatozoïde non traité par USP21. Cette hétérogénéité observée est similaire aux expériences de transfert nucléaire (Hörmanseder et al. 2017) ou d'injection intracytoplasmique de spermatozoïde (Teperek et al. 2016). Dans nos cas, cette hétérogénéité de l'effet de USP21 pourrait provenir de la déubiquitination, partielle de certains spermatozoïdes, ou de l'œuf qui aurait la capacité de reprogrammer un noyau dont la configuration épigénétique ne serait pas optimale, comme on peut l'observer avec 1% des embryons clonés à partir de cellules d'intestin qui arrivent jusqu'au stade de la grenouille (Pasque et al. 2011). Enfin, l'utilisation d'embryons haploïdes permet de mesurer la transcription de l'embryon dont la chromatine est dérivée uniquement du spermatozoïde, ce qui permet d'identifier un effet du traitement USP21 sur le spermatozoïde sans confusion avec la partie maternelle inactivée au préalable. Néanmoins, une perspective serait de vérifier ce qui se passe dans les embryons diploïdes dans un premier temps, puis d'évaluer indépendamment la contribution maternelle chez *Xenopus laevis* en déubiquitinant l'ovocyte à l'aide de CHIP anti-H2AK119Ub1. Enfin, j'ai mis en évidence que les gènes dérégulés par le traitement USP21 ont

une composante transcriptomique à la fois maternelle et zygotique. Cependant, il est difficile de démontrer si USP21 affecte préférentiellement la composante maternelle et/ou zygotique de ces gènes. L'étude des introns à partir de la vélocité des données RNA-Seq ne nous a pas permis de conclure. Une alternative serait de refaire des expériences RNA-Seq des nouveaux transcrits en marquant les nouveaux transcrits avec le 5-ethynyl uridine (EU) (H. Chen et Good 2022) ou de récupérer le transcriptome des embryons à la ZGA, qui a lieu avant le stade de gastrula où les embryons ont été séquencés dans notre étude.

## La distribution de H2AK119Ub1 dans le spermatozoïde et au cours du développement embryonnaire

Dans la littérature, H2AK119Ub1 est très peu référencé dans le spermatozoïde. Dans les espèces où les histones ne sont pas remodelées comme le zébrafish, il n'y aurait pas de H2AK119Ub1 (Hickey et al. 2022). Dans les espèces avec un grand remodelage des histones comme la souris, H2AK119Ub1 est détectée dans le spermatozoïde, mais n'est pas exploitée, car considérée comme non importante notamment suite à l'effacement global (9150/10164 régions) de H2AK119Ub1 du spermatozoïde après fécondation (Z. Chen, Djekidel, et Zhang 2021). Malgré tout, nous avons mis en évidence que dans un contexte de fort remodelage (51437/67233 régions) de H2AK119Ub1 du spermatozoïde après fécondation chez *Xenopus laevis*, l'information H2AK119Ub1 conservée semblait importante pour le développement de l'embryon. De plus, notre étude montre que H2AK119Ub1 dans le spermatozoïde d'un vertébré a une distribution similaire à celle décrite pour chez les mammifères (Z. Chen, Djekidel, et Zhang 2021). Une limite de notre modèle expérimental est que nous n'avons pas l'information de la contribution maternelle, des expériences de ChIP-Sequencing dans l'ovocyte permettrait de contrôler si on observe une dissymétrie pour H3K4me3 comme chez les mammifères, avec un signal très large sur l'ovocyte, et très localisé sur les spermatozoïdes (Dahl et al. 2016). Enfin, l'utilisation de RECOGNICER est une nouveauté dans cette étude alors que H2AK119Ub1 est très souvent étudié en regardant la région de fixation du complexe PRC1, RING1B (Blackledge et al. 2020) - (Tamburri et al. 2020), ou plus récemment par détection de l'enrichissement du signal avec MACS2 (Hickey et al. 2022). L'utilisation de RECOGNICER dans notre cas permet de caractériser le signal H2AK119Ub1 sur l'épigénome en identifiant des régions similaires qu'avec MACS2 ainsi que de nouvelles régions candidates que MACS2 n'identifie pas. Mon analyse du signal par RECOGNICER met en évidence l'association de la

bivalence H2AK119Ub1/H3K4me3 avec les gènes sensibles au traitement USP21. Une perspective d'étude serait d'analyser la proportion de spermatozoïdes qui contiennent cette information épigénétique H2AK119Ub1/H3K4me3. Or, il a été mis en évidence chez *Xenopus laevis* que seulement une fraction des pics H3K4me3 seraient présents dans tous les spermatozoïdes (Oikawa et al. 2020). Un spike-in de nucléosome H2AK119Ub1 a été ajouté dans certaines expériences en cette perspective, notamment pour déterminer si la marque H2AK119Ub1 était présente sur tous les spermatozoïdes. Une autre perspective serait d'identifier les régions où l'histone H2A est conservée sur tous les spermatozoïdes en employant un modèle de markov caché dit HMM (Oikawa et al. 2020). Ces analyses complémentaires permettraient d'identifier les régions contenant H2AK119Ub1 dans tous les spermatozoïdes, dans un cas où H2A est forcément retenue. Enfin, il est possible que d'autres marques d'histones non investiguées pourraient jouer un rôle de relai de l'information épigénétique dans l'embryon après la fécondation ou que l'effet de l'effacement de H2AK119Ub1 se manifeste uniquement dans des conditions spécifiques comme par la présence des facteurs de transcription des gènes régulés par H2AK119Ub1.

## L'effet sur le développement de l'embryon par USP21 est-il spécifique de son activité deubiquitylase ?

Notre approche de traitement par USP21 élimine les effets indirects qui peuvent se produire lorsqu'on altère l'épigénome pendant la spermatogénèse ou durant l'embryogenèse précoce, puisque nous altérons uniquement l'épigénome du spermatozoïde. Cependant, nous n'avons pas de preuve formelle que l'effet de USP21 est lié à son activité d'effacement de H2AK119Ub1 comme décrit précédemment. L'effet observé pourrait être lié à une autre activité, ou la déubiquitination d'une autre protéine par USP21. Néanmoins, plusieurs arguments suggèrent un effet dépendant de l'activité d'effacement de H2AK119Ub1 par USP21. Premièrement, on observe que ~75% des gènes sont up-régulés lors du traitement par USP21, ce qui est compatible avec l'élimination de la marque répressive H2AK119Ub1. On observe aussi une association des régions régulatrices des gènes affectés par le traitement USP21 avec la présence de H2AK119Ub1. Enfin, nous avons identifié que les gènes dérégulés par USP21 sont enrichis en H2AK119Ub1 lorsqu'ils sont également marqués avec H3K4me3 et qu'ils maintiennent H2AK119Ub1 après réplication dans l'embryon. L'effet de USP21 pourrait ainsi se manifester que lorsque H2AK119Ub1 est colocalisé avec H3K4me3 et

maintenu dans l'embryon. Dans cette perspective, il serait important de vérifier dans un embryon pré-ZGA si l'épigénome H2AK119Ub1/H3K4me3 est similaire à celui observé après la première réplication dans des extraits d'œufs. La littérature montre que le signal H3K4me3 après réplication chez *Xenopus laevis* est similaire à l'embryon pré-ZGA (Oikawa et al. 2020) ou bien que H2AK119Ub1 est maintenu avant la ZGA (Mei et al. 2021). Néanmoins, seulement une fraction des gènes présentant les caractéristiques décrites ci-dessus est dérégulée par le traitement USP21. Par exemple, on observe de façon surprenante que les gènes à contribution maternelle et zygotique ont un marquage H2AK119Ub1 plus élevé après réplication et qu'ils sont H3K4me3 positifs, mais qu'ils ne sont pas tous dérégulés par le traitement USP21. D'autres caractéristiques épigénétique présentent sur ces gènes pourraient empêcher leur surexpression lorsque H2AK119Ub1 est éliminé, comme la présence de H3K9me3, des variants d'histones, ou la méthylation de l'ADN. Alternativement, il est possible que les facteurs de transcriptions capables d'activer ces gènes en absence de H2AK119Ub1 ne soient pas présents au stade de gastrula où le transcriptome a été évalué. Il serait ainsi intéressant de mesurer l'expression à différents stades de développement dans les embryons traités par USP21. Malgré tout, nos données permettent d'envisager des tests fonctionnels de l'effet de la déubiquitination de H2AK119Ub1 dans le spermatozoïde, basés sur les sites H2AK119Ub1/H3K4me3 identifiés sur les gènes sensibles à USP21. Nous pourrions envisager de réaliser une déubiquitination ciblée de ces gènes (USP21-dcas9) pour évaluer l'effet sur le développement de l'embryon et son transcriptome de leur déubiquitination.

## La transmission de H2AK119Ub1 du spermatozoïde à l'embryon

Durant ma thèse, j'ai évalué dans quelle mesure la déposition de H2AK119Ub1 après la fécondation dépendait de H2AK119Ub1 provenant du spermatozoïde. Par exemple, il a été démontré que la marque H3K27me3 du spermatozoïde n'était pas maintenue chez la souris (Mei et al. 2021) (Z. Chen, Djekidel, et Zhang 2021) et chez l'Homme (Xia et al. 2019). Ainsi, soit l'information épigénétique est perdue, soit une autre marque prend le relais. Chez *Xenopus laevis*, j'ai trouvé que seulement une fraction de H2AK119Ub1 du spermatozoïde était transmise à l'embryon (**Figure 54**). L'expérience de calibration du signal ChIP-Seq nous a permis de distinguer les scénarios du devenir de la marque dans l'embryon, mettant en évidence que la majorité des régions H2AK119Ub1 et H3K4me3 sont diluées. Néanmoins, pour valider cette conclusion, nous avons besoin de contrôle additionnel pour valider les corrections de signal appliqué pour pallier les différences entre les échantillons des ratios xénope/drosophile

engendrés par la réplication. Enfin, cette étude permet des perspectives intéressantes pour évaluer dans quelle mesure le profil H2AK119Ub1 après réplication est dicté par le profil avant réplication. En effet, il est possible que l'œuf dépose H2AK119Ub1 de façon indépendante même si des zones semblent transmises du spermatozoïde. Nous pouvons évaluer ce processus en comparant le profil, après incubation dans l'extrait d'œuf, de spermatozoïdes contrôle ou traités par USP21.

## H2AK119Ub1 dans un contexte de transmission intergénérationnelle

Nous avons mis en évidence un rôle nécessaire de H2AK119Ub1 pour le développement embryonnaire. Il est possible que ces mécanismes soient détournés pour transmettre des informations en réponse à l'histoire et l'environnement du père. En effet, une étude démontre par exemple qu'une souris mâle soumise à un stress transmet son caractère anxieux à sa descendance, qui est plus sujet à la dépression (Cunningham, Walker, et Nestler 2021). Cette modification comportementale a été associée à une modification de l'état de méthylation des îlots CpG dans les cellules germinales des souris mâles (Franklin et al. 2010). Par ailleurs, l'effacement de H3K4me3 du spermatozoïde chez le père par une déméthylase KDM1A surexprimée dans les spermatogonies chez la souris a mis en évidence des défauts de développement de la génération suivante avec un échappement de la reprogrammation de H3K4me3 du spermatozoïde, ainsi transmise à la descendance (Lismer et al. 2020). Tous ces résultats peuvent ainsi laisser penser que la perturbation de H2AK119Ub1 pourrait suivre un modèle similaire, et qu'une meilleure compréhension des mécanismes de transmission intergénérationnelle est nécessaire pour mieux comprendre certaines causes d'infertilité dans le futur.

## Publication et éveil scientifique

P. Colombier *et al.*, « NOTO Transcription Factor Directs Human Induced Pluripotent Stem Cell-Derived Mesendoderm Progenitors to a Notochordal Fate », *Cells*, vol. 9, n° 2, p. 509, févr. 2020, doi: 10.3390/cells9020509.

D. Meistermann *et al.*, « Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification », *Cell Stem Cell*, vol. 28, n° 9, p. 1625-1640.e6, sept. 2021, doi: 10.1016/j.stem.2021.04.027.

G. Castel *et al.*, « Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells », *Cell Reports*, vol. 33, n° 8, p. 108419, nov. 2020, doi: 10.1016/j.celrep.2020.108419.

## Références

- Agarwal, Ashok, Saradha Baskaran, Neel Parekh, Chak-Lam Cho, Ralf Henkel, Sarah Vij, Mohamed Arafa, Manesh Kumar Panner Selvam, et Rupin Shah. 2021. « Male Infertility ». *The Lancet* 397 (10271): 319-33. [https://doi.org/10.1016/S0140-6736\(20\)32667-2](https://doi.org/10.1016/S0140-6736(20)32667-2).
- Alegría-Torres, Jorge Alejandro, Andrea Baccarelli, et Valentina Bollati. 2011. « Epigenetics and Lifestyle ». *Epigenomics* 3 (3): 267-77. <https://doi.org/10.2217/epi.11.22>.
- Anders, S., P. T. Pyl, et W. Huber. 2015. « HTSeq--a Python Framework to Work with High-Throughput Sequencing Data ». *Bioinformatics* 31 (2): 166-69. <https://doi.org/10.1093/bioinformatics/btu638>.
- Andrea Scelfo, Daniel Fernández-Pérez, Simone Tamburri, Marika Zanotti, Elisa Lavarone, Monica Soldi, Tiziana Bonaldi, Karin Johanna Ferrari, et Diego Pasini. 2019. « Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/B-Independent-Specific Activities ». *Molecular Cell* 74 (5): 1037-1052.e7. <https://doi.org/10.1016/j.molcel.2019.04.002>.
- Andrews S. 2010. « FastQC: A Quality Control Tool for High Throughput Sequence Data ». 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. « Gene Ontology: Tool for the Unification of Biology ». *Nature Genetics* 25 (1): 25-29. <https://doi.org/10.1038/75556>.
- Baedke, Jan. 2013. « The Epigenetic Landscape in the Course of Time: Conrad Hal Waddington's Methodological Impact on the Life Sciences ». *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 756-73. <https://doi.org/10.1016/j.shpsc.2013.06.001>.
- Bannister, Andrew J, et Tony Kouzarides. 2011. « Regulation of Chromatin by Histone Modifications ». *Cell Research* 21 (3): 381-95. <https://doi.org/10.1038/cr.2011.22>.
- Baruzzo, Giacomo, Katharina E Hayer, Eun Ji Kim, Barbara Di Camillo, Garret A FitzGerald, et Gregory R Grant. 2017. « Simulation-Based Comprehensive Benchmarking of RNA-Seq Aligners ». *Nature Methods* 14 (2): 135-39. <https://doi.org/10.1038/nmeth.4106>.
- Beacon, Tasnim H., Geneviève P. Delcuve, Camila López, Gino Nardocci, Igor Kovalchuk, Andre J. van Wijnen, et James R. Davie. 2021. « The Dynamic Broad Epigenetic (H3K4me3, H3K27ac) Domain as a Mark of Essential Genes ». *Clinical Epigenetics* 13 (1): 138. <https://doi.org/10.1186/s13148-021-01126-1>.
- Bendandi, Artemi, Alessandro S. Patelli, Alberto Diaspro, et Walter Rocchia. 2020. « The Role of Histone Tails in Nucleosome Stability: An Electrostatic Perspective ». *Computational and Structural Biotechnology Journal* 18: 2799-2809. <https://doi.org/10.1016/j.csbj.2020.09.034>.

- Bird, Adrian, Mary Taggart, Marianne Frommer, Orlando J. Miller, et Donald Macleod. 1985. « A Fraction of the Mouse Genome That Is Derived from Islands of Nonmethylated, CpG-Rich DNA ». *Cell* 40 (1): 91-99. [https://doi.org/10.1016/0092-8674\(85\)90312-5](https://doi.org/10.1016/0092-8674(85)90312-5).
- Blackledge, Neil P., Nadezda A. Fursova, Jessica R. Kelley, Miles K. Huseyin, Angelika Feldmann, et Robert J. Klose. 2020. « PRC1 Catalytic Activity Is Central to Polycomb System Function ». *Molecular Cell* 77 (4): 857-874.e9. <https://doi.org/10.1016/j.molcel.2019.12.001>.
- Blackledge, Neil P., et Robert J. Klose. 2021. « The Molecular Principles of Gene Regulation by Polycomb Repressive Complexes ». *Nature Reviews Molecular Cell Biology* 22 (12): 815-33. <https://doi.org/10.1038/s41580-021-00398-y>.
- Blanco, Enrique, Luciano Di Croce, et Sergi Aranda. 2021. « SpikChIP: A Novel Computational Methodology to Compare Multiple ChIP-Seq Using Spike-in Chromatin ». *NAR Genomics and Bioinformatics* 3 (3): lqab064. <https://doi.org/10.1093/nargab/lqab064>.
- Blanco, Enrique, Mar González-Ramírez, et Luciano Di Croce. 2021. « Productive Visualization of High-Throughput Sequencing Data Using the SeqCode Open Portable Platform ». *Scientific Reports* 11 (1): 19545. <https://doi.org/10.1038/s41598-021-98889-7>.
- Blum, Martin, et Tim Ott. 2018. « *Xenopus*: An Undervalued Model Organism to Study and Model Human Genetic Disease ». *Cells Tissues Organs* 205 (5-6): 303-13. <https://doi.org/10.1159/000490898>.
- Boyer, Laurie A., Kathrin Plath, Julia Zeitlinger, Tobias Brambrink, Lea A. Medeiros, Tong Ihn Lee, Stuart S. Levine, et al. 2006. « Polycomb complexes repress developmental regulators in murine embryonic stem cells ». *Nature* 441 (7091): 349-53. <https://doi.org/10.1038/nature04733>.
- Breton-Larrivé, Mélanie, Elizabeth Elder, et Serge McGraw. 2019. « DNA methylation, environmental exposures and early embryo development ». *Animal Reproduction* 16 (3): 465-74. <https://doi.org/10.21451/1984-3143-AR2019-0062>.
- Broad Institute of MIT and Harvard. s. d. « Picard ». <http://broadinstitute.github.io/picard>.
- Brykczynska, Urszula, Mizue Hisano, Serap Erkek, Liliana Ramos, Edward J Oakeley, Tim C Roloff, Christian Beisel, Dirk Schübeler, Michael B Stadler, et Antoine H F M Peters. 2010. « Repressive and Active Histone Methylation Mark Distinct Promoters in Human and Mouse Spermatozoa ». *Nature Structural & Molecular Biology* 17 (6): 679-87. <https://doi.org/10.1038/nsmb.1821>.
- Carone, Benjamin R., Jui-Hung Hung, Sarah J. Hainer, Min-Te Chou, Dawn M. Carone, Zhiping Weng, Thomas G. Fazzio, et Oliver J. Rando. 2014. « High-Resolution Mapping of Chromatin Packaging in Mouse Embryonic Stem Cells and Sperm ». *Developmental Cell* 30 (1): 11-22. <https://doi.org/10.1016/j.devcel.2014.05.024>.
- Chen, Hui, et Matthew C. Good. 2022. « Nascent Transcriptome Reveals Orchestration of Zygotic Genome Activation in Early Embryogenesis ». *Current Biology*, août, S0960982222012325. <https://doi.org/10.1016/j.cub.2022.07.078>.

