Formation et devenir de l'empreinte parentale chez la levure S. pombe

Célia Raimondi

▶ **To cite this version:**

Célia Raimondi. Formation et devenir de l'empreinte parentale chez la levure S. pombe. Génétique. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066187 . tel-01663600

HAL Id: tel-01663600
https://theses.hal.science/tel-01663600

Submitted on 14 Dec 2017

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.

Université Pierre et Marie Curie

Ecole doctorale : Complexité du Vivant

Laboratoire Dynamique du génome

Formation et devenir de l'empreinte parentale chez la levure *S. pombe*

Thèse de doctorat en Complexité du Vivant

Présentée par

Célia Raimondi

Dirigée par Benoit Arcangioli et Serge Gangloff

Présentée et soutenue publiquement le

Devant un jury composé de :

Dr Gilles Fischer	Président
Dr Javerzat Jean-Paul	Rapporteur
Dr Thon Geneviève	Rapporteur
Dr Wu Pei-Yun Jenny	Examinatrice
Dr Arcangioli Benoit	Directeur de thèse
Dr Gangloff Serge	Directeur de thèse

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Résumé

Des études moléculaires et génétiques ont montré qu'une empreinte épigénétique, située au niveau du locus sexuel (*mat1*) de la levure *Schizosaccharomyces pombe*, initie le changement de type sexuel. La réplication unidirectionnelle du locus *mat1* permet la formation de l'empreinte sur le brin Watson. Les éléments moléculaires qui forment et protègent l'empreinte durant le cycle cellulaire restent peu connus. Afin de mieux comprendre le mécanisme de formation et de maintien de l'empreinte, j'ai caractérisé le recrutement au niveau du locus *mat1* d'acteurs précoces dans le changement de type sexuel. J'ai montré que la protéine Sap1 (*switch activating protein 1*) est préférentiellement recrutée à l'intérieur de la séquence *SS13*, une séquence qui stabilise l'empreinte. Les protéines Lsd1/2 (*lysine specific demethylases*) qui contrôlent la pause de la fourche de réplication à *mat1* et la formation de l'empreinte sont spécialement recrutées au niveau de *mat1* indépendamment de l'allèle présent. La protéine Abp1 (homologue de CENP-B) est enrichie à côté de *mat1* mais n'est pas impliquée dans la formation/maintien de l'empreinte. De plus, j'ai établi la signature de l'empreinte par séquençage à haut débit. En utilisant cette signature, j'ai mis en évidence que les protéines Lsd1/2 et Sap1 immunoprécipitent les deux côtés de la chromatide qui porte l'empreinte ce qui suggère la formation d'une structure protective définie comme l'imprintosome.

Abstract

Genetic and molecular studies have indicated that an epigenetic imprint at *mat1*, the sexual locus of fission yeast, initiates mating type switching. The polar DNA replication of *mat1* generates an imprint on the Watson strand. The process by which the imprint is formed and maintained through the cell cycle remains unclear. To understand better the mechanism of imprint formation and stability, we characterized the recruitment of early players of mating type switch at the *mat1* region. We found that the switching activating protein 1 (Sap1) is preferentially recruited inside the *mat1M* allele on a sequence (*SS13*) that enhances the imprint. The lysine specific demethylases, Lsd1/2, that control the replication fork pause at *MPS1* and the formation of the imprint are specifically drafted inside of *mat1*, regardless of the allele. The CENP-B homolog, Abp1, is highly enriched next to *mat1* but it is not required in the process. Additionally, we established the computational signature of the imprint. Using this signature, we show that both sides of the imprinted molecule are bound by Lsd1/2 and Sap1, suggesting a nucleoprotein protective structure defined as imprintosome.

Table des matières

LISTE DES FIGURES ET DES TABLEAUX 9

LISTE DES ABREVIATIONS 11

INTRODUCTION 12

1. LA LEVURE *S. POMBE* 13

1.1 CYCLE DE VIE 13
1.1.1 CYCLE VEGETATIF 13
1.1.2 CYCLE SEXUEL 13
1.2 LE GENOME DE *S. POMBE* 14
1.4 L'EPIGENOME DE *S. POMBE* ET SON ORGANISATION TRIDIMENSIONNELLE (3D) 16
1.4.1 L'EPIGENOME DE *S. POMBE* : LES MECANISMES EPIGENETIQUES CONNUS CHEZ *S. POMBE* 16
1.4.2 L'ORGANISATION 3D DU NOYAU 17

2. LE MECANISME DE CHANGEMENT DE TYPE SEXUEL 19

2.1 CONCEPTS GENERAUX 19
2.2 LA REGION *MAT* CONTROLE LA DIVISION ASYMETRIQUE 20
2.3 LE MECANISME 23

3. L'INITIATION DU MECANISME : UNE ASYMETRIE REPLICATIVE 25

3.1 UNE EMPREINTE CHROMOSOMIQUE FORMEE ASYMETRIQUEMENT A *MAT1* 25
3.1.1 L'ASYMETRIE EST PORTEE PAR LE CHROMOSOME 25
3.1.2 LA POSITION ET LA NATURE DE L'EMPREINTE 27
3.2 LA REPLICATION ASYMETRIQUE DE *MAT1* ET SON CONTROLE 28
3.2.1 LA REPLICATION UNIDIRECTIONNELLE PERMET LA FORMATION DE L'EMPREINTE 28
3.2.2 LE COMPLEXE SWI1-SWI3 PROTEGE LA FOURCHE A *MPS1* 30
3.2.3 LES LYSINES DEMETHYLASES LSD1-LSD2 AGISSENT SUR LA FOURCHE A *MPS1* 31
3.2.4 LE CONTROLE DE RTS1 33
3.2.5 LE MODELE DE LA REPLICATION DU LOCUS SEXUEL AUJOURD'HUI 33
3.3 LES ARRETS DE FOURCHES ET LES ACTEURS DU CTS 34
3.3.1 LES PAUSES AU NIVEAU DES LTRS ET DE L'ADNR : LES PROTEINES SAP1 ET ABP1 35
3.3.2 LA COLLISION REPLICATION / TRANSCRIPTION : UN AUTRE EXEMPLE D'ARRET DE FOURCHE 36
3.4 QU'ADVIENT-IL D'UNE FOURCHE BLOQUEE ? 37

4. L'EMPREINTE ET SON POTENTIEL RECOMBINOGENE 38

4.1 LA FORMATION/MAINTENANCE DE L'EMPREINTE 38
4.1.1 LES ACTEURS FORMANT ET STABILISANT L'EMPREINTE 38
4.1.2 *MPS1* EST ACTIF AVEC ET SANS EMPREINTE 42
4.2 LE RESULTAT DE LA REPLICATION A TRAVERS L'EMPREINTE 42
4.3 LA REPARATION DE LA CDB 43
4.3.1 DANS UN CONTEXTE *H90* 43
4.3.2 EN ABSENCE DES DONNEURS 44
4.4 LA CHROMATINE DANS LA REGION *MAT* 45
4.4.1 L'HETEROCHROMATINE AU NIVEAU DES DONNEURS 45
4.5.2 LES SITES *IR-L* ET *IR-R* AINSI QUE *CENH* PARTICIPENT A LA FORMATION DE L'HETEROCHROMATINE 47
4.5 LE CHOIX DU DONNEUR 49
4.5.1 LA NATURE CHROMATINIENNE PARTICIPE A LA DIRECTIONNALITE 50

4.5.2 LE COMPLEXE SWI5-SWI2 CONTROLE LE CHOIX DU DONNEUR 50
4.5.3 SRE2 ET SRE3, DES ELEMENTS DE SEQUENCES CONTROLANT LA DIRECTIONNALITE 52

5. LE PROJET : LA FORMATION ET LE DEVENIR DE L'EMPREINTE CHEZ LA LEVURE *S. POMBE* 54

MATERIELS ET METHODES 55

1. COMMENT TESTER L'EFFICACITE DES DIFFERENTES ETAPES DU CTS 56

1.1 QUANTIFICATION DE L'EFFICACITE D'UNE PAUSE DE LA FOURCHE DE REPLICATION : LES GELS 2D 56
1.2 OBSERVATION DE L'EMPREINTE A *MAT1* PAR SOUTHERN-BLOT 60
1.3 UTILISATION DE LA QPCR POUR ETABLIR LE POURCENTAGE DE CELLULES M ET P DANS UNE POPULATION 61

2. DETERMINATION D'UN SITE DE RECRUTEMENT PROTEIQUE 62

2.1 *IN VITRO* : LA TECHNIQUE DE GEL RETARD 62
2.2 CHROMATINE IMMUNOPRECIPITATION CHEZ *S. POMBE* 64
2.3 LA CHIP-SEQUENÇAGE 66
2.3.1 PREPARATION DE L'ADN EN VUE DU SEQUENÇAGE 66
2.3.2 ANALYSE DES FICHIERS DE SEQUENCE OBTENUS 67
2.3.3 COMMENT OBTENIR DES PICS STATISTIQUEMENT VALIDES ? 69
2.3.4 COMPARAISON DES PICS ET VALIDATION STATISTIQUE 71

3. CARTOGRAPHIE DE L'EMPREINTE PAR SEQUENÇAGE A HAUT DEBIT 73

4. LES FONDS GENETIQUES UTILISEES 74

RESULTATS 76

1. CARTOGRAPHIE DES PROTEINES LSD1/LSD2, SAP1 ET ABP1 DANS LA REGION *MAT* 77

1.1 LES FONDS GENETIQUES ET LA TECHNIQUE UTILISEES 78
1.2 CARTOGRAPHIE DES PROTEINES LSD1/2, SAP1 ET D'ABP1 DANS LA REGION *MAT* 79
1.3 LSD1/2, SAP1 ET D'ABP1 SONT RECRUTEES A *MAT1* INDEPENDAMMENT DES DONNEURS 85
1.4 SAP1 EST RECRUTEE ENTRE 211 ET 111 PB DU COTE CENTROMERE PROXIMAL DU SITE DE L'EMPREINTE 87
1.5 CARACTERISATION DU SITE DE RECRUTEMENT DE LSD1 ET DE SAP1 PAR CHIP-SEQ 88
1.6 SANS EMPREINTE, LA FOURCHE EN PAUSE CASSE SPECIFIQUEMENT 92
1.7 CONCLUSIONS 96

2. ETUDE DU ROLE DE LA PROTEINE ABP1 SUR LE CHANGEMENT DE TYPE SEXUEL CHEZ *S. POMBE* 97

2.1 LE DOMAINE DE LIAISON A L'ADN EST NECESSAIRE AU MECANISME DE CHANGEMENT DE TYPE SEXUEL 98
2.2 COMMENT ABP1 EST-ELLE RECRUTEE DU COTE CENTROMERE DISTAL DE *MAT1* ? 99
2.2.1 QUANTIFICATION DU RECRUTEMENT D'ABP1 99
2.2.2 ETUDE DU RECRUTEMENT D'ABP1 EN FONCTION DE MUTATIONS EN TRANS ET EN CIS 100
2.2.3 L'ABSENCE DE SWI1 NE PERTURBE PAS GLOBALEMENT LE RECRUTEMENT D'ABP1 102
2.3 L'ACTIVITE DE *MPS1* ET LA FORMATION DE L'EMPREINTE SONT INDEPENDANTES D'ABP1 103
2.4 ABP1 EST CRUCIALE POUR L'EQUILIBRE ENTRE LES DEUX TYPES SEXUELS 106
2.4.1 LSD1 ET SAP1 NE RESTAURENT PAS L'EQUILIBRE M/P EN ABSENCE D'ABP1 106
2.4.2 SWI1 AGIT EN AMONT D'ABP1 108
2.4.3 ABP1 SE LIE A *SWI2* SEULEMENT DANS LES CELLULES M 109
2.5 ABP1 ET CBH1 NE SONT PAS REDONDANT AU NIVEAU DU CONTROLE DU CTS 110
2.6 L'INSERTION DU PROMOTEUR *NMT1* A DROITE DE *MAT1* DIMINUE LE RECRUTEMENT D'ABP1 111
2.7 CONCLUSIONS 114

3. ANALYSE DU RECRUTEMENT DE LSD1/2, SAP1 ET ABP1 SUR L'ENSEMBLE DU GENOME 115

3.1 ANALYSE DES SITES DE RECRUTEMENTS DE LSD1/LSD2, SAP1 ET ABP1 SUR L'ENSEMBLE DU GENOME 115
3.2 LSD1/LSD2, SAP1 ET ABP1 PARTAGENT DES SITES AVEC RAD52 120
3.3 CES PROTEINES SONT RECRUTEES AU NIVEAU DE GENES FORTEMENT TRANSCRITS 123
3.4 CONCLUSIONS 125

4. SIGNATURE DE L'EMPREINTE PAR SEQUENÇAGE HAUT DEBIT 126

4.1 L'UTILISATION DU SEQUENÇAGE HAUT DEBIT POUR L'ANALYSE D'UN SITE FRAGILE 126
4.2 OBSERVATION DU COMMENCEMENT DES LECTURES SUR TOUT LE GENOME 127
4.3 ANALYSE DE L'EMPREINTE A *MAT1* 135
4.5 LE 5' COMPTE A L'EMPREINTE EST ENRICHI DANS LES IP 139
4.5.1 ABP1 IMMUNOPRECIPITE LA MOLECULE QUI PORTE L'EMPREINTE 139
4.5.2 LSD1, LSD2 ET SAP1 LIENT LA CHROMATIDE QUI PORTE L'EMPREINTE 143
4.4 CONCLUSIONS 149

5. CONTROLE DE LA REPLICATION DE *WTF9* 150

5.1 CARACTERISATION DE LA DUPLICATION DU BRAS GAUCHE DU CHROMOSOME 3 150
5.2 LSD1 CONTROLE-T-ELLE LA REPLICATION DE *WTF9* ? 151
5.3 CONCLUSIONS 153

6. COMPUTATIONAL SIGNATURE OF THE IMPRINTOSOME COMPLEX AT THE *MAT1* LOCUS IN FISSION YEAST 154

ABSTRACT 155
INTRODUCTION 155
RESULT 158
CHIP-SEQUENCING MAPPING OF ABP1, LSD1/2 AND SAP1 TO THE H^{90} MATING TYPE REGION 158
ABP1 IS RECRUITED TO *SWI2* WHEN *M* IS PRESENT AT *MAT1* 158
LSD1/2, SAP1 AND ABP1 MAP AT *MAT1* 159
CHARACTERIZATION OF THE SAP1 BINDING IN THE *M* ALLELE 159
DETECTION OF AN IMPRINTED SITE BY HIGH THROUGHPUT SEQUENCING 160
IMMUNOPRECIPITATION OF THE IMPRINTED CHROMATID 161
DISCUSSION 161
MATERIALS AND METHODS 163
FISSION YEAST STRAINS AND GENETIC PROCEDURES: 163
2D GEL ANALYSIS: 163
GEL SHIFT ASSAY: 164
CHROMATIN IMMUNOPRECIPITATION (CHIP): 164
ILLUMINA SEQUENCING OF DNA: 164
ILLUMINA READ PROCESSING AND ALIGNMENT OF DNA LIBRARIES: 164
CHIP-SEQUENCING ANALYSIS: 164
5' COUNT ANALYSIS: 165
ACKNOWLEDGMENTS 165
REFERENCES 165
FIGURE AND TABLE LEGENDS 168
FIGURES AND TABLES 171
SUPPLEMENTARY TABLES MATERIALS AND METHODS 182
SUPPLEMENTARY MATERIALS AND METHODS 184
PROCESSING OF SEQUENCING DATA 184
STATISTICAL VALIDATION OF CHIP PEAKS 185
5' COUNT ANALYSIS AND STATISTICAL VALIDATION 186

CONCLUSIONS ET DISCUSSION **187**

UN REPLIEMENT CHROMOSOMIQUE AUTOUR DU SITE DE L'EMPREINTE ? **188**
LA DIRECTIONNALITE ET L'HETEROCHROMATINE A *MAT2P* ET A *MAT3M* **193**
LA COLOCALISATION DE LSD1-LSD2, SAP1 ET ABP1 AVEC RAD52 **195**

PERSPECTIVES **197**

BIBLIOGRAPHIE **200**

Liste des figures et des tableaux

Introduction

FIGURE 1 : LES CYCLES DE VIES DE *S. POMBE*. 14
FIGURE 2 : REPRESENTATION SCHEMATIQUE DES CHROMOSOMES DE *S. POMBE*. 15
FIGURE 3 : ANALYSE DES REGIONS BIOLOGIQUES DU GENOME DE *S. POMBE.* 16
FIGURE 4 : MODELE DU CHANGEMENT DE TYPE SEXUEL. 20
FIGURE 5 : ORGANISATION DE LA REGION *MAT.* 22
FIGURE 6 : LE MECANISME DE DIVISION ASYMETRIQUE CHEZ *S. POMBE.* 24
FIGURE 7 : LE FRAGMENT DE DIGESTION CONTENANT *MAT1* EST CASSE DANS 50% DES CAS. 26
FIGURE 8 : MODELE PROPOSE ET VALIDE PAR KLAR. L'EMPREINTE EST PORTEE PAR UN BRIN D'ADN. 27
FIGURE 9 : POSITION DE L'EMPREINTE EN FONCTION DE L'ALLELE A *MAT1.* 28
FIGURE 10 : LA REPLICATION UNIDIRECTIONNELLE DU LOCUS *MAT1.* 30
FIGURE 11 : LES ELEMENTS CONTROLANT L'ACTIVITE DE *MPS1.* 31
FIGURE 12 : LES DOMAINES DES PROTEINES LSD1 ET LSD2 33
FIGURE 13 : LE MODELE ET LES ACTEURS DE LA PREMIERE REPLICATION DU LOCUS SEXUEL 34
FIGURE 14 : LES DOMAINES DES PROTEINES ABP1 ET SAP1 36
FIGURE 15 : LES ELEMENTS CONTROLANT LA FORMATION/STABILISATION DE L'EMPREINTE. 41
FIGURE 16 : L'EMPREINTE EST CONVERTIE EN CASSURE DOUBLE BRIN POLAIRE PAR LA FOURCHE DE REPLICATION. 43
FIGURE 17 : L'ASSEMBLAGE DE L'HETEROCHROMATINE CHEZ *S. POMBE.* 47
FIGURE 18 : LA NUCLEATION ET LES BARRIERES DE LA CHROMATINE DANS LA REGION *MAT.* 49
FIGURE 19 : MODELE DE LA REPARTITION DU COMPLEXE SWI2-SWI5 EN FONCTION DU TYPE SEXUEL. 52
FIGURE 20 : LES ELEMENTS *SRE2* ET *SRE3.* 53

Matériels et Méthodes

FIGURE 1 : PHOTOGRAPHIE DU GEL APRES MIGRATION DES FRACTIONS. 58
FIGURE 2 : RESULTAT SCHEMATIQUE D'UN GEL 2-D. 59
FIGURE 3 : L'EMPREINTE CREE DEUX SOUS FRAGMENTS. 60
FIGURE 4 : SCHEMA DE LA POSITION DES OLIGONUCLEOTIDES UTILISES EN QPCR. 62
FIGURE 5 : SCHEMA DE LA POSITION DES OLIGONUCLEOTIDES UTILISES POUR LE GEL SHIFT. 63
TABLEAU 1 : LISTE DES OLIGONUCLEOTIDES UTILISES POUR L'EXPERIENCE DE GEL RETARD. 64
FIGURE 6 : PHOTOGRAPHIE DU GEL DE VERIFICATION DE LA PURIFICATION DU DOMAINE HMG MARQUE AU BLEU DE COOMASSIE. 64
TABLEAU 2 : CONDITIONS TECHNIQUES DE LA CHIP EN FONCTION DE CHAQUE PROTEINE. 65
TABLEAU 3 : LISTE DES OLIGONUCLEOTIDES UTILISES POUR LES EXPERIENCES DE CHIP-QPCR. 66
FIGURE 7 : REPRESENTATION SCHEMATIQUE DES SEQUENCES UTILISEES POUR L'ANALYSE. 68
TABLEAU 4 : STATISTIQUES DES BANQUES DE SEQUENÇAGE. 69
FIGURE 8 : RESULTAT DE L'ANALYSE DE LA COHERENCE STATISTIQUE. 70
FIGURE 9 : DISTRIBUTION DES VALEURS DES INTERSECTIONS OBSERVEES ET ALEATOIRES. 73
TABLEAU 5 : VALEURS OBSERVEES ET THEORIQUES DES DISTRIBUTIONS DES 5' COMPTES DU BRIN WATSON. 74
TABLEAU 6 : LISTE DES SOUCHES UTILISEES DANS CETTE ETUDE. 75

Résultats

FIGURE 1 : SCHEMA DE LA REGION *MAT* DANS LES SOUCHES UTILISEES. 79
FIGURE 2 : ANALYSE DE LA REGION *MAT.* 81
FIGURE 3 : LSD1, LSD2 ET ABP1 NE SONT RECRUTEES QU'A *MAT1*. 84
FIGURE 4 : LSD1, LSD2, SAP1 ET ABP1 SONT RECRUTEES A *MAT1* INDEPENDAMMENT DES DONNEURS. 86
FIGURE 5 : IDENTIFICATION DE LA SEQUENCE LIEE PAR SAP1 AU NIVEAU DE L'ALLELE *M*. 88