- Chen, Zhiyuan, Mohamed Nadhir Djekidel, et Yi Zhang. 2021. « Distinct Dynamics and Functions of H2AK119ub1 and H3K27me3 in Mouse Preimplantation Embryos ». *Nature Genetics* 53 (4): 551-63. <https://doi.org/10.1038/s41588-021-00821-2>.
- Conway, Eric, Emilia Jerman, Evan Healy, Shinsuke Ito, Daniel Holoch, Giorgio Oliviero, Orla Deevy, et al. 2018. « A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities ». *Molecular Cell* 70 (3): 408-421.e8. <https://doi.org/10.1016/j.molcel.2018.03.005>.
- Conway, Eric, Federico Rossi, Daniel Fernandez-Perez, Eleonora Ponzio, Karin Johanna Ferrari, Marika Zanotti, Daria Manganaro, Simona Rodighiero, Simone Tamburri, et Diego Pasini. 2021. « BAP1 Enhances Polycomb Repression by Counteracting Widespread H2AK119ub1 Deposition and Chromatin Condensation ». *Molecular Cell* 81 (17): 3526-3541.e8. <https://doi.org/10.1016/j.molcel.2021.06.020>.
- Costa, Fabrício F. 2008. « Non-Coding RNAs, Epigenetics and Complexity ». *Gene* 410 (1): 9-17. <https://doi.org/10.1016/j.gene.2007.12.008>.
- Cunningham, Ashley M., Deena M. Walker, et Eric J. Nestler. 2021. « Paternal Transgenerational Epigenetic Mechanisms Mediating Stress Phenotypes of Offspring ». *European Journal of Neuroscience* 53 (1): 271-80. <https://doi.org/10.1111/ejn.14582>.
- Cutter, Amber R., et Jeffrey J. Hayes. 2015. « A Brief Review of Nucleosome Structure ». *FEBS Letters* 589 (20PartA): 2914-22. <https://doi.org/10.1016/j.febslet.2015.05.016>.
- Dahl, John Arne, Inkyung Jung, Håvard Aanes, Gareth D. Greggains, Adeel Manaf, Mads Lerdrup, Guoqiang Li, et al. 2016. « Broad Histone H3K4me3 Domains in Mouse Oocytes Modulate Maternal-to-Zygotic Transition ». *Nature* 537 (7621): 548-52. <https://doi.org/10.1038/nature19360>.
- Decombe, Sheldon, François Loll, Laura Caccianini, Kévin Affannoukoué, Ignacio Izeddin, Julien Mozziconacci, Christophe Escudé, et Judith Lopes. 2021. « Epigenetic Rewriting at Centromeric DNA Repeats Leads to Increased Chromatin Accessibility and Chromosomal Instability ». *Epigenetics & Chromatin* 14 (1): 35. <https://doi.org/10.1186/s13072-021-00410-x>.
- Diaz, Aaron, Kiyoub Park, Daniel A. Lim, et Jun S. Song. 2012. « Normalization, bias correction, and peak calling for ChIP-seq ». *Statistical Applications in Genetics and Molecular Biology* 11 (3). <https://doi.org/10.1515/1544-6115.1750>.
- Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, et Thomas R. Gingeras. 2013. « STAR: Ultrafast Universal RNA-Seq Aligner ». *Bioinformatics* 29 (1): 15-21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Egan, Brian, Chih-Chi Yuan, Madeleine Lisa Craske, Paul Labhart, Gulfer D. Guler, David Arnott, Tobias M. Maile, et al. 2016. « An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition ». Édité par Zhaohui Qin. *PLOS ONE* 11 (11): e0166438. <https://doi.org/10.1371/journal.pone.0166438>.

- Elurbe, Dei M., Sarita S. Paranjpe, Georgios Georgiou, Ila van Kruijsbergen, Ozren Bogdanovic, Romain Gibeaux, Rebecca Heald, et al. 2017. « Regulatory Remodeling in the Allo-Tetraploid Frog *Xenopus Laevis* ». *Genome Biology* 18 (1): 198. <https://doi.org/10.1186/s13059-017-1335-7>.
- Erkek, Serap, Mizue Hisano, Ching-Yeu Liang, Mark Gill, Rabih Murr, Jürgen Dieker, Dirk Schübeler, Johan van der Vlag, Michael B Stadler, et Antoine H F M Peters. 2013. « Molecular Determinants of Nucleosome Retention at CpG-Rich Sequences in Mouse Spermatozoa ». *Nature Structural & Molecular Biology* 20 (7): 868-75. <https://doi.org/10.1038/nsmb.2599>.
- Evans, Ciaran, Johanna Hardin, et Daniel M Stoebel. 2018. « Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions ». *Briefings in Bioinformatics* 19 (5): 776-92. <https://doi.org/10.1093/bib/bbx008>.
- Ewels, Philip, Måns Magnusson, Sverker Lundin, et Max Källner. 2016. « MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report ». *Bioinformatics* 32 (19): 3047-48. <https://doi.org/10.1093/bioinformatics/btw354>.
- Faust, C., A. Schumacher, B. Holdener, et T. Magnuson. 1995. « The Eed Mutation Disrupts Anterior Mesoderm Production in Mice ». *Development* 121 (2): 273-85. <https://doi.org/10.1242/dev.121.2.273>.
- Felsenfeld, Gary, et Mark Groudine. 2003. « Controlling the Double Helix ». *Nature* 421 (6921): 448-53. <https://doi.org/10.1038/nature01411>.
- Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, et al. 2017. « GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards ». *Database* 2017 (janvier). <https://doi.org/10.1093/database/bax028>.
- Fortriede, Joshua D, Troy J Pells, Stanley Chu, Praneet Chaturvedi, DongZhuo Wang, Malcom E Fisher, Christina James-Zorn, et al. 2019. « Xenbase: Deep Integration of GEO & SRA RNA-Seq and ChIP-Seq Data in a Model Organism Database ». *Nucleic Acids Research*, novembre, gkz933. <https://doi.org/10.1093/nar/gkz933>.
- Franklin, Tamara B., Holger Russig, Isabelle C. Weiss, Johannes Gräff, Natacha Linder, Aubin Michalon, Sandor Vizi, et Isabelle M. Mansuy. 2010. « Epigenetic Transmission of the Impact of Early Stress Across Generations ». *Biological Psychiatry* 68 (5): 408-15. <https://doi.org/10.1016/j.biopsych.2010.05.036>.
- Fursova, Nadezda A., Neil P. Blackledge, Manabu Nakayama, Shinsuke Ito, Yoko Koseki, Anca M. Farcas, Hamish W. King, Haruhiko Koseki, et Robert J. Klose. 2019. « Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression ». *Molecular Cell* 74 (5): 1020-1036.e8. <https://doi.org/10.1016/j.molcel.2019.03.024>.
- Gaggioli, Vincent, Barbara Le Viet, Thomas Germe, et Olivier Hyrien. 2013. « DNA Topoisomerase II $\alpha$  Controls Replication Origin Cluster Licensing and Firing Time in *Xenopus* Egg Extracts ». *Nucleic Acids Research* 41 (15): 7313-31. <https://doi.org/10.1093/nar/gkt494>.

- Gaidatzis, Dimos, Lukas Burger, Maria Florescu, et Michael B Stadler. 2015. « Analysis of Intronic and Exonic Reads in RNA-Seq Data Characterizes Transcriptional and Post-Transcriptional Regulation ». *Nature Biotechnology* 33 (7): 722-29. <https://doi.org/10.1038/nbt.3269>.
- Gardiner-Garden, M., et M. Frommer. 1987. « CpG Islands in Vertebrate Genomes ». *Journal of Molecular Biology* 196 (2): 261-82. [https://doi.org/10.1016/0022-2836\(87\)90689-9](https://doi.org/10.1016/0022-2836(87)90689-9).
- Gel, Bernat, Anna Díez-Villanueva, Eduard Serra, Marcus Buschbeck, Miguel A. Peinado, et Roberto Malinverni. 2015. « RegioneR: An R/Bioconductor Package for the Association Analysis of Genomic Regions Based on Permutation Tests ». *Bioinformatics*, septembre, btv562. <https://doi.org/10.1093/bioinformatics/btv562>.
- Gillespie, Marc, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, et al. 2022. « The Reactome Pathway Knowledgebase 2022 ». *Nucleic Acids Research* 50 (D1): D687-92. <https://doi.org/10.1093/nar/gkab1028>.
- Goldberg, Aaron D., C. David Allis, et Emily Bernstein. 2007. « Epigenetics: A Landscape Takes Shape ». *Cell* 128 (4): 635-38. <https://doi.org/10.1016/j.cell.2007.02.006>.
- Gramates, L Sian, Julie Agapite, Helen Attrill, Brian R Calvi, Madeline A Crosby, Gilberto dos Santos, Joshua L Goodman, et al. 2022. « FlyBase: A Guided Tour of Highlighted Features ». Édité par V Wood. *Genetics* 220 (4): iyac035. <https://doi.org/10.1093/genetics/iyac035>.
- Grzybowski, Adrian T., Zhonglei Chen, et Alexander J. Ruthenburg. 2015. « Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide ». *Molecular Cell* 58 (5): 886-99. <https://doi.org/10.1016/j.molcel.2015.04.022>.
- Gu, Zuguang, Roland Eils, et Matthias Schlesner. 2016. « Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data ». *Bioinformatics* 32 (18): 2847-49. <https://doi.org/10.1093/bioinformatics/btw313>.
- Halbach, Rebecca, Pascal Miesen, Joep Joosten, Ezgi Taşköprü, Inge Rondeel, Bas Pennings, Chantal B. F. Vogels, et al. 2020. « A Satellite Repeat-Derived PiRNA Controls Embryonic Development of *Aedes* ». *Nature* 580 (7802): 274-77. <https://doi.org/10.1038/s41586-020-2159-2>.
- Hammoud, Saher Sue, David A. Nix, Haiying Zhang, Jahnvi Purwar, Douglas T. Carrell, et Bradley R. Cairns. 2009. « Distinctive Chromatin in Human Sperm Packages Genes for Embryo Development ». *Nature* 460 (7254): 473-78. <https://doi.org/10.1038/nature08162>.
- He, Yu-Fei, Bin-Zhong Li, Zheng Li, Peng Liu, Yang Wang, Qingyu Tang, Jianping Ding, et al. 2011. « Tet-Mediated Formation of 5-Carboxylecytosine and Its Excision by TDG in Mammalian DNA ». *Science* 333 (6047): 1303-7. <https://doi.org/10.1126/science.1210944>.
- Heasman, Janet. 2006. « Patterning the Early *Xenopus* Embryo ». *Development* 133 (7): 1205-17. <https://doi.org/10.1242/dev.02304>.

- Henikoff, Steven, Jorja G. Henikoff, Akiko Sakai, Gabriel B. Loeb, et Kami Ahmad. 2009. « Genome-Wide Profiling of Salt Fractions Maps Physical Properties of Chromatin ». *Genome Research* 19 (3): 460-69. <https://doi.org/10.1101/gr.087619.108>.
- Hickey, Graham JM, Candice L Wike, Xichen Nie, Yixuan Guo, Mengyao Tan, Patrick J Murphy, et Bradley R Cairns. 2022. « Establishment of Developmental Gene Silencing by Ordered Polycomb Complex Recruitment in Early Zebrafish Embryos ». *ELife* 11 (janvier): e67738. <https://doi.org/10.7554/eLife.67738>.
- Hörmanseder, Eva, Angela Simeone, George E. Allen, Charles R. Bradshaw, Magdalena Figlmüller, John Gurdon, et Jerome Jullien. 2017. « H3K4 Methylation-Dependent Memory of Somatic Cell Identity Inhibits Reprogramming and Development of Nuclear Transfer Embryos ». *Cell Stem Cell* 21 (1): 135-143.e6. <https://doi.org/10.1016/j.stem.2017.03.003>.
- Huseyin, Miles K., et Robert J. Klose. 2021. « Live-Cell Single Particle Tracking of PRC1 Reveals a Highly Dynamic System with Low Target Site Occupancy ». *Nature Communications* 12 (1): 887. <https://doi.org/10.1038/s41467-021-21130-6>.
- Inoue, Azusa, Zhiyuan Chen, Qiangzong Yin, et Yi Zhang. 2018. « Maternal *Eed* Knockout Causes Loss of H3K27me3 Imprinting and Random X Inactivation in the Extraembryonic Cells ». *Genes & Development* 32 (23-24): 1525-36. <https://doi.org/10.1101/gad.318675.118>.
- Jung, Yoon Hee, Michael E.G. Sauria, Xiaowen Lyu, Manjinder S. Cheema, Juan Ausio, James Taylor, et Victor G. Corces. 2017. « Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes ». *Cell Reports* 18 (6): 1366-82. <https://doi.org/10.1016/j.celrep.2017.01.034>.
- Kanehisa, M. 2000. « KEGG: Kyoto Encyclopedia of Genes and Genomes ». *Nucleic Acids Research* 28 (1): 27-30. <https://doi.org/10.1093/nar/28.1.27>.
- Kannan, Manoj, et Bhavya Ravi. 2013. « Epigenetics in the Nervous System: An Overview of Its Essential Role ». *Indian Journal of Human Genetics* 19 (4): 384. <https://doi.org/10.4103/0971-6866.124357>.
- Kasinath, Vignesh, Curtis Beck, Paul Sauer, Simon Poepsel, Jennifer Kosmatka, Marco Faini, Daniel Toso, Ruedi Aebersold, et Eva Nogales. 2021. « JARID2 and AEBP2 Regulate PRC2 in the Presence of H2AK119ub1 and Other Histone Modifications ». *Science* 371 (6527): eabc3393. <https://doi.org/10.1126/science.abc3393>.
- Kassis, Judith A, James A Kennison, et John W Tamkun. 2017. « Polycomb and Trithorax Group Genes in *Drosophila* ». *Genetics* 206 (4): 1699-1725. <https://doi.org/10.1534/genetics.115.185116>.
- Katagiri, C, et K Ohsumi. 1994. « Remodeling of sperm chromatin induced in egg extracts of amphibians. » *The International journal of developmental biology* vol. 38,2: 209-16. 38,2: 209-16.
- Kim, Daehwan, Ben Langmead, et Steven L Salzberg. 2015. « HISAT: A Fast Spliced Aligner with Low Memory Requirements ». *Nature Methods* 12 (4): 357-60. <https://doi.org/10.1038/nmeth.3317>.

- Kitaoka, Maiko, Rebecca Heald, et Romain Gibeaux. 2018. « Spindle Assembly in Egg Extracts of the Marsabit Clawed Frog, *Xenopus Borealis* ». *Cytoskeleton* 75 (6): 244-57. <https://doi.org/10.1002/cm.21444>.
- Klungland, Arne, et Adam B. Robertson. 2017. « Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine ». *Oxidative DNA Damage & Repair* 107 (juin): 62-68. <https://doi.org/10.1016/j.freeradbiomed.2016.11.038>.
- Kumar, Suresh, Viswanathan Chinnusamy, et Trilochan Mohapatra. 2018. « Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond ». *Frontiers in Genetics* 9 (décembre): 640. <https://doi.org/10.3389/fgene.2018.00640>.
- Kuznetsov, Jeffim N., Tristan H. Agüero, Dawn A. Owens, Stefan Kurtenbach, Matthew G. Field, Michael A. Durante, Daniel A. Rodriguez, Mary Lou King, et J. William Harbour. 2019. « BAP1 Regulates Epigenetic Switch from Pluripotency to Differentiation in Developmental Lineages Giving Rise to BAP1-Mutant Cancers ». *Science Advances* 5 (9): eaax1738. <https://doi.org/10.1126/sciadv.aax1738>.
- La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, et al. 2018. « RNA Velocity of Single Cells ». *Nature* 560 (7719): 494-98. <https://doi.org/10.1038/s41586-018-0414-6>.
- Lacal, Irene, et Rossella Ventura. 2018. « Epigenetic Inheritance: Concepts, Mechanisms and Perspectives ». *Frontiers in Molecular Neuroscience* 11 (septembre): 292. <https://doi.org/10.3389/fnmol.2018.00292>.
- Landt, Stephen G., Georgi K. Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E. Bernstein, et al. 2012. « ChIP-Seq Guidelines and Practices of the ENCODE and ModENCODE Consortia ». *Genome Research* 22 (9): 1813-31. <https://doi.org/10.1101/gr.136184.111>.
- LeRoy, Gary, Peter A DiMaggio, Eric Y Chan, Barry M Zee, M Andres Blanco, Barbara Bryant, Ian Z Flaniken, et al. 2013. « A Quantitative Atlas of Histone Modification Signatures from Human Cancer Cells ». *Epigenetics & Chromatin* 6 (1): 20. <https://doi.org/10.1186/1756-8935-6-20>.
- Li, H., et R. Durbin. 2009. « Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform ». *Bioinformatics* 25 (14): 1754-60. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, et 1000 Genome Project Data Processing Subgroup. 2009. « The Sequence Alignment/Map Format and SAMtools ». *Bioinformatics* 25 (16): 2078-79. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Haojie, Robert Liefke, Junyi Jiang, Jesse Vigoda Kurland, Wei Tian, Pujuan Deng, Weidi Zhang, et al. 2017. « Polycomb-like Proteins Link the PRC2 Complex to CpG Islands ». *Nature* 549 (7671): 287-91. <https://doi.org/10.1038/nature23881>.
- Liang, Kun, et Sündüz Keleş. 2012. « Normalization of ChIP-Seq Data with Control ». *BMC Bioinformatics* 13 (1): 199. <https://doi.org/10.1186/1471-2105-13-199>.

- Liao, Yuxing, Jing Wang, Eric J Jaehnig, Zhiao Shi, et Bing Zhang. 2019. « WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs ». *Nucleic Acids Research* 47 (W1): W199-205. <https://doi.org/10.1093/nar/gkz401>.
- Lismer, Ariane, Keith Siklenka, Christine Lafleur, Vanessa Dumeaux, et Sarah Kimmins. 2020. « Sperm Histone H3 Lysine 4 Trimethylation Is Altered in a Genetic Mouse Model of Transgenerational Epigenetic Inheritance ». *Nucleic Acids Research* 48 (20): 11380-93. <https://doi.org/10.1093/nar/gkaa712>.
- Love, Michael I, Wolfgang Huber, et Simon Anders. 2014. « Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2 ». *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Macadangdang, Benjamin R, Amit Oberai, Tanya Spektor, Oscar A Campos, Fang Sheng, Michael F Carey, Maria Vogelauer, et Siavash K Kurdistani. 2014. « Evolution of Histone 2A for Chromatin Compaction in Eukaryotes ». *ELife* 3 (juin): e02792. <https://doi.org/10.7554/eLife.02792>.
- Martin, Marcel. 2011. « Cutadapt removes adapter sequences from high-throughput sequencing reads ». *EMBnet.journal* 17 (1): 10. <https://doi.org/10.14806/ej.17.1.200>.
- Mei, Hailiang, Chisayo Kozuka, Ryoya Hayashi, Mami Kumon, Haruhiko Koseki, et Azusa Inoue. 2021. « H2AK119ub1 Guides Maternal Inheritance and Zygotic Deposition of H3K27me3 in Mouse Embryos ». *Nature Genetics* 53 (4): 539-50. <https://doi.org/10.1038/s41588-021-00820-3>.
- Mercer, Tim R., Marcel E. Dinger, et John S. Mattick. 2009. « Long Non-Coding RNAs: Insights into Functions ». *Nature Reviews Genetics* 10 (3): 155-59. <https://doi.org/10.1038/nrg2521>.
- Moore, Lisa D, Thuc Le, et Guoping Fan. 2013. « DNA Methylation and Its Basic Function ». *Neuropsychopharmacology* 38 (1): 23-38. <https://doi.org/10.1038/npp.2012.112>.
- Moosavi, Azam, et Ali Motevalizadeh Ardekani. 2016. « Role of Epigenetics in Biology and Human Diseases ». *Iranian Biomedical Journal*, n° 5 (novembre): 246-58. <https://doi.org/10.22045/ibj.2016.01>.
- Morris, Kevin V., Simon W.-L. Chan, Steven E. Jacobsen, et David J. Looney. 2004. « Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells ». *Science* 305 (5688): 1289-92. <https://doi.org/10.1126/science.1101372>.
- Nenni, Mardi J., Malcolm E. Fisher, Christina James-Zorn, Troy J. Pells, Virgilio Ponferrada, Stanley Chu, Joshua D. Fortriede, et al. 2019. « Xenbase: Facilitating the Use of Xenopus to Model Human Disease ». *Frontiers in Physiology* 10 (février): 154. <https://doi.org/10.3389/fphys.2019.00154>.
- Netea, Mihai G., Jorge Domínguez-Andrés, Luis B. Barreiro, Triantafyllos Chavakis, Maziar Divangahi, Elaine Fuchs, Leo A. B. Joosten, et al. 2020. « Defining Trained Immunity and Its Role in Health and Disease ». *Nature Reviews Immunology* 20 (6): 375-88. <https://doi.org/10.1038/s41577-020-0285-6>.

- Nishimura, Hitoshi, et Steven W. L'Hernault. 2017. « Spermatogenesis ». *Current Biology* 27 (18): R988-94. <https://doi.org/10.1016/j.cub.2017.07.067>.
- O'Brien, Jacob, Heyam Hayder, Yara Zayed, et Chun Peng. 2018. « Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation ». *Frontiers in Endocrinology* 9 (août): 402. <https://doi.org/10.3389/fendo.2018.00402>.
- O'Carroll, Dónal, Sylvia Erhardt, Michaela Pagani, Sheila C. Barton, M. Azim Surani, et Thomas Jenuwein. 2001. « The *Polycomb* -Group Gene *Ezh2* Is Required for Early Mouse Development ». *Molecular and Cellular Biology* 21 (13): 4330-36. <https://doi.org/10.1128/MCB.21.13.4330-4336.2001>.
- Oikawa, Mami, Angela Simeone, Eva Hormanseder, Marta Teperek, Vincent Gaggioli, Alan O'Doherty, Emma Falk, et al. 2020. « Epigenetic Homogeneity in Histone Methylation Underlies Sperm Programming for Embryonic Transcription ». *Nature Communications* 11 (1): 3491. <https://doi.org/10.1038/s41467-020-17238-w>.
- Orlando, David A., Mei Wei Chen, Victoria E. Brown, Snehakumari Solanki, Yoon J. Choi, Eric R. Olson, Christian C. Fritz, James E. Bradner, et Matthew G. Guenther. 2014. « Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome ». *Cell Reports* 9 (3): 1163-70. <https://doi.org/10.1016/j.celrep.2014.10.018>.
- Park, Peter J. 2009. « ChIP-Seq: Advantages and Challenges of a Maturing Technology ». *Nature Reviews Genetics* 10 (10): 669-80. <https://doi.org/10.1038/nrg2641>.
- Pasini, Diego, Adrian P Bracken, Michael R Jensen, Eros Lazzerini Denchi, et Kristian Helin. 2004. « Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity ». *The EMBO Journal* 23 (20): 4061-71. <https://doi.org/10.1038/sj.emboj.7600402>.
- Pasque, Vincent, Jerome Jullien, Kei Miyamoto, Richard P. Halley-Stott, et J.B. Gurdon. 2011. « Epigenetic Factors Influencing Resistance to Nuclear Reprogramming ». *Trends in Genetics* 27 (12): 516-25. <https://doi.org/10.1016/j.tig.2011.08.002>.
- Potok, Magdalena E., David A. Nix, Timothy J. Parnell, et Bradley R. Cairns. 2013. « Reprogramming the Maternal Zebrafish Genome after Fertilization to Match the Paternal Methylation Pattern ». *Cell* 153 (4): 759-72. <https://doi.org/10.1016/j.cell.2013.04.030>.
- Probst, Aline V., Elaine Dunleavy, et Geneviève Almouzni. 2009. « Epigenetic Inheritance during the Cell Cycle ». *Nature Reviews Molecular Cell Biology* 10 (3): 192-206. <https://doi.org/10.1038/nrm2640>.
- Quinlan, Aaron R., et Ira M. Hall. 2010. « BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features ». *Bioinformatics* 26 (6): 841-42. <https://doi.org/10.1093/bioinformatics/btq033>.
- Ramírez, Fidel, Devon P Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S Richter, Steffen Heyne, Friederike Dündar, et Thomas Manke. 2016. « DeepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis ». *Nucleic Acids Research* 44 (W1): W160-65. <https://doi.org/10.1093/nar/gkw257>.

- Raplee, Isaac D., Alexei V. Evsikov, et Caralina Marín de Evsikova. 2019. « Aligning the Aligners: Comparison of RNA Sequencing Data Alignment and Gene Expression Quantification Tools for Clinical Breast Cancer Research ». *Journal of Personalized Medicine* 9 (2): 18. <https://doi.org/10.3390/jpm9020018>.
- Raudvere, Uku, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, et Jaak Vilo. 2019. « G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update) ». *Nucleic Acids Research* 47 (W1): W191-98. <https://doi.org/10.1093/nar/gkz369>.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, et Gordon K. Smyth. 2015. « Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies ». *Nucleic Acids Research* 43 (7): e47-e47. <https://doi.org/10.1093/nar/gkv007>.
- Robinson, M. D., D. J. McCarthy, et G. K. Smyth. 2010. « EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data ». *Bioinformatics* 26 (1): 139-40. <https://doi.org/10.1093/bioinformatics/btp616>.
- Saksouk, Nehmé, Elisabeth Simboeck, et Jérôme Déjardin. 2015. « Constitutive Heterochromatin Formation and Transcription in Mammals ». *Epigenetics & Chromatin* 8 (1): 3. <https://doi.org/10.1186/1756-8935-8-3>.
- Session, Adam M., Yoshinobu Uno, Taejoon Kwon, Jarrod A. Chapman, Atsushi Toyoda, Shuji Takahashi, Akimasa Fukui, et al. 2016. « Genome Evolution in the Allotetraploid Frog *Xenopus Laevis* ». *Nature* 538 (7625): 336-43. <https://doi.org/10.1038/nature19840>.
- Shao, Zhaohui, Florian Raible, Ramin Mollaaghababa, Jeffrey R. Guyon, Chao-ting Wu, Welcome Bender, et Robert E. Kingston. 1999. « Stabilization of Chromatin Structure by PRC1, a Polycomb Complex ». *Cell* 98 (1): 37-46. [https://doi.org/10.1016/S0092-8674\(00\)80604-2](https://doi.org/10.1016/S0092-8674(00)80604-2).
- Siklenka, Keith, Serap Erkek, Maren Godmann, Romain Lambrot, Serge McGraw, Christine Lafleur, Tamara Cohen, et al. 2015. « Disruption of Histone Methylation in Developing Sperm Impairs Offspring Health Transgenerationally ». *Science* 350 (6261): aab2006. <https://doi.org/10.1126/science.aab2006>.
- Silva, Willian T. A. F. 2018. « Methylation Dynamics during the Maternal-to-Zygotic Genome Transition in Dioecious Species ». Édité par Arnar Palsson. *PLOS ONE* 13 (7): e0200028. <https://doi.org/10.1371/journal.pone.0200028>.
- Sitbon, David, Ekaterina Boyarchuk, Florent Dingli, Damarys Loew, et Geneviève Almouzni. 2020. « Histone Variant H3.3 Residue S31 Is Essential for *Xenopus* Gastrulation Regardless of the Deposition Pathway ». *Nature Communications* 11 (1): 1256. <https://doi.org/10.1038/s41467-020-15084-4>.
- Skvortsova, Ksenia, Nicola Iovino, et Ozren Bogdanović. 2018. « Functions and Mechanisms of Epigenetic Inheritance in Animals ». *Nature Reviews Molecular Cell Biology* 19 (12): 774-90. <https://doi.org/10.1038/s41580-018-0074-2>.

- Supek, Fran, Matko Bošnjak, Nives Škunca, et Tomislav Šmuc. 2011. « REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms ». Édité par Cynthia Gibas. *PLoS ONE* 6 (7): e21800. <https://doi.org/10.1371/journal.pone.0021800>.
- Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A. Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, et al. 2009. « Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1 ». *Science* 324 (5929): 930-35. <https://doi.org/10.1126/science.1170116>.
- Tamburri, Simone, Elisa Lavarone, Daniel Fernández-Pérez, Eric Conway, Marika Zanotti, Daria Manganaro, et Diego Pasini. 2020. « Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression ». *Molecular Cell* 77 (4): 840-856.e5. <https://doi.org/10.1016/j.molcel.2019.11.021>.
- Tarasov, Artem, Albert J. Vilella, Edwin Cuppen, Isaac J. Nijman, et Pjotr Prins. 2015. « Sambamba: Fast Processing of NGS Alignment Formats ». *Bioinformatics* 31 (12): 2032-34. <https://doi.org/10.1093/bioinformatics/btv098>.
- Teperek, Marta, Angela Simeone, Vincent Gaggioli, Kei Miyamoto, George E. Allen, Serap Erkek, Taejoon Kwon, et al. 2016. « Sperm Is Epigenetically Programmed to Regulate Gene Transcription in Embryos ». *Genome Research* 26 (8): 1034-46. <https://doi.org/10.1101/gr.201541.115>.
- Thorvaldsdottir, H., J. T. Robinson, et J. P. Mesirov. 2013. « Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration ». *Briefings in Bioinformatics* 14 (2): 178-92. <https://doi.org/10.1093/bib/bbs017>.
- Tian, Shulan, Shuxia Peng, Michael Kalmbach, Krutika S. Gaonkar, Aditya Bhagwate, Wei Ding, Jeanette Eckel-Passow, Huihuang Yan, et Susan L. Slager. 2019. « Identification of Factors Associated with Duplicate Rate in ChIP-Seq Data ». Édité par Luis David Alcaraz. *PLOS ONE* 14 (4): e0214723. <https://doi.org/10.1371/journal.pone.0214723>.
- Tolstorukov, Michael Y., Peter V. Kharchenko, Joseph A. Goldman, Robert E. Kingston, et Peter J. Park. 2009. « Comparative Analysis of H2A.Z Nucleosome Organization in the Human and Yeast Genomes ». *Genome Research* 19 (6): 967-77. <https://doi.org/10.1101/gr.084830.108>.
- Waddington, C. H. 1968. « Towards a Theoretical Biology ». *Nature* 218 (5141): 525-27. <https://doi.org/10.1038/218525a0>.
- Waddington, C.H. 1957. *The Strategy of the Genes*. 0 éd. Routledge. <https://doi.org/10.4324/9781315765471>.
- Wang, Tong, Hui Gao, Wei Li, et Chao Liu. 2019. « Essential Role of Histone Replacement and Modifications in Male Fertility ». *Frontiers in Genetics* 10 (octobre): 962. <https://doi.org/10.3389/fgene.2019.00962>.
- Watson, J. D., et F. H. C. Crick. 1953. « Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid ». *Nature* 171 (4356): 737-38. <https://doi.org/10.1038/171737a0>.

- Wei, Jian-Wei, Kai Huang, Chao Yang, et Chun-Sheng Kang. 2017. « Non-Coding RNAs as Regulators in Epigenetics ». *Oncology Reports* 37 (1): 3-9. <https://doi.org/10.3892/or.2016.5236>.
- Wu, Kou-Juey. 2020. « The Epigenetic Roles of DNA N6-Methyladenine (6mA) Modification in Eukaryotes ». *Cancer Letters* 494 (décembre): 40-46. <https://doi.org/10.1016/j.canlet.2020.08.025>.
- Wu, Shan-Fu, Haiying Zhang, et Bradley R. Cairns. 2011. « Genes for Embryo Development Are Packaged in Blocks of Multivalent Chromatin in Zebrafish Sperm ». *Genome Research* 21 (4): 578-89. <https://doi.org/10.1101/gr.113167.110>.
- Xia, Weikun, Jiawei Xu, Guang Yu, Guidong Yao, Kai Xu, Xueshan Ma, Nan Zhang, et al. 2019. « Resetting Histone Modifications during Human Parental-to-Zygotic Transition ». *Science* 365 (6451): 353-60. <https://doi.org/10.1126/science.aaw5118>.
- Yu, Guangchuang, Li-Gen Wang, et Qing-Yu He. 2015. « ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization ». *Bioinformatics* 31 (14): 2382-83. <https://doi.org/10.1093/bioinformatics/btv145>.
- Yuen, Benjamin T. K., Kelly M. Bush, Bonnie L. Barrilleaux, Rebecca Cotterman, et Paul S. Knoepfler. 2014. « Histone H3.3 Regulates Dynamic Chromatin States during Spermatogenesis ». *Development* 141 (18): 3483-94. <https://doi.org/10.1242/dev.106450>.
- Zang, Chongzhi, Dustin E. Schones, Chen Zeng, Kairong Cui, Keji Zhao, et Weiqun Peng. 2009. « A Clustering Approach for Identification of Enriched Domains from Histone Modification ChIP-Seq Data ». *Bioinformatics* 25 (15): 1952-58. <https://doi.org/10.1093/bioinformatics/btp340>.
- Zang, Chongzhi, Yiren Wang, et Weiqun Peng. 2020. « RECOGNICER: A Coarse-Graining Approach for Identifying Broad Domains from ChIP-Seq Data ». *Quantitative Biology* 8 (4): 359-68. <https://doi.org/10.1007/s40484-020-0225-2>.
- Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, et al. 2008. « Model-Based Analysis of ChIP-Seq (MACS) ». *Genome Biology* 9 (9): R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
- Zhou, Keda, Guillaume Gaullier, et Karolin Luger. 2019. « Nucleosome Structure and Dynamics Are Coming of Age ». *Nature Structural & Molecular Biology* 26 (1): 3-13. <https://doi.org/10.1038/s41594-018-0166-x>.

# Abréviations

**ADN:** Acide DésoxyriboNucléique

**Reads:** Nom des séquences nucléotidique récupérées après séquençage

**Lane:** Couloir de séquençage spécifique à un échantillon

**BAM:** Binary Alignment Map

**SAM:** Sequence Alignment Map

**GTF:** Gene Transfer Format

**GFF3:** Gene Feature Format 3

**Count:** Nom donné au nombre de reads associé à un gène dans la matrice de comptage

**FDR:** False Discovery Rate

**ORA:** Over-representation Analysis

**GSEA:** Gene Set Enrichment Analysis

**NTA:** Network topology-based Analysis

**H2Aub:** H2AK119Ub1

**USP21:** Ubiquitin Specific Peptidase 21

**ZGA:** Zygote Genome Activation

**Scaffold :** Séquences non alignées sur le génome

**scRNA-Seq :** Single-cell RNA-Sequencing

**HMM :** Hidden Markov Model

## Article

1

2 **Sperm derived H2AK119ub1 is required for embryonic gene expression**  
3 **regulation in *Xenopus Laevis***

4 Valentin Francois--Campion<sup>1,^</sup>, Florian Berger<sup>1,^</sup>, Mami Oikawa<sup>2</sup>, Brinda Hountondji<sup>1</sup>,  
5 Maissa Goumeidane<sup>1</sup>, Romain Guibeaux<sup>3</sup> and Jérôme Jullien<sup>1,\*</sup>

6

7 **AFFILIATIONS**

8 <sup>1</sup> Nantes Université, CHU Nantes, INSERM, Center for Research in Transplantation  
9 and Translational Immunology, UMR 1064, F-44000 Nantes, France

10

11 <sup>2</sup> Laboratory of Regenerative Medicine, Tokyo University of Pharmacy and Life  
12 Science, Tokyo, Japan

13 <sup>3</sup> UMR 6290 CNRS-UR1, ERL Inserm U1305, Campus de Beaulieu, Bât. 13, 263  
14 avenue du Général Leclerc, CS 74205 / 35042 Rennes Cedex, France

15

16 <sup>^</sup>equal contribution

17 <sup>\*</sup>corresponding author email: jerome.jullien@inserm.fr

18

19 **ABSTRACT [141words]**

20

21 Deposition of H2AK119ub1 by the polycomb repressive complexe-1 play a key role in  
22 the initiation of facultative heterochromatin formation in somatic cells. Here we  
23 evaluate the contribution of sperm derived H2AK119ub1 to embryonic gene expression.  
24 In *Xenopus Laevis* we found that H2AK119ub1 is retained during spermiogenesis and  
25 into early embryonic development, highlighting its credential for a role in the  
26 transmission of epigenetic information from the sperm to the embryo. Sperm treated  
27 with USP21, a H2A deubiquitylase, just prior to injection to egg, results in  
28 developmental defects associated with gene upregulation. Regulatory regions of USP21  
29 sensitive genes are characterized by a bivalent H2AK119ub1/H3K4me3 configuration.  
30 Upon replication of sperm chromatin, preferential retention of USP21 sensitive genes-  
31 associated H2AK119ub1 is observed, while the bulk of sperm H2AK19ub1 is diluted.  
32 We conclude that the repression of some embryonic genes requires sperm derived  
33 H2AK119ub1 transmission to the developing embryo.

34

35 **INTRODUCTION**

36 Several studies point towards an epigenetic contribution of sperm for embryonic  
37 development. First, in several species paternal exposure to various environmental  
38 insults (diet, stress, pollutant) results in changes to the offspring<sup>1-5</sup>. In addition to  
39 such environmentally induced epigenetic changes, sperm is also proposed to  
40 naturally harbour epigenetic cues that are required for development. Histones

41 have been considered as possible vectors for such epigenetic information. Indeed,  
42 in several species including human, modified histones such as H3K4me3 and  
43 H3K27me3 are found in the sperm chromatin, in particular around developmental  
44 genes and in a pattern reminiscent of that found in embryonic stem cells<sup>6-9</sup>. In a  
45 previous work we provided evidence that in *Xenopus Laevis* some of these  
46 modified histones are homogenously distributed in a population of sperm cells,  
47 hence indicating that they have the required attributes for a role in the regulation  
48 of embryonic development<sup>10</sup>. Lastly, interference with histone variant or post-  
49 translational modifications during the formation of the gametes, or after  
50 fertilisation, lead to embryonic gene misregulation<sup>9,11,12</sup>. Although these later  
51 experiments support a functional role for sperm-derived modified histone in  
52 embryonic development, indirect effect of the interference on the process of  
53 spermiogenesis or early embryos development cannot be ruled out. A direct assay  
54 to test for the requirement of sperm modified histones in embryo development is  
55 still lacking. This is due in part to the difficulty to access the chromatin in the  
56 highly compacted sperm nucleus. In this work we aim to fill this gap and devised  
57 a strategy to alter the sperm epigenome just prior to its introduction to the egg.  
58 We specifically considered H2AK119ub1 as a possible carrier of necessary epigenetic  
59 information from the sperm to the embryos. Indeed, recent works points towards a key  
60 role of polycomb repressive complexe-1 (PRC1), upstream of PRC2, in the  
61 establishment of facultative heterochromatin, both in cultured cells and in embryos.  
62 The H2AK119 ubiquitylation activity of the PRC1 complex appears to be integral part  
63 of this mechanism<sup>13,14</sup>. Here we first found that H2AK119ub1 dynamic during  
64 spermiogenesis and early embryonic development is compatible with a role in the  
65 epigenetic programming of sperm for embryonic development. Then, using the H2A  
66 deubiquitylase USP21<sup>15-18</sup> we applied our epigenetic editing assay to erase  
67 H2AK119ub1 from sperm. Embryos derived from such epigenetically edited sperm are  
68 developmentally impaired. Transcriptome and epigenome analysis indicates that  
69 H2AK119ub, in combination with H3K4me3, is enriched on a set of embryonic genes  
70 misregulated in embryos derived from USP21 treated sperm. Furthermore USP21  
71 sensitive genes preferentially retain H2AK119ub upon sperm chromatin replication.  
72 This enable us to conclude that sperm H2AK119ub is required for the regulation of  
73 embryonic gene expression. To our knowledge this work is the first to document that  
74 modified histone erasure in a mature sperm leads to embryonic defect.