FIGURE 6 : SCHEMA DE L'ORGANISATION DE LA REGION *MAT* DES SOUCHES UTILISEES. 90
FIGURE 7 : LSD1 ET SAP1 SONT RECRUTEES DANS *MAT1*. 91
FIGURE 8 : LA FOURCHE CASSE DURANT LA PAUSE. 95
FIGURE 9 : SCHEMA DU MODELE DE LA REPLICATION DE *MAT1* DANS UNE SOUCHE PORTANT LA DELETION *SMT-0*. 96
FIGURE 10 : LE DOMAINE DE LIAISON A L'ADN EST NECESSAIRE AU CTS. 98
FIGURE 11 : ABP1 EST RECRUTEE A *MAT1* INDEPENDAMMENT DE L'ALLELE PRESENT. 100
FIGURE 12 : LE SITE DE LIAISON D'ABP1 EST DANS LA DELETION *SMT–0*. 101
FIGURE 13 : LE NIVEAU D'EXPRESSION D'ABP1 EST CONSTANT DANS LES DIFFERENTES SOUCHES. 102
FIGURE 14 : COMPARAISON DE L'ENRICHISSEMENT D'ABP1 DANS UN CONTEXTE SAUVAGE OU EN ABSENCE DE SWI1. 103
FIGURE 15 : MESURE DE L'ACTIVITE DE *MPS1* ET DU NIVEAU DE L'EMPREINTE. 105
FIGURE 16 : LA PROTEINE ABP1 EST NECESSAIRE POUR LE MAINTIEN DE L'EQUILIBRE ENTRE LES CELLULES M ET LES CELLULES P D'UNE POPULATION. 107
FIGURE 17 : LE MUTANT *SAP1-C* FAVORISE LA CROISSANCE DU MUTANT *ABP1Δ*. 108
FIGURE 18 : SWI1 AGIT EN AMONT D'ABP1. 109
FIGURE 19 : ABP1 EST RECRUTEE A *SWI2* EN FONCTION DE L'ALLELE *M*. 110
FIGURE 20 : LA PROTEINE CBH1 NE JOUE PAS LE MEME ROLE QU'ABP1 AU NIVEAU DU CTS. 111
FIGURE 21 : LE SYSTEME *NMT1* NE PERMET PAS D'ETUDIER LE RECRUTEMENT D'ABP1. 113
TABLEAU 1 : ANALYSE DU NOMBRE DE PICS. 116
TABLEAU 2 : COMPARAISON AVEC LES DONNEES PUBLIEES. 116
FIGURE 22 : ANALYSE DE LA TAILLE DES PICS. 117
FIGURE 23 : ANALYSE DE LA REPARTITION DES PICS SUR L'ENSEMBLE DU GENOME. 118
FIGURE 24 : ANALYSE DES REGIONS ENRICHIES. 120
FIGURE 25 : ANALYSE DES INTERSECTIONS DES REGIONS ENRICHIES. 122
TABLEAU 3 : COMPARAISON DES INTERSECTIONS AVEC LES DONNEES PUBLIEES 123
FIGURE 26 : ANALYSE DES PICS EN FONCTION DE LA TRANSCRIPTION. 124
FIGURE 27 : SCHEMA DE L'ANALYSE DES 5' COMPTES. 127
FIGURE 28 : LES 5' WATSON ET CRICK SONT CONSERVES AU COURS DE LA PREPARATION DES BANQUES. 129
FIGURE 29 : DISTRIBUTION DU NOMBRE DE 5' COMPTES PAR POSITIONS GENOMIQUES. 130
FIGURE 30 : DISTRIBUTION DES 5' COMPTES SUR L'ENSEMBLE DU GENOME. 132
TABLEAU 4 : SITES QUI PRESENTENT UN ENRICHISSEMENT EN 5' COMPTE. 133
FIGURE 31 : DISTRIBUTION DES 5' COMPTES SUR L'ENSEMBLE DU GENOME DANS QUATRE SOUCHES *MAT1M Δ2-3*. 134
FIGURE 32 : ANALYSE DES 5' COMPTES DANS LA REGION *MAT*. 136
FIGURE 33 : ANALYSE DES 5' COMPTES DANS LES 10 KBP AUTOUR DE *MAT1*. 137
FIGURE 34 : ANALYSE DES 5' COMPTES AU NIVEAU DES 40 PB AUTOUR DE L'EMPREINTE. 138
FIGURE 35 : DISTRIBUTIONS THEORIQUES DES 5' COMPTES DE L'IP ABP1 EN FONCTION DE LA PRESENCE DE L'EMPREINTE. 140
FIGURE 36 : ANALYSE DES 5' COMPTES AU NIVEAU DE L'EMPREINTE DANS L'IP D'ABP1. 142
FIGURE 37 : DISTRIBUTIONS THEORIQUES DES 5' COMPTES DES IP LSD1-2 ET SAP1 EN FONCTION DE LA PRESENCE DE L'EMPREINTE. 144
FIGURE 38 : ANALYSE DES 5' COMPTES AU NIVEAU DE L'EMPREINTE DANS LES IP DE LSD1/2 ET SAP1. 146
FIGURE 39 : ANALYSE DES 5' COMPTES AU NIVEAU DE L'EMPREINTE DANS LES IP DE LSD1, LSD2 ET SAP1. 148
FIGURE 40 : SCHEMA DE LA REGION DUPLIQUEE. 151
FIGURE 41 : LSD1 NE CONTROLE PAS LA REPLICATION AU NIVEAU DE *WTF9*. 152
FIGURE 42 : LSD1 NE CONTROLE PAS DIRECTEMENT LA REPLICATION AU NIVEAU DE *WTF9*. 153

Conclusions et Discussion

FIGURE 1 : MODELE DU REPLIEMENT AUTOUR DE L'EMPREINTE. 193

Liste des abréviations

(Abréviations utilisées plus de 5 fois dans le manuscrit)

S. pombe	*Schizosaccharomyces pombe*
5' compte	Nombre de début de lecture d'ADN
Abp1	ARS binding protein 1
ADN	Acide DésoxyriboNucléique
ADN db	Acide DésoxyriboNucléique double brin
ADN sb	Acide DésoxyriboNucléique simple brin
ADNr	Acide DésoxyriboNucléique ribosomal
ARN	Acide RiboNucléique
ARNi	ARN interférent
ARNt	ARN de transfert
CDB	Cassure Double Brin
CEN	Centromère
Cf	Concentration finale
ChIP	Chromatine Immunoprécipitation
ChIP-qPCR	ChIP couplée à la PCR quantitative
ChIP-séq	ChIP couplée au séquençage à haut débit
Ci	Concentration initiale
CTS	Changement de Type Sexuel
H	Histone
IP	ImmunoPrécipitation
IR	Intermédiaire de Réplication
K	Lysine
Kb	Kilo base
Lsd1	Lysine specific demethylase 1
Lsd2	Lysine specific demethylase 2
LTR	Long Terminal Repeat
Mb	Méga base
MPS1	Mat1 Pause Site 1
pb	paire de base
Pol II, Pol III	ARN polymérase II et III
Pol α, Pol δ et Pol ε	ADN polymérase α, δ et ε
qPCR	Réaction de polymérisation en chaine quantitative
RH	Recombinaison Homologue
RpKM	Lectures (Read) par Kilobase par million
RPM	Reads Per Million
RTS1	Replication Terminator Site 1
Sap1	Switch-activating protein 1
SPB	Sindle Pole Body
Tel	Télomère
WCE	Whole Cell Extract

Introduction

1. La levure *S. pombe*

La levure *Schizosaccharomyces pombe* (*S. pombe*) a été découverte par Paul Linder en 1893 dans un millet de bière provenant d'Afrique de l'Est. Son nom rend compte du fait qu'elle se divise par fission "*Schizo*" et qu'elle ne donne pas de mycélium comme les levures appartenant au genre taxonomique "*saccharomyces*". L'addition du mot "*pombe*" vient de l'endroit où elle a été isolée, puisque "*pombe*" signifie bière en Swahili, un groupe de langue d'Afrique orientale. La levure aujourd'hui utilisée en laboratoire dérive d'un isolat de Urs Leupold entre 1946-1947. Cet isolat de *S. pombe* a été utilisé comme organisme modèle dans différents domaines. Il fut utilisé tout d'abord par Urs Leupold dans le domaine de la génétique et par Murdock Mitchison dans l'étude du cycle cellulaire. Aujourd'hui c'est un modèle d'étude pour la biologie moléculaire et cellulaire eucaryote.

1.1 Cycle de vie

1.1.1 Cycle végétatif

Schizosaccharomyces pombe est de forme cylindrique de 4 µm de diamètre et croit par extension des extrémités jusqu'à une longueur de 14 µm. Une fois cette taille atteinte, la cellule arrête de grandir et entre en division cellulaire. *S. pombe* a un cycle de vie haplophasique. En milieu riche, cette levure se divise par fission et donne deux cellules filles identiques phénotypiquement. Le cycle de vie mitotique de cette levure est comparable aux autres eucaryotes. Il est séparé en quatre phases consécutives : la phase S durant laquelle la réplication de l'ADN a lieu ; la phase G2, durant laquelle chaque chromosome contient deux chromatides ; la phase M qui permet la ségrégation des chromosomes et la séparation des deux cellules, et la phase G1 au cours de laquelle la cellule est haploïde et prépare la prochaine réplication. Le mécanisme de division cellulaire est finement régulé par de nombreux "*checkpoints*" ou points de contrôle permettant qu'il soit hautement fidèle. Comme illustré par la figure 1, la phase G2 occupe 70 % du cycle cellulaire alors que les trois autres phases se répartissent les 30 % du cycle restant.

1.1.2 Cycle sexuel

En milieu carencé en azote, la cellule n'ayant pas de partenaire de conjugaison diminue son métabolisme s'arrête en phase G1 et entre en phase de quiescence nommée G0. Par ailleurs, les cellules en milieu carencé en azote expriment les fonctions sexuelles et deux cellules de type sexuel opposé peuvent conjuguer et former un diploïde transitoire. Ce diploïde sporule et forme

un asque à quatre spores. Les asques persistent jusqu'à ce que l'environnement devienne favorable et permette la germination des spores qui entrent alors dans un nouveau cycle végétatif (figure 1, cycle méiotique).

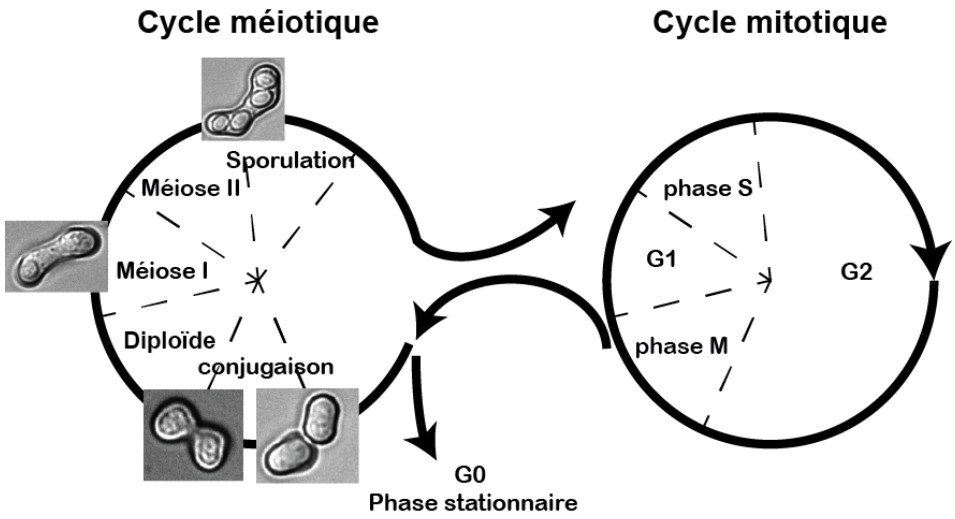


Figure 1 : Les cycles de vie de *S. pombe*.
À gauche, le cycle de vie en milieu pauvre. Deux cellules de type sexuel opposé conjuguent et entrent en méiose pour former un asque à quatre spores. À droite, le cycle de vie en milieu riche. La cellule se divise par fission pour donner deux cellules filles.

La souche sauvage a la capacité de changer de type sexuel au cours des divisions mitotiques. Elle donne donc des colonies composées d'une population hétérogène de cellules *P* (plus) et *M* (minus). Cette souche est homothallique et autofertile. Les levures homothalliques sont aussi appelées h^{90}, car les colonies qu'elles forment en milieu carencé sont composées à 90 % de spores. Les levures ne sont dotées d'aucun élément pour se déplacer. Ainsi, la capacité de changer leur type sexuel leur permet de pouvoir se croiser avec leur voisin(e) en milieu pauvre en azote et de déclencher la méiose pour former des spores résistantes. Le changement de type sexuel (CTS) fait intervenir une étape de recombinaison qui réarrange rarement la région contenant les loci sexuels (10^{-3} à 10^{-4} des cellules). À la suite de ces réarrangements, la cellule perd sa capacité de changer de type sexuel et devient hétérothallique *M* ou *P*. Ces souches hétérothalliques ne sont donc fertiles que par croisement.

1.2 Le génome de *S. pombe*

La séquence du génome de *S. pombe* a été publiée par Wood et al en 2002 (Wood et al., 2002). Il est composé de 12,7 Mpb distribuées sur 3 chromosomes (figure 2). Les trois chromosomes contiennent chacun un centromère noté *CENI*, *CENII* et *CENIII* qui ont des tailles variant de

35 kpb à 110 kpb. À leurs extrémités, chaque chromosome, portent des télomères d'une taille de 300 pb. Les télomères sont composés de la séquence répétée dégénérée ($G_{2\text{-}6}$TTACA) (Szostak and Blackburn, 1982). La région contrôlant le CTS que j'appellerai la région *mat*, est présente sur le bras droit du chromosome 2 et s'étend dans une souche sauvage sur approximativement 35 kpb (Beach et al., 1982; Egel and Gutz, 1981). Le chromosome 3 contient deux répétitions en tandem de l'ADN ribosomal (ADNr) qui flanquent les télomères (Schaak et al., 1982). Ces régions répétées ne peuvent être assemblées correctement sur le génome, et ne sont placées sur la figure 2 que pour indication. Ainsi, le génome de référence utilisé pour les analyses à haut débit ne contient pas les télomères ni la totalité des répétitions de l'ADNr.

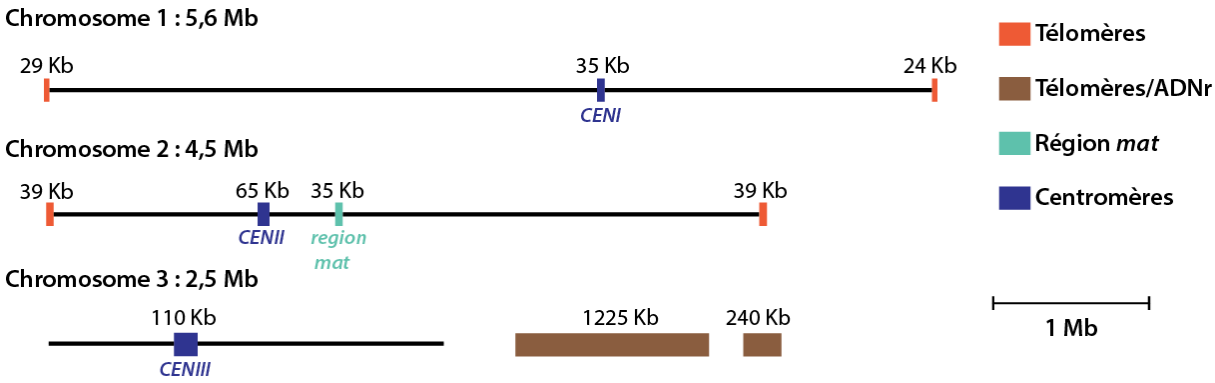
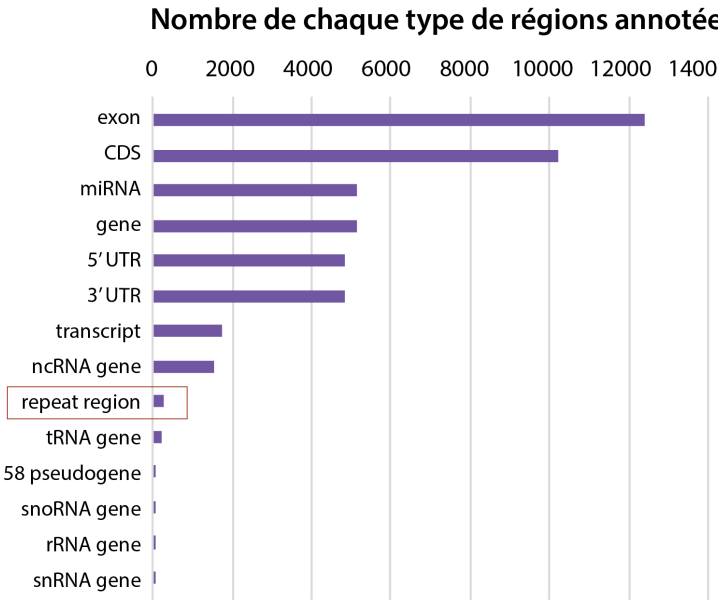


Figure 2 : Représentation schématique des chromosomes de *S. pombe*.
Les trois chromosomes sont représentés en noir. Les centromères sont annotés en bleu foncé. La région *mat* est représentée en bleu clair. Les télomères des chromosomes 1 et 2 sont représentés en orange. Pour le chromosome 3, les régions télomériques entrecoupées des répétitions de l'ADNr sont représentées en marron.

Le génome de cette levure contient 5118 gènes codants pour des protéines (figure 3). Ils sont répartis en moyenne tous les 2,470 pb sur les chromosomes 1 et 2 et tous les 2,790 pb sur le chromosome 3. La concentration en gènes diminue au niveau des télomères et des centromères. En plus de ces gènes, plusieurs autres types de loci codant pour des ARN non traduits sont présents. Par exemple, le génome de *S. pombe* contient 171 loci codant pour des ARN de transfert (ARNt). L'organisation de ces ARNt est particulière puisqu'ils sont distribués en majorité de chaque côté des centromères. Les éléments transposables, de type Tf2 et solo LTRs (appelés aussi LTRs), représentent 1,1 % du génome. Il y a 13 éléments transposables intacts de type Tf2 dans le génome de la souche séquencée (figure 3). À la différence d'autres isolats, cette souche séquencée ne contient pas de rétrotransposons de type Tf1 (Levin et al., 1990).

D'autre part, de nombreux LTRs sont également retrouvés seuls (solo LTRs) reflétant d'anciennes transpositions. Parmi les solos LTRs, 35 sont issus des éléments Tf2 alors que 135 viennent d'autres éléments Tf. Le chromosome 3 a la particularité de contenir deux fois plus de marques d'anciennes insertions. De plus, une autre classe d'éléments répétés spécifique à *S. pombe* a été découverte lors du séquençage de ce génome : ces éléments répétés sont appelés *wtf* (*With TF*). Des 25 *wtf* annotés, 23 sont retrouvés au niveau du chromosome 3 (Bowen et al., 2003; Wood et al., 2002). Cette distribution suggère soit que ce chromosome provient d'une autre population de *S. pombe,* soit que ces éléments s'étendent spécifiquement sur ce chromosome (Bowen et al., 2003).



Régions répétées	nombre
Tf-type transposons	13
Long terminal repeats (LTRs)	249
wtf	25
Centromères	3
Télomères	6
Répétitions ADNr	NA

Figure 3 : Analyse des régions biologiques du génome de *S. pombe.*
À gauche, un graphique représentant le nombre d'éléments génomiques répartis en classe déterminé à partir du fichier d'annotation (gff). À droite, le nombre et le type de régions répétées sont indiqués (Wood et al., 2002).

1.4 L'épigénome de *S. pombe* et son organisation tridimensionnelle (3D)

1.4.1 L'épigénome de *S. pombe* : Les mécanismes épigénétiques connus chez *S. pombe*

La molécule d'ADN est enroulée autour d'un complexe protéique composé de huit histones. Cette structure (ADN / Histones) est appelée le nucléosome, et forme l'unité fondamentale de la chromatine. Les nucléosomes sont composés d'environ 146 pb d'ADN enroulées autour d'un octamère d'histone qui est composé de deux copies de chaque histone (H2A, H2B, H3 et H4). Cette structure permet d'une part, la compaction de la molécule d'ADN, et d'autre part de réguler de nombreux mécanismes cellulaires. En effet, au-delà de la séquence d'ADN, la chromatine

influence l'activité des régions d'ADN. Ce contrôle passe notamment par la modification post-traductionnelle des histones. Le fait que la séquence d'ADN soit identique dans toutes les cellules d'un organisme, mais que le phénotype de ces dernières varie a conduit à l'introduction du terme "épigénétique". De nombreuses définitions de ce terme existent et Adrian Bird en 2007 rassembla différentes notions pour le définir : " *The following could be a unifying definition of epigenetic events: the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states* " (Bird, 2007). Il est aujourd'hui admis que l'épigénétique participe, au même titre que la séquence d'ADN, à la régulation des phénomènes cellulaires comme la transcription ou la réplication. Deux structures épigénétiques majeures ont été définies : l'hétérochromatine et l'euchromatine. Les régions dites hétérochromatiques sont défavorables à la transcription et à la recombinaison. Les régions euchromatiques sont plus vastes et sont favorables à la transcription. Cette définition des deux régions majeures est cependant très schématique, puisque localement nous observons des différences notables en termes de transcription ou de recombinaison. La compréhension de l'épigénome passe par la mise en évidence des mécanismes qui permettent l'établissement et le maintien de l'environnement répressif créé par l'hétérochromatine. La levure *S. pombe* est utilisée comme organisme modèle pour l'étude de la régulation épigénétique, car elle partage avec d'autres eucaryotes tel que l'humain, de nombreux mécanismes utilisés pour créer l'hétérochromatine.

Chez *S. pombe*, les régions hétérochromatiques sont situées principalement au niveau des télomères, des centromères, de l'ADNr et de la région *mat* (Allshire et al., 1994; Cooper et al., 1997; Nimmo et al., 1994; Thon and Klar, 1992). Tout comme chez la drosophile, la découverte du concept de PEV (*Position Effect Variegation*) qui consiste à insérer un gène rapporteur à différents endroits dans le génome et d'analyser son niveau d'expression, a permis de définir des régions hétérochromatiniennes. Grâce à cette technique utilisée dans différents contextes génétiques, la communauté scientifique a mis en évidence des acteurs responsables du contrôle du paysage épigénétique. La formation de l'hétérochromatine chez *S. pombe* nécessite la machinerie des ARN interférents (ARNi) ainsi que des enzymes modifiant chimiquement la queue des histones (Cam et al., 2005; Noma et al., 2001; Thon et al., 2002; Thon and Klar, 1992; Thon and Verhein-Hansen, 2000). À la différence d'autres eucaryotes, la méthylation de l'ADN dans des cellules de *S. pombe* en croissance n'est pas détectable (Wilkinson et al., 1995). Pour revue, se référer à (Allshire and Ekwall, 2015).

1.4.2 L'organisation 3D du noyau

L'organisation 3D de la chromatine est un domaine d'étude en pleine expansion. Il a permis de

mettre en évidence la relation entre la structure et la fonction. Le concept étant que des régions ayant le même statut épigénétique se retrouvent proches l'une de l'autre dans l'espace du noyau. Le regroupement de ces régions permettrait une structuration de la chromatine de plus haut niveau. Nous pouvons imaginer que cette organisation participe à maintenir les statuts épigénétiques. Plusieurs techniques ont été utilisées pour étudier l'association des différentes régions génomiques. L'utilisation de la microscopie et du système Cre/loxP ont permis de définir de façon globale l'organisation du noyau (Funabiki et al., 1993; Molnar and Kleckner, 2008; Noma et al., 2006). Dans le noyau d'une cellule en interphase, un pôle proche de l'enveloppe nucléaire est défini par le centre d'organisation des microtubules, le *spindle-pole body* (SPB). Chez *S. pombe,* les centromères sont regroupés près du SPB ainsi que les gènes transcrits par la ARN polymérase III (Pol III), tels que les loci codant pour les ARNt et l'ARN 5s (Funabiki et al., 1993; Iwasaki et al., 2010). La région *mat* est également localisée à proximité du SPB durant l'interphase (Alfredsson-Timmins et al., 2007). Les télomères des chromosomes 1 et 2 se retrouvent ensemble à la périphérie du noyau, mais ne sont pas associés avec le SPB (Funabiki et al., 1993). Les répétitions de l'ADNr fusionnent de façon opposée au SPB et forment le nucléole au niveau duquel elles sont transcrites et les ribosomes préassemblés (Umesono et al., 1983). Enfin, le statut transcriptionnel des gènes est corrélé avec leur association. Par exemple, les gènes hautement exprimés sont retrouvés associés tout comme les gènes co-régulés (Tanizawa et al., 2010). D'autre part, les régions répétées de types Tf2 et LTRs sont retrouvées préférentiellement regroupées dans le noyau (Cam et al., 2008; Tanizawa et al., 2010). Ce regroupement est corrélé à la répression transcriptionnelle de ces derniers par la protéine Abp1 (*ARS Binding Protein 1*) qui est homologue de la protéine CENP-B des mammifères (Lorenz et al., 2012). Ces différentes caractéristiques d'organisation ont permis de définir des régions nucléaires chez *S. pombe* :

- les corps-Tf sont là où se regroupent les répétitions de type Tf2 (Lorenz et al., 2012)
- les usines de transcription sont l'endroit où les gènes hautement transcrits sont associés (Tanizawa et al., 2010)
- les usines de réplication sont des foci discrets dans le noyau en phase S au niveau desquels la réplication a lieu (Meister et al., 2007)
- les centromères sont au niveau du SPB (Funabiki et al., 1993; Iwasaki et al., 2010)
- les télomères sont attachés à la membrane nucléaire (Funabiki et al., 1993)
- l'ADNr est au niveau du nucléole (Umesono et al., 1983).