75

## 76 RESULTS

77

78 We first evaluated overall levels of H2AK119ub1 in chromatin during  
79 spermiogenesis and early embryonic development. In frog the last stage of sperm  
80 differentiation is associated with a remodelling of nucleosomes entailing a partial loss  
81 of H2A/H2B while a full complement of H3/H4 is retained (**Fig. 1A** and <sup>10</sup>). WB  
82 analysis captures this loss of H2A and reveal an overall decrease in H2AK119ub1 in  
83 sperm compared to spermatid (**Fig. 1B**, see H2A/H4 and H2AK119ub/H4 ratio,  
84 respectively). Importantly however, the H2AK119ub1 to H2A ratio is higher in sperm  
85 compared to spermatid, indicating that in this context of spermiogenesis-associated  
86 H2A loss, H2AK119ub1 appears selectively retained (**Fig. 1B**). We next assessed  
87 H2AK119ub1 occurrence after fertilisation: in early frog embryos prior to ZGA

88 (blastulae stage) and later during gastrulation and neurulation. Interestingly contrary to  
89 what is seen with H3K27me3, a PRC2 associated modification, H2AK119ub1 is  
90 detected throughout early embryogenesis (**Fig. 1C**). Altogether these observations  
91 highlight H2AK119ub1 credentials for a potential role in the transmission of epigenetic  
92 information from sperm to the embryos. To directly evaluate sperm H2AK119ub1  
93 functional relevance to embryos development we sought to erase this histone  
94 modification, after completion of spermatogenesis, and prior to the formation of  
95 embryos. To that end we incubated mature sperm in an oocyte extract containing  
96 USP21, a H2A deubiquitylase<sup>15,19</sup> (**Fig.1D**). In such extract the sperm nucleus is  
97 partially decondensed, allowing efficient deubiquitylation of H2AK119ub while the  
98 levels of trimethylated H3K4, H3K9 and H3K27 remain unchanged (**Fig.1E**).  
99 Importantly, MNase digestion of oocyte extract treated sperm yields nucleosomal and  
100 subnucleosomal size DNA fragments typical to that of naïve xenopus sperm<sup>9</sup>. (**Fig.1F**,  
101 **lane 1&2**). This indicates that the overall sperm chromatin structure is maintained after  
102 oocyte extract treatment and does not convert to a somatic type of chromatin  
103 organisation as triggered by egg extract treatment<sup>20</sup> (**Fig.1F, lane 3&4**). To then  
104 determine the contribution of sperm-derived chromatin to the development of embryos  
105 we took advantage of the availability of paternally derived embryos in frog. As  
106 described previously, such embryos develop solely on chromatin derived from sperm  
107 and provide a direct measure of the paternal contribution to early development<sup>9</sup>. We  
108 generated paternally derived embryos from sperm pre-incubated in control or USP21  
109 containing oocyte extracts and monitored their development (**Fig.1G**). From  
110 gastrulation onwards, we observed a significant decrease in the developmental potential  
111 of USP21 treated- compared to control treated- sperm derived embryos (**Fig.1H, Table**  
112 **S1**). We conclude that besides its DNA content, sperm also contributes to the embryo  
113 an USP21 sensitive signal necessary for development.

114 The observed defects occur around the time of zygotic gene activation  
115 suggesting that USP21 treatment associated developmental failure might relate to  
116 defective regulation of embryonic gene expression. To evaluate this possibility, we  
117 carried out transcriptome analysis of early gastrulae generated from pool of untreated,  
118 control extract-treated or USP21 extract-treated haploid embryos. We performed RNA-  
119 seq analysis using sperm, oocyte extract, and egg originating from different  
120 preparations for each of the biological replicates. Differential gene expression analysis  
121 indicates that the treatment of sperm with a control oocyte extract has minor effect on  
122 the resulting embryo transcriptome when compared to embryos generated from  
123 untreated sperm (**Fig.2A, right; Fig.S1A**). By contrast USP21 extract treatment leads  
124 to embryonic misregulation of 269 genes (**Fig.2A, left**), the majority of which are  
125 upregulated in embryos derived from USP21 extract treated sperm compared to  
126 embryos derived from control extract treated sperm (74%, 200/269, **Fig.2B, Fig.S1B**)  
127 GO enrichment analysis highlight misregulation in genes related to cell cycle processes.  
128 We then interrogated the developmental dynamic of genes found differentially  
129 expressed in embryos derived from sperm treated with USP21 containing oocyte extract  
130 (thereafter referred to as USP21 sensitive genes). We observe a clear distinction  
131 between up- and down regulated USP21 sensitive genes (**Fig. 2C**). During development  
132 of embryos generated by *in vitro* fertilisation<sup>21</sup> upregulated genes exhibit a strong  
133 maternal contribution followed by clearing prior to gastrulation and re-expression at  
134 tadpole stage while downregulated gene show increased expression around gastrulation.  
135 A recent study indicates that at ZGA, transcription occurs on genes with (“MZ genes”)  
136 or without (“Z genes”) a strong maternal contribution of transcripts<sup>22,23</sup>. Interestingly

137 we observe that USP21 sensitive genes are enriched for MZ genes, indicating that these  
138 genes exhibit zygotic transcription in the context of a strong load of maternally derived  
139 transcripts (**Fig.2C** and **Fig.S1-C**). Taken together these analyses support the  
140 hypothesis whereby USP21 mediated erasure of the repressive histone modifications  
141 H2AK119ub1 from sperm leads to precocious embryonic expression of a subset of  
142 genes that could in turn impact critical embryonic process such as cell division.

143 To better delineate the mechanism by which sperm USP21 treatment mediates  
144 its effect on embryonic gene expression we focused on H2AK119ub1, its cognate  
145 chromatin target. We performed H2AK119ub1 ChIP-seq analysis to map the  
146 distribution of this histone PTM on the sperm genome and evaluate if the presence of  
147 this histone modification on sperm genes is associated with sensitivity to USP21. Data  
148 were generated from biological triplicates and mapped to the X. Laevis 9.2 genome.  
149 The samples were well correlated (**Fig. 3A** and **Fig.S2-A**) and subsequently pooled for  
150 downstream analysis. We first assessed the distribution of H2AK119ub1 around the  
151 transcription start of all sperm genes. H2AK119ub1 accumulates around the TSS of a  
152 fraction of sperm genes (**Fig. 3B**), similar to what is observed in somatic cells<sup>24</sup> and  
153 embryos<sup>14,25</sup>. Noticeably, we observed in the top 10% TSS harbouring H2AK119ub1 an  
154 enrichment for genes associated with developmental function (**Fig. 3B**), and  
155 reminiscent of the association reported with H3K27me3 in many species<sup>6-9</sup>. We next  
156 sought to more generally identify regions associated with H2AK119ub1 genome-wide.  
157 H2AK119ub1 is a pervasive histone modification typically present on ~10% of nuclear  
158 H2A<sup>26,27</sup>, and it accumulates over region covering larger part of the genome than other  
159 modified histones with broad distribution such as H3K27me3 (**Fig.S2-B**). To identify  
160 region of H2AK119ub1 genomic enrichment we therefore used RECOGNICER, a  
161 coarse-graining approach designed to capture enrichment over broad scale<sup>28</sup>. Compared  
162 to standard peak caller such as MACS2, RECOGNICER enable the capture of larger  
163 peaks with a greater overlap with H3K27me2 sites of enrichment (**FigS2**). Although  
164 we observed peaks of accumulation around genes, such as illustrated for USP21  
165 sensitive genes pacs2.L or trib2.S (**Fig. 3C**), a large proportion of H2AK119ub1 also  
166 accumulates on the sperm genome in distal region away from genes (**Fig. 3D**).  
167 Noticeably we observe that H2AK119ub1 is significantly associated with sperm TSSs  
168 and enhancers, further strengthening its potential for gene regulation across generation  
169 (**Fig. 3E**). We therefore next tested if various genomic features associated with USP21  
170 sensitive genes are enriched for H2AK119ub1 peaks and detected an enrichment in  
171 their gene body and enhancers (**Fig. 3F**). Altogether this survey of H2AK119ub1  
172 distribution in sperm reveals its enrichment on gene regulatory regions and identify an  
173 enrichment of H2AK119ub1 on USP21 sensitive genes.

174 We observed an association of USP21 sensitive genes with H2AK119ub1 in sperm, but  
175 many more genes harbour this modification without being affected by USP21  
176 treatment. Epigenetic control of gene expression often entails a combination of  
177 epigenetic cues such as those observed in X inactivation<sup>29</sup> or in transcriptional  
178 reprogramming<sup>30</sup>. We therefore next investigated if USP21 sensitivity could arise from  
179 H2AK119ub1 removal from sperm genes harbouring additional epigenetic  
180 modifications to core histones. We considered H3K27me3 as it is often combined with  
181 H2AK119ub1 in polycomb complex mediated repression of gene expression. We also  
182 considered H3K4me3 which conversely associates with gene activity and is linked with  
183 poised status of transcription when co-existing with repressive marks<sup>31</sup>. As expected  
184 from sperm nucleoparticles composition, INPUT fractions yield fragments of  
185 nucleosomal and subnucleosomal size (**suppl S3**). Similarly, we observe an expected

186 depletion of sub-nucleosomal sized DNA fragment in H2AK119ub1 ChIP sample,  
187 given that the subnucleosome are devoid/depleted of H2A. Surprisingly however,  
188 observe the opposite trend with H3K4me3 and H3K27me3 ChIP samples, both  
189 showing enrichment for sub-nucleosomal sized DNA fragment. These data indicate  
190 that globally, H2AK119ub1 and H3K4me3/H3K27me3 modifications are generally  
191 present on distinct nucleoproteic particles: H2AK119ub1 on nucleosomes and  
192 H3K4me3/H3K27me3 on subnucleosomes. We then used H3K4me3<sup>10</sup>, H3K27me3<sup>10</sup>  
193 and H2AK119ub1 ChIP-seq data to focus on genomic regions enriched for those  
194 modifications in sperm (Peaks, **Fig. 4A**). We then focused on the association of  
195 H2AK119ub1 with methylated K4 and/or K27 on histone H3. While most  
196 H2AK119ub1 happens independently of these marks, ~24% of H2AK119ub1 occurs  
197 in combination with H3K4me3 and/or H3K27me3 (**Fig. 4B**). We observed a very  
198 strong promoter bias for peaks of H2AK119ub1 co-occurring with methylated  
199 H3K4me3, an expected trend given that H3K4me3 is known to strongly associate with  
200 gene promoter (**Fig. 4C**). By contrast when occurring in the absence of H3K4me3,  
201 H2AK119ub1 is located away from promoter and in intergenic regions. We next  
202 interrogated the association of USP21 sensitive genes with these different categories of  
203 H2AK119ub1 peaks. We observe a slight association of USP21 sensitive gene features  
204 with either H2AK119ub1 only peaks (H2AK119ub1+; H3K4me3-; H3K27me3-) or  
205 trivalent H2AK119ub1 peaks (H2AK119ub1+; H3K4me3+; H3K27me3+) (**Fig. 4D**).  
206 Strikingly however, USP21 sensitive genes show an enrichment for bivalent  
207 H2AK119ub1/H3K4me3 and depletion for bivalent H2AK119ub1/H3K27me3  
208 chromatin configuration. This analysis suggests that the effect on embryonic gene  
209 expression of sperm USP21 treatment arise from erasure of H2AK119ub1 in the context  
210 of a bivalent H2AK119ub1/H3K4me3 chromatin status.

211 To further evaluate how sperm H2AK119ub1/H3K4me3 could impact  
212 embryonic gene expression we next sought to follow their fate after fertilisation.  
213 Following delivery to the egg, protamines are removed from sperm chromatin and a  
214 canonical nucleosomal structure is re-instated. In conjunction with these events the  
215 sperm chromatin is replicated. Therefore, the extensive changes to sperm chromatin  
216 that follow fertilization have the potential to profoundly alter the sperm derived  
217 modified histones landscape. To determine the precise fate of sperm modified histones  
218 we turned to interphase egg extract that can mimic the sperm chromatin remodelling  
219 events that follow fertilization<sup>10</sup>. We use either untreated sperm or sperm having gone  
220 through egg extract mediated replication and carried out calibrated H2AK119ub1 and  
221 H3K4me3 ChIP (**Fig. 5A**). In such calibrated ChIP a known amount of drosophila cells  
222 was spiked-in the frog sperm sample just prior to chromatin fragmentation<sup>24,32,33</sup>. The  
223 drosophila spike-in enables to normalise and then directly compare ChIP signal  
224 intensity between frog sperm samples. In that way we can not only determine whether  
225 replication leads to maintenance or loss of local modified histones accumulation as in  
226 standard ChIP analysis (presence/absence of a peak), but we can also monitor any  
227 decrease in signal that could arise from replicative dilution (signal intensity of the peak).  
228 To first gain a global view of replication associated changes to the genomic distribution  
229 of sperm modified histone we compared H2AK119ub1 and H3K4me3 peaks in  
230 untreated *versus* egg extract treated sperm. We observed a marked disparity in  
231 replication dynamics between sperm H3K4me3, mostly stable (77% of sperm peaks  
232 conserved after replication (19345/24982) and 51% stable peaks overall  
233 (19345/37747)), and sperm H2AK119ub1, mostly changing (23.5% of sperm peaks  
234 maintained after replication (15796/67233) and 15% stable peaks overall

235 (15796/104956)) (**Fig. 5B**). Most of the replication-induced redistribution of  
236 H2AK119ub1 happens in intergenic region, while H2AK119ub1 peaks stable through  
237 this process are mostly located on promoter (**Fig. 5C**). In this moving landscape, the  
238 subset of H2AK119ub1 genomic location associated with H3K4me3 in sperm appears  
239 selectively conserved (93% of sperm bivalent peaks are maintained after replication  
240 (5222/5597) (**Fig. 5B**). Importantly, the significant enrichment of USP21 sensitive  
241 genes promoter and gene body for a bivalent H2AK119ub1/H3K4me3 chromatin  
242 configuration reported in sperm holds true post-replication (**Fig. 5D**). We conclude that  
243 replication profoundly remodel H2AK119ub1 genomic distribution and that in this  
244 dynamic epigenetic landscape USP21 sensitive genes are characterised by a stable  
245 association with bivalent H2AK119ub1/H3K4me3 chromatin domains. We next  
246 considered the possibility that while a local accumulation of H2AK119ub1 could be  
247 maintained post-replication, sperm modified histones could have undergone replicative  
248 dilution. In other word, during replication the local accumulation of histone along the  
249 genome could be maintained albeit at a reduced level, for example through  
250 redistribution of sperm derived modified histones between the replicated strands of  
251 DNA. Mechanistically this is important as in frog ~8 rounds of replication separate  
252 fertilisation from the main wave of ZGA and could therefore leads to complete loss of  
253 sperm derived modified histones. Therefore, using ChIP signal calibrated with  
254 drosophila spike-in we directly compared the intensity of modified histone peaks pre-  
255 replication and post-replication. Focusing on gene promoters, where bivalent  
256 H2AK119ub1/H3K4me3 configuration is mainly located (see **Fig. 4C**), we observe a  
257 global decrease of both H2AK119ub1 and H3K4me3 signal upon replication. This is  
258 suggesting that replication induces dilution of modified histones derived from the  
259 sperm chromatin (blue line, **Fig. 5E**). When considering USP21 sensitive gene however  
260 we observe a different trend. While an overall decrease of H2AK119ub1&H3K4me3  
261 signal is also detected upon replication, H2AK119ub1 is clearly less pronounced on the  
262 TSSs of this subset of gene than seen genome wide (**Fig 5.E**), a feature shared with the  
263 broader set of gene that have maternal-zygotic contribution. We conclude that bivalent  
264 H2AK119ub1/H3K4me3 chromatin associated with USP21 sensitive gene in sperm  
265 partially resist replication induced dilution that affect the bulk of TSSs.

## 266 **DISCUSSION**

267 In mouse embryonic stem cell, polycomb repressive domain requires ring1A/B  
268 histone ubiquitylase activity of the PRC1 complex<sup>13,14,24,32</sup>. The establishment of such  
269 domain in ESC occurs in a stepwise manner with a variant PRC1 complex initiating  
270 their formation, followed by recruitment of the PRC2 complex and subsequent  
271 methylation of H3K27me3. Such sequential events have also been described during  
272 NPC differentiation<sup>34</sup> and X chromosome inactivation<sup>35</sup>. Recently, this sequence of  
273 events have been investigated in the mouse and zebrafish embryos<sup>25,36,37</sup>. In these  
274 embryos H2AK119ub1 is also required for the deposition of H3K27me3 and  
275 transcriptional repression around the time of zygotic gene activation. Our observation  
276 that H2AK119ub1 is present in pre-ZGA embryo in the absence of H3K27me3 argues  
277 for the existence of a similar mechanism in frog<sup>38,39</sup>. Interference with H2AK119ub1  
278 either by knocking down PRC1 component<sup>36</sup> in oocyte or erasing H2AK119ub1 from  
279 embryos<sup>25,37</sup> leads to misregulation of genes at ZGA and is associated with embryonic  
280 lethality. While clearly demonstrating a role for H2AK119ub1 in regulation of  
281 transcription around ZGA, these studies have not clearly delineated to which extent  
282 embryonic H2AK119ub1 is instructed/inherited from sperm and/or oocyte  
283 H2AK119ub1. Importantly, in contrast to previous studies, in this work the sperm

284 epigenome is edited without affecting the oocyte or zygote epigenetic  
285 machinery<sup>9,11,12,25,36,37</sup>. Therefore, here we can rule out that any observed effects  
286 transcription stem from an indirect consequence of H2AK119ub1 depletion on  
287 oogenesis or pre-ZGA processes. Additionally, our study uses haploid embryos so that  
288 all measured transcription arises from an embryonic template derived only from an  
289 epigenetically edited paternal chromatin. Embryos derived from H2AK119ub1 edited  
290 sperm exhibit transcriptional misregulation, mostly upregulation, compatible with a  
291 repressive function of sperm derived H2AK119ub. Taken together this provides strong  
292 evidence that sperm derived H2AK119ub1 is required for proper embryonic genes  
293 regulation. Such role imply that H2AK119ub1 is present on sperm chromatin and is  
294 maintained in the embryos to exert its repressive effect. We identify such an association  
295 of misregulated genes with H2AK119ub1 in frog sperm and find that this association  
296 is preferentially maintained on these genes upon sperm chromatin replication, further  
297 supporting a direct role for this histone post-translational modification in the epigenetic  
298 programming of sperm for regulation of embryonic transcription. A contrasting picture  
299 emerges regarding sperm H2AK119ub1 across species. In mouse sperm, despite a loss  
300 of 99% of histones during spermiogenesis, H2AK119ub1 domains are present and  
301 partially retained in one cell stage embryos<sup>37</sup>, raising the possibility that sperm derived  
302 H2AK119ub1 could exert a similar role in mammals as found here in frogs.  
303 Surprisingly however Zebrafish sperm, which fully maintain histones during  
304 spermiogenesis, seems to lack this modification altogether<sup>25</sup>. Spermiogenesis in  
305 *Xenopus Laevis* entails differential remodelling of histones with removal of H2A/H2B  
306 while retaining a full complement of H3/H4<sup>10,40</sup>. Associated with H2A loss, we detect  
307 a preferential retention of H2AK119ub1 during spermiogenesis, further suggesting that  
308 this histone PTM is not merely a leftover of transcriptional history but represent an  
309 epigenetic programming of sperm for the next stage of development. Additionally, we  
310 found that sperm post-translationally modified histones are generally present in  
311 different nucleoprotein complex: H2AK119ub1 is retained on nucleosomes, while  
312 H3K4me3 and H3K27me3 are retained on subnucleosomes depleted of H2A/H2B. It  
313 would be interesting to evaluate if histones PTM fate in embryos is influenced by their  
314 location on nucleosomes or subnucleosomes. We find that genes misregulated by sperm  
315 H2AK119ub1 editing are associated with H2AK119ub/H3K4me3. Of note such  
316 bivalency is not assumed to be a colocalization of these histone modifications on the  
317 same nucleoparticles but instead inferred from their co-occurrence on misregulated  
318 gene regulatory regions. Additionally, not all affected genes harbour this configuration.  
319 We cannot rule out that part of the genes are indirectly affected, for example as  
320 downstream target of a “primary” H2AK119ub1 target. Alternatively, it is possible to  
321 envisage that additional combination of epigenetic cues besides H3K4me3 associate  
322 with gene sensitivity to H2AK119ub1 erasure from sperm. Lastly, we observe that most  
323 affected genes are normally expressed late in embryogenesis (tadpole stage) a feature  
324 also seen in mouse with a large proportion of gene sensitive to H2AK119ub1 depletion  
325 expressed post-implantation<sup>8</sup>. It is therefore possible that the effect of H2AK119ub1  
326 expression is only partially revealed at early embryonic stage since transcription factors  
327 targeting these genes might not be present. Identifying such candidate TF would be  
328 important in future studies. Owing to the very large size of H2AK119ub1 domains,  
329 standard motif discovery methods are proving insufficient and alternative approach  
330 would be needed.

331 Similarly to what is found in the mouse, we observe a very dynamic landscape  
332 of H2Ak119ub1 around the time of replication of sperm chromatin. Our work advance

333 knowledge from previous studies by showing that part of H2AK119ub1 found in  
334 embryos is paternally instructed. This necessary function of sperm derived modified  
335 histones has important bearing on our understanding of inter/trans generational  
336 transmission of epigenetic information. One could indeed envisage that such a naturally  
337 occurring mechanism of epigenetic information transmission to the embryos could be  
338 hijacked to transmit environmentally induced sperm epigenetic cues. Based on the very  
339 dynamic nature of histone PTMs around fertilisation it is likely that transmission to the  
340 embryos will be context (combination of marks) and location depends. Future studies  
341 will aim at evaluating to which sperm histones PTMs differ between individuals and to  
342 which extent they are maintained in embryos.

343

#### 344 **FIGURE LEGENDS**

345 **Figure 1: Decreased developmental potential of embryos derived from sperm treated**  
346 **with USP21, a H2AK119 deubiquitylase.** (A) Schematic representation of *Xenopus*  
347 *laevis* nucleosome remodelling during spermiogenesis<sup>10,40</sup>. (B) Left, Western blot  
348 analysis of H2AK119Ub1, H2A and H4 histones. Right signal ratio of H2AK119Ub1  
349 to H2A, H2A to H4, and H2AK119Ub1 to H4 in spermatid and sperm. (C) Western  
350 blot analysis of H2AK119Ub1, H3K4me3, H3K9me3, H3K27me3 and H4 in *Xenopus*  
351 *laevis* embryo pre-ZGA (blastulae, stage 7), post- ZGA (gastrulae (stage 11<sup>1/2</sup>), and  
352 neurula (stage 21)) (staging according to [https://www.xenbase.org/entry/landmarks-](https://www.xenbase.org/entry/landmarks-table.do)  
353 [table.do](https://www.xenbase.org/entry/landmarks-table.do)). (D) Schematic representation of sperm treatment with USP21 oocyte extract.  
354 (E) Left, Western blot of H2AK119Ub1, H3K4me3, H3K9me3, H3K27me3 and H4  
355 histone on oocyte extract treated sperm, USP21 treated sperm, or untreated sperm.  
356 Right, signal ratio of H2AK119Ub1 to H4 in oocyte extract treated sperm, USP21  
357 treated sperm, or untreated sperm. (F) DNA fragments generated by MNase digestion  
358 of *Xenopus laevis* sperm, oocyte extract treated sperm, egg extract treated sperm and  
359 somatic cell. (G) Experimental design of paternally derived embryos from sperm or  
360 USP21 treated sperm. (H) Developmental progression to blastula, gastrula or  
361 swimming tadpole stage of haploid embryos derived from control or a USP21 oocyte  
362 extract treated sperm. Chi-squared test, \* p-value<0.05.]

363 **Figure 2: Gene misregulation occurs in gastrulae derived from sperm treated with**  
364 **USP21.** (A) MA-Plot showing as y-axis the log foldchange (LogFC) of gene expression  
365 between USP21 oocyte extract treated sperm derived embryo versus untreated sperm  
366 (left) and between control oocyte extract treated sperm and untreated sperm (right). X-  
367 axis values are average gene expression across samples. (B) Heatmap showing  
368 expression of differentially expressed gene in gastrulae derived from USP21 oocyte  
369 extract treated sperm versus that derived from control oocyte extract treated sperm. (C)  
370 GO enrichment of sensitive genes to USP21 treatment. (D) Heatmap showing the  
371 expression of USP21 sensitive genes during *Xenopus laevis* early embryo development.  
372 Data from<sup>21</sup>.

373 **Figure 3: H2AK119ub1 is associated with gene regulatory regions in sperm.**

374 (A) Genome browser view of the log ratio of H2AK119ub1 ChIP to input signal over  
375 the *hoxa.L* cluster on 3 sperm replicates. (B) Heat map of H2AK119ub1 signal (log  
376 (ChIP/Input)) on all *Xenopus Laevis* 9.2 genes TSS +/-5kB. GO enrichment is shown  
377 for the top 10% genes. (C) Genome browser view of the log ratio of sperm  
378 H2Ak119ub1 ChIP to input signal on two USP21 sensitive genes, *pacs2.l* and *trib2.s*.

379 Regions of enrichment (peaks) are shown as black boxes. **(D)** Genomic distribution of  
380 sperm H2AK119Ub1 peaks. **(E)** Fold enrichment of H2AK119Ub1 peaks  
381 (observed/random, over 1000 randomisations) on different genomic features, \*\*\*  
382  $p < 0.001$ , randomisation test. **(F)** Fold enrichment of H2AK119Ub1 peaks  
383 (observed/random, over 10000 permutations) on promoter, enhancer, and gene body of  
384 USP21 sensitive genes. \*  $p\text{-value} < 0.05$ ; permutation test.

385 **Figure 4:** *USP21 sensitive genes are characterised by a bivalent*  
386 *H2AK119ub1/H3K4me3 chromatin status in sperm.*

387 **(A)** Genome browser view of log ratio of ChIP to input signal of sperm H2AK119Ub,  
388 H3K4me3, and H3K27me3 on *pacs2.L*, a USP21 sensitive gene. Regions of enrichment  
389 (peaks) are shown underneath the gene track. **(B)** Pie chart indicating the proportion of  
390 H2AK119Ub1 peaks alone or in combination with H3K4me3 and/or H3K27me3 on the  
391 sperm chromatin. **(C)** Genomic distribution of sperm H2AK119Ub1 peaks according  
392 to association with H3K4me3 and/or H3K27me3 peaks. **(D)** Fold enrichment of  
393 different H2AK119Ub1 peaks categories (observed/random, over 10000 permutations)  
394 on promoter, enhancer, and gene body of USP21 sensitive genes. \*  $p\text{-value} < 0.05$ ; \*\*  
395  $p\text{-value} < 0.01$ , permutation test

396 **Figure 5:** *USP21 sensitive genes escape H2AK119ub1 replicative dilution*

397 **(A)** Experimental design of the use of egg extract mediated sperm replication and  
398 drosophila melanogaster spike-in to assess modified histones fate by calibrated ChIP-  
399 seq. **(B)** Venn diagram illustrating the change in H3K4me3, H2AK119Ub, or  
400 H3K4me3/H2AK119ub loci during sperm chromatin replication. Left, all H2AK119ub  
401 peaks; Middle all H3K4me3 peaks; Right, overlapping H2AK119ub/ H3K4me3 peaks.  
402 **(C)** Genomic distribution of sperm peaks before or after replication as well as the subset  
403 of peaks lost, maintained, or gained during replication. **(D)** Fold enrichment of lost,  
404 maintained, or gained H2AK119Ub1 peaks (observed/random, over 10000  
405 permutations) on promoter, enhancer, and gene body of USP21 sensitive genes. \*  $p\text{-}$   
406  $\text{value} < 0.05$ ; \*\*  $p\text{-value} < 0.01$ , \*\*\* $p\text{-value} < 0.001$ , permutation test. **(E)** Calibrated ChIP  
407 signal on TSS +/- 5kB for all genes (blue), USP21 genes (red), Zygotic genes (orange),  
408 and maternal zygotic gene (green)<sup>22,23</sup>. \*\*  $p\text{-value} < 0.01$ , Wilcoxon-Mann-Whitney test

409 **Supplemental Figure S1:** *Gene misregulation occurs in gastrulae derived from*  
410 *sperm treated with a H2AK119 deubiquitylase*

411 **(A)** Principal Component analysis of replicate samples used in this study (USP21,  
412 USP21 extract treated sperm derived gastrulae; Control, control extract treated sperm  
413 derived gastrulae; Untreated, Untreated sperm derived blastulae. **(B)** Dendrogram from  
414 hierarchical clustering of replicate samples used in this analysis. **(C)** Heat map  
415 depicting expression of USP21 sensitive genes identified when comparing USP21  
416 extract treated sperm derived gastrulae to control extract treated sperm derived  
417 gastrulae. Gene expression is shown also for blastulae derived from untreated sperm.  
418 **(D)** USP21 sensitive genes are enriched for genes with maternal and zygotic expression  
419 in *Xenopus Laevis*. The table shows p-Value resulting from an hypergeometric test to  
420 evaluate enrichment of USP21 sensitive gene for either maternal-zygotic (MZ) or  
421 zygotic only (Z) genes. MZ and Z dataset taken from<sup>22,23</sup>

422 **Supplemental Figure S2.** *Sperm H2AK119Ub1 ChIP-seq analysis.* **(A)** Spearman  
423 correlation of sperm or replicated sperm ChIP-seq replicates for H2AK119Ub1,

424 H3K27me3 and H3K4me3. Bins of 10kb are used to assess the genome wide correlation.  
425 (B) Fingerprintplot analysis of Input and H3K4me3, H3K27me3, H2AK119ub1 ChIP  
426 samples. (C) Heat map displaying H2AK119Ub1 ChIP/Input signal +/- 5kb from  
427 RECOGNICER peaks center (top) versus an equivalent number of genomic region  
428 randomly selected (bottom). On the right RECOGNICER peaks are spitted between  
429 those identified by RECOGNICER only versus those detected in both RECOGNICER  
430 and MACS2. (D) Overlap between RECOGNICER H2AK119Ub1 peaks, MACS2  
431 H2AK119Ub1 peaks, and H3K7me3 peaks. (E) Box plot indicating sizes of  
432 RECOGNICER H2AK119Ub1 peaks (all RECOGNICER peaks, peaks overlapping  
433 with those of MACS2, or peaks specific to RECOGNICER), as well as MACS2 peaks.  
434 (F) Genome track showing log ratio of ChIP to input signal for H2AK119Ub1 and  
435 H3K27me3 as well as the region where peaks are detected by MACS2 and  
436 RECOGNICER (boxed area).

437 **Supplemental Figure S3. Sperm nucleoparticules composition** (A) Distribution of the  
438 fragment size for H2AK119ub1 in spermatid, (B) H2AK119ub1 in sperm, (C)  
439 H2AK119ub1 in replicated sperm, (D) H3K27me3 in sperm, (E) H3K4me3 in sperm  
440 and (F) H3K4me3 in replicated sperm. Spearman correlation of sperm or replicated  
441 sperm.

442

443 **Table S1: Development of embryos derived from oocyte treated sperm**

444 **Table S2: Lists of H2AK119ub1 associated genes**

## 445 MATERIALS AND METHODS

### 446 Contact for Reagent and Resource Sharing

447 Further information and requests for resources and reagents should be directed  
448 to and will be fulfilled by the Lead Contact, Dr J. Jullien ([jerome.jullien@inserm.fr](mailto:jerome.jullien@inserm.fr)).

### 449 Experimental Model and Subject Details

450 For embryo work, mature *Xenopus laevis* females were obtained from Nasco (901  
451 Janesville Avenue, PO Box 901, Fort Atkinson, WI 53538-  
452 0901; <https://www.enasco.com/xenopus>). Our work with *Xenopus* complied  
453 with all relevant ethical regulations for animal testing and research. Work with  
454 frog was covered under the Home Office Project License PPL 70/8591 and frog  
455 husbandry and all experiments were performed according to the relevant  
456 regulatory standards. Animals were kept in a Marine Biotech recirculating system  
457 at a density of one adult/3l, with 10% water change per day. Water was  
458 sequentially filtered with mechanical pad sump filter, nitrifying bacteria filter,  
459 mechanical canister filter, carbon filter, and UV sterilized. Water quality  
460 parameters were as follow: conductivity 1500us; temperature 17-22°C; PH 6-8.  
461 Photoperiod was set to 12h ON/12h OFF. Frogs are fed twice per week with Royal  
462 Horizon 4.5 mm pellets (skretting, <https://www.skrettingfishfeeds.co.uk/>).  
463 Unconsumed food was removed 10 min after the start of feeding. Oocytes and  
464 testis were obtained from Université de Rennes 1 CRB Xénope Bâtiment 13 -  
465 Campus de Beaulieu F - 35042 Rennes cedex 35042 RENNES CEDEX

466 **Oocyte extract.** Fresh ovaries were manually dissociated in 15-25 oocytes clumps.  
467 3mL of dissociated ovary fragments were transferred into 12.5mL 1X Modified  
468 Barth's Saline (MBS, 88mM NaCl, 1 mM KCl, 10 mM Hepes, 2.4 mM NaHCO<sub>3</sub>, 1.7  
469 mM MgSO<sub>4</sub>, 1.3 mM Ca(NO<sub>3</sub>)<sub>2</sub>, 0.8mM CaCl<sub>2</sub>) supplemented with 250uL of  
470 5mg/mL Liberase (Roche, 05 401 127 001) and incubated for 2h30 at room  
471 temperature (RT) with gentle rocking. Defolliculated oocytes were washed three  
472 times with 1X MBS and healthy stage 5-6 oocytes manually sorted. Oocytes were  
473 then injected in the vegetal pole with 25ng of USP21-NLS-HA mRNA, or control  
474 GFP mRNA and incubated for 3 days at 18°C in 1X MBS, renewing medium and  
475 discarding dead cells every day. ~500 oocytes were collected in an Eppendorf tube  
476 on ice, the excess 1X MBS discarded and replaced with 1X MBS supplemented with  
477 Protease Inhibitor Cocktail (Sigma, P2714-1BTL). The oocytes were then stacked  
478 up by centrifugation at 100g for 1min at RT and excess medium carefully  
479 discarded. Oocytes were then thoroughly vortexed and then centrifuged at  
480 21 000g for 10min at 4°C. The clear intermediate phase between yolk (top) and  
481 pigment (bottom) was retrieved. The collected fraction was further cleared by an  
482 additional centrifugation at 21 000g for 10min at 4°C. The oocyte extract was then  
483 aliquoted and stored at -80°C.

484 **Oocyte extract treatment.** Permeabilized sperm stored in SSB at -80°C were  
485 thawed, washed with two volumes of ice cold 1XPBS and centrifuged at 800g for  
486 8min at 4°C. 2M sperm in 1X PBS were centrifuged at 21 000g for 1min at 4°C. The  
487 supernatant was discarded and the pelleted sperm resuspended in 10uL of oocyte  
488 extract on ice. The sperm incubated in the oocyte extract for 20-40min at 23°C on  
489 a heating block. Treatment was stopped by addition of 90uL of ice cold 1X PBS and  
490 centrifugation at 21 000g for 1min at 4°C. The supernatant discarded and the  
491 remaining pellets resuspended in SSB for -80°C long term storage.

492 **Protein extraction and Western Blot analysis.** Sperm samples stored in SSB were  
493 washed in 1XPBS and 1M sperm pellets were resuspended in 30uL 2X Blue Eye  
494 (600mM NaCl, 1% Triton X100, 0.5% Na-Deocycholate, 4% Sodium Dodecyl  
495 Sulfate, 20% Glycerol, 0.0004% Bromophenol blue, 175mM Tris-HCl, 10%  
496 Betamercapto ethanol), boiled at 96°C for 5min and centrifuged at 21 000g for 5min  
497 at 4°C. The supernatant was transferred to a new tube and directly use for analysis or  
498 stored at -20°C. 4-20% mini-PROTEAN® TGX (Biorad, #4561094) were loaded with  
499 10uL of sample and western blot analysis were performed according to standard  
500 protocols. For blotting nitrocellulose membranes, a semi-dry transfer was used (Biorad,  
501 Trans-Blot® SD Semi-Dry Transfer Cell) 30-40min at 25V. The following primary  
502 antibodies were used: anti-H2A (1:2500, Cell Signaling, 3636), anti-H2AK119ub1  
503 (1:2500, Cell Signaling, 8240), anti-H3K4me3 (1:2500, Abcam, ab8580), anti  
504 H3K9me3 (1:2500, Abcam, ab8898), anti-H3K27me3(1:2500, Cell Signaling, 9735S),  
505 anti-H4 (1:2500, Abcam, ab31830) and anti-HA (1:2500, Sigma, H9658). Membranes  
506 were incubated with primary antibodies 2h at RT or overnight (ON) at 4 °C. For the  
507 secondary reactions, anti-rabbit (1:25000, Thermo Fisher ,A21076) , anti-mouse IgG  
508 Alexa Fluor 680 (1:25000, Thermo Fisher ,A21058),Goat anti-rabbit Alexa 800  
509 (1:25,000, Thermo Fisher, A32735), or IRDye 800CW Goat Anti-Mouse IgG (H+L)  
510 (LI-COR, 926-32210) were used. The signal was captured either on a ChemiDoc MP  
511 Imaging System (biorad) or an Odyssey scanner (LI-COR).