2. Le mécanisme de changement de type sexuel

2.1 Concepts généraux

Le CTS chez *S. pombe* est un exemple de division asymétrique qui a permis des avancées considérables dans la compréhension de différents mécanismes, comme le contrôle de la réplication, la formation de l'hétérochromatine ou encore la structuration du noyau. En prenant comme appui ce mécanisme, les éléments de bases connus grâce à ce dernier seront développés au cours de l'introduction.

Une population de cellules sauvages homothalliques, appelée h^{90}, est composée de 50 % de cellule de type sexuel *M* (Minus) et de 50 % de type sexuel *P* (Plus). Cette capacité de former une colonie composée des deux types sexuels en même proportion démontre le potentiel de changement de type sexuel d'une cellule. En absence d'azote, les cellules de type sexuel opposé sont capables de conjuguer et de donner un asque composé de quatre spores (Leupold en 1950). Cette propriété a été utilisée pour l'analyse du pédigrée d'une seule cellule. Miyata et Miyata en 1981 ont réalisé des pédigrées au niveau d'une seule cellule haploïde déterminant ainsi différentes règles du CTS chez *S. pombe*. La première observation fut qu'un seul zygote est formé parmi les quatre cellules petites filles d'une même cellule (figure 4C). Cette observation détermina une des lois du CTS qui est la règle "une sur quatre". Cette règle indique que deux divisions asymétriques sont nécessaires pour qu'une cellule change de type sexuel. La première division cellulaire donne deux cellules avec le même type sexuel, mais avec des potentiels de changer de type sexuel différents. Ces potentiels sont notés "u" pour "*unswitchable*" (qui ne peut pas changer) et "s" pour "*switchable*" (qui peut changer) (figure 4B). La cellule "s" produira à la division suivante, une cellule ayant interverti son type sexuel et une cellule identique à elle-même (figure 4C). Cette technique utilisée ne permet pas l'étude de la troisième division cellulaire puisque les cellules petites filles auront conjugué. Afin de pallier ce problème, l'analyse de pédigrée a été conduite sur des cellules diploïdes. Dans cette cellule, seul un chromosome est capable de changer de type sexuel. Si les deux chromosomes sont de même type sexuel, le diploïde se divise. Si les chromosomes portent des types sexuels opposés, alors le diploïde sporule. Ainsi, les deux cellules diploïdes filles peuvent être suivies indépendamment. En analysant le pédigrée de ces cellules, il a été montré que la cellule sœur d'une cellule qui vient de changer de type sexuel est compétente pour changer de type sexuel à la réplication suivante. Cette démonstration définit la seconde règle du CTS chez *S. pombe*, à savoir, la loi "consécutive" (figure 4C et 4D) (Egel and Eie, 1987; Klar, 1990). Cette loi résulte du fait qu'une cellule capable de changer de type sexuel est formée à chaque division, formant

ainsi une lignée de cellules. Cette règle montre que le potentiel de changer de type sexuel est réparti de façon asymétrique entre les deux cellules filles.

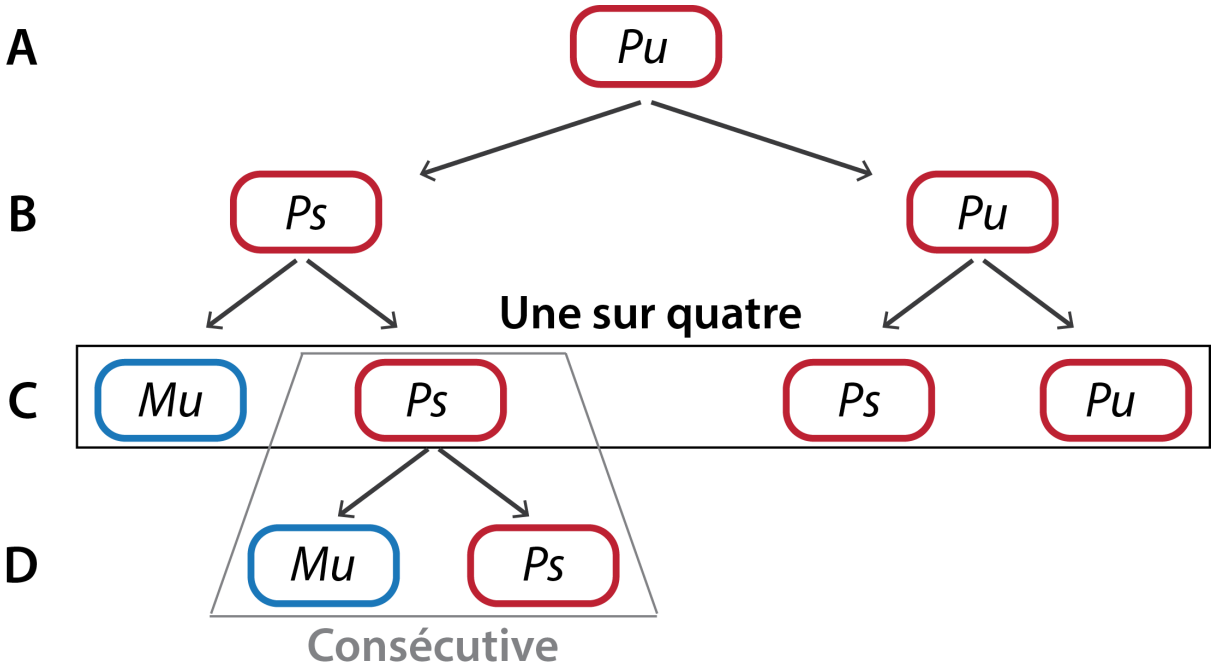


Figure 4 : Modèle du changement de type sexuel.
P et M représentent les deux types sexuels. Les suffixes "u" (*unswitchable*) et "s" (*switchable*) indiquent le potentiel de CTS ou non pour une cellule. Deux divisions asymétriques consécutives sont nécessaires pour produire une cellule de type sexuel opposé (règle "Une sur quatre"). La sœur de la cellule qui vient de changer de type sexuel est capable de changer de type sexuel à la prochaine division (règle "consécutive").

2.2 La région *mat* contrôle la division asymétrique

La région qui contrôle le CTS est la région *mat*. Elle est située sur le bras droit du chromosome 2 à environ 500 kb de *CENII*. Cette région est composée de trois loci : *mat1*, *mat2P* et *mat3M* (figure 5). Le locus *mat1* peut contenir l'allèle *M* (noir) ou l'allèle *P* (gris), et il détermine le type sexuel de la cellule. Le locus *mat2* contient l'allèle *P* uniquement et le locus *mat3* l'allèle *M*. Les loci *mat2P* et *mat3M* sont silencieux et sont utilisés comme donneurs lors du changement de l'information à *mat1*. Les allèles *P* et *M* sont composés chacun de deux gènes divergents : *Pc*, *Pi* (ou *Pm*) et *Mc*, *Mi* (ou *Mm*) (figure 5). *Mc* et *Pc* contrôlent le type sexuel de la cellule alors que *Pi* et *Mi* contrôlent l'entrée en méiose. L'expression concomitante de *Pi* et de *Mi* est nécessaire pour activer la méiose (Willer et al., 1995). Cette double nécessité restreint l'entrée en méiose au seul diploïde. Les boites d'homologies H1 et H2 sont présentes au niveau des bordures des trois loci. La boite H1 (59 pb) est située du côté centromère distal de *mat1*, *2* et *3* alors que la boite H2 (135 pb) est située du côté centromère proximal de ces trois loci

(Kelly et al., 1988). Les loci donneur *mat2P* et *mat3M* partagent en plus la boite H3 (57 pb) du côté centromère proximal (Kelly et al., 1988). Les boites H1 et H2 sont probablement nécessaires lors de la recherche d'homologie pendant la conversion de l'allèle à *mat1*. En ce qui concerne la boite H3, son rôle dans le CTS semble mineur puisque, par exemple, des mutations introduites dans cette boite n'affectent pas le mécanisme (Fleck et al., 1994).

Deux régions séparant les loci impliqués dans le CTS ont été définies. La région L (17 kb) située entre le locus *mat1* et *mat2P* qui contient au moins un gène essentiel (*rpt6* ou *let1*) *(Michael et al., 1994)*. La délétion de cette région lors de réarrangements spontanés est létale (Beach, 1983; Beach and Klar, 1984). D'autre part, la région K, séparant les loci *mat2P* et *mat3M,* contient le locus *CENH* qui est composé de séquences homologues aux centromères (Grewal and Klar, 1997). Enfin, une caractéristique intéressante de la région *mat* est la présence de deux séquences répétées inversées entourant les loci donneurs : *IR-L* et *IR-R* (Noma et al., 2001). Ces répétitions participent au contrôle de l'état chromatinien de cette région et déterminent les limites de l'hétérochromatine (Thon et al., 2002).

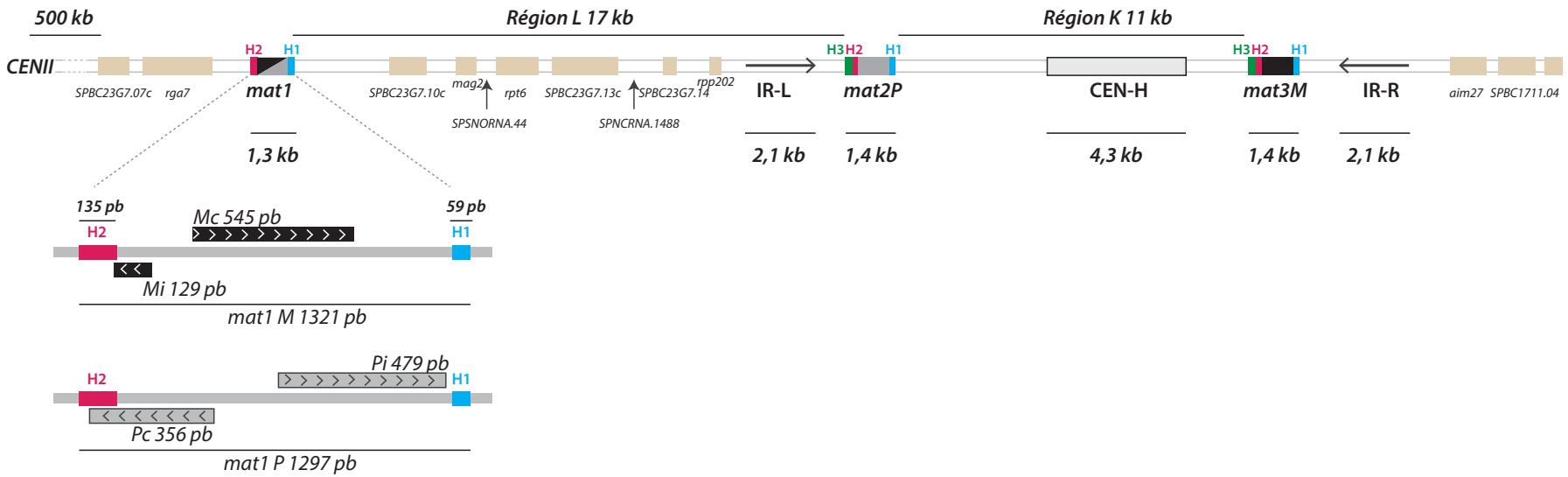


Figure 5 : Organisation de la région *mat.*

La région *mat* est située sur le bras droit du chromosome 2. Les allèles *M* et *P* sont représentés en noir et gris, respectivement. Les boites d'homologies H1 (bleu), H2 (rose) et H3 (vert) sont également indiquées. Les régions répétées *IR-L* et *IR-R* sont représentées par des flèches. La région *CEN-H* de 4,3 kb présentant des homologies avec les régions centromériques est représentée par un rectangle gris aux bordures noires. Les gènes présents dans la région sont indiqués en beige, leur nom est aussi annoté. Les régions L et K séparant les loci impliqués dans le CTS sont annotées. Les régions codantes les gènes *Mi, Mc, Pc* et *Pi* sont également indiquées ainsi que le sens de la transcription.

2.3 Le mécanisme

Depuis la découverte de ce modèle de division asymétrique, les scientifiques ont réalisé de nombreuses expériences qui ont permis de déterminer des éléments moléculaires responsables du CTS. Le modèle actuel peut être divisé en plusieurs étapes qui seront approfondies dans la suite de l'introduction (figure 6). De façon générale, nous pouvons décrire le mécanisme de la façon suivante : tout d'abord la réplication d'un locus particulier, *mat1,* est unidirectionnelle et permet la modification de la molécule d'ADN sans en affecter sa séquence. Cette modification est appelée empreinte épigénétique. L'empreinte n'est formée que sur le brin retardé néo-synthétisé et détermine l'asymétrie du mécanisme de CTS. La cellule contient deux chromatides différentes à la fin de la réplication, dont une seule porte l'empreinte. À la suite de la division cellulaire, une seule des deux cellules filles acquiert le potentiel de changer de type sexuel. L'empreinte est maintenue durant tout le cycle cellulaire jusqu'à la réplication suivante. Lors de la seconde réplication, l'empreinte présente au niveau du brin Watson sera convertie en une cassure double brin (CDB) recombinogène, au passage de la fourche à *mat1*. Au niveau du brin retardé néo-synthétisé, à nouveau, une empreinte sera déposée. La CDB est réparée par recombinaison homologue avec un des deux loci silencieux, *mat2P* ou *mat3M* en utilisant les boites d'homologies H1 et H2 pour l'initiation et la résolution de la recombinaison, respectivement. De plus, dans 90 % des cas, la réparation utilise le locus donneur portant l'information opposée à celle présente à *mat1*. Ce choix particulier, nommé directionnalité, permet à une population cellulaire d'établir rapidement une répartition égale entre les types sexuels *M* et *P*.

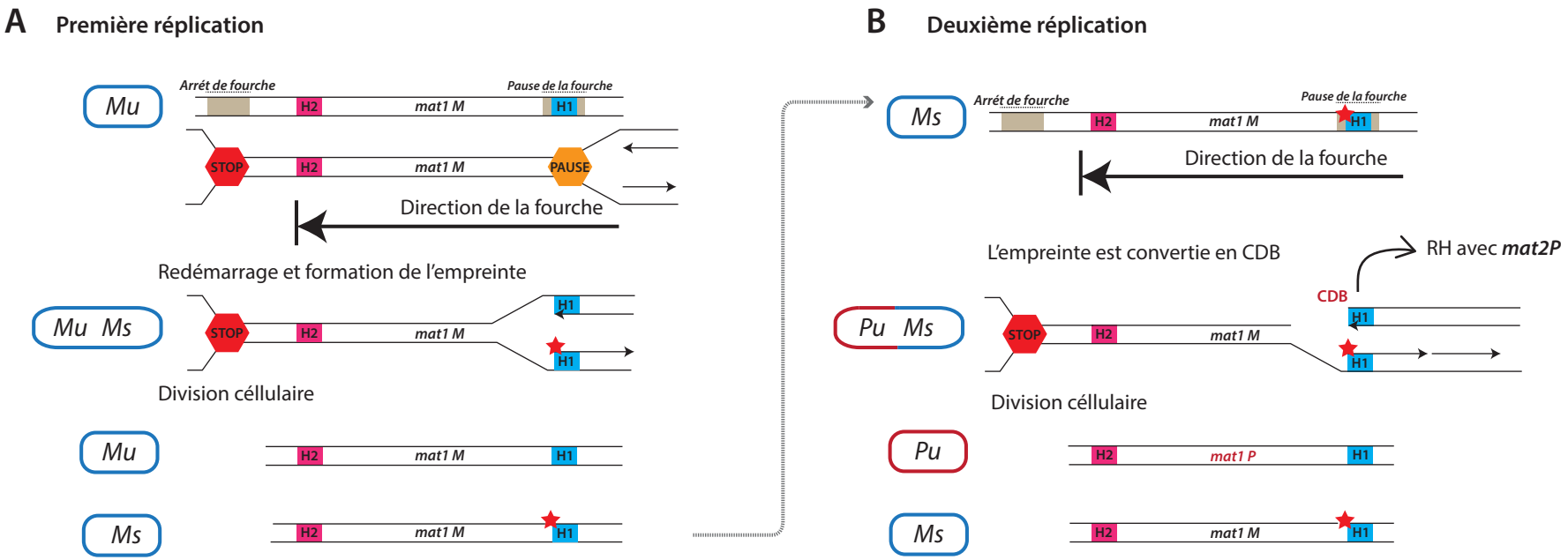


Figure 6 : Le mécanisme de division asymétrique chez *S. pombe.*
A) Schéma représentatif de la réplication du locus sexuel *mat1M* ne portant pas d'empreinte (*Mu*). Les sites contrôlant la fourche de réplication sont indiqués. Les boites d'homologies H1 (bleu) et H2 (rouge) sont représentées. La fourche de réplication fait une pause au niveau de la boite H1 et permet la formation de l'empreinte (étoile rouge). À la fin de la première réplication et de la division cellulaire, une des deux cellules filles contient le chromosome qui porte l'empreinte. B) Schéma représentant la réplication à travers la molécule qui porte l'empreinte. La fourche convertit l'empreinte en une cassure double brin (CDB) qui sera réparée en utilisant l'information d'un des donneurs. Dans le cas représenté, le donneur *mat2P* sera préférentiellement utilisé, car l'allèle à *mat1* est *M*. Le redémarrage de la fourche permet la formation de l'empreinte au niveau du brin retardé néo-synthétisé. À la fin de la deuxième réplication, une des cellules a changé de type sexuel (ici la cellule est devenue *mat1P*).

3. L'initiation du mécanisme : une asymétrie réplicative

Dans cette partie, les éléments moléculaires impliqués dans l'initiation du changement de type sexuel seront définis. De plus, les acteurs protéiques et les sites qui influencent cette initiation seront également décrits.

3.1 Une empreinte chromosomique formée asymétriquement à *mat1*

3.1.1 L'asymétrie est portée par le chromosome

Lors de l'analyse de la région *mat* par Southern-Blot, David Beach montra que le fragment de restriction HindIII de 10,4 kb contenant le locus *mat1* est divisé en deux sous fragments dans approximativement 50 % des cas, un fragment centromère proximal (5,4 kb) et un centromère distal (5 kb)(Beach, 1983) (figure 7). Cette observation suggéra qu'une cassure double brin détectable par Southern-Blot est à l'origine de la transposition de l'allèle à *mat1* (Beach, 1983), à l'image de ce qui est observé chez *Saccharomyces cerevisiae* (Kostriken et al., 1983). Au vu du caractère asymétrique du CTS, la question sous-jacente était de déterminer quels éléments sont ségrégés asymétriquement dans les deux cellules filles, influençant la formation de cette cassure. Deux hypothèses peuvent être prises en considération : la première étant qu'une répartition asymétrique d'éléments cytoplasmiques permet de donner le potentiel de CTS à une seule des cellules filles ; la seconde hypothèse est qu'un seul des chromosomes homologues est porteur de l'asymétrie. L'une de ces deux hypothèses fut validée par une expérience de pédigrée réalisée par Richard Egel en 1984. En effet, il a suivi l'évolution du CTS dans des cellules diploïdes et a mis en évidence que les deux chromosomes sont capables d'alterner leur type sexuel indépendamment l'un de l'autre (Egel, 1984a). Par la suite, Amar Klar montra que le potentiel de changer de type sexuel est acquis par une marque épigénétique définie comme empreinte présente au niveau du locus *mat1*. Pour ce faire, il développa un système génétique dans lequel l'allèle *mat1* est dupliqué dans deux orientations (figure 8A et 8B). Il observa que l'orientation des cassettes dupliquées détermine la présence de l'empreinte (de nature inconnue à cette époque). Il montra qu'il est nécessaire que les deux cassettes soient en orientations directes pour simultanément porter l'empreinte. En effet, dans ce contexte, il détecte par Southern-Blot le fragment d'ADN entre les deux cassettes (figure 8C, puits 2). Ce fragment commence au niveau de l'empreinte de l'une des cassettes et finit au niveau de l'empreinte de l'autre. La présence de ce fragment montre que deux empreintes sont présentes simultanément sur la même molécule d'ADN. À l'inverse, les loci *mat1*, placés en orientation inverse, ont individuellement le même niveau d'empreinte, mais n'ont pas l'empreinte au même moment sur

la même molécule d'ADN, puisque le fragment les séparant n'est pas détectable (figure 8C, puits 3) (Klar, 1987, 1990). Ces expériences démontrent qu'une seule des deux chaines d'ADN est porteuse de l'empreinte et contrôle le potentiel asymétrique de ce mécanisme. À cette époque, l'empreinte analysée par Southern-Blot a été décrite comme étant une CDB à la jonction de la boite H1 et du locus *mat1*. Bien qu'Amar Klar en 1987 supposa que seul l'un des deux brins d'ADN est porteur de l'empreinte, il fallut attendre 1998 pour que la preuve soit apportée.