512 **Sperm Chromatin Immunoprecipitation (ChIP)** Chromatin fractionation and  
513 chromatin immunoprecipitation (ChIP) were performed as described before<sup>10,41</sup>  
514 with slight modifications. 80  $\mu$ l of Magnetic beads (M280 Sheep anti-rabbit IgG,  
515 Invitrogen, #11204D) were used per reaction and all wash steps were carried out  
516 with a magnet. Beads were washed with 1 ml of TE pH 8.0, lysis buffer (LB)(1:1  
517 mixture of Buffer 1 and MNase buffer, see below), and incubated in 1 ml of LB with  
518 100  $\mu$ l of 10 mg/ml BSA at 4C for 30 min (“pre-blocking”). Pre-blocked beads were  
519 washed with 1 ml of LB twice and resuspend in 80  $\mu$ l of LB. Half of beads were  
520 incubated with antibody overnight and leftovers were stored in 4C for “pre-clear  
521 chromatin” step. For ChIP, the following antibodies were used: anti-H3K4me3  
522 (Abcam, ab8580), anti-H2AK119ub1 (Cell Signaling, 8240). Digitonin  
523 permeabilised sperm nuclei or egg extract treated sperm nuclei were resuspended  
524 in 50  $\mu$ l of Buffer 1 (0.3 M Sucrose, 15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 5  
525 mM MgCl<sub>2</sub>, 0.1 mM EGTA, 0.5 mM DTT) and added 50  $\mu$ l of Buffer 1 with detergent  
526 (Buffer 1 including 0.5% NP40 and 1% NaDOC). Samples were incubated for 10  
527 min on ice. 100  $\mu$ l of MNase buffer (0.3 M Sucrose, 85 mM Tris, 3 mM MgCl<sub>2</sub>, 2  
528 mM CaCl<sub>2</sub>, 2.5U of micrococcal nuclease: Roche 10107921001) was added in to  
529 each tube (0.5 million of cells per tube). Tubes were incubated at 37C for 30 min  
530 in pre-warmed water bath. Reaction was stopped by adding 2  $\mu$ l of 0.5 M EDTA  
531 pH8.0 in the same order as started. Tubes were vortexed and placed on ice for at  
532 least 5 min. Supernatant and pellet were separated by centrifugation at 13,000  
533 rpm for 10 min at room temperature. Beads (stored at 4C for pre-clear chromatin)  
534 were washed with 1 ml of LB twice. Supernatant and EDTA-free Protease Inhibitor  
535 Cocktail were added into the beads and rotating on the wheel at 4C for 60 min.  
536 10% of volume of pre-cleared chromatin was taken as “input” and stored at 4C.  
537 Antibody conjugated beads were washed with 1 ml of LB twice. Pre-cleared  
538 chromatin were added to washed beads and incubated at 4C for 6 h. After  
539 incubation, beads were washed with 1 ml of Washing buffer A (50 mM Tris-HCl  
540 pH 7.5, 10 mM EDTA and 75 mM NaCl) on the wheel at 4C for 5 min and buffer was  
541 discarded carefully. Subsequently, beads were washed with 1 ml of Washing  
542 buffer B (50 mM Tris-HCl pH 7.5, 10 mM EDTA and 125 mM NaCl) and buffer was  
543 discarded carefully. Another 1 ml of Washing buffer B was added and transferred  
544 to 1.5 ml Protein lo-bind tubes (Eppendorf, 0030108116) with beads. Buffer was  
545 discarded carefully. 150  $\mu$ l of elution buffer (1:9 mixture of 10% SDS and TE) was  
546 added. Tubes were placed at 25C on the Eppendorf thermomixer and shaken at  
547 800 rpm for 15 min. Supernatant were taken and transferred to the new tube  
548 using magnetic stand. Another 150  $\mu$ l of elution buffer was added and repeated  
549 this step (in total 300  $\mu$ l). Input samples were diluted to 300  $\mu$ l total by adding TE.  
550 2  $\mu$ l of 2 mg/ml RNase A (heat-inactivated or DNase-free) was added and  
551 incubated at 37C for 30 min. 4  $\mu$ l of 10 mg/ml Proteinase K was added and  
552 incubated at 55C overnight. After incubation, tubes were shortly spun. 300  $\mu$ l of  
553 Phenol/chloroform were added and vortexed for 30 sec. Samples were  
554 centrifuged at 13,000 rpm for 10 min at room temperature. 1  $\mu$ l of Glycogen, 1/10  
555 volume of NaAcetate and 2 volume of 100% EtOH were added and vortexed. Tubes  
556 were stored on dry ice for 20-30 min. After freezing, tubes were centrifuged at  
557 13,000 rpm for 30 min at 4C. Supernatant was discarded carefully and 500  $\mu$ l of  
558 70% EtOH was added. Tubes were centrifuged at 13,000 rpm for 10 min at 4C.  
559 Pellets were dried and suspended in 30  $\mu$ l of ddH<sub>2</sub>O.

## 560 **Egg extract treatment**

561 Extracts were generated as previously described<sup>10</sup>. For the quantification of DNA  
562 replication in individual nuclei in egg extract, 33  $\mu$ M rhodamine dUTP (Roche,  
563 11534378910) was added at the indicated times (20, 40, 60, 80 mins after  
564 incubation) for 5 min as indicated, and the reaction was stopped with 1x PBS and  
565 fixed in 4% paraformaldehyde. The reaction was spun through a 30% sucrose  
566 cushion in PBS. The sperm cells were rinsed twice in 1x PBS for 10 min and  
567 mounted in Vectashield mounting medium with DAPI (Vector laboratories, H-  
568 1200). Counting of rhodamine positive cells were performed on an  
569 epifluorescence microscope.

570 ***Xenopus* sperm collection** *Xenopus* sperm collection  
571 was performed as described before<sup>10,41</sup>. For each round of sperm purification,  
572 testes from six adult *Xenopus laevis* males were isolated and manually cleaned  
573 from blood vessels and fat bodies in 1  $\times$  Marc's Modified Ringers (MMR, 100 mM  
574 NaCl, 2 mM KCl, 1 mM MgSO<sub>4</sub>, 2 mM CaCl<sub>2</sub>, 3 mM HEPES, pH 7.4) using forceps and  
575 paper tissues. It is crucial to clean the testes well from any non-testicular tissues,  
576 as otherwise the cells released from the tissues may negatively affect the final  
577 purity of isolated cells. Subsequently, testes were torn into small pieces with  
578 forceps and homogenized with 2–3 strokes of a dounce homogenizer (tissue from  
579 one testis at a time). The cell suspension was then filtered to remove tissue debris  
580 and cell clumps (CellTrics, cat. 04-0042-2317) and spun down at 800 rcf, 4°C, for  
581 20 min. Supernatant was discarded and the cell pellet was resuspended in 12 mL  
582 of 1  $\times$  MMR. If any red blood cells were visible at the bottom of the pellet (a result  
583 of incomplete removal of blood vessels), only the uncontaminated part of the  
584 pellet was recovered, taking extreme care not to disturb the red blood cells.  
585 Subsequently, step gradients of iodixanol (Optiprep; Sigma, D1556; 60% iodixanol  
586 in water) in 1 $\times$ MMR final were manually prepared in pre-chilled 14 mL ultra- clear  
587 centrifuge tubes (Beckman Coulter, #344060) in the following order from the  
588 bottom to the top of the tube: 4 mL of 30% iodixanol, 1 mL of 20% iodixanol, 5 mL  
589 of 12% iodixanol (all in 1  $\times$  MMR), and 2 mL of cell suspension in 1  $\times$  MMR on top.  
590 Gradients were spun down in a pre-chilled SW40Ti rotor at 7500 rpm (10,000g),  
591 4°C, for 15 min, deceleration without brake (Beckman Coulter Ultracentrifuge,  
592 Optima L-100XP). The pelleted fraction, containing mature sperm, was collected.  
593 Collected fractions were diluted six times with 1 $\times$  MMR and collected by spinning  
594 first at 805 rcf, 4°C, for 20 min and respinning at 3220 rcf, 4°C, for 20 min to pellet  
595 remaining cells. Pelleted cells were subsequently permeabilized with Digitonin  
596 (10 mg/mL as a final concentration, Sigma, D141) for 5 min at room temperature  
597 as detailed before (Smith et al., 2006)

598 **Library preparation and sequencing** ChIP samples were subjected to ChIP-seq  
599 library preparation using QIAseq Ultralow Input Lib UDI-A Kit (Qiagen, 180497) and  
600 Agencourt AMPure XP beads (Beckman coulter, A63881). Samples pre- and post-  
601 library protocol were analyzed on a LabChip GX (PerkinElmer) using DNA 1KHS and  
602 1K Assay LabChIP, respectively. Libraries were quantitated using NEBNext Library  
603 Quant Kit (NEB,E7630,). Paired-end sequencing was carried out either on a NOVA-  
604 seq 6000 or HiSeq 1500 sequencer (Illumina).

## 605 **RNA-Seq analysis**

606 Adapters were trimmed and low-quality reads filtered using cutadapt (v1.18).  
607 Sample quality was confirmed with fastqc (v0.11.8). fastq were aligned to *Xenopus*  
608 *laevis* genome (v9.2) using STAR aligner (v2.7.3.a). Count table was generated  
609 using HTSeq-count (v0.12.3) and exon as genes reference. MultiQC (v1.9)  
610 summarized data pre-processing. Differential gene expression analysis of the  
611 count table was performed on R (v3.6) using DESeq2tools. Details of the pipeline  
612 used can be found at <https://gitlab.univ-nantes.fr/E114424Z/BulkRNAseq>. Gene  
613 ontology analysis were carried out with gprofiler using human orthologs of  
614 *Xenopus laevis* development RNA-seq data were from published  
615 dataset GSE73430.

## 616 **ChIP-Seq analysis**

617 Adapters were trimmed and low-quality reads filtered using cutadapt (v1.18).  
618 Sample quality was confirmed with fastqc (v0.11.9). Alignment was performed  
619 with bwa mem against *Xenopus laevis* (v9.2) or against a - *Xenopus laevis* (v9.2)  
620 *Drosophila melanogaster* (v6.40) hybrid genome generated with bwa index  
621 (v0.7.17-r1188). Reads extraction and analysis of each species was performed  
622 with samtools view and samtools flagstat (v1.12). ChIP-Seq sample quality was  
623 assessed by Deeptools (v3.5.1) using plotFingerprint function and  
624 bamPEFragmentSize. Replicates quality was visualised with plotCorrelation using  
625 sperman correlation after genome fragmentation into bins of 10Kb. Replicates  
626 were pooled by subsampling the same amount of reads per replicates using  
627 sambamba (v0.6.6) and samtools. PCR duplicates were marked by PICARD mark  
628 duplicates tools. Peak calling was processed with MACS2 (-q 0.01 --nomodel --  
629 extsize 147) for H3K4me3, with MACS2 broad option for H3K27me3 (--broad --  
630 broad-cutoff 0.1 -q 0.01 --nomodel --extsize 147) and RECOGNICER (qvalue 0.01)  
631 /MACS2 (--broad --broad-cutoff 0.1 -q 0.01 --nomodel --extsize 147) for  
632 H2AK119ub1. *Xenopus laevis* signal was normalized using *Drosophila*  
633 *melanogaster* spike-in following SpikChIP algorithm.  
634 (<https://github.com/eblancoga/spikChIP/blob/master/spikChIP.pl>). Wilcoxon  
635 test was performed on signal 2Kb around TSS of genes to compare H2AK119Ub  
636 and H3K4me3 signal between USP21 sensitive genes and all genes. ChIPSeeker  
637 and ChIPPeakAnno were used to annotate peaks on R. RegioneR was used to  
638 perform peaks permutation. Signal was computed using Deeptools with  
639 BamCompare. Heatmap was produced with computeMatrix and plotHeatmap. IGV  
640 was used to visualize output.

## 641 **Data Availability**

642 The *Xenopus laevis* ChIP-seq and RNA-seq data will be deposited on GEO. Data set  
643 used in this studies are GSE126599 ,GSE73430, GSE125982.

644

## 645 **AUTHORS CONTRIBUTION**

646 Conceptualization: JJ; Methodology: JJ, VF-C, FB; Validation: JJ, VF-C, FB; Formal  
647 analysis: VF-C, FB, MO, MG; Investigation: JJ, VF-C, FB, RG, MO, BH; Writing original  
648 draft: JJ; Writing review and editing: JJ, FB, VFC. Visualisation: FB, VFC;  
649 Supervision: JJ; Project administration: JJ; Funding acquisition: JJ.

650

651 **COMPETING INTERESTS**

652 The authors declare no competing interests

653

654 **ACKNOWLEDGEMENTS**

655 J.J is funded from an ANR grant. J.J., F.B., V.F-C, M.G. are funded from a Connect  
656 Talent grant from Ville de Nantes and Region pays de la Loire. We would like to  
657 acknowledge the support of Xenbase.

658

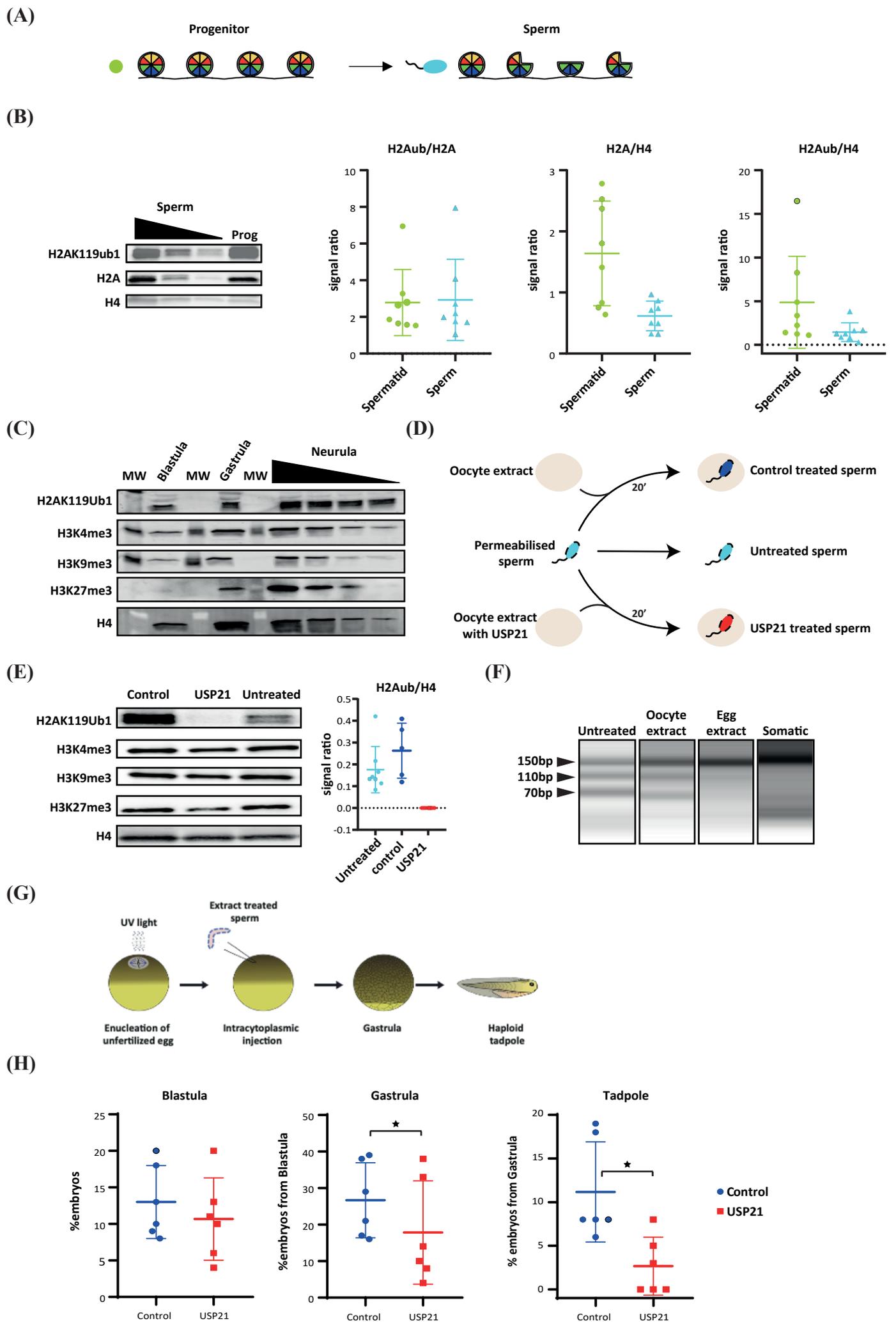
659 **REFERENCE**

- 660 1. Öst, A. *et al.* Paternal diet defines offspring chromatin state and  
661 intergenerational obesity. *Cell* (2014) doi:10.1016/j.cell.2014.11.005.
- 662 2. Dias, B. G. & Ressler, K. J. Parental olfactory experience influences behavior  
663 and neural structure in subsequent generations. *Nat. Neurosci.* (2014)  
664 doi:10.1038/nn.3594.
- 665 3. Radford, E. J. *et al.* In utero undernourishment perturbs the adult sperm  
666 methylome and is linked to metabolic disease transmission. *Science* (80-. ).  
667 **345**, 1–26 (2015).
- 668 4. Ben Maamar, M., Beck, D., Nilsson, E. E., Kubsad, D. & Skinner, M. K.  
669 Epigenome-wide association study for glyphosate induced transgenerational  
670 sperm DNA methylation and histone retention epigenetic biomarkers for  
671 disease. *Epigenetics* **16**, (2021).
- 672 5. Lismer, A., Siklenka, K., Lafleur, C., Dumeaux, V. & Kimmins, S. Sperm  
673 histone H3 lysine 4 trimethylation is altered in a genetic mouse model of  
674 transgenerational epigenetic inheritance. *Nucleic Acids Res.* **48**, (2020).
- 675 6. U., B. *et al.* Repressive and active histone methylation mark distinct promoters  
676 in human and mouse spermatozoa. *Nature Structural and Molecular Biology*  
677 (2010).
- 678 7. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes  
679 for embryo development. *Nature* (2009) doi:10.1038/nature08162.
- 680 8. Wu, S. F., Zhang, H. & Cairns, B. R. Genes for embryo development are  
681 packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res.*  
682 (2011) doi:10.1101/gr.113167.110.
- 683 9. Teperek, M. *et al.* Sperm is epigenetically programmed to regulate gene  
684 transcription in embryos. *Genome Res.* **26**, (2016).
- 685 10. Oikawa, M. *et al.* Epigenetic homogeneity in histone methylation underlies  
686 sperm programming for embryonic transcription. *Nat. Commun.* (2020)  
687 doi:10.1038/s41467-020-17238-w.

- 688 11. Siklenka, K. *et al.* Disruption of histone methylation in developing sperm  
689 impairs offspring health transgenerationally. *Science* (80-. ). (2015)  
690 doi:10.1126/science.aab2006.
- 691 12. Murphy, P. J., Wu, S. F., James, C. R., Wike, C. L. & Cairns, B. R. Placeholder  
692 Nucleosomes Underlie Germline-to-Embryo DNA Methylation  
693 Reprogramming. *Cell* (2018) doi:10.1016/j.cell.2018.01.022.
- 694 13. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation  
695 drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, (2014).
- 696 14. Tamburri, S. *et al.* Histone H2AK119 Mono-Ubiquitination Is Essential for  
697 Polycomb-Mediated Transcriptional Repression. *Mol. Cell* **77**, (2020).
- 698 15. Nakagawa, T. *et al.* Deubiquitylation of histone H2A activates transcriptional  
699 initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes*  
700 *Dev.* (2008) doi:10.1101/gad.1609708.
- 701 16. Joo, H. Y. *et al.* Regulation of cell cycle progression and gene expression by  
702 H2A deubiquitination. *Nature* (2007) doi:10.1038/nature06256.
- 703 17. Cao, J. & Yan, Q. Histone ubiquitination and deubiquitination in transcription,  
704 DNA damage response, and cancer. *Frontiers in Oncology* (2012)  
705 doi:10.3389/fonc.2012.00026.
- 706 18. Chen, D., Dai, C. & Jiang, Y. Histone H2A and H2B Deubiquitinase in  
707 Developmental Disease and Cancer. *Cancer Transl. Med.* (2015)  
708 doi:10.4103/2395-3977.168578.
- 709 19. Jin, J. *et al.* The deubiquitinase USP21 maintains the stemness of mouse  
710 embryonic stem cells via stabilization of Nanog. *Nat. Commun.* (2016)  
711 doi:10.1038/ncomms13594.
- 712 20. Gaggioli, V., Le Viet, B., Germe, T. & Hyrien, O. DNA topoisomerase II $\alpha$   
713 controls replication origin cluster licensing and firing time in *Xenopus* egg  
714 extracts. *Nucleic Acids Res.* **41**, (2013).
- 715 21. Session, A. M. *et al.* Genome evolution in the allotetraploid frog *Xenopus*  
716 *laevis*. *Nature* **538**, (2016).
- 717 22. Chen, H. & Good, M. C. Imaging nascent transcription in wholemount  
718 vertebrate embryos to characterize zygotic genome activation. in *Methods in*  
719 *Enzymology* vol. 638 (2020).
- 720 23. Chen, H., Einstein, L. C., Little, S. C. & Good, M. C. Spatiotemporal  
721 Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. *Dev.*  
722 *Cell* **49**, (2019).
- 723 24. Conway, E. *et al.* BAP1 enhances Polycomb repression by counteracting  
724 widespread H2AK119ub1 deposition and chromatin condensation. *Mol. Cell*  
725 **81**, (2021).
- 726 25. Hickey, G. J. M. *et al.* Establishment of developmental gene silencing by  
727 ordered polycomb complex recruitment in early zebrafish embryos. *Elife* **11**,

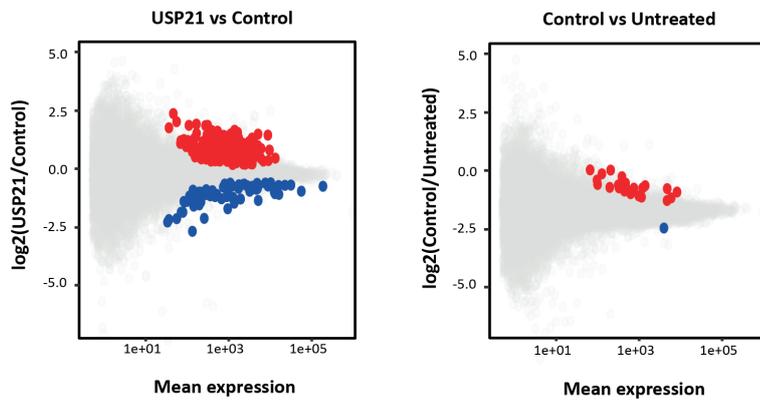
- 728 (2022).
- 729 26. Albright, S. C., Nelson, P. P. & Garrard, W. T. Histone molar ratios among  
730 different electrophoretic forms of mono- and dinucleosomes. *J. Biol. Chem.*  
731 **254**, (1979).
- 732 27. Matsui, S., Seon, B. K. & Sandberg, A. A. Disappearance of a structural  
733 chromatin protein A24 in mitosis: Implications for molecular basis of  
734 chromatin condensation. *Proc. Natl. Acad. Sci. U. S. A.* **76**, (1979).
- 735 28. Zang, C., Wang, Y. & Peng, W. RECOGNICER: A coarse-graining approach  
736 for identifying broad domains from ChIP-seq data. *Quant. Biol.* **8**, (2020).
- 737 29. Heard, E. & Disteché, C. M. Dosage compensation in mammals: Fine-tuning  
738 the expression of the X chromosome. *Genes and Development* vol. 20 (2006).
- 739 30. Jullien, J. *et al.* Gene Resistance to Transcriptional Reprogramming following  
740 Nuclear Transfer Is Directly Mediated by Multiple Chromatin-Repressive  
741 Pathways. *Mol. Cell* **65**, (2017).
- 742 31. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key  
743 Developmental Genes in Embryonic Stem Cells. *Cell* (2006)  
744 doi:10.1016/j.cell.2006.02.041.
- 745 32. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines  
746 Polycomb-Mediated Gene Repression. *Mol. Cell* **74**, (2019).
- 747 33. Blanco, E., Di Croce, L. & Aranda, S. SpikChIP: A novel computational  
748 methodology to compare multiple ChIP-seq using spike-in chromatin. *NAR*  
749 *Genomics Bioinforma.* **3**, (2021).
- 750 34. Tsuboi, M. *et al.* Ubiquitination-Independent Repression of PRC1 Targets  
751 during Neuronal Fate Restriction in the Developing Mouse Neocortex. *Dev.*  
752 *Cell* **47**, (2018).
- 753 35. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X  
754 Chromosome Inactivation. *Cell* **176**, (2019).
- 755 36. Mei, H. *et al.* H2AK119ub1 guides maternal inheritance and zygotic deposition  
756 of H3K27me3 in mouse embryos. *Nat. Genet.* **53**, (2021).
- 757 37. Chen, Z., Djekidel, M. N. & Zhang, Y. Distinct dynamics and functions of  
758 H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nat. Genet.*  
759 **53**, (2021).
- 760 38. Van Heeringen, S. J. *et al.* Principles of nucleation of h3k27 methylation  
761 during embryonic development. *Genome Res.* (2014)  
762 doi:10.1101/gr.159608.113.
- 763 39. Akkers, R. C. *et al.* A Hierarchy of H3K4me3 and H3K27me3 Acquisition in  
764 Spatial Gene Regulation in *Xenopus* Embryos. *Dev. Cell* (2009)  
765 doi:10.1016/j.devcel.2009.08.005.
- 766 40. Mann, M., Risley, M. S., Eckhardt, R. A. & Kasinsky, H. E. Characterization

- 767 of spermatid/sperm basic chromosomal proteins in the genus *Xenopus* (Anura,  
768 Pipidae). *J. Exp. Zool.* (1982) doi:10.1002/jez.1402220209.
- 769 41. Teperek, M. *et al.* Sperm and spermatids contain different proteins and bind  
770 distinct egg factors. *Int. J. Mol. Sci.* **15**, (2014).
- 771

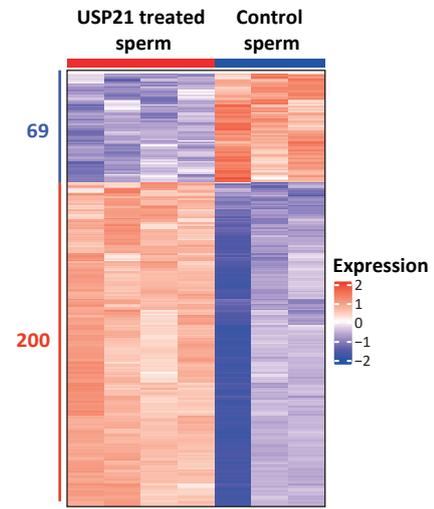


**FIGURE 1**

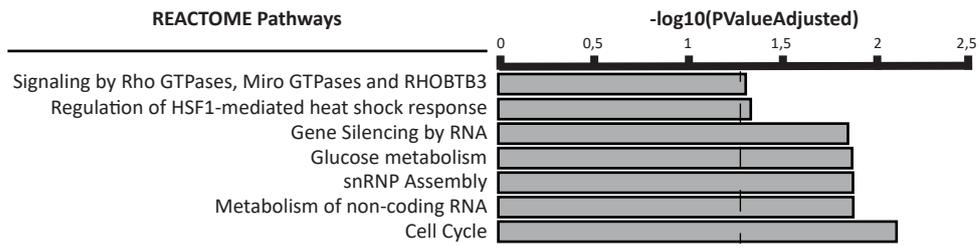
(A)



(B)



(C)



(D)

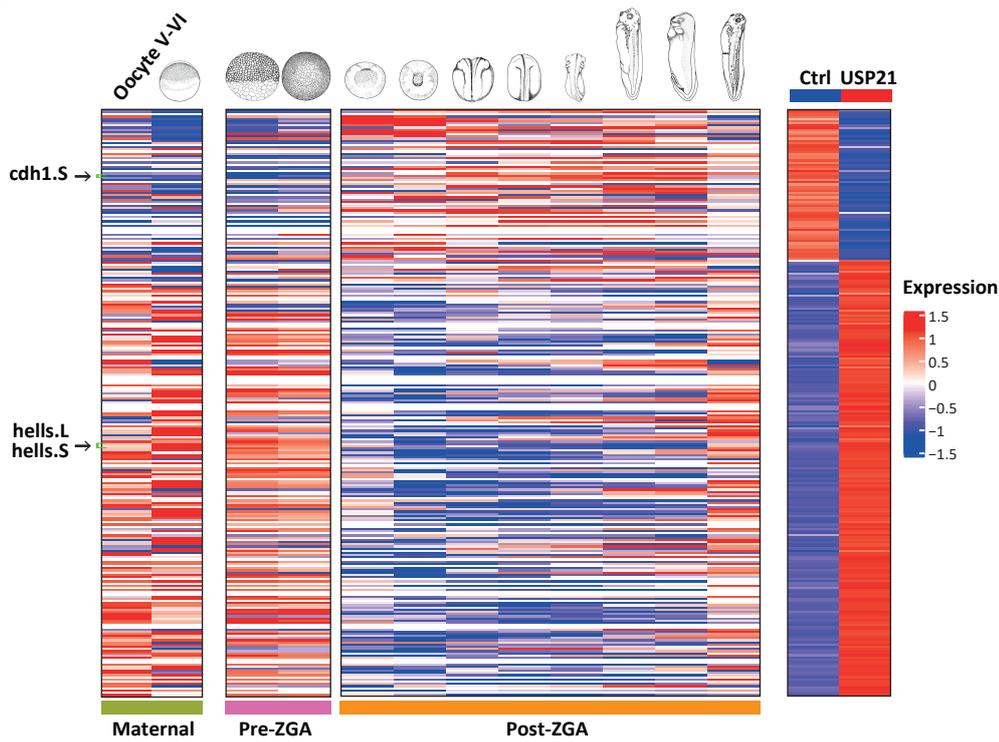
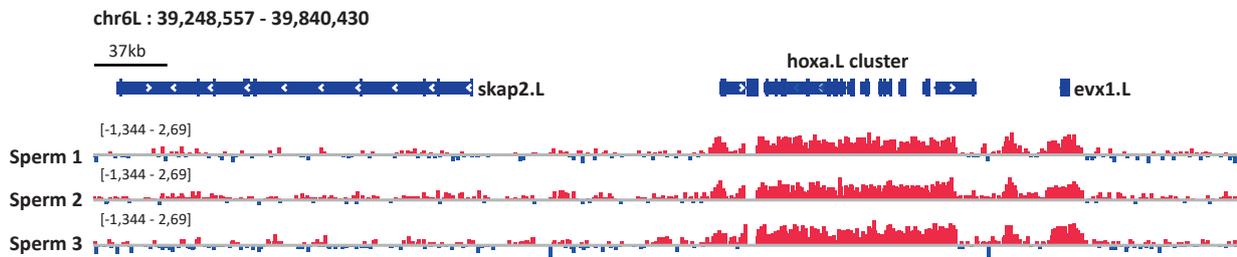
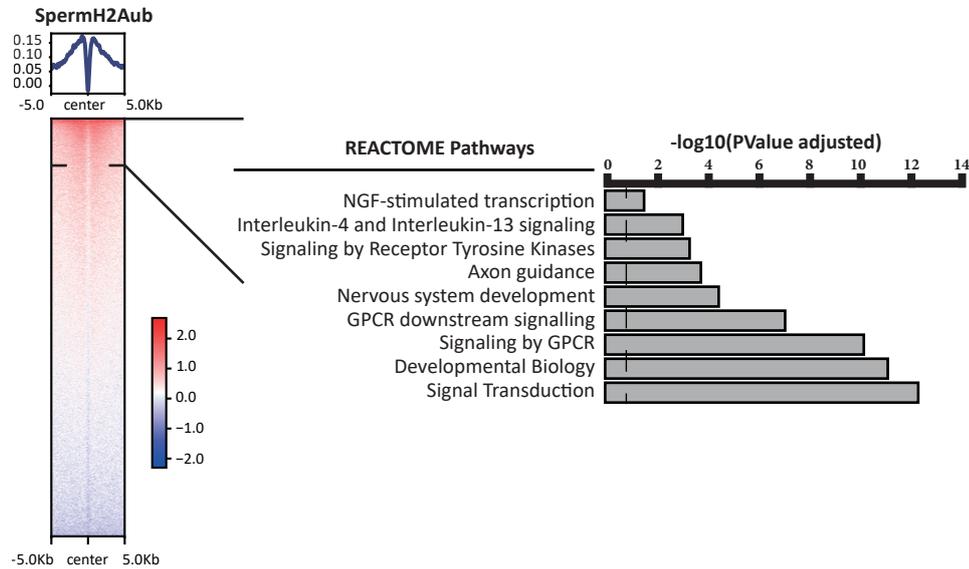


FIGURE 2

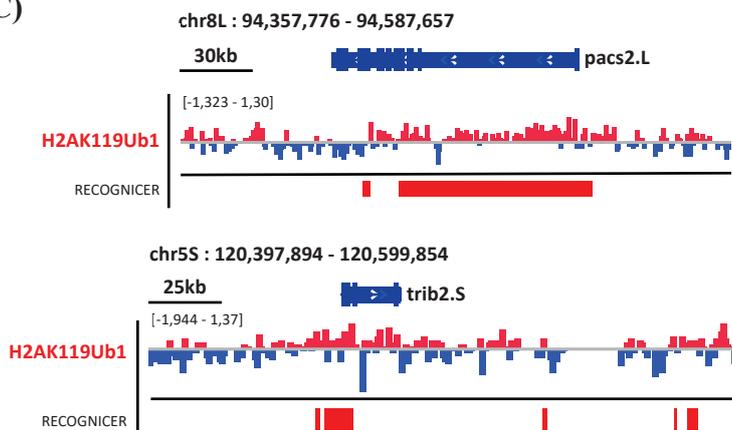
(A)



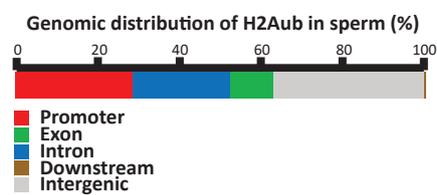
(B)



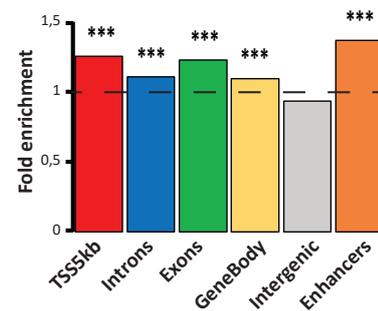
(C)



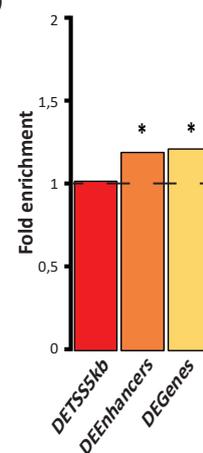
(D)

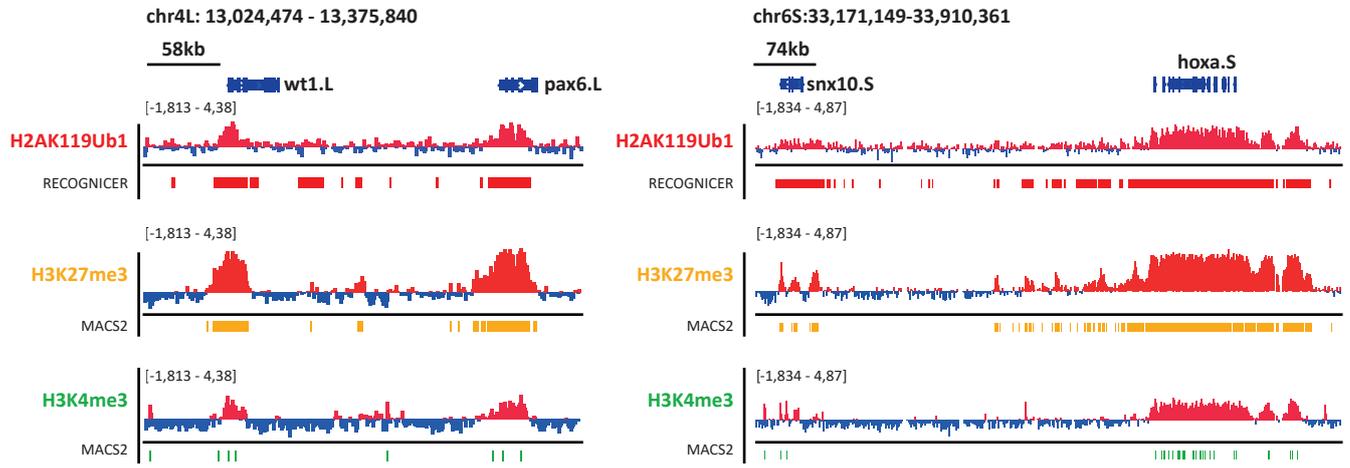
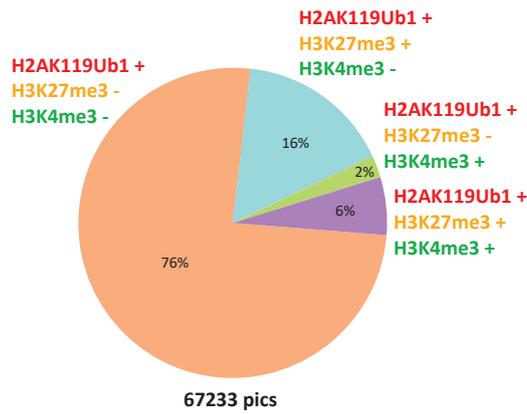
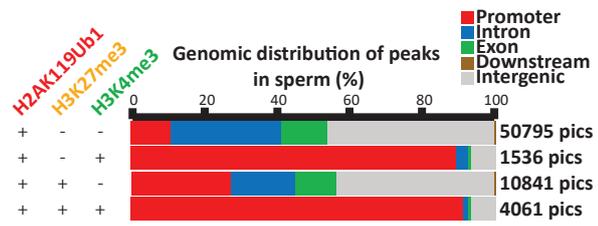
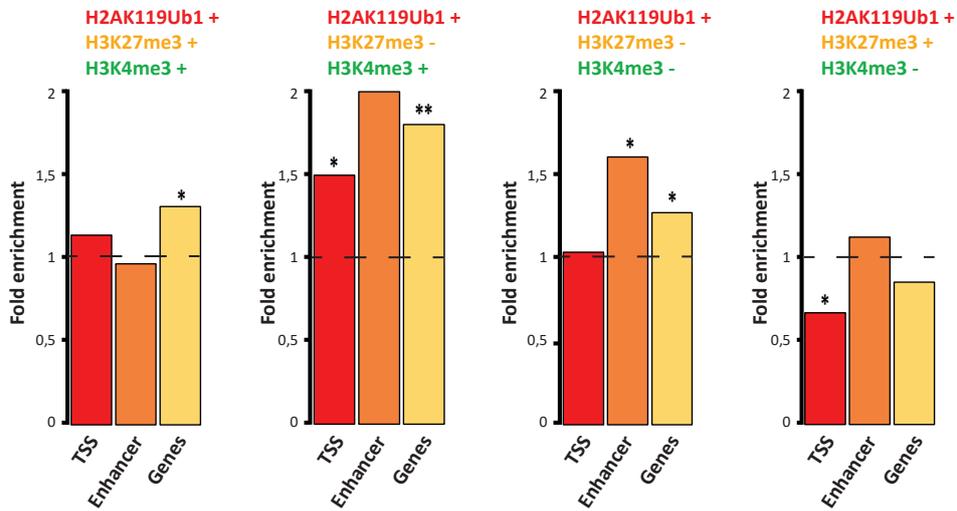