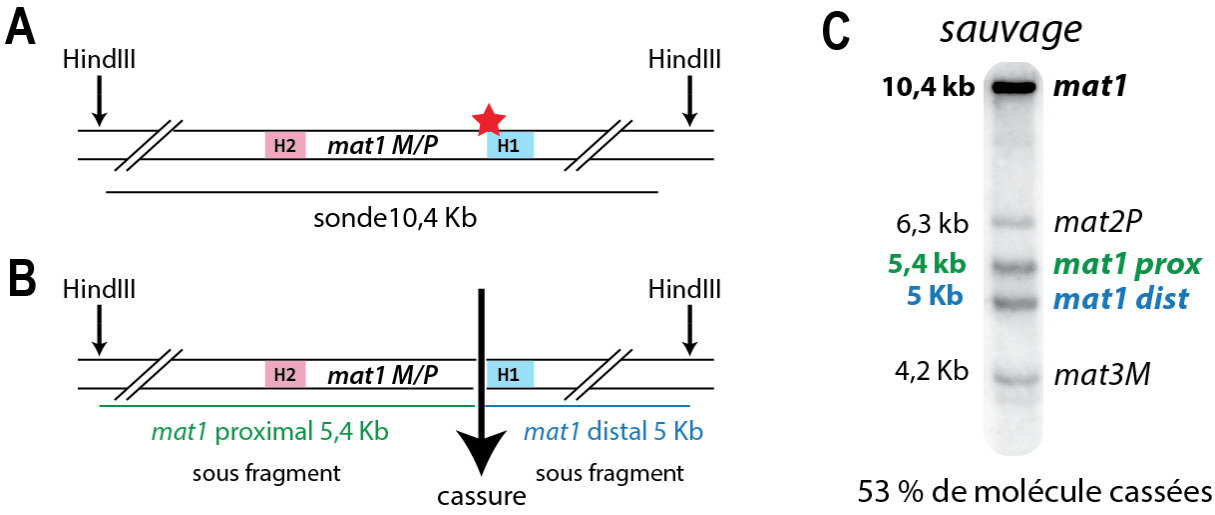


Figure 7 : Le fragment de digestion contenant *mat1* est cassé dans 50% des cas.
A) Schéma représentant le fragment de digestion HindIII contenant le locus *mat1*. La sonde utilisée pour le Southern-Blot est représentée par un trait noir. B) Schéma représentant les deux sous fragments obtenus après la préparation de l'ADN. C) Résultat d'un Southern-Blot réalisé à partir d'ADN génomique d'une souche sauvage *h90* révélée avec la sonde indiquée en A). Le pourcentage de molécules de 10,4 kb générant les deux sous-fragments est indiqué.

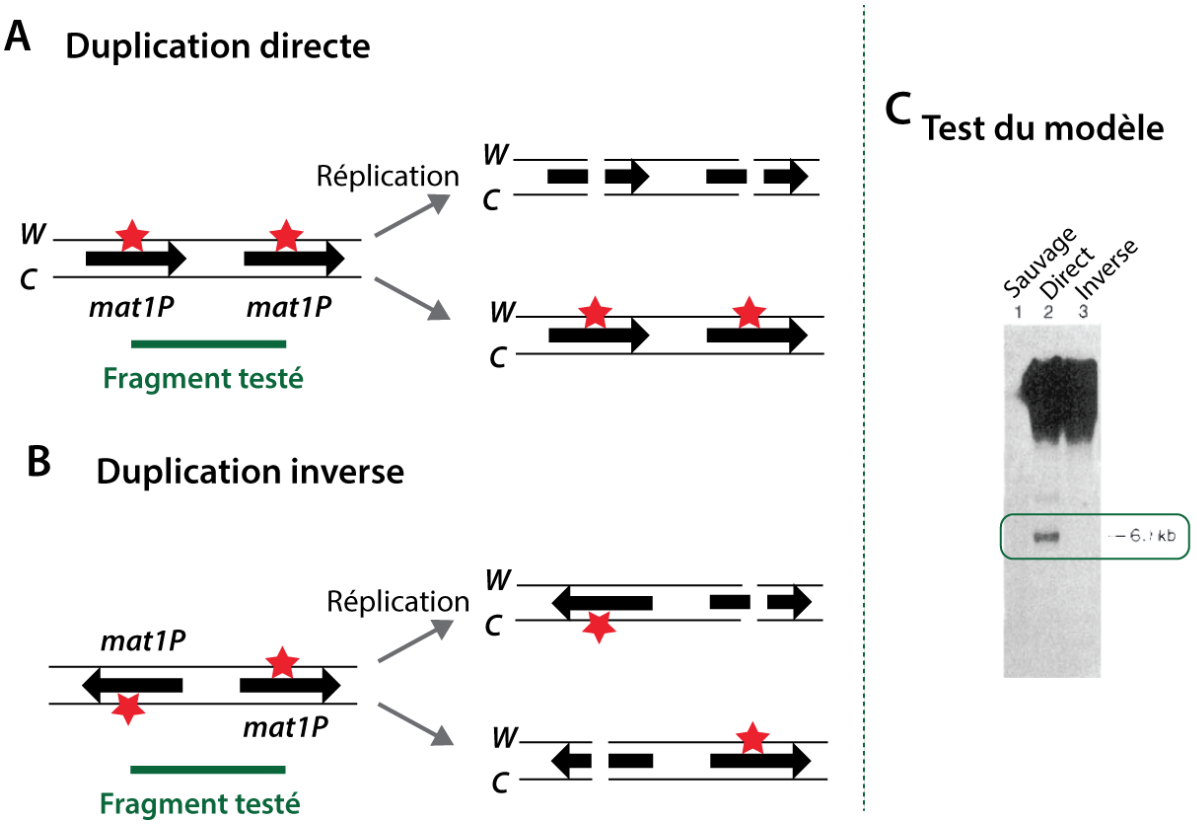


Figure 8 : Modèle proposé et validé par Klar. L'empreinte est portée par un brin d'ADN.
A-B) Prédiction du modèle de ségrégation des brins tirée de (Klar, 1987). A) Les répétitions du locus *mat1* sont placées en orientation directe. Les flèches noires représentent les cassettes *mat1*. L'étoile rouge représente l'empreinte. Les brins Watson (W) et Crick (C) sont indiqués. Le fragment testé par Southern-Blot est marqué en vert. B) Même schéma avec les duplications placées en orientation inverse. C) Résultat du Southern-Blot (Klar, 1987) testant le modèle proposé. L'ADN génomique de la souche sauvage (puits 1), portant la duplication en orientation directe (puits 2) et indirecte (puits 3) a été préparé puis analysé par Southern-Blot. La sonde utilisée permet de révéler le fragment vert présent entre les duplications. Ce fragment contient le gène *LEU2* de *S. cerevisiae*.

3.1.2 La position et la nature de l'empreinte

En utilisant le fait que l'empreinte soit détectable par Southern-Blot, sa position a pu être déterminée. Un des éléments observés est qu'elle diffère en fonction de l'allèle présent à *mat1* (Nielsen and Egel, 1989) (figure 9). De plus, dans cette étude le site de l'empreinte a été déterminé au niveau du brin Watson (W), mais pas au niveau du brin Crick (C). Cette observation suggère que sur le brin Crick, le site de l'empreinte n'est pas site spécifique ou moins stable. Cet élément sera étayé par des résultats que nous avons également obtenus. De façon intéressante, il fut ensuite montré que seul le brin Watson portait l'empreinte. C'est en fait la méthode de préparation d'ADN qui génère la cassure double brin (Arcangioli, 1998). Cette empreinte crée un site fragile qui sera cassé par tension mécanique (vortex, sonication) ou par

digestion enzymatique de l'ADN ou l'ARN simple brin (Mung Bean). Cette première étude suggère que l'empreinte est une discontinuité simple chaine située à la jonction de la boite H1 et du locus *mat1*. Une autre étude confirma cette nature en utilisant une technique de "*ligation-mediated*" PCR (Kaykov and Arcangioli, 2004). Cependant, deux publications ont montré que l'empreinte est sensible à la RNAse T2 (une endonucléase que digère l'ARN), suggérant que l'empreinte est constituée de ribonucléotides (Vengrova and Dalgaard, 2004; Vengrova and Dalgaard, 2006). Bien que la nature de l'empreinte soit toujours sujette à débat, il existe un consensus sur le fait que l'empreinte crée un site fragile qui initie la formation de la CDB utilisée pour la conversion génique. L'empreinte étant formée sur le brin Watson, il est probable que la cassure du brin Crick ne soit pas exactement en face de cette dernière et soit aléatoire. Ce phénomène peut expliquer la difficulté qu'a eu Olaf Nielsen pour détecter l'extrémité du brin Crick. Par ailleurs, l'empreinte n'est pas spécifique de la séquence puisque des mutations introduites au site de l'empreinte n'abolissent pas sa formation (Kaykov et al., 2004; Sayrac et al., 2011).

Aujourd'hui, l'acteur initiateur de l'empreinte est toujours sujet à questionnement. L'analyse de la réplication du locus *mat1* a permis de déterminer une étape précoce du mécanisme, à savoir la pause de la fourche de réplication à la jonction de la boite H1 et du locus *mat1*.

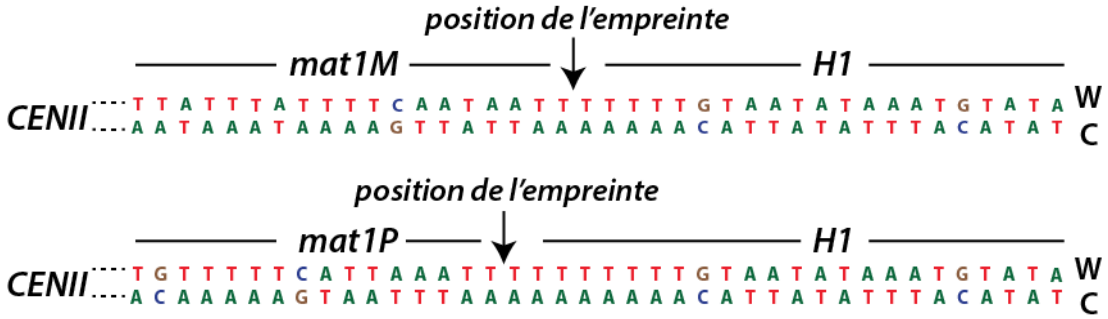


Figure 9 : Position de l'empreinte en fonction de l'allèle à *mat1*.
L'empreinte est décalée de 3 nucléotides dans les deux allèles. La boite H1 du côté centromère distal est identique dans les deux contextes ainsi que les deux allèles. Les séquences des brins Watson (W) et Crick (C) sont indiquées pour chaque allèle. La position relative du centromère *CENII* est indiquée.

3.2 La réplication asymétrique de *mat1* et son contrôle

3.2.1 La réplication unidirectionnelle permet la formation de l'empreinte

Un des premiers mutants isolés déficients pour le CTS a été le gène *swi7* (Egel et al., 1984). Ce gène code pour la sous-unité catalytique de l'ADN polymérase α (Pol α) qui initie la réplication

(Singh and Klar, 1993). Cet élément est le premier indice corrélant l'efficacité du mécanisme et la réplication. Par la suite, l'étude des intermédiaires de réplication par gel bidimensionnel d'agarose a révélé que la fourche se déplace de façon unidirectionnelle à travers le locus *mat1* (Dalgaard and Klar, 1999). La fourche de réplication progresse du côté centromère distal de *mat1* vers le côté centromère proximal. Cette caractéristique détermine l'efficacité de la formation de l'empreinte, puisque l'introduction d'une origine de réplication du côté centromère proximal de *mat1* ou l'inversion du locus *mat1* inhibe la formation de l'empreinte (Dalgaard and Klar, 1999). La caractérisation fine de la réplication de *mat1* a mis en évidence deux sites importants : le site *RTS1* (*Replication Termination Sequence 1*) et le site *MPS1* (*Mating Type Pause Site 1*) (Dalgaard and Klar, 1999, 2001) et (Figure 10).

Le site *RTS1*, situé à environ 800 pb de la boite H2, bloque les fourches de réplication provenant du côté centromère proximal de *mat1*. La délétion de ce dernier entraine une diminution de 30 % de la formation de l'empreinte (Dalgaard and Klar, 2001).

D'autre part, du côté centromère distal de *mat1,* une origine de réplication a été mise en évidence. Cependant, sa participation au CTS est mineure et suppose que d'autres origines de réplication sont également capables d'initier la réplication de ce côté du locus (Dalgaard and Klar, 2001; Maundrell et al., 1988). Les délétions combinées de cette origine et de *RTS1* entrainent une réplication bidirectionnelle du locus *mat1*. Dans ce contexte, le brin Watson néo-synthétisé n'est plus uniquement le brin retardé.

Enfin, l'étude de la réplication a également montré que la fourche provenant du côté centromère distal, fait une pause avant de répliquer *mat1* (figure 10) (Dalgaard and Klar, 1999). Ce site de pause nommé *MPS1* (*Mat Pause Site 1*) est localisé au niveau de l'empreinte indépendamment de l'allèle présent à *mat1* (Sayrac et al., 2011). De plus, l'activité de *MPS1* est absolument requise pour la formation de l'empreinte. En effet, plusieurs types de mutations impactant cette pause entrainent une diminution très forte de l'empreinte. Ainsi, deux observations suggèrent que la pause de la fourche de réplication à *MPS1* est l'étape initiatrice du CTS : premièrement, jamais un mutant dépourvu de pause et possédant une empreinte n'a été isolé ; deuxièmement, la dissection des étapes du mécanisme a montré que c'est précisément au redémarrage de la fourche en pause que l'empreinte est détectable à la jonction du locus *mat1* et de la boite H1 (Holmes et al., 2005).

La réplication hautement contrôlée de ce locus est donc un élément fondamental pour le CTS. Ces observations suggèrent que l'asymétrie intrinsèque à la réplication est utilisée pour la formation de l'empreinte sur le brin retardé (Watson en l'occurrence).

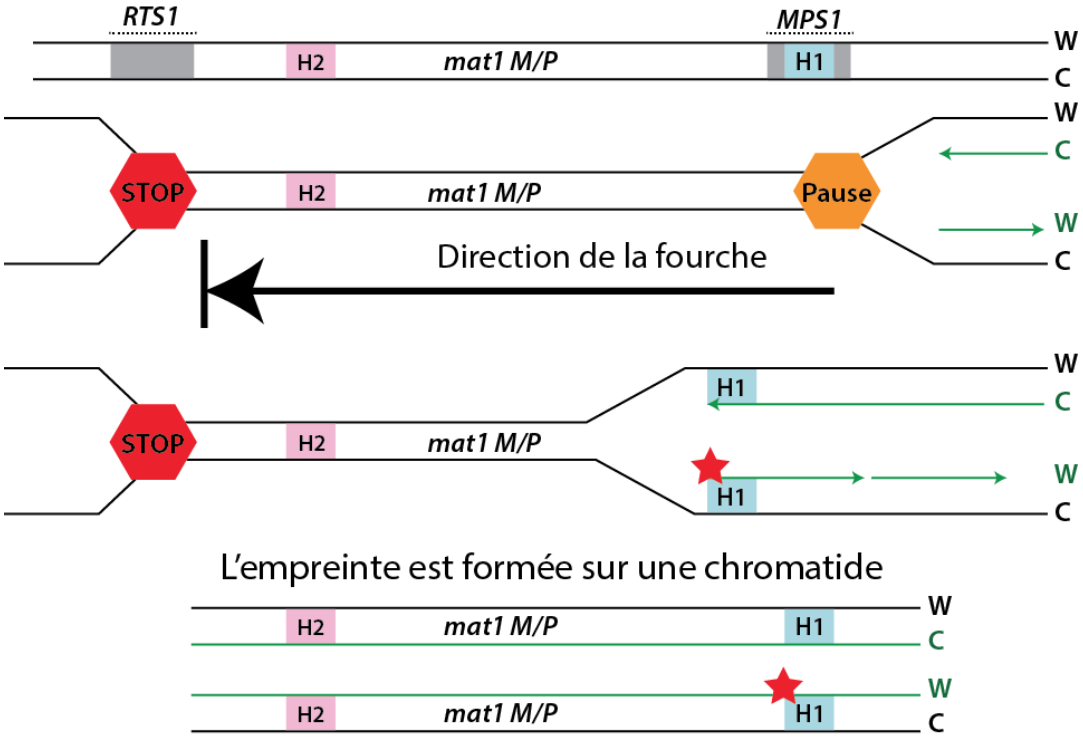


Figure 10 : La réplication unidirectionnelle du locus *mat1*.
Schéma représentant la direction de la réplication du locus *mat1*. Les boites d'homologie H2 (rose) et H1 (bleu) sont indiquées ainsi que les sites *RTS1* et *MPS1*. La fourche de réplication provenant du côté centromère proximal est arrêtée par *RTS1*. La fourche de réplication provenant du côté centromère distal fait une pause à *MPS1* puis redémarre et permet la formation asymétrique de l'empreinte à *mat1*. Les brins d'ADN Watson (W) et Crick (C) sont indiqués. Les brins néo-synthétisés sont en vert. L'empreinte est indiquée par une étoile rouge. Ainsi, une seule chromatide porte l'empreinte sur le brin Watson néo-synthétisé.

3.2.2 Le complexe Swi1-Swi3 protège la fourche à *MPS1*

L'activité du site *MPS1* a été analysée par gel bidimensionnel dans des souches contenant différentes délétions/substitutions autour de l'empreinte. Ces analyses ont permis d'établir les séquences nécessaires et/ou suffisantes au blocage réversible de la fourche de réplication. De plus, des acteurs protéiques impliqués dans l'activité du site *MPS1* ont également été isolés.

Tout d'abord, l'activité de *MPS1* est abolie quand les gènes *swi1* (*Switch gene 1)* ou *swi3* (*Switch gene 3*) sont mutés (Dalgaard and Klar, 2000). La protéine Swi1 est retrouvée en complexe avec la protéine Swi3 (Noguchi et al., 2004). Initialement, des mutants de ces deux gènes ainsi que du gène *swi7* ont été isolés comme présentant un défaut d'empreinte (Egel et al., 1984). L'analyse de la réplication dans ces mutants a permis de déterminer plus précisément leur rôle. Le complexe Swi1-Swi3 (homologue de TIMELESS-TIPIN chez l'humain), à la différence de Swi7, participe à l'activité du site *MPS1* (Dalgaard and Klar, 2000). Swi1 et Swi3 appartiennent au complexe RPC (*Replication Fork Complex*) qui maintient les fourches de réplications stables

en vue de leur redémarrage (Noguchi et al., 2003; Noguchi et al., 2004). Ce complexe se déplace avec la fourche de réplication et permet l'activation du point de contrôle réplicatif : la kinase Cds1 (*Checking DNA Synthesis*) (Noguchi et al., 2003). À *mat1*, l'arrivée de la fourche de réplication est concomitante avec l'arrivée de Swi1-Swi3, permettant la stabilisation de la structure réplicative formée à *MPS1* (Holmes et al., 2005; Noguchi et al., 2004). D'autre part, à 15 pb de l'empreinte du côté centromère distal de *mat1,* une substitution de 5 pb, appelée *mut3* diminue le recrutement de Swi1 ainsi que le niveau de *MPS1* (Kaykov et al., 2004 ; Lee et al., 2004) (figure 11).

Ainsi, du côté centromère distal, nous pouvons établir que deux éléments en *trans* (Swi1-Swi3) et un élément en *cis* (*mut3*) contrôlent l'activité de la pause et par ce fait la formation de l'empreinte.

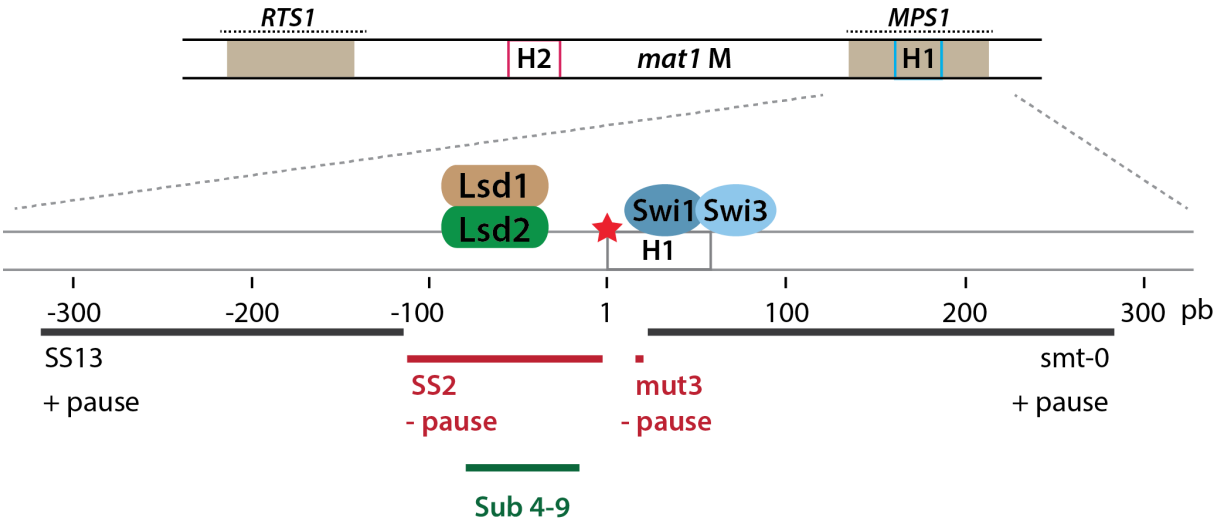


Figure 11 : Les éléments contrôlant l'activité de *MPS1.*
Le locus *mat1M* n'est représenté qu'avec les sites *RTS1*, *MPS1* et les boites d'homologie H1/H2. Le panneau du dessous est un zoom au niveau du site *MPS1*. Les distances en pb sont indiquées par rapport à la position de l'empreinte représentée par une étoile rouge. Les sites de recrutement des deux complexes Lsd1/Lsd2 et Swi1/Swi3 sont représentés. Différentes délétions sont notées, en noir, celles n'impactant pratiquement pas la pause et en rouge, celles abolissant l'activité de *MPS1*. La région Sub 4-9, en vert est également représentée.

3.2.3 Les lysines déméthylases Lsd1-Lsd2 agissent sur la fourche à *MPS1*

Du côté centromère proximal de *mat1M*, la délétion de la séquence *SS2* présente à l'intérieur du locus *mat1M* abolit l'activité de *MPS1* (Sayrac et al., 2011) (figure 11). Cette séquence de 111 pb est liée par le complexe Lsd1-Lsd2 (*Lysine specific demetylase 1 - 2)* (Holmes et al., 2012). De façon surprenante, ce complexe d'histone déméthylases contrôle l'activité de *MPS1* (Gordon et al., 2007; Holmes et al., 2012; Lan et al., 2007; Nicolas et al., 2006; Shi et al., 2004).

Lsd1 et Lsd2 sont en complexe avec deux autres protéines, Phf1 et Phf2 (*PHD finger 1* et *2*) (Nicolas et al., 2006). Lsd1 et Lsd2 sont composées de trois domaines identifiés : le domaine SWIRM, qui est présent dans des protéines qui régule la chromatine, le domaine amine oxydase (ao) qui permet à Lsd1 de déméthyler les lysines K4 et K9 de l'histone H3 et le domaine HMG (*High-Mobility Group*) qui pourrait permettre à Lsd1 et Lsd2 de lier l'ADN (figure 12). Les deux autres protéines retrouvées en complexe avec LSD1 et Lsd2, Phf1 et Phf2 ont un domaine PHD finger qui permet de reconnaitre les nucléosomes (Gordon et al., 2007). Les délétions de *lsd2*, *phf1* ou *phf2* sont létales, tandis que la délétion de *lsd1* entraine un fort retard de croissance (Gordon et al., 2007; Lan et al., 2007; Opel et al., 2007). De façon globale, plusieurs éléments suggèrent que Lsd1 régule la transcription au travers de la déméthylation de la lysine 4 et 9 mono- et di-méthylées de l'histone H3 (Gordon et al., 2007; Lan et al., 2007; Opel et al., 2007). Cependant, une souche mutante pour les deux activités enzymatiques de Lsd1 et de Lsd2 est viable (*lsd1 ao lsd2 ao*), indiquant que ces protéines jouent également un rôle non enzymatique requis pour la survie de cette levure (Gordon et al., 2007; Lan et al., 2007; Opel et al., 2007). À *mat1*, les activités enzymatiques et non enzymatiques de ce complexe contrôlent l'efficacité de *MPS1* (Holmes et al., 2012). En effet, les mutants *lsd1-18* et le double mutant *lsd1 ao lsd2 ao* ont un CTS peu efficace (figure 12). La mise en évidence de ce complexe comme acteur du CTS soutient la notion que la structure chromatinienne à *mat1* influence la formation de l'empreinte.