(E)



(F)



**(A)****(B)****(C)****(D)****FIGURE 4**

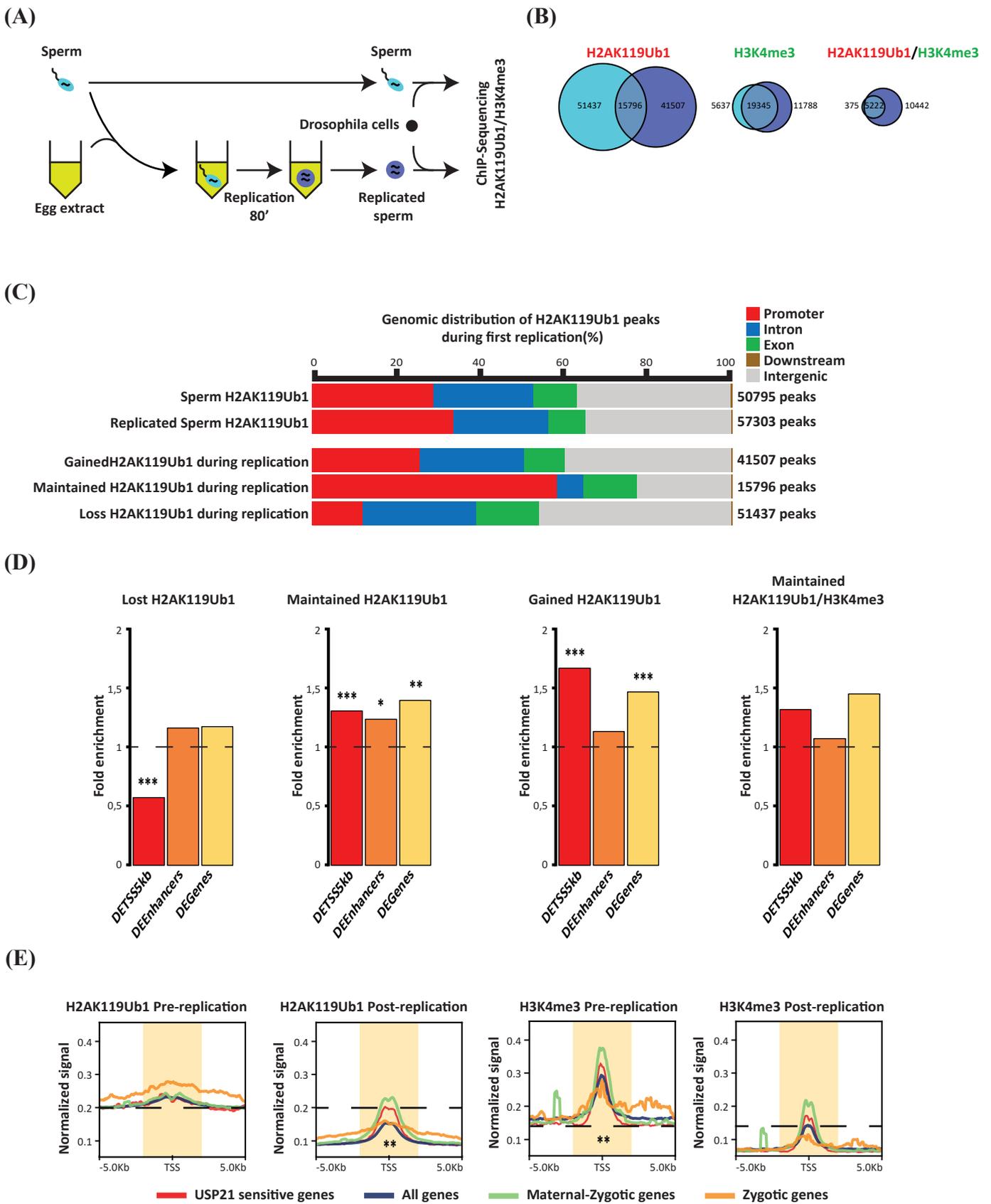


FIGURE 5

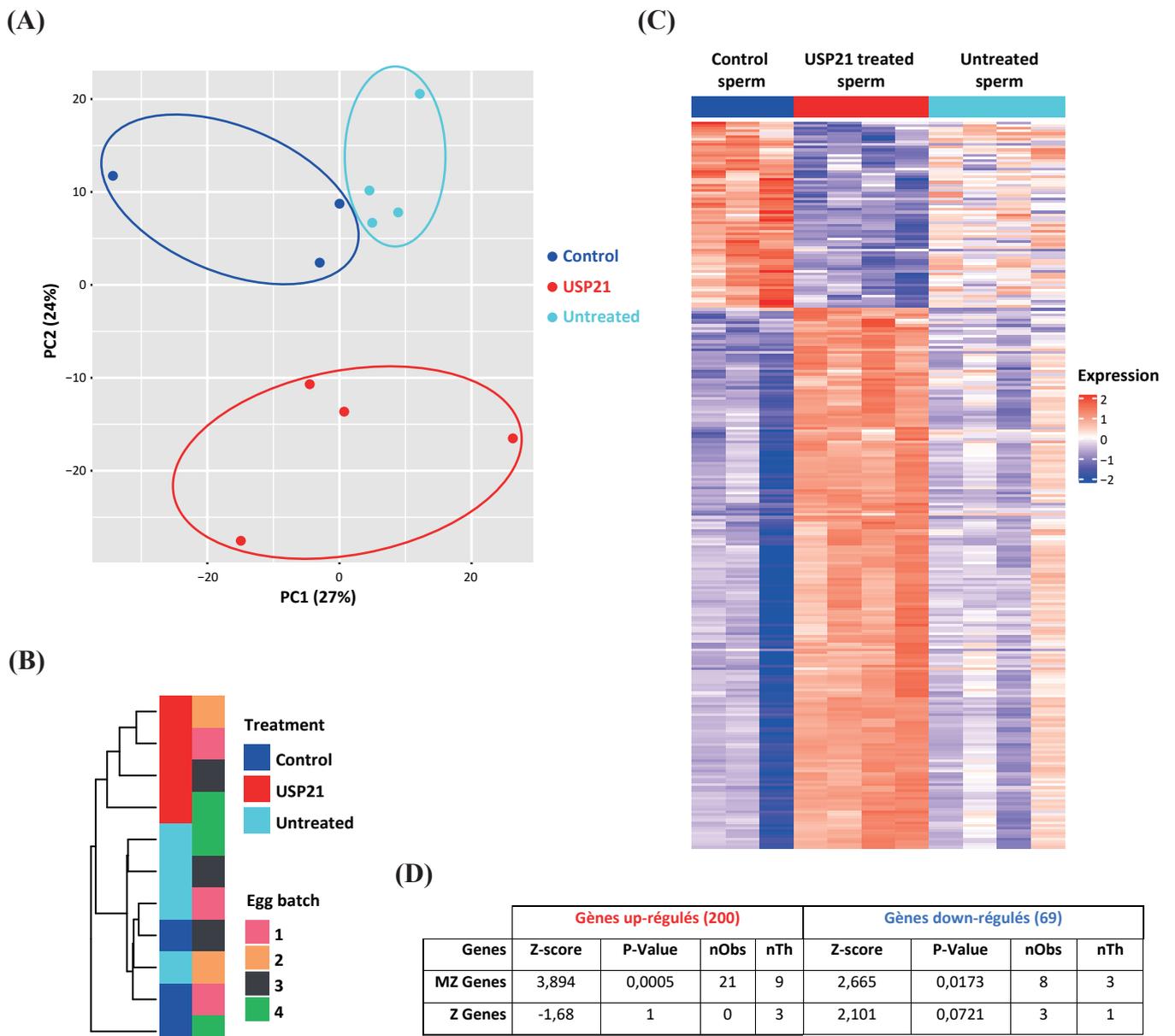


FIGURE S1

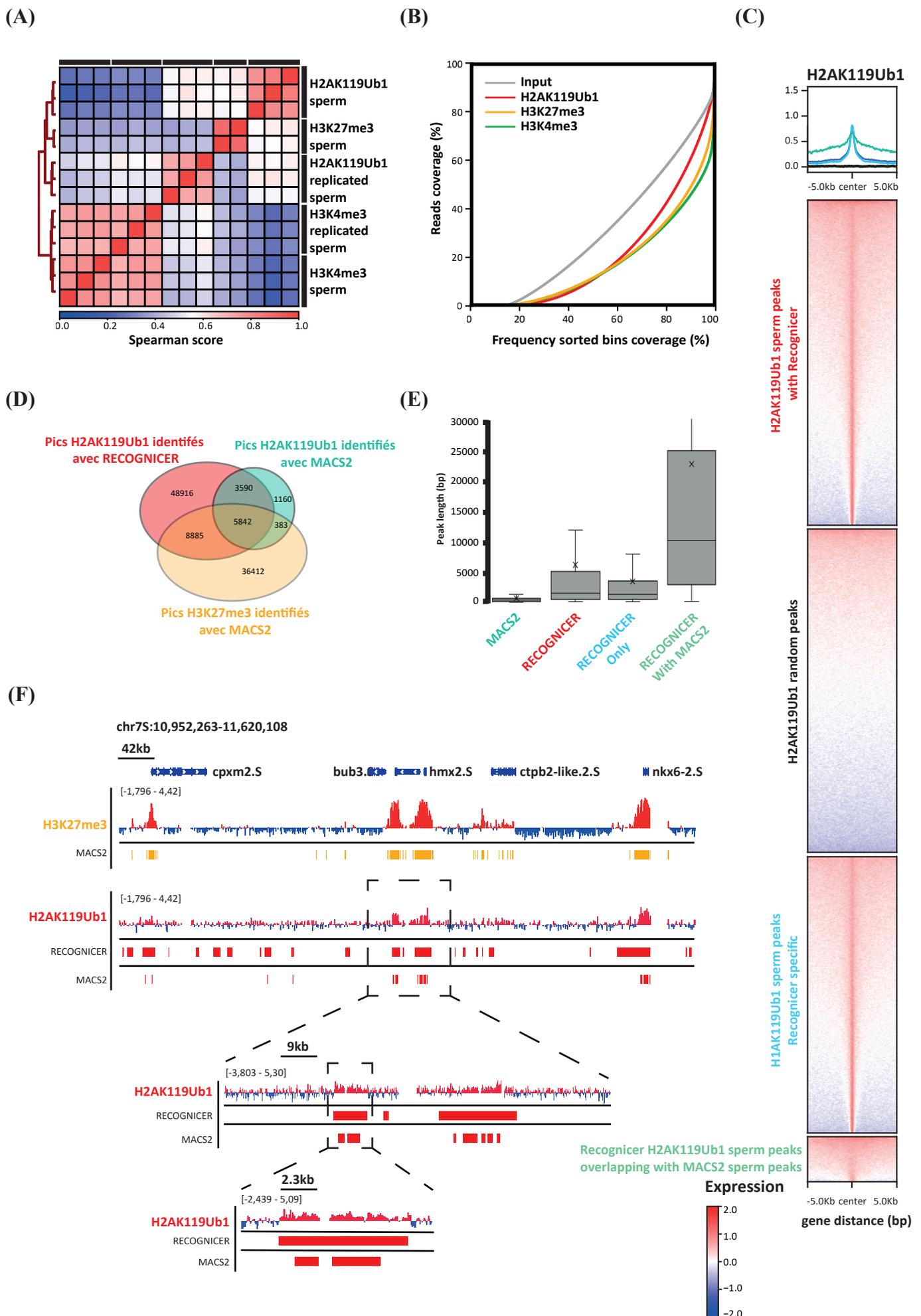
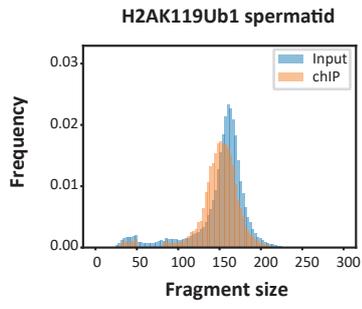
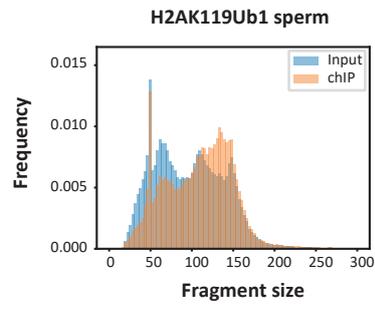


FIGURE S2

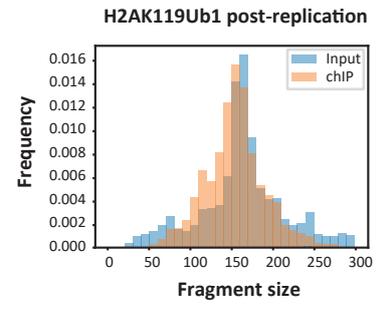
(A)



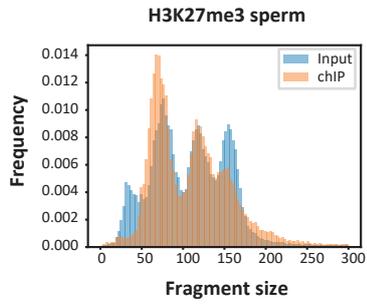
(B)



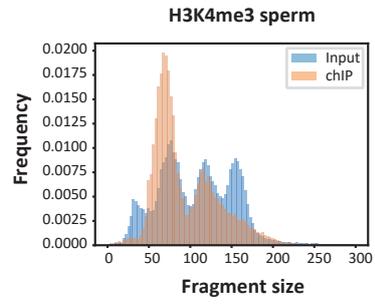
(C)



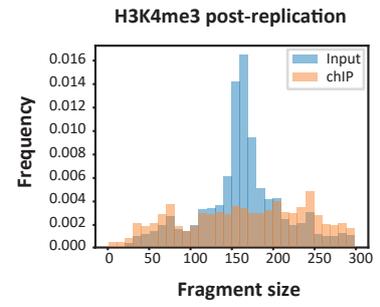
(D)



(E)



(F)





**Titre :** Contribution de H2AK119Ub1 dérivée du spermatozoïde au développement embryonnaire précoce du Xénope

**Mots clés :** Embryogénèse, Spermatogénèse, Analyses multi-omics, *Xenopus laevis*, Epigénétique, Histones, H2AK119Ub1

**Résumé :** La déposition de H2AK119Ub1 par le complexe répressif Polycomb joue un rôle clé dans la formation de l'hétérochromatine facultative dans les cellules somatiques. Ici, nous évaluons la contribution de la marque H2AK119Ub1 du spermatozoïde sur l'expression des gènes embryonnaires. Chez *Xenopus laevis*, nous avons constaté que H2AK119Ub1 est conservé pendant la spermiogénèse et au début du développement embryonnaire, soulignant son potentiel pour un rôle dans la transmission d'informations épigénétiques du spermatozoïde à l'embryon. Le spermatozoïde traité avec USP21, une deubiquitylase H2A, juste avant l'injection dans l'œuf, entraîne des anomalies du développement. Des analyses RNA-Seq dans les embryons correspondants ont démontré une dérégulation

de l'expression génique après traitement par USP21. Des analyses ChIP-Seq ont permis de mettre en évidence un enrichissement de la configuration H2AK119Ub1/H3K4me3 sur les régions régulatrices des gènes sensibles à l'USP21. Enfin, une étude ChIP-Seq quantitative complémentaire indique que lors de la réplication de la chromatine du spermatozoïde, H2AK119Ub1 est retenue préférentiellement sur les régions associées aux gènes sensibles à USP21, dans un contexte où la majeure partie de H2AK119Ub1 est diluée. Nous concluons que la transmission à l'embryon de H2AK119Ub1 dérivée du spermatozoïde est requise pour un bon déroulement du développement embryonnaire précoce et la répression de certains gènes embryonnaires.

---

**Title :** Contribution of sperm derived H2AK119Ub1 to early embryo development in *Xenopus laevis*

**Keywords :** Embryogenesis, Spermatogenesis, Multi-omics analyses, *Xenopus laevis*, Epigenetics, Histones, H2AK119Ub1

**Abstract :** Deposition of H2AK119Ub1 by the Polycomb repressive complex plays a key role in the formation of facultative heterochromatin in somatic cells. Here, we assess the contribution of sperm H2AK119Ub1 to embryonic gene expression. In *Xenopus laevis*, we found that H2AK119Ub1 is conserved during spermiogenesis and early embryonic development, highlighting its credential for a role in transmitting epigenetic information from sperm to embryo. Sperm treated with USP21, an H2A deubiquitylase, just prior to injection into the egg, results in developmental abnormalities. RNA-Seq analyzes in the corresponding embryos demonstrated dysregulation of gene expression after treatment with USP21.

ChIP-Seq analyzes revealed an enrichment of the H2AK119Ub1/H3K4me3 bivalent configuration on the regulatory regions of genes sensitive to USP21. Finally, a complementary quantitative ChIP-Seq study indicates that during sperm chromatin replication, H2AK119Ub1 is preferentially retained on regions associated with genes sensitive to USP21, in a context where most of H2AK119Ub1 is diluted. We conclude that the transmission to the embryo of sperm derived H2AK119Ub1 is required for proper early embryonic development and the repression of certain embryonic genes.