Enfin, nous pouvons ajouter qu'une sous-région de *SS2* : la région Sub 4-9 (figure 11), est la séquence minimale nécessaire à la formation de la pause *MPS1*. Cette région est à l'intérieur de l'allèle *M*. Une région similaire a aussi été isolée à l'intérieur de l'allèle *P* : Sub 4-9 P (Sayrac et al., 2011). Ces régions (Sub 4-9 M et Sub 4-9 P) placées dans un plasmide sont capables d'arrêter la fourche de réplication (Sayrac et al., 2011). Il est assez remarquable que les séquences responsables de la pause soient situées à l'intérieur des allèles *M* et *P*, et donc différentes. Cela suppose qu'un élément commun à ces séquences permet de déclencher la pause de la fourche. À ce jour, la liaison du complexe Lsd1-Lsd2 sur le plasmide contenant les substitutions (Sub 4-9 M et Sub 4-9 P) ou sur l'allèle *mat1P* n'a pas été analysée.

Ainsi, deux complexes protéiques contrôlent *MPS1* : Swi1-Swi3 progressant avec la fourche de réplication, et Lsd1-Lsd2 se liant à la séquence minimale nécessaire pour la pause. Par ailleurs, de part et d'autre du site de l'empreinte des sites précis contrôlent la réplication (*SS2* et *mut3*), tandis que la délétion de séquences plus éloignées du site de l'empreinte (*SS13* et *smt-0*) n'impacte pas l'activité de *MPS1* (figure 11) (Dalgaard and Klar, 2000; Sayrac et al., 2011).

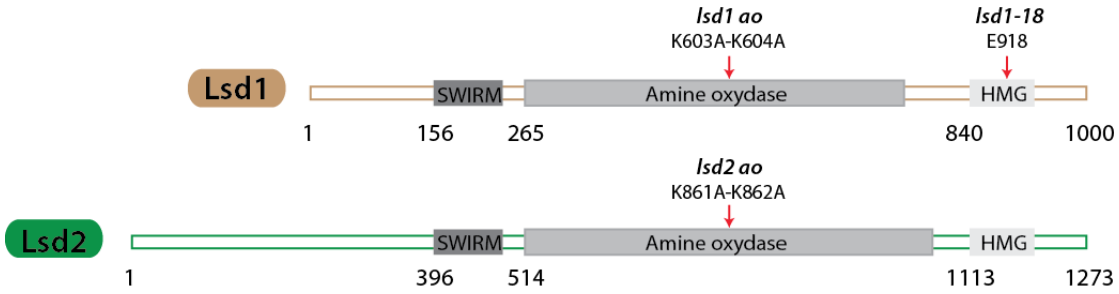


Figure 12 : Les domaines des protéines Lsd1 et Lsd2
Représentation schématique des protéines Lsd1 (en marron), Lsd2 (en vert) et de leurs domaines. Les positions des acides aminés sont indiquées. Les positions des mutations *lsd1 ao*, *lsd2 ao* et *lsd1-18* sont indiquées.

3.2.4 Le contrôle de RTS1

Comme présenté précédemment, le site *RTS1* bloque les fourches de réplication provenant du côté centromère proximal de *mat1* optimisant la formation de l'empreinte (Dalgaard and Klar, 2001). Tout comme à *MPS1*, le complexe Swi1 et Swi3 stabilise la fourche bloquée à *RTS1* (Dalgaard and Klar, 2000). Un crible permettant d'isoler des mutants affectant *RTS1* a révélé qu'un des allèles mutant du gène *swi1* impacte *RTS1* et non *MPS1*. Cette observation montre que *swi1* joue deux fonctions séparées dans le contrôle de la réplication à *mat1* (Dalgaard and Klar, 2000). D'autre part, les protéines Rtf1 et Rtf2 (*Replication Termination Factors 1 et 2*) participent à la terminaison de la réplication à *RTS1* (Eydmann et al., 2008; Inagawa et al., 2009). Il est proposé que Rtf1, en se liant à plusieurs sites à l'intérieur des répétitions définissant *RTS1*, bloque la progression de la fourche (Eydmann et al., 2008). Puis, en aval de cette réaction, Rtf2 convertirait la barrière de réplication en une terminaison de la réplication en stabilisant la fourche (Inagawa et al., 2009).

3.2.5 Le modèle de la réplication du locus sexuel aujourd'hui

En absence d'empreinte à *mat1,* la réplication de ce locus peut être imaginée comme représentée dans la figure 13. Le complexe protéique Lsd1-Lsd2 est recruté à l'intérieur de l'allèle *mat1M* et permet la pause de la fourche de réplication à *MPS1*. Swi1-Swi3 arrivant avec la machinerie de réplication stabilisent la fourche en pause à *MPS1*. Au redémarrage de la réplication, l'empreinte est formée à la jonction de la boite H1 et de *mat1M*. La fourche du côté centromère proximal est bloquée à *RTS1* et stabilisée par deux complexes, Swi1-Swi3 et Rtf1-Rtf2. À l'issue de la réplication à travers le locus sans empreinte, une des deux cellules filles porte la marque épigénétique. Les acteurs en *trans* contrôlant la pause quand l'allèle *P* est à *mat1* ne sont pas définis.

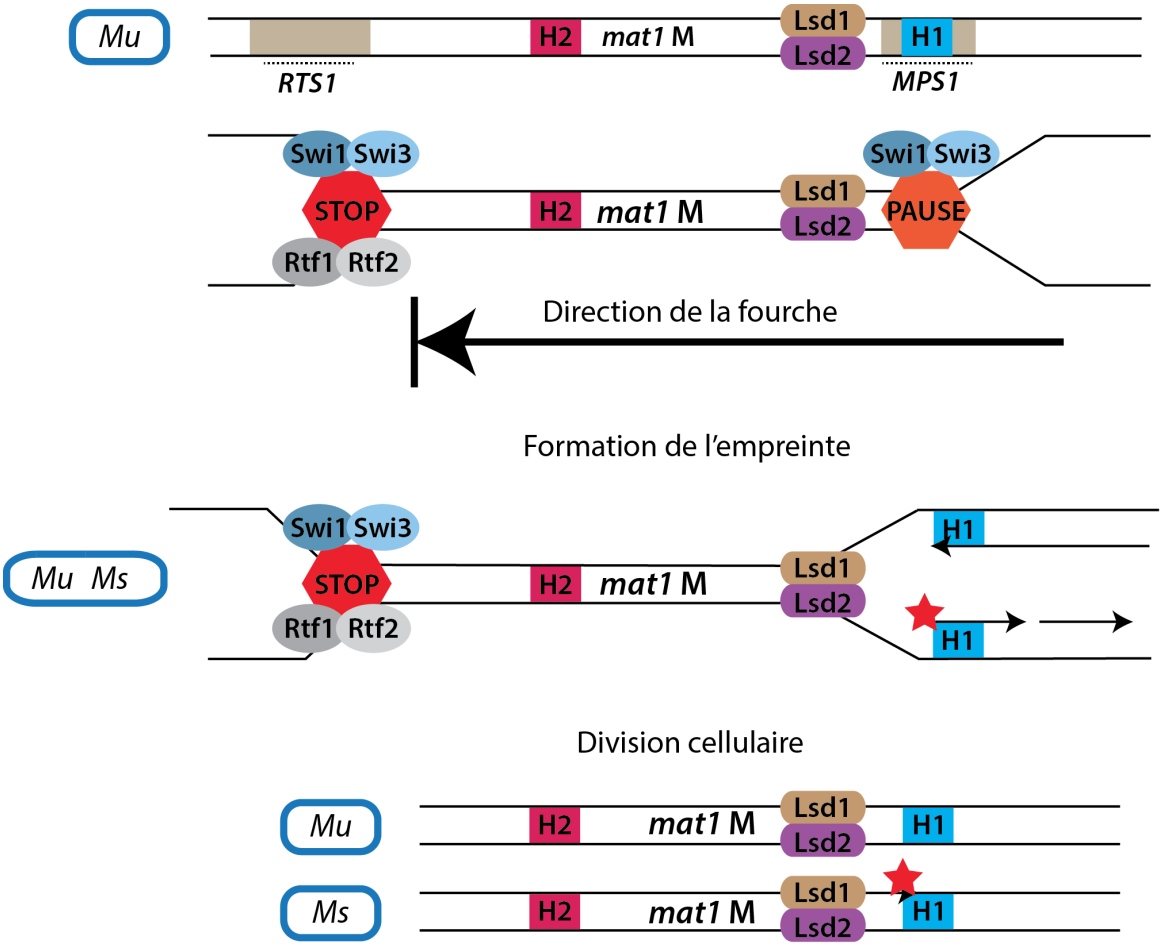


Figure 13 : Le modèle et les acteurs de la première réplication du locus sexuel
La réplication du locus *mat1* ne portant pas d'empreinte est schématisée. Les sites de contrôle de la fourche, *MPS1* et *RTS1* sont indiqués. Les acteurs agissant en *trans* sont également annotés. Le potentiel de CTS des cellules est schématisé sur la gauche.

3.3 Les arrêts de fourches et les acteurs du CTS

La fourche de réplication progresse le long du génome jusqu'à ce qu'elle rencontre une autre fourche ou un obstacle. Différents types d'obstacles présents sur la chromatine impactent la progression de la fourche, comme des protéines liées à la chromatine, des structures secondaires de la molécule d'ADN, des lésions dans l'ADN ou encore la transcription active de certains loci. Chez *S. pombe*, des blocages de fourche indépendant du CTS ont été observés. Leur contrôle a mis en cause des protéines importantes pour le CTS, soulevant la question de leur rôle au niveau de la réplication à *mat1*.

3.3.1 Les pauses au niveau des LTRs et de l'ADNr : les protéines Sap1 et Abp1

La réplication de chaque unité de l'ADN ribosomique est unidirectionnelle, tout comme à *mat1* (Kim and Huberman, 1998; Sanchez et al., 1998). Dans chaque unité de l'ADNr, une origine de réplication est présente ainsi qu'une série de barrières pour la fourche de réplication. Les fourches de réplication sont arrêtées de façon polaire au niveau de l'ADNr. La direction de la fourche est de par ce fait la même que la direction de la machinerie de transcription. Il est supposé que la polarité de la réplication de ce locus empêche la collision de la fourche réplication avec la machinerie de transcription. Plusieurs protéines contrôlent les arrêts de fourche au niveau de l'ADNr, et notamment une protéine recrutée à *mat1* : Sap1 (*Switch Activating Protein 1*) (Krings and Bastia, 2005; Mejia-Ramirez et al., 2005). Isolée comme un facteur du CTS, Sap1 est une protéine spécifique de *S. pombe* (Arcangioli et al., 1994a). Par ailleurs, Sap1 participe également à la pause des fourches de réplication au niveau des LTRs (Zaratiegui et al., 2011). En outre, la protéine Abp1 (*ARS Binding Protein 1*), homologue de la protéine CENP-B (*Centromeric Protein B*), permet le redémarrage de la fourche de réplication à travers la barrière imposée par Sap1 au niveau des LTRs (Zaratiegui et al., 2011). La protéine CENP-B contient trois homologues chez *S. pombe* : Abp1, Cbh1 et Cbh2 (*CENP-B Homologue 1 and 2*). Le double mutant *abp1*Δ *cbh1*Δ a une croissance lente et une mutation ponctuelle dans le gène *sap1* (*sap1-c*) supprime en partie ce phénotype (Zaratiegui et al., 2011). Les protéines Abp1 et Sap1 sont toutes deux composées de domaines de liaison à l'ADN et semblent jouer un rôle antagoniste sur le contrôle de la réplication (figure 14).

Abp1 et Sap1 sont recrutées à *mat1* (Arcangioli and Klar, 1991; Zaratiegui et al., 2011). Sap1 se lie du côté centromère distal de *mat1* au niveau du site *SAS1* à environ 160 pb du site de l'empreinte *in vitro* (Arcangioli and Klar, 1991). Le site consensus, TAACG, de recrutement de Sap1 à *mat1* est retrouvé dans la séquence de Ter1 qui bloque la fourche à l'ADNr (Arcangioli et al., 1994b; Krings and Bastia, 2005, 2006). Bien que Sap1 soit recrutée au niveau de deux loci imposant des arrêts de fourche, son rôle est distinct puisqu'à contrario de *Ter1*, le site *SAS1* à *mat1* ne semble pas important pour l'activité de *MPS1*. En effet, des délétions comprenant se site n'affectent pas l'efficacité de la pause (*smt-0* et Δ*17*). Abp1 est également observée au niveau du locus *mat1* et son rôle au niveau de ce locus sera détaillé dans la partie résultat (Zaratiegui et al., 2011).

Ainsi, deux autres acteurs, Sap1 et Abp1, impliqués de façon plus ou moins directe dans le CTS, impactent la réplication d'autres régions génomiques. À *mat1*, bien que le site de fixation de Sap1 ne soit pas nécessaire à l'activité de *MPS1*, il est probable que son recrutement optimise le CTS. Quant à la protéine Abp1, son action sur la réplication à *mat1* n'a pas été étudiée.

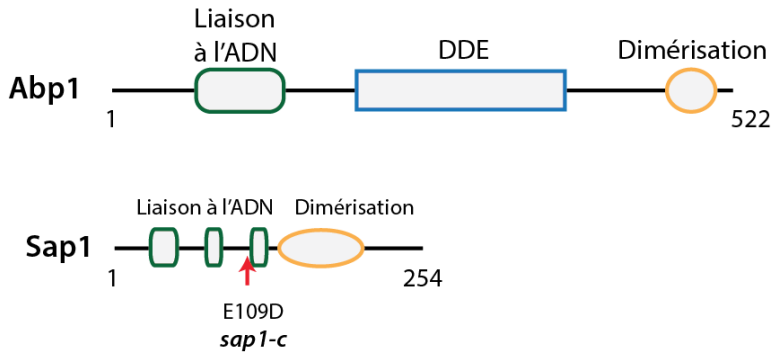


Figure 14 : Les domaines des protéines Abp1 et Sap1
Représentation schématique des protéines Abp1 et Sap1 et de leurs domaines. Les domaines de liaison à l'ADN sont représentés en vert et ceux de dimérisation en jaune. Le domaine transposase d'Abp1 est inactif et est dessiné en bleu. Les positions des résidus sont indiquées. La position de la mutation *sap1-c* est également annotée.

3.3.2 La collision réplication / transcription : un autre exemple d'arrêt de fourche

Souvent décrit comme "obstacles naturels" la collision entre la machinerie réplicative et transcriptionnelle entraine un stress réplicatif bloquant la fourche (Alzu et al., 2012; Azvolinsky et al., 2009; Deshpande and Newlon, 1996; Dutta et al., 2011; Merrikh et al., 2011). Plus particulièrement, chez *S. pombe* il a été montré qu'au niveau des loci sujets à cette collision, l'ARN polymérase II (Pol II) s'accumule. Cette accumulation de la Pol II combinée à l'arrivée de la machinerie de réplication peut résulter en des contraintes topologiques (Bermejo et al., 2012). Ces contraintes pourront entrainer la formation de structures impactant la stabilité génomique comme la régression de la fourche de réplication ou l'accumulation de R-Loop (hybride ADN:ARN). Le déplacement de la Pol II est donc requis pour le redémarrage de la réplication. Plusieurs mécanismes existent pour déplacer la Pol II entrainant la terminaison de la transcription. Par exemple, des signaux de séquence au niveau du Poly(A) ont été montrés comme participant à la terminaison (Birse et al., 1997). Par ailleurs, le facteur d'élongation de transcription TFIIS déclenche l'activité nucléase intrinsèque de la Pol II permettant la coupure de l'ARN messager (Izban and Luse, 1992; Reines, 1992; Wang and Hawley, 1993). D'autre part, l'étude de la Pol II a révélé que la phosphorylation du son domaine C-terminal permet le recrutement de facteurs de terminaison. Chez *S. pombe*, il a été montré récemment que la protéine Dicer permet le déplacement de la Pol II au niveau des gènes hautement transcrits sujets à la collision entre la machinerie de réplication et de transcription (Castel et al., 2014; Zaratiegui et al., 2011).

De façon générale, les gènes hautement transcrits seront donc des lieux où ce type de collisions sera le plus souvent observé. À ce jour, la régulation des fourches de réplication au niveau de ces gènes n'est que peu étayée. Il est possible qu'un mécanisme de contrôle des fourches soit requis à ces loci par analogie à ce qui est observé au niveau de l'ADNr. En effet, la direction de la fourche de réplication et de la transcription est la même au niveau de l'ADNr empêchant une collision. Ce phénomène a probablement été favorisé au cours de l'évolution pour empêcher de perturber la fourche de réplication qui pourrait réarranger la région et entrainer la perte des copies de l'ADNr.

3.4 Qu'advient-il d'une fourche bloquée ?

Plusieurs mécanismes permettent de redémarrer une fourche bloquée afin d'achever la réplication. Pour une vision exhaustive de ces mécanismes, se référer à : (Berti and Vindigni, 2016; Brewer, 1988; Deshpande and Newlon, 1996; Magdalou et al., 2014; Neelsen and Lopes, 2015).

Au niveau d'un locus répliqué bidirectionnellement, l'arrivée d'une fourche en sens opposé de celle bloquée permet de terminer la réplication. D'autre part, au niveau des sites de collision entre la réplication et la transcription, il a été montré que la Pol II doit être relâchée pour permettre à la fourche de réplication de poursuivre son avancée. La persistance de la liaison de la Pol II est un signal de dommage de l'ADN entrainant le recrutement de Rad52, une protéine importante pour la réparation de l'ADN via la recombinaison homologue (Castel et al., 2014; Svejstrup, 2010; Zaratiegui et al., 2011a).
Au niveau de loci répliqués unidirectionnellement comme l'ADNr ou *mat1*, aucune fourche en direction opposée ne peut terminer la réplication. Le mécanisme utilisé pour pallier la terminaison précoce de la réplication observée à l'ADNr est la recombinaison homologue (RH). Au niveau du locus sexuel en absence d'empreinte, la fourche bloquée à *MPS1* redémarre et permet le dépôt de l'empreinte sur le brin retardé néo-synthétisé (Holmes et al., 2005). Les acteurs impliqués dans ce redémarrage ne sont pas décrits à l'heure actuelle. Ce redémarrage ne nécessite pas les enzymes importantes pour la RH, suggérant que la pause n'est pas recombinogène. A contrario, en présence de l'empreinte, la fourche de réplication à *mat1* sera réparée par RH. Les détails de cette étape seront décrits dans la partie suivante.

4. L'empreinte et son potentiel recombinogène

L'empreinte créée à *mat1* en phase S est maintenue durant le cycle cellulaire. Cette caractéristique suggère que cette empreinte est protégée. La réplication du locus *mat1* portant l'empreinte entraine la formation d'une CDB polaire. Cette cassure permet l'initiation de la RH qui conduit au changement de l'allèle à *mat1*. Les éléments connus qui déstabilisent l'empreinte ainsi que le mécanisme de conversion génique seront détaillés dans cette partie.

4.1 La formation/maintenance de l'empreinte

4.1.1 Les acteurs formant et stabilisant l'empreinte

Précédemment, des mutants impactant la pause à *MPS1* ont été introduits. Pour résumer, ces mutants sont : les complexes protéiques Lsd1-Lsd2, Swi1-Swi3 ainsi que les séquences *SS2* et *mut3* (figure 11). Cette classe de mutants impacte également l'empreinte et est épistatique aux autres mutants du CTS. Différentes études ont permis de caractériser un groupe de mutant n'impactant que l'empreinte (Beach, 1983; Egel et al., 1984; Egel and Gutz, 1981; Engelke, 1987; Kaykov et al., 2004; Klar et al., 1991; Sayrac et al., 2011). À l'intérieur de ce groupe, je ne pourrai que très rarement faire la distinction entre les acteurs agissant sur la formation de l'empreinte de ceux agissant sur son maintien. En effet, la façon de tester un mutant affectant le niveau de l'empreinte consiste à faire un Southern-Blot et d'analyser le niveau de cassure sur une population en croissance. Afin de distinguer les acteurs impliqués dans la formation de l'empreinte de ceux impliqués dans sa maintenance, il faudrait par exemple suivre au cours du cycle cellulaire le niveau de l'empreinte. D'un côté, les mutants déficients pour la formation de l'empreinte devraient avoir un niveau d'empreinte constant et faible durant tout le cycle. D'un autre côté, les mutants impliqués dans le maintien de l'empreinte devraient avoir un niveau d'empreinte qui augmente en phase S, puis qui diminue. Ce type d'expérience n'a été réalisé qu'avec quelques-uns des mutants isolés. Je parlerai donc dans la plupart des cas de mutants impliqués dans la formation et/ou maintenance de l'empreinte.

Tout d'abord, une mutation dans le gène *swi7* (sous unité catalytique de la Pol α) abolit la formation/maintien de l'empreinte (Egel et al., 1984). Cet élément ajouté au caractère unidirectionnel de la réplication du locus soulève l'importance que le brin Watson soit le brin retardé pour l'établissement de l'empreinte. Le rôle de la sous-unité catalytique de la Pol α dans la formation de l'asymétrie n'a cependant pas été étayé. La cartographie des sites initiateurs de la réplication a montré que l'ARN synthétisé par la primase de la Pol α ne semble pas être à l'origine de l'empreinte. Deux études ont analysé l'extrémité 5' du dernier fragment d'Okazaki

dans la région de l'empreinte. La première a estimé la position du 5' à 342-345 pb de l'empreinte (Vengrova and Dalgaard, 2004) et la seconde l'a estimé à 30 pb (Sayrac et al., 2011). Les conclusions tirées sont que l'empreinte serait formée durant la polymérisation ou de façon post-réplicative.

À ce jour, le mutant de la sous-unité catalytique de la Pol α est le seul connu impactant la formation de l'empreinte sans influencer la pause de la fourche à *MPS1*. Au locus *mat1*, une autre protéine est recrutée du côté centromère distal : la protéine Sap1 (*Switch Activating Protein 1*) (figure 15). Le site de recrutement de Sap1, *SAS1*, est impliqué dans la maintenance de l'empreinte. Sap1 pourrait au travers de *SAS1* participer à la formation/maintenance de l'empreinte. La délétion de *sap1* étant létale, son rôle dans le CTS n'a pas été clairement déterminé. Comme décrit précédemment, Sap1 participe à la pause de la fourche de réplication au niveau des LTRs et de l'ADNr. Au contraire, la délétion ou la mutation de son site de recrutement n'altère pas l'activité de *MPS1* (*smt-0* et Δ*17*) à *mat1*. Aujourd'hui, nous ne savons toujours pas pourquoi l'activité de Sap1 à *mat1* diffère de celle observée aux LTRs ou à l'ADNr.

Par différentes techniques, des sites ont été décrits comme affectant le niveau de l'empreinte (Arcangioli and Klar, 1991; Engelke, 1987; Kaykov et al., 2004; Styrkarsdottir et al., 1993).

Nous pouvons extraire de ces études quatre éléments aidant à la compréhension de la formation et de la maintenance de l'empreinte : (figure 15)

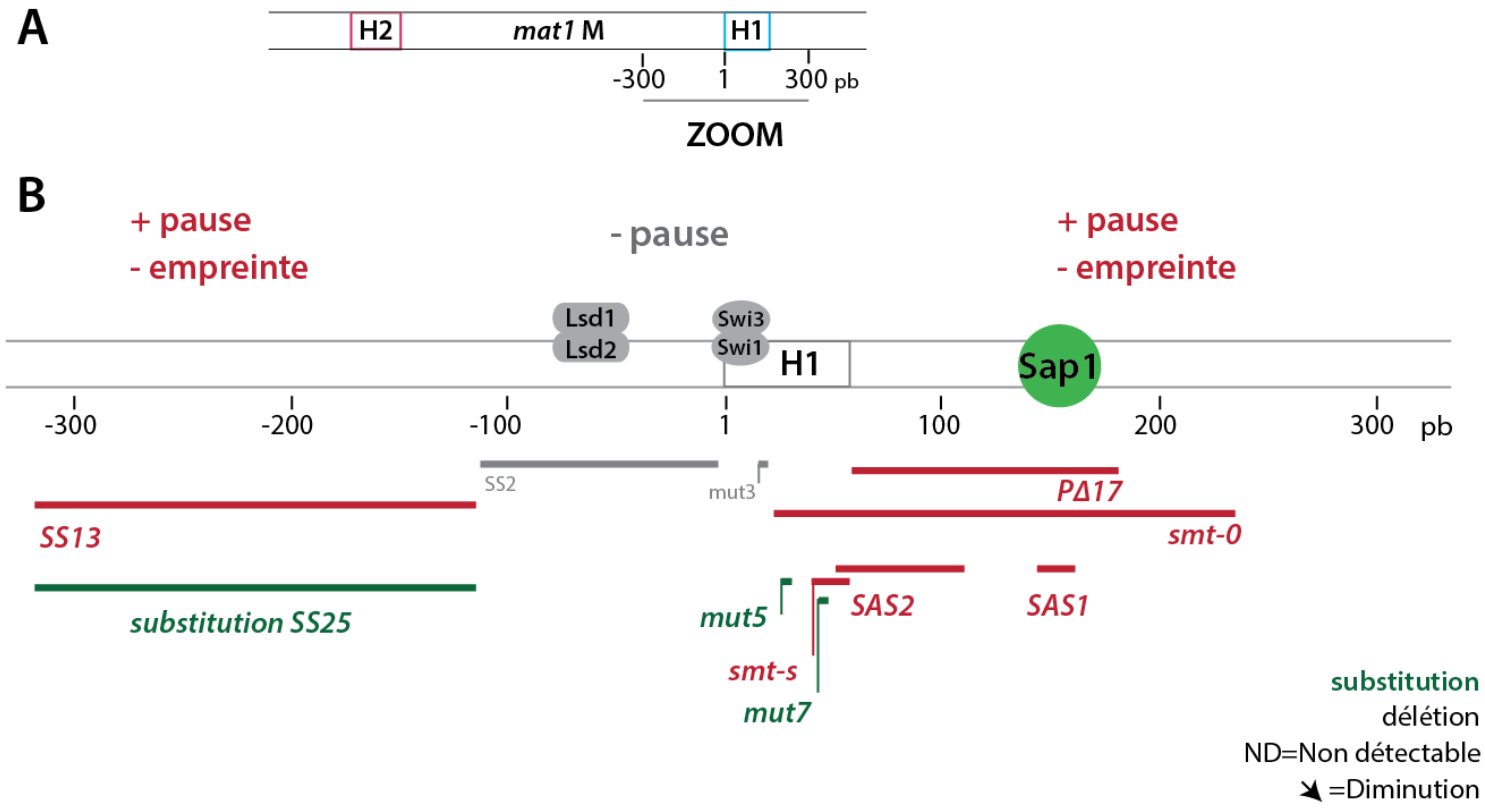
Premièrement, deux larges délétions du côté centromère distal de *mat1* inhibent la formation de l'empreinte et le CTS : *smt-0* et Δ*17*. Ces délétions ont été établies avec l'allèle *M* à *mat1* et avec l'allèle *P* à *mat1*, respectivement. Elles sont ainsi appelées *M smt-0* et *P* Δ*17*. Des éléments de contrôle de la formation de l'empreinte sont donc présents du côté centromère distal de l'empreinte, indépendamment de l'allèle présent à *mat1*.

Deuxièmement, des mutations/délétions plus courtes présentes dans la délétion *smt-0* (*mut5*, *mut7*, *smt-s*, *SAS1* et *SAS2*) diminuent le niveau de l'empreinte. Ces mutations sont décrites comme ayant un impact sur la stabilisation de l'empreinte. Plus particulièrement, quand la substitution *mut7* est présente, l'empreinte n'est pratiquement pas détectable durant le cycle cellulaire, excepté en phase S où son niveau est multiplié par deux (Kaykov et al., 2004). La substitution *mut7* consiste à l'échange de 5 pb situées à 55 pb du site de l'empreinte par le site de restriction *Pst1*.

Troisièmement, à l'intérieur de *mat1M*, la délétion *SS13* impacte la formation/maintenance de l'empreinte (figure 115). De façon intrigante, la substitution de la séquence *SS13* par une séquence sans homologie permet de récupérer 72 % du niveau de l'empreinte comparé à une

souche sauvage (figure 15 *SS25*). En absence de cette région, Sayrac et ses collaborateurs montrèrent qu'un site initiateur de la réplication normalement présent à 30 pb du côté centromère distal n'était plus utilisé (Sayrac et al., 2011). L'absence de cette séquence empêcherait le positionnement correct de l'initiation de la réplication et donc la formation/stabilité de l'empreinte.

Quatrièmement, nous pouvons noter que les séquences impliquées dans la formation/maintien sont situées de part et d'autre de celle importante pour la pause (figure 15). Les séquences les plus proches du site de l'empreinte contrôlent la réplication ainsi que l'empreinte alors que celles plus distantes contrôlent la formation/maintenance de l'empreinte. Des deux côtés du site de l'empreinte, des séquences participent à sa stabilisation, ce qui suggère qu'une structure particulière autour de celle-ci est formée. Cette structure pourrait participer à son maintien au cours du cycle cellulaire.



C

	SS13	SS25	SS2	mut3	smt-0	mut5	PΔ17	smt-s	mut7	SAS2	SAS1
Position	-318/-114	-318/-114	-111/-1	+16/+21	+23/+285	+36/+41	+59/+77	+51/+77	+56/+61	+60/+123	+140/+168
Taille	204	204	110	5	263	5	122	27	5	64	4 et 11
Pause MPS1	64%	80%	N D	N D	100%	100%		100%	100%		
Empreinte	N D	72%	N D	N D	N D	↘	N D	↘	↘	↘	↘
Recrutement			Lsd1-Lsd2	Swi1-Swi3							Sap1

Figure 15 : Les éléments contrôlant la formation/stabilisation de l'empreinte.
A) Schéma du locus *mat1M*. B) Zoom sur la région autour du site de l'empreinte. Les acteurs en *cis* et en *trans* impliqués dans l'activité de *MPS1* sont en gris. Les sites des séquences supprimées, importantes pour la formation et maintenance de l'empreinte, sont en rouge. Les sites des séquences substituées sont indiqués en vert. La protéine Sap1 et son site de recrutement sont annotés. C) Tableau résumant les phénotypes des différents sites. "ND" signifie non détectable et la flèche descendante signifie qu'il y a une diminution. Des cases ne sont pas remplies, car les données ne sont pas disponibles.

4.1.2 *MPS1* est actif avec et sans empreinte

Par Southern-Blot, 40 à 50 % des molécules à *mat1* cassent durant la préparation d'ADN. Ce pourcentage reflète globalement le pourcentage de cellules avec une empreinte. Une population de *S. pombe* sauvage est donc composée d'un mélange de cellules avec et sans empreinte en quantité équivalente. Ainsi, la fourche de réplication passant au locus sexuel rencontre dans environ 50 % des cas une empreinte. La découverte de mutants dans lesquels la pause est active, mais l'empreinte est absente indique que l'empreinte n'est pas à l'origine de l'activité de *MPS1*. Par exemple, dans le mutant *M smt-0*, la pause est effectuée, mais l'empreinte n'est pas détectable. La fourche fait une pause à *MPS1* indépendamment de la présence de l'empreinte.

4.2 Le résultat de la réplication à travers l'empreinte

Le modèle de la réplication du locus *mat1* portant l'empreinte peut être illustré comme le montre la figure 16. La fourche de réplication arrive au locus *mat1* qui porte l'empreinte et la convertit en CDB polaire, puisqu'une seule extrémité double brin est présente (figure 16). Cette CDB polaire, toxique pour la cellule, est réparée en utilisant l'un des deux donneurs silencieux comme matrice. Plusieurs observations valident ce modèle : La première est que lorsque des mutations sont introduites dans l'un des deux donneurs, elles se retrouvent au niveau du locus *mat1* (Beach, 1983; Egel and Gutz, 1981). La seconde est qu'un intermédiaire de réplication formé entre *mat1* et l'un ou l'autre des donneurs peut être capturé en phase S par PCR (Arcangioli and de Lahondes, 2000). La troisième est que les deux brins d'ADN du locus *mat1* qui change d'allèle sont néo-synthétisés (Arcangioli, 2000). À la différence du reste du génome, le locus *mat1* échappe donc à la réplication semi-conservative. D'autre part, bien que la réplication à travers l'empreinte entraine la formation d'une CDB polaire, elle permet à nouveau la formation de la marque chromosomique sur le brin retardé. Ce fait explique la "loi consécutive" qui montre que la cellule sœur de celle qui vient de changer de type sexuel a acquis le potentiel de changer de type sexuel à son tour.

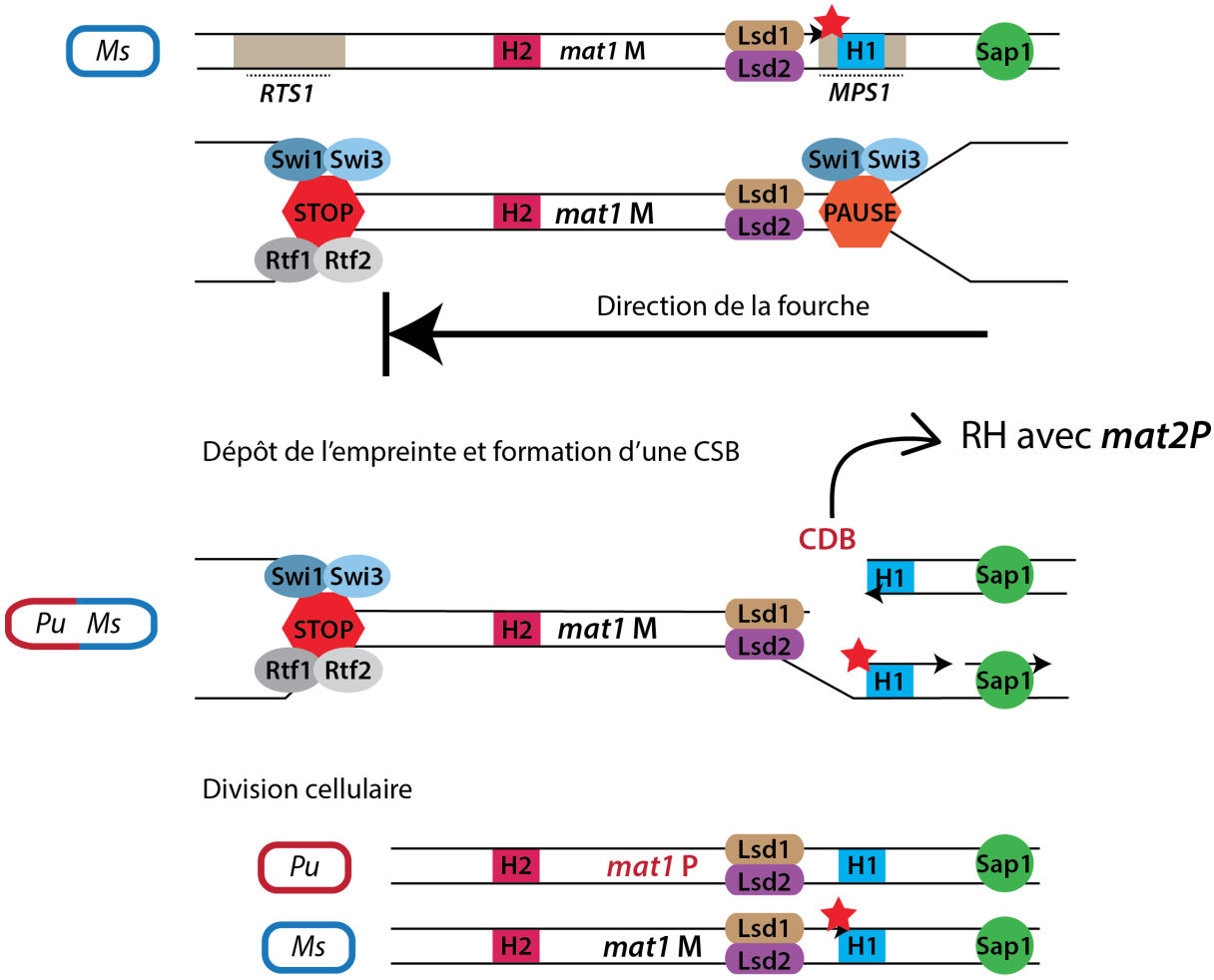


Figure 16 : L'empreinte est convertie en cassure double brin polaire par la fourche de réplication.
Schéma représentant la réplication du locus *mat1* portant l'empreinte. La fourche de réplication est unidirectionnelle. À son passage, la CDB est formée ainsi que l'empreinte sur le brin retardé. La CBD est réparée par RH avec l'un des donneurs. À la fin de la division cellulaire, deux cellules de type sexuel opposé sont formées. La cellule qui n'a pas changé de type sexuel porte l'empreinte.

4.3 La réparation de la CDB

4.3.1 Dans un contexte *h90*

Une autre classe de mutants affectant le CTS a été isolée lors du crible initial réalisé par Richard Egel en 1984 (Egel et al., 1984). Ces mutants ne présentent ni défaut de pause ni défaut d'empreinte, mais produisent des réarrangements de la région *mat* avec une fréquence plus élevée que le sauvage. Les gènes mutés codent pour des protéines impliquées dans l'étape de résolution de la conversion génique responsable du CTS.

Tout d'abord, la protéine Rad52 de *pombe* (synonyme de Rad22) est essentielle pour les souches *h90* (homothallique, changeant de type sexuel) et non pour les souches hétérothalliques (ne changeant plus de type sexuel) (Ostermann et al., 1993; Schmidt et al., 1989). Cette observation montre que l'empreinte est réparée via la RH, et plus particulièrement via Rad52. D'autre part, quatre protéines sont importantes pour la réparation de l'empreinte. Premièrement, le complexe Swi4-Swi8, (*MutS homologue 3* et *MutS homologue 2*) impliqué dans le mécanisme de réparation des mésappariements permet de stabiliser les intermédiaires de réplication formés lors de l'initiation et de la résolution de la RH (Fleck et al., 1994; Hanada et al., 2007; Roseaulin et al., 2008; Vagin et al., 2006).

Deuxièmement, le complexe Swi9-Swi10 (homologue des protéines mammifères XPF et ERCC1) participe aux étapes de clivage des intermédiaires de RH formés. Ce complexe appartient à la famille des endonucléase flap. Il clive le squelette de l'ADN à la jonction spécifique entre l'ADN double brin et l'ADN simple brin. Swi9-Swi10 participe à l'étape de résolution de la RH entre *mat1* et l'un des donneurs. Le modèle est basé sur le fait que le brin d'ADN à *mat1* envahit l'un des donneurs en utilisant la boite d'homologie H1. Ce brin est stabilisé par le complexe Swi4-Swi8. L'ADN est ensuite synthétisé jusqu'à la première homologie, à savoir la boite d'homologie H2. L'intermédiaire de recombinaison formé est stabilisé par Swi4-Swi8 et clivé par le complexe Swi9-Swi10, permettant l'invasion de la boite H2 de *mat1* (Roseaulin et al., 2008). Ce modèle permet d'expliquer la transposition d'un des allèles donneurs à *mat1*.

4.3.2 En absence des donneurs

En absence des donneurs, l'empreinte est formée et réparée par RH en utilisant la chromatide sœur (Klar and Miglio, 1986; Nielsen and Egel, 1989; Roseaulin et al., 2008). Cette observation permet notamment d'utiliser des souches présentant un type sexuel stable, mais sans défaut de pause ou d'empreinte. Dans ces fonds génétiques, les donneurs *mat2P* et *mat3M* ainsi que la région K ont été remplacés par le gène *LEU2* de *S. cerevisae* (Klar and Miglio, 1986). Ces souches sont appelées *mat1M* Δ*2-3* et *mat1P* Δ*2-3*. En absence des donneurs, il a été montré que la chromatide sœur est utilisée pour la réparation de la CDB. À la différence de ce qui est observé dans une souche *h90*, les endonucléases Mus81-Eme1 sont utilisées pour résoudre l'intermédiaire de RH au lieu du complexe Swi9-Swi10 (Hanada et al., 2007; Roseaulin et al., 2008). Dans un contexte sauvage, un mécanisme permet que la RH soit préférentiellement initiée au niveau du donneur portant l'information opposée à celle présente à *mat1*. Nous verrons par la suite que la structure chromatinienne de la région des donneurs participe à ce choix.

4.4 La chromatine dans la région *mat*

Bien que dans une cellule sauvage les deux allèles *M* et *P* soient présents au niveau des donneurs, seul l'allèle présent à *mat1* est exprimé (Egel, 1984b). Par ailleurs, un gène rapporteur inséré dans la région des donneurs n'est pas exprimé (Grewal and Klar, 1996; Thon and Klar, 1992). Les loci *mat2P* et *mat3M* sont séparés physiquement par 11 kbp, ce qui en moyenne, devrait correspondra à une distance génétique d'environ 3 cM (Nakaseko et al., 1986). Or, sur 17 000 clones sélectionnés, aucun ne présente de conversion génique entre des marqueurs adjacents à ces loci (Egel, 1984b). La région contenant les donneurs est donc un point froid de recombinaison. En utilisant ces caractéristiques, des acteurs en *cis* et en *trans* qui modulent le CTS ont été isolés. Ces acteurs ont ensuite été montrés comme participant de façon globale à la spécification de la chromatine chez *S. pombe*. Ces expériences ont déterminé le rôle de la structure chromatinienne dans le CTS.

4.4.1 L'hétérochromatine au niveau des donneurs

La nature de l'hétérochromatine au niveau des donneurs est similaire à celle retrouvée au niveau des télomères, des centromères et de l'ADNr (Grewal, 2000; Thon and Verhein-Hansen, 2000). Les modifications épigénétiques communes aux régions hétérochromatiques sont présentes au niveau des donneurs, notamment une hypoacétylation des lysines 9 et 14 de l'histone H3, une hypométhylation de la lysine 4 de l'histone H3, ainsi qu'une méthylation spécifique de la lysine 9 de l'histone H3 (Noma et al., 2001) (figure 17).

La formation de l'hétérochromatine chez *S. pombe* nécessite l'action de différentes enzymes. Cela inclut, des histones déacétylases (HDACs) comme Clr3, Clr6 et Sir2 ainsi que la méthyle transférase Clr4 qui méthyle H3K9. En outre, les protéines à chromodomaine Swi6, Chp2 (homologues de HP1) et Chp1 qui se lient spécifiquement à H3K9 di- ou triméthylée sont aussi nécessaires. Par ailleurs, la machinerie d'ARN interférence (ARNi) participe à la formation et à la maintenance de l'hétérochromatine. Les petits ARN (ARNi) sont présents dans le complexe RITS (*RNA-Induced Transcriptional Silencing*) avec les protéines de liaison à Argonaute. Ce complexe utilise la complémentarité des bases de l'ARNi pour guider son association à des ARN naissants. Cette association est stabilisée par la liaison de RITS à H3K9me. La présence de ces deux activités dans le complexe RITS, la reconnaissance des ARN naissants complémentaires et la liaison à H3K9me, permet la formation de l'hétérochromatine en collaboration avec les autres facteurs énoncés précédemment. Pour revue sur la formation de l'hétérochromatine chez *S. pombe* se référer à : (Allshire and Ekwall, 2015; Martienssen and

Moazed, 2015; Meller et al., 2015; Mizuguchi et al., 2015; Nikolov and Taddei, 2016) (figure 17).

De surcroit, cette identité chromatinienne est nécessaire au CTS puisque des mutants impactant la formation de cette structure diminuent également l'efficacité du CTS. Par exemple, une des premières mutations isolées affectant le CTS touche le gène *swi6* (Egel et al., 1984). Le gène *swi6* produit une protéine homologue à HP1 (*Heterochromatin Protein 1*) qui lie H3K9me et participe à la répression transcriptionnelle (Bannister et al., 2001; Lachner et al., 2001). Cette protéine est conservée de l'Homme à la levure. Par exemple, l'allèle *M31* de la souris peut remplacer l'allèle *swi6* de *S. pombe* (Eissenberg and Elgin, 2000; Ma et al., 2001; Wang et al., 2000). Chez *S. pombe* en absence de *swi6* les donneurs *mat2P* et *mat3M* sont exprimés et la recombinaison dans cette région n'est plus réprimée (Klar and Bonaduce, 1991; Lorentz et al., 1992; Lorentz et al., 1994; Nakayama et al., 2000). En outre, en absence des protéines Clr3, Clr6 et Sir2 (HDACs) les donneurs sont exprimés (Bjerling et al., 2002; Ekwall and Ruusala, 1994; Olsson et al., 1999; Shankaranarayana et al., 2003; Thon et al., 1994). Par ailleurs, la méthyle transférase Clr4 ainsi que la protéine Rik1 contrôlent la localisation de Swi6 et par conséquent la répression transcriptionnelle de cette région (Ekwall et al., 1996; Ekwall and Ruusala, 1994; Ivanova et al., 1998; Thon et al., 1994; Zhang et al., 2008a). Enfin, la machinerie d'ARNi participe aussi à formation de l'hétérochromatine au niveau des donneurs (Hall et al., 2002; Noma et al., 2004).

Sachant que la pause de la fourche de réplication à *MPS1* est contrôlée par des séquences internes à *mat1M* et *mat1P*, il serait attendu que la fourche pause également au niveau des deux donneurs. Or, aucune pause de la fourche de réplication n'est observée au niveau de *mat2P* et *mat3M* (Sayrac et al., 2011; Vengrova and Dalgaard, 2004). Cet élément suppose que soit l'hétérochromatine soit l'incapacité de recruter des acteurs protéiques réprime l'activité de pause de la fourche au niveau des donneurs. La réplication dans des mutants qui affectent la formation de l'hétérochromatine, par exemple Clr4 ou Swi6, n'a pas été étudiée. D'autre part, il est aussi possible que la direction de la fourche de réplication au niveau des donneurs ne permette pas la formation de la pause.

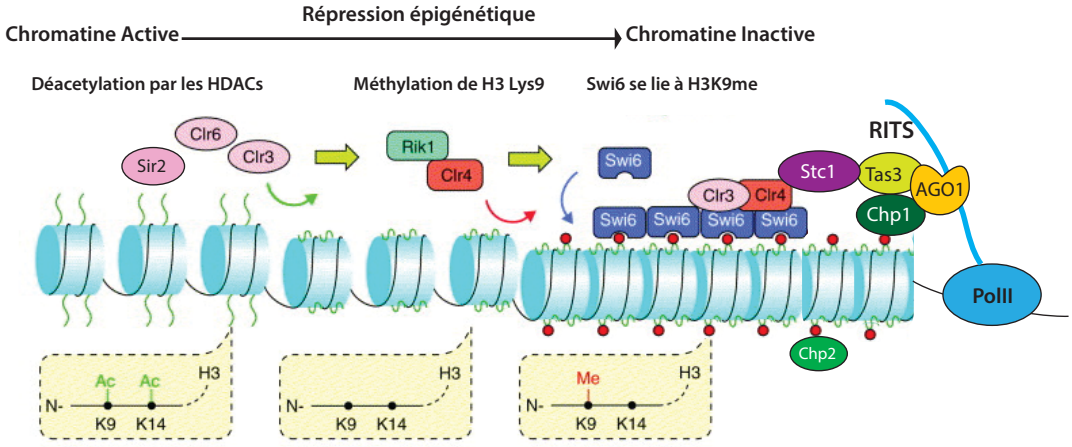


Figure 17 : L'assemblage de l'hétérochromatine chez *S. pombe.*
Schéma adapté de (Grewal and Elgin, 2002) et de (Allshire and Ekwall, 2015). Les protéines Clr3, Clr6 et Sir2 participent à la déacétylation de H3K9 et H3K4. La protéine Clr4 en complexe avec Rik1 permet de méthyler H3K9. Les protéines Swi6 et Chp2 sont recrutées au niveau de H3H9me. Le complexe RITS se lie via AGO1 (Argonaute) à l'ARN naissant synthétisé par la Pol II participant ainsi à la formation de l'hétérochromatine.

4.5.2 Les sites *IR-L* et *IR-R* ainsi que *CENH* participent à la formation de l'hétérochromatine

Deux questions sous-jacentes dans le domaine de l'épigénétique ont été adressées au travers de la structure de la région *mat*. La première étant de définir le site et les acteurs responsables de la formation de l'hétérochromatine. La seconde étant de mettre en évidence les éléments bloquant la propagation de l'hétérochromatine pour qu'elle n'affecte pas l'expression de loci adjacents.

Tout d'abord, la région K séparant les loci *mat2P* et *mat3M* contient le locus *CENH* qui est composé de séquences homologues à celle des centromères (Grewal and Klar, 1997). Cette région insérée de façon ectopique dans le génome de *S. pombe* permet la nucléation de l'hétérochromatine (Hall et al., 2002). De surcroit, une mutation du gène *swi6* est épistatique à la délétion de la région K (Grewal and Klar, 1997; Thon and Friis, 1997). Plusieurs laboratoires ont ensuite montré que *CENH* promeut la formation de l'hétérochromatine au travers de la machinerie ARNi et via des protéines de la famille ATF/CREB (Hall et al., 2002; Jia et al., 2004a; Kim et al., 2004; Noma et al., 2004). *CENH* potentialise le recrutement des acteurs hétérochromatiques décrits précédemment (figure 17 et 18). Par ailleurs, deux autres éléments de séquence participent à la formation de l'hétérochromatine : *REII* et *REM3* (Ekwall et al., 1991; Thon et al., 1994; Ayoub et al., 1999; Ayoub et al., 2000). Ces sites sont situés

respectivement du côté centromère proximal de *mat2P* et de *mat3M* (figure 18). La délétion de ces sites perturbe la formation de l'hétérochromatine (Ekwall et al., 1991; Thon et al., 1994; Ayoub et al., 1999; Ayoub et al., 2000).

Enfin, une caractéristique intéressante de cette région est la présence de deux séquences répétées inversées entourant les loci donneurs : *IR-L* et *IR-R* (Noma et al., 2001). Une des observations les plus parlantes est le fait que la nature de la chromatine change brutalement à la bordure de ces régions. En effet, entre ces éléments IR, H3K9 est méthylée et H3K4,14 sont hypoacétylées alors qu'à l'extérieur de cette région, l'inverse est observé (Noma et al., 2001). Ces répétitions marquent les limites de la région réprimée. Par la suite, d'autres études ont montré que la délétion de ces répétitions entraine une perturbation de l'hétérochromatine (Thon et al., 2002) et que cette délétion combinée à la surexpression de la protéine Swi6 entraine une propagation de l'hétérochromatine (Noma et al., 2006; Noma, 2001). En outre, les facteurs décrits comme participant aux barrières hétérochromatiques sont recrutés au niveau de ces éléments. Cela inclut, le facteur TFIIIC (facteur de transcription), la protéine Bdf2 (*BET family double bromodomain protein* 2, homologue de BRD3 et BRD4 chez l'homme) via Epe1/Kdm2 (*Jmjc domain chromatin associated protein,* homologue des protéines de la famille des lysines déméthylases chez l'homme) ainsi que les deux composants du complexe de protection des télomères, Taz1 (*human TRF ortholog*) et Raf1 (*Rik1-associated factor*) (Noma et al., 2006; Toteva et al., 2017; Wang et al., 2013). De surcroit, le complexe Lsd1-Lsd2 agissant au niveau de la réplication à *mat1* a aussi été montré comme participant à l'établissement de ces barrières au niveau de la région des donneurs (Lan et al., 2007) (figure 18). Enfin, il a été montré que la localisation au niveau de la membrane nucléaire de la région *mat* dépend des éléments *IR-L* et *IR-R* combinés à l'expression de la protéine Clr4 (Alfredsson-Timmins et al., 2007; Jakociunas et al., 2013a). Ces observations, suggèrent que ces deux séquences participent au maintien du passage d'une région hétérochromatique à euchromatique.

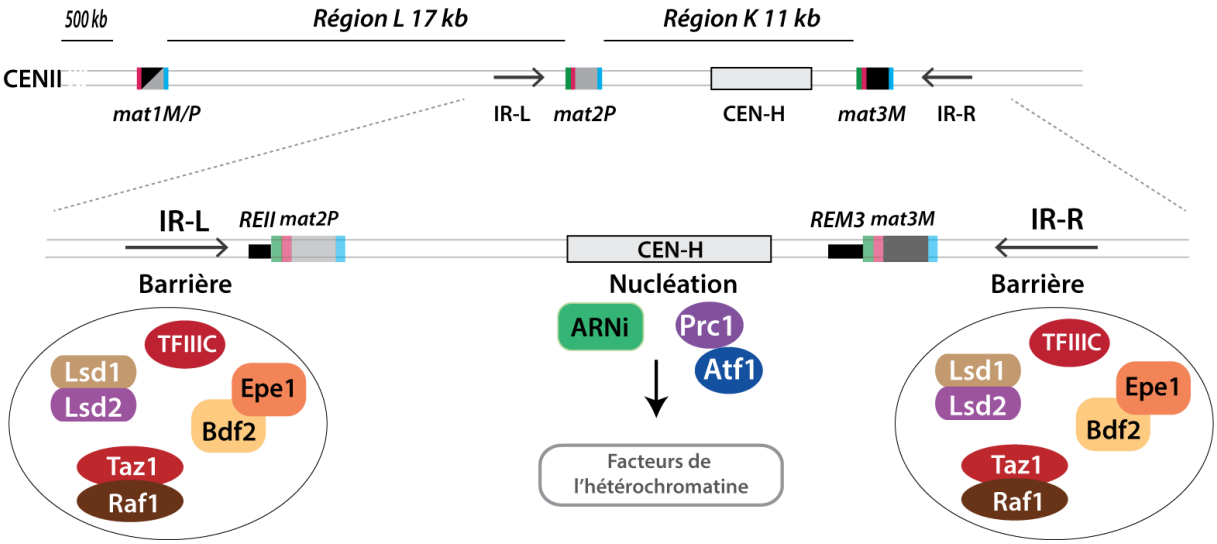


Figure 18 : La nucléation et les barrières de la chromatine dans la région *mat.*
La région *mat* est schématisée en haut de la figure. Un agrandissement sur la région des donneurs est représenté dans le panneau du milieu. Les acteurs décrits dans le texte comme participant à la nucléation et la délimitation de l'hétérochromatine sont représentés.

4.5 Le choix du donneur

Les expériences de pédigrée réalisées chez *S. pombe* ont révélé que dans 90 % des cas, la cellule choisit le donneur portant l'allèle opposé à *mat1* pour réparer la CDB polaire (Miyata and Miyata, 1981; Egel and Eie, 1987; Klar, 1990). Le mécanisme contrôlant cette décision non aléatoire est appelé "la directionnalité" (Thon and Klar, 1993). Dans des mutants qui affectent la directionnalité, ni la pause de la réplication à *MPS1*, ni le niveau de l'empreinte, ni la réparation ne seront affectés. La première expérience clef qui a permis d'envisager un modèle expliquant ce mécanisme est l'étude de l'efficacité du CTS dans une souche contenant une inversion des cassettes donneuses. L'allèle *mat3M* est placé à la place de l'allèle *mat2P* et inversement. Dans ce contexte, l'efficacité du CTS est réduite sans que le niveau de l'empreinte ne soit affecté. Ce phénotype montre, d'une part, que la localisation plutôt que le contenu des cassettes donneuses dirige la réparation, et, d'autre part, que l'allèle présent à *mat1* détermine ce choix (Thon and Klar, 1993). Le second résultat important est que la présence de mutations dans le gène *swi6* ou *clr4* entraine un déséquilibre entre les deux types sexuels dans une colonie (Thon and Klar, 1993). Les colonies *swi6*Δ et *clr4*Δ sont en majorité composées de cellules de type sexuel *M*. Ces résultats suggèrent que le type sexuel de la cellule ainsi que des acteurs protéiques en *trans* contrôlent la directionnalité.

4.5.1 La nature chromatinienne participe à la directionnalité

Initialement, trois mutations qui conduisent à un phénotype attendu s'il affecte la directionnalité ont été isolées : *swi6*, *swi2* et *swi5* (*switch gene 6*, *2* et *5*) (Egel et al., 1984). L'implication de la protéine Swi6 (homologue de la protéine HP1) dans ce mécanisme suppose que l'hétérochromatine participe à la directionnalité. D'autres éléments ont permis d'étayer cette hypothèse. Par exemple, la disruption de la région K, contenant *CENH* entraine une utilisation préférentielle de *mat3M* lors du CTS. Les colonies contenant cette délétion sont composées en majorité de cellules de type sexuel *M* (Grewal and Klar, 1997). Par ailleurs, les protéines Clr4, Rik1, Clr7 et Clr8 participant à la méthylation de H3K9 et à la localisation de Swi6 ont aussi été décrites comme participant à la directionnalité (Ivanova et al., 1998; Nakayama et al., 2001; Thon et al., 2005; Tuzon et al., 2004).

Enfin, un autre élément est à souligner, à savoir que la structure de l'hétérochromatine de la région du donneur varie en fonction du type sexuel. En particulier, des gènes rapporteurs introduits dans cette région sont plus fortement réprimés dans une cellule de type *M* que *P* (Ayoub et al., 1999; Thon et al., 1999). En outre, la protéine Swi6 est enrichie dans le type cellulaire M (Noma et al., 2001). Ces observations suggèrent que le type sexuel participe à la structure de la chromatine dans la région des donneurs et par ce fait guide la RH.

4.5.2 Le complexe Swi5-Swi2 contrôle le choix du donneur

Un deuxième type de protéine impactant la directionnalité a aussi été décrit : le complexe Swi2-Swi5. Le phénotype de mutants des gènes *swi2* et *swi5* ressemble à celui du gène *swi6,* suggérant que ce complexe joue un rôle dans la directionnalité au travers de l'hétérochromatine. Premièrement, la protéine Swi5 peut former deux complexes différents : le premier est composé de Swi5, Sfr1 (*Swi five-dependent recombination mediator*) et Rad51 et le deuxième de Swi5, Swi2 et Rad51 (Akamatsu et al., 2003). Le premier complexe contenant Sfr1 participe de façon générale à la RH, alors que le deuxième participe spécifiquement au CTS (Akamatsu et al., 2007; Haruta et al., 2008; Jia et al., 2004).

Trois études montrent que le complexe Swi2-Swi5 est recruté au niveau des donneurs en fonction du type sexuel de la cellule (Aguilar-Arnal et al., 2008; Jia et al., 2004; Matsuda et al., 2011). Ce complexe est positionné au niveau de *mat3M* constitutivement, mais se lie à *mat2P* seulement dans les cellules *M* (Aguilar-Arnal et al., 2008; Jia et al., 2004b; Matsuda et al., 2011). Ce recrutement spécifique est dépendant de la protéine Swi6 ainsi que d'un élément de séquence *SRE3* situé du côté centromère distal de *mat3M* (Jia et al., 2004) (figure 19). En outre, la délétion de *swi2* ou de *swi5* entraine un déséquilibre entre les deux types sexuels (Aguilar-

Arnal et al., 2008; Jakociunas et al., 2013b; Matsuda et al., 2011). L'étude de plusieurs colonies indépendantes a montré qu'en absence de *swi2* les colonies sont soit biaisées vers *M* soit biaisées vers *P*. Tandis qu'en absence de *swi5*, les colonies sont en majorité biaisées vers *M* (Jakociunas et al., 2013). Ces observations supposent que la répartition de ce complexe sur l'hétérochromatine participe au mécanisme de directionnalité comme le décrit la figure 19. Une des questions restant en suspens est de déterminer le mécanisme moléculaire contrôlant la répartition spécifique de ce complexe. Un des éléments de réponse est que la liaison de Swi2-Swi5 au niveau de *mat2P* est dépendante du facteur Mc produit dans des cellules de type *M* et de la protéine Abp1 (Aguilar-Arnal et al., 2008; Matsuda et al., 2011). Finalement, il a été montré que le facteur Mc, en collaboration avec la protéine Abp1, permet l'expression d'un transcrit alternatif du gène *swi2* (Yu et al., 2012). Il a été supposé que ce transcrit forme une protéine courte de Swi2 (Swi2S) qui en complexe avec la protéine Swi5 serait présente au niveau de *mat2P* pour favoriser le CTS avec celui-ci (Yu et al., 2012) (figure 19). Le recrutement de ce complexe spécifiquement au locus *mat2P* n'a cependant pas encore été étudié. La technique utilisée pour localiser Swi2 ne permet pas de discriminer la forme de Swi2.

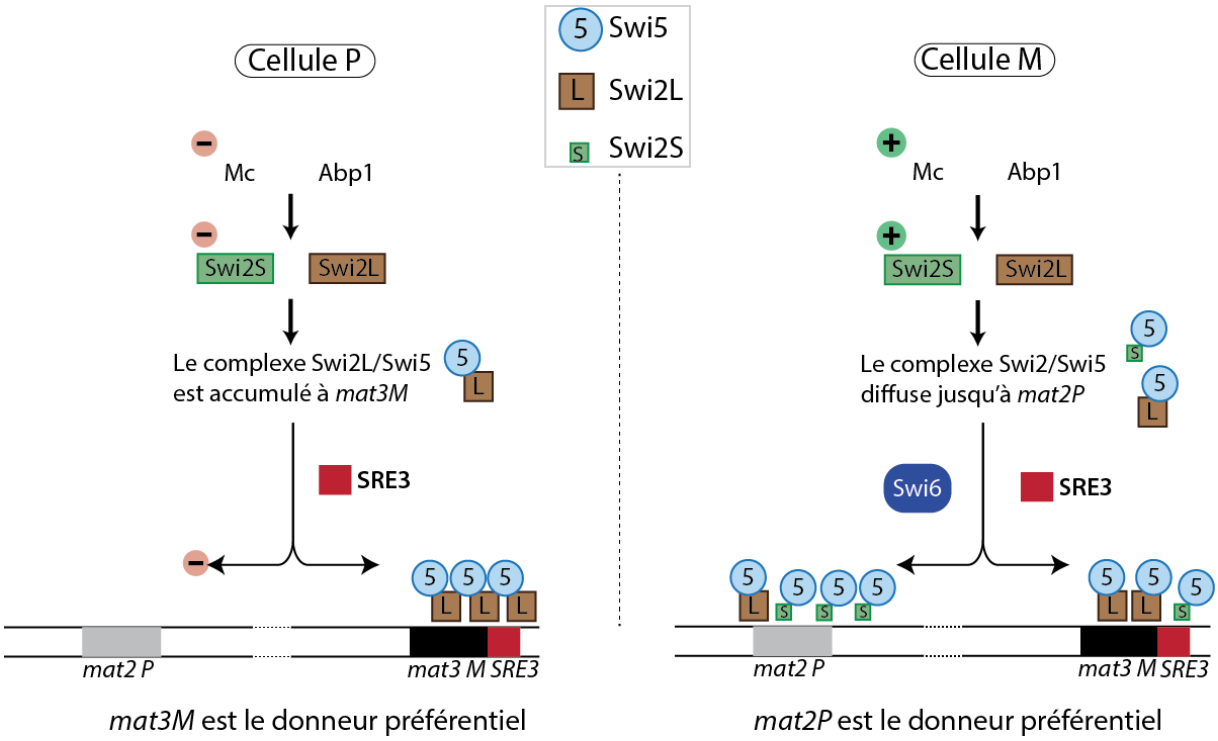


Figure 19 : Modèle de la répartition du complexe Swi2-Swi5 en fonction du type sexuel. Modifié de (Yu et al., 2012a). Le panneau de gauche schématise ce qui se passerait dans une cellule *P*. Seul le complexe Swi5-Swi2L (Large) est présent au niveau de *mat3M* via *SRE3*. Le panneau de droite représente ce qui se passerait dans une cellule *M*. Le facteur Mc ainsi que la protéine Abp1 permettrait l'expression de la protéine Swi2S (Small). Les complexes Swi5-Swi2S et Swi5-Swi2L diffuseraient jusqu'au donneur *mat2P* permettant qu'il soit favorisé pour la recombinaison.

4.5.3 SRE2 et SRE3, des éléments de séquences contrôlant la directionnalité

Deux séquences qui participent au choix du donneur ont été caractérisées finement : *SRE2* et *SRE3* (Jakociunas et al., 2013b; Jia et al., 2004; Yu et al., 2012). Ces deux éléments sont situés du côté centromère distal de *mat2P* et de *mat3M*, respectivement (figure 20). Un des résultats clefs montre qu'une souche présentant une inversion des cassettes donneuses combinée avec une inversion des éléments *SRE2* et *SRE3* change de type sexuel de façon sauvage (Jakociunas et al., 2013). De plus, la délétion de *SRE2* résulte en des colonies de type sexuel *M* et la délétion de *SRE3* entraine la formation de colonies de type sexuel *P* (Jakociunas et al., 2013). Ainsi, *SRE2* et *SRE3* en collaboration avec Swi2 et Swi6, plus abondantes dans une cellule de type sexuel M, contrôlent la directionnalité chez *S. pombe* (figure 20).

L'action exacte des différents acteurs protéiques et génétiques décrits reste encore à être précisée. Cela permettra de proposer un modèle qui intègre tous les résultats générés jusqu'alors.

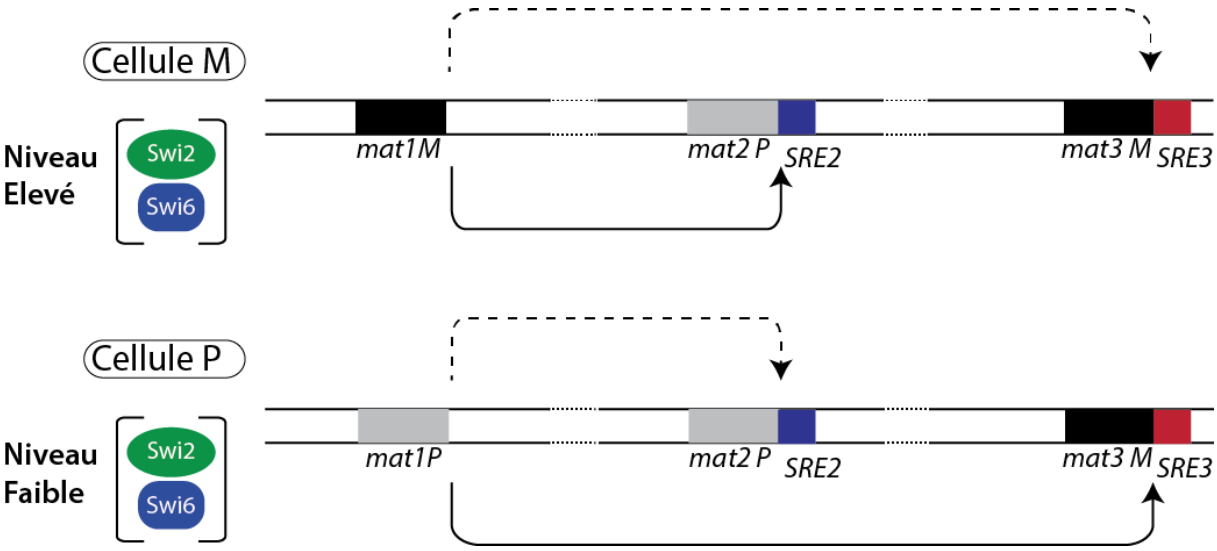


Figure 20 : Les éléments *SRE2* et *SRE3*.
Tiré de (Jakociunas et al., 2013b). Modèle de l'action des éléments *SRE2* et *SRE3* en fonction du type sexuel de la cellule. Dans une cellule *M*, les protéines Swi6 et Swi2 favoriseraient l'utilisation de *SRE2* comme promoteur de la recombinaison. Dans les cellules *P*, Swi2 et Swi6 étant moins abondantes, *SRE3* serait utilisé comme promoteur de recombinaison.

5. Le projet : la formation et le devenir de l'empreinte chez la levure *S. pombe*

Les différentes étapes du mécanisme de CTS ont été étudiées depuis de nombreuses années, mais les éléments responsables de la formation et du maintien de l'empreinte restent encore peu connus. Le contrôle de la formation de l'empreinte à *mat1* est un modèle d'étude pour disséquer les mécanismes qui participent à la stabilité génétique. Plus particulièrement, l'analyse des éléments qui contrôlent la pause programmée à *MPS1* puis la formation d'une empreinte recombinogène permettra de mieux comprendre les relations qui existent entre la réplication et l'épigénétique. A l'instar de ce qu'a apporté l'étude de la fourche de réplication au niveau de l'ADNr ou au niveau du site *RTS1*, l'analyse de l'activité de *MPS1* a mis en évidence de nouveaux acteurs qui influencent la progression de la fourche, comme les protéines Lsd1 et Lsd2.

Les protéines qui influencent les étapes initiatrices du CTS sont multiples. Cette pluralité suggère que des mécanismes distincts contrôlent cette division asymétrique. Par exemple, une des protéines nécessaires à la RH, Rad52, interagit avec le locus sexuel et est nécessaire ou CTS. D'autre part, Lsd1-Lsd2, des protéines qui modifient les histones, contrôlent l'activité de *MPS1*. La question est de comprendre comment l'empreinte chromosomique à *mat1* est formée et est maintenue.

Mon projet a été de déterminer la localisation des protéines Lsd1, Lsd2, Sap1 et Abp1 au niveau de la région *mat*. En outre, j'ai caractérisé finement les sites de recrutement des protéines Lsd1 et Sap1. Par ailleurs, j'ai également réalisé la première analyse du rôle de la protéine Abp1 au niveau du locus sexuel.
L'analyse à haut débit des sites préférentiellement enrichis en Lsd1, Lsd2, Sap1 et Abp1 m'ont permis de suggérer un rôle plus général de ces protéines sur le contrôle de la réplication.
De plus, en cartographiant l'empreinte par séquençage à haut débit, j'ai pu analyser la structure chromatinienne formée au locus sexuel.
Enfin, j'ai également participé à un projet visant à identifier l'implication de la protéine Lsd1 dans la stabilité génétique. J'ai analysé de nouveaux sites de pause pour la fourche de réplication potentiellement contrôlés par la protéine Lsd1.

Matériels et Méthodes

1. Comment tester l'efficacité des différentes étapes du CTS

Comme décrit précédemment le mécanisme de CTS comporte trois étapes majeures qui sont la pause de la fourche de réplication à *MPS1* (*Mat1 Pause Site 1*), la formation/maintenance de l'empreinte et le remplacement de l'allèle à *mat1* par conversion génique. Durant ma thèse, j'ai analysé l'efficacité du CTS en testant l'efficacité de ces étapes dans différents fonds génétiques. Les techniques utilisées, ont plus la plupart, été mises au point par plusieurs laboratoires travaillant dans le domaine au cours des trente dernières années. Je ne présenterai ici que les étapes spécifiques à cette étude.

Les gels bidimensionnels d'ADN permettent de mesurer l'efficacité du site *MPS1*. L'empreinte est, quant à elle, analysée par Southern Blot. Enfin, l'équilibre dans la population des deux types sexuels est analysé par PCR quantitative (qPCR).

1.1 Quantification de l'efficacité d'une pause de la fourche de réplication : les Gels 2D

La réplication du locus *mat1* est à l'origine du CTS puisque c'est au redémarrage de la fourche de réplication en pause au niveau du site *MPS1* que l'empreinte est déposée (Dalgaard and Klar, 1999, 2000, 2001; Kaykov et al., 2004). J'ai analysé les intermédiaires de réplications (IR) au niveau de *MPS1* par la technique de gel 2-D (Brewer and Fangman, 1987, 1988; Kaykov et al., 2004). Sachant que seulement 20 % des cellules sont en réplication dans une culture en phase exponentielle, une partie minime des cellules répliquera le locus *mat1* à un temps t. De plus, les structures d'ADN formées au cours de cette étape du cycle sont fragiles et facilement perdues pendant la préparation. Le laboratoire a mis au point un protocole qui permet de préserver et de sélectionner les IR.

Une pré-culture en milieu riche (YES) à 33 °C de *S. pombe* est réalisée afin d'obtenir 250 mL d'une culture cellulaire en phase exponentielle à une densité comprise entre 5.10^6 et 1.10^7 cellules par mL. Rapidement, la réplication est bloquée en ajoutant du l'azide de sodium à une concentration finale (Cf) de 0,1 % et un glaçon de 40 mL d'EDTA (pH 8, Cf 0,04 M) à la culture. Cette étape est réalisée à froid (4 °C). En vue de préserver les IR, l'ADN est préparé et digéré enzymatiquement en plug d'agarose (Arcangioli, 1998; Schwartz and Cantor, 1984). 2.10^8 cellules sont utilisées par plug d'agarose de 150 µL. Pour un gel 2-D, 10 plugs d'agarose (2.10^9 cellules) sont utilisés. Les plugs contenant les cellules subissent différents traitements qui permettent la purification de l'ADN (Brewer and Fangman, 1988; Kaykov et al., 2004). Une fois l'ADN purifié, il est digéré enzymatiquement en plug. Les enzymes de restrictions sont

choisies en fonction de leur capacité à digérer l'ADN dans ces conditions. Je me réfère aux recommandations du site du fournisseur NEB suivant :
https://www.neb.com/tools-and-resources/usage-guidelines/digestion-of-agarose-embedded-dna-info-for-specific-enzymes
Pour l'analyse de *MPS1*, nous utilisons l'enzyme NdeI qui génère un fragment de 3,1 kb.
L'ADN digéré est moins fragile et peut être extrait des plugs. L'agarose est dissous pendant 10 min à 70 °C puis digéré avec 2 unités/plug de ß-agarase I. Les ARN sont digérés pendant cette étape avec 20 µg/plug de RNase A, pendant 1 h à 37 °C.
Le surnageant est collecté puis dilué 2 fois dans le même tampon que celui utilisé pour l'équilibrage de la colonne de BND-cellulose (10 mM Tris-HCL pH 7,5 / 1 mM EDTA pH 8 / 300 mM NaCl). Le surnageant est ensuite doucement chargé sur la colonne de BND-cellulose. Cette étape permet l'enrichissement en IR. En effet, la BND-cellulose possède de l'affinité pour l'ADN simple brin. Les IR de réplication seront plus favorablement liés à la colonne que les ADN double brin. Pour décrocher les IR, la caféine, un analogue des bases purines, est utilisée comme compétiteur (Ebrahimi et al., 2013).
Les colonnes de BND-cellulose (Benzoylated Naphthoylated DEA-cellulose, Sigma) sont préparées de la façon suivante :
Pour une colonne, 2 g de BND-cellulose sont lavés dans 10 mL d'un tampon hautement concentré en sel afin d'éliminer les impuretés (10 mM Tris-HCL pH 7,5 / 1 mM EDTA pH 8 / 5 M NaCl). Ensuite, une centrifugation de 2 min à 3000 rpm est effectuée. La phase flottante de BND est éliminée avec la totalité du tampon.
Enfin, la colonne est équilibrée toute la nuit dans le tampon suivant 10 mM Tris-HCL pH 7,5 / 1 mM EDTA pH 8 / 300 mM NaCl. La BND-cellulose est ensuite chargée sur une colonne de chromatographie dans le but d'avoir 1 mL de BND-cellulose compacte utilisable.
L'ADN est chargé sur la colonne. Puis, la colonne est lavée avec 3 mL d'un tampon ayant une concentration plus élevée en sel (10 mM Tris-HCL pH 7,5 / 1 mM EDTA pH 8 / 800 mM NaCl) cette étape permet d'éliminer l'ADN double brin. Les IR sont ensuite élués avec un tampon à base de caféine (10 mM Tris-HCL pH 7,5 / 1 mM EDTA pH 8 / 1 M NaCl / 1,8% Caféine (sigma C-8960)). Des fractions de 500 µL sont récupérées.

Chaque fraction est analysée sur gel d'agarose pour déterminer lesquelles seront utilisées pour le gel 2D. 10 µL de chacune des fractions sont chargés sur un gel d'agarose de 0,8 % et migrés pendant 20 min à 100 V (figure 1).

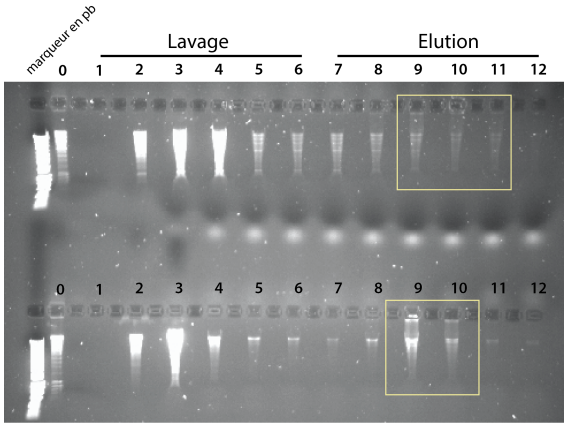


Figure 1 : Photographie du gel après migration des fractions.
Les résultats de 2 colonnes de purification sont représentés. La piste 0 représente l'ADN avant le passage sur colonne. Les pistes 1 à 6 représentent les fractions de lavage et celles de 7 à 12 les fractions d'élution. Les carrés jaunes représentent les fractions récupérées.

Les IR n'ont pas les mêmes propriétés de migration que l'ADN double brin. Comme le montre la figure 1, les fractions contenant les IR présentent une trainée qui s'étend du puits jusqu'à la première bande la plus intense. Nous remarquons que la plupart du matériel se retrouve dans la fraction de lavage, d'où la nécessité d'enrichir en IR.

Les fractions d'intérêt sont rassemblées, puis l'ADN est précipité et repris dans 40 µL de tampon de migration (10 mM Tris-HCL pH 7,5 / 0,1 mM EDTA pH 8 / 10 mM NaCl).

Les 40 µL d'ADN sont déposés et séparés pendant 18 h à 1 V/cm à température ambiante dans un gel d'agarose de 300 mL à 0,4%. La première dimension permet de séparer l'ADN en fonction de sa taille et non de sa structure.

Pour la seconde dimension, une bande du premier gel d'agarose contenant les fragments d'intérêt est découpée puis placée orthogonalement sur un nouveau support. Pour connaitre la taille du bloc d'agarose à découper, la piste du marqueur de poids moléculaire est découpée, colorée au Bromure d'Ethydium (BET) puis photographiée sous UVs. La taille du bloc d'ADN dépend du fragment de restriction à analyser. Ici, l'enzyme NdeI donne un fragment de restriction d'intérêt de 3,1 kb. Ce fragment en fin de réplication aura une taille maximale de 6,2 kb. Nous découpons donc la piste comprise entre 1,5 kb et 9 kb. Enfin, un gel d'agarose à 1,1 % avec 0,3 µg/mL de BET préalablement refroidit à 45 °C est coulé sur la bande du premier gel. Les ADN sont ensuite séparés pendant 5 h à 5 V/cm en chambre froide. Cette migration permet de dissocier l'ADN en fonction de sa structure.

Le gel est ensuite traité comme un Southern-Blot en utilisant comme protocole celui décrit par le volume 2 du livre "Molecular Cloning a laboratory manual" (Sambrook *et al,* 1989). Le transfert est effectué par capillarité.

La membrane obtenue est hybridée avec une sonde radio-marquée par le fragment de Klenow de l'ADN polymérase I, en utilisant de l'ATP marqué en α-P32. La sonde utilisée est obtenue par PCR sur de l'ADN génomique d'une souche de *S. pombe h90* (préparé par une méthode classique) en utilisant les oligonucléotides suivant : 5' TTCGGTATTTAAGTCTGGCG 3' et 5' CCAATTATGCTGTTCGTGTC 3'.

La figure 2 présente le résultat schématique d'un gel 2-D. Un signal fort sur l'arc représente une accumulation d'IR. Cela signifie que dans une population cellulaire, une partie des cellules auront une fourche bloquée à ce site particulier. Ainsi, sur l'arc de réplication ce signal résultant de cette accumulation sera quantifié pour déterminer l'efficacité du site de pause.

Dans une autre partie de ma thèse, j'ai aussi analysé la réplication du locus *wtf9*. Pour ce faire, j'ai utilisé le même protocole que celui décrit précédemment en changeant les enzymes de restriction et la sonde. Les différentes enzymes utilisées seront indiquées dans la partie d'intérêt. La sonde utilisée est obtenue par PCR sur de l'ADN génomique d'une souche de *S. pombe h90* (préparé par une méthode classique) en utilisant les oligonucléotides suivant : 5' GTGACTGAGTAAATAGATCG 3' et 5' TTATTCTATGGACAGCAGTG 3'.

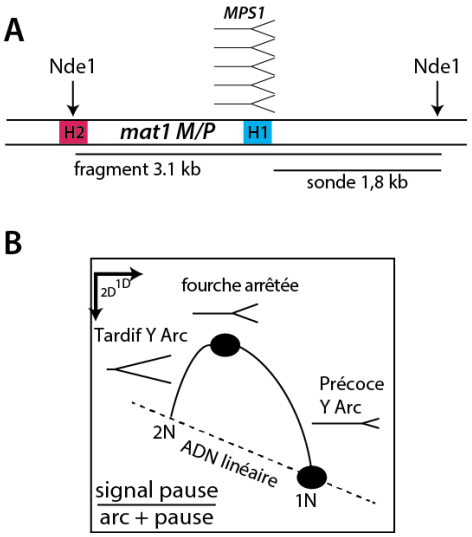


Figure 2 : Résultat schématique d'un gel 2-D.
A) L'enzyme de restriction ainsi que la taille du fragment obtenu sont indiqués. La taille et la position de la sonde sont indiqués. B) Schéma représentant une pause de la fourche de réplication après analyse par gel 2-D. La fourche en pause est reconnaissable car les IR s'accumule en un point au niveau de l'arc. La position du point dépend de l'endroit où se trouve le site de pause sur le fragment de digestion. L'ADN double brin linéaire est retrouvé sur la diagonale. La quantification de la pause d'effectue en utilisant le rapport : signal du point / (signal de l'arc + signal du point)

1.2 Observation de l'empreinte à *mat1* par Southern-blot

L'empreinte à *mat1* engendre une cassure double brin durant une préparation classique d'ADN. Ce phénomène nous permet d'analyser la quantité d'empreinte par Southern-Blot.

L'ADN génomique est préparé selon la méthode décrite dans l'expérience 17 du livre "Experiments with fission yeast: A laboratory course manual" (Alfa et al,. 1993).

5 µg de l'ADN ainsi préparé sont digérés par 60 unités de l'enzyme HindIII (NEB).

L'ADN est ensuite séparé dans un gel 0,8 % d'agarose en TAE à 50 V (2 V/cm) pendant 14 h. Le gel est traité comme le décrit le protocole de Southern-Blot dans le volume 2 du livre "Molecular Cloning a laboratory manual" (Sambrook et al,. 1989). La membrane est hybridée avec une sonde radio-marquée par le fragment de Klenow en utilisant de l'ATP marquée en α-P32 (figure 3). La sonde est obtenue par PCR sur de l'ADN génomique d'une souche de *S. pombe h90* (préparé par une méthode classique) en utilisant les oligonucléotides suivant : 5' TTCGGTATTTAAGTCTGGCG 3' et 5' CCAATTATGCTGTTCGTGTC 3'. Le résultat schématique est présenté en figure 3. Le pourcentage de molécules d'ADN portant l'empreinte est calculé en utilisant le rapport suivant : signal du fragment distal + signal du fragment proximal / somme des signaux des trois fragments. Cette technique permet d'estimer le nombre de molécules qui possèdent l'empreinte, sachant que dans une population de cellules de *S. pombe* selon le modèle de CTS, environ 50% des cellules portent une empreinte.

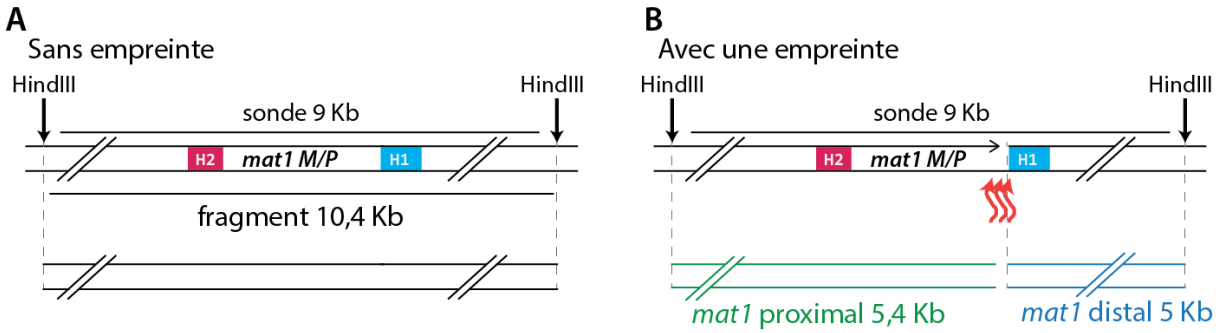


Figure 3 : L'empreinte crée deux sous fragments.
A) Sans empreinte, la digestion enzymatique produit un fragment de 10,4 kb contenant le locus *mat1*. B) Si l'empreinte est présente le fragment de digestion HindIII de 10,4 kb sera divisé en deux fragments de 5,4 kb (proximal) et de 5 kb (distal). Le rapport entre le signal du brin proximal + distal / signal des trois fragments nous donne le pourcentage de molécules d'ADN ayant l'empreinte.

1.3 Utilisation de la qPCR pour établir le pourcentage de cellules M et P dans une population

Le passage de la fourche de réplication au travers de l'empreinte à *mat1* crée une cassure double brin. Cette cassure est réparée en utilisant le donneur portant l'allèle opposé à celui de *mat1*. Dans un contexte sauvage, le donneur ayant l'allèle opposé est choisi dans 90 % des cas (Miyata and Miyata, 1981). Un des phénotypes de mutants affectant le CTS est un déséquilibre entre la proportion de chaque type sexuel dans une colonie. Dans le but de quantifier la proportion de chaque type sexuel, j'ai utilisé la PCR quantitative (qPCR). Cette technique permet de déterminer la quantité d'ADN double brin présente à chaque cycle de PCR. Cette quantification est basée sur le fait qu'un fluorochrome est ajouté au mélange de la réaction de PCR. Ce fluorochrome a la particularité d'émettre un signal quand il se lie à de l'ADN double brin. Le signal émis est donc proportionnel à la quantité d'ADN double brin dans le puits de PCR. La quantité de signal fluorescent est mesuré à chaque cycle. Le mélange utilisé dans ce travail est "eurobio Green qPCR Mix Lo-Rox". La machine de qPCR utilisée est la qTower Jenna A6.
Afin de mesurer la proportion de chaque type sexuel dans une colonie de *S. pombe*, j'ai extrait l'ADN avec la méthode dite "extraction par billes". Cette méthode consiste à casser les cellules en utilisant un broyage mécanique avec des billes de verre. L'ADN est ensuite purifié, précipité et resuspendu en suivant une méthode classique. La qPCR est réalisée en utilisant deux couples d'oligonucléotides, un permettant de doser la quantité de molécule d'ADN contenant l'allèle P à *mat1* et l'autre l'allèle P à *mat1* (figure 4). La séquence d'oligonucléotide distal est OL198 5' AGGGAAAATTGGAGGGAG 3'. La séquence du primer 1M est OL199 5' TGTTTTAGAGGGTGATGCTT 3' et celle de 1P est OL200 5' AGCGGCTACTGGATTAAC 3'. La méthode des ΔCq est utilisée pour la normalisation. La comparaison entre les quantifications des deux couples d'amorces me permet de calculer la proportion de chaque allèle présent à *mat1*. Les barres d'erreurs représentent les déviations standards de la moyenne. Un t-test non apparié est utilisé pour le calcul des p values.

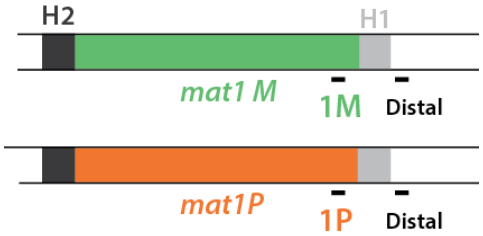


Figure 4 : Schéma de la position des oligonucléotides utilisés en qPCR.
La position des couples d'oligonucléotides est indiquée pour chaque allèle à *mat1*. L'oligonucléotide nommé distal, hybride indépendamment de l'allèle présent à *mat1*.

2. Détermination d'un site de recrutement protéique

Durant ma thèse, j'ai utilisé différentes techniques afin de déterminer le site de recrutement des protéines Lsd1, Lsd2, Sap1 et Abp1 sur la chromatine. Dans cette partie je vais décrire spécifiquement chacune de ces techniques en mettant l'accent sur les points différents de ce qui est publié.

2.1 *In vitro* : La technique de gel retard

La technique de gel retard permet d'analyser *in vitro* le recrutement d'une ou plusieurs protéines sur une séquence d'ADN double brin (ADN db). Le principe est que le complexe ADN/protéine migre dans un gel de polyacrylamide plus doucement que de l'ADN seul.

Pour tester le recrutement de Sap1 et Lsd1 à *mat1M*, j'ai utilisé des séquences d'ADN double brin recouvrant la région *mat1M* (figure 5). La liste des séquences des ADN db utilisés est indiquée dans le tableau 1. 50 ng d'un des brins d'ADN est marqué radio-activement en 5' par l'enzyme T4 polynucleotide kinase (NEB) et de l'ATP γ-P32 en suivant les instructions du fabricant. L'ADN marqué est ensuite purifié sur des colonnes de sépharose G50 afin de retirer la radioactivité non incorporée. Le fragment marqué et purifié est hybridé avec son complémentaire.

1 ng de chacun des ADN db radio-marqué est incubé avec 5 µg d'extrait protéique total préparé selon la méthode décrite précédemment (Arcangioli and Klar, 1991).

Le gel retard est ensuite réalisé comme décrit précédemment (Arcangioli and Lescure, 1985).

Dans le but d'analyser Sap1 et Lsd1, j'ai utilisé des protéines purifiées. Pour ce faire, j'ai utilisé la protéine Sap1-6His exprimée chez *E. coli* suivant une méthode décrite précédemment (Arcangioli et al., 1994a). L'expression de la protéine d'intérêt est induite pendant 3 h à 25 °C en présence de 1 mM d'IPTG. L'extrait protéique total de *E. coli* est ensuite purifié sur une résine chargée de Nickel (colonne d'agarose Ni-NTA His-tag) selon les instructions du

fabricant. Cette colonne a de l'affinité pour l'histidine et retient donc la protéine Sap1-6His. La protéine obtenue est ensuite analysée sur gel polyacrylamide avec une coloration au Bleu de Coomassie. Un dosage selon la méthode de Bradford est utilisé pour la quantification. Pour chaque réaction, 40 ng de protéine purifiée seront utilisés.

De la même façon, le domaine HMG de la protéine Lsd1 a été exprimé dans *E. coli*. La souche BL21 de *E. coli* a été transformée avec un plasmide PET portant l'insert HMG. La purification de ce domaine a été faite en utilisant une colonne de Ni Sépharose (HisTrap 1 mL GE Healthcare Life Sciences) en suivant les recommandations du fabriquant. Comme le montre la figure 6, nous remarquons que le domaine est élué dans la première fraction d'élution (piste 8). La quantité de protéine dans cette fraction est estimée à 0,8 mg/mL. Un gel retard utilisant 100 ng du domaine HMG par réaction a été effectué, sans succès. Nous supposons que le domaine HMG n'est pas actif dans les conditions utilisées.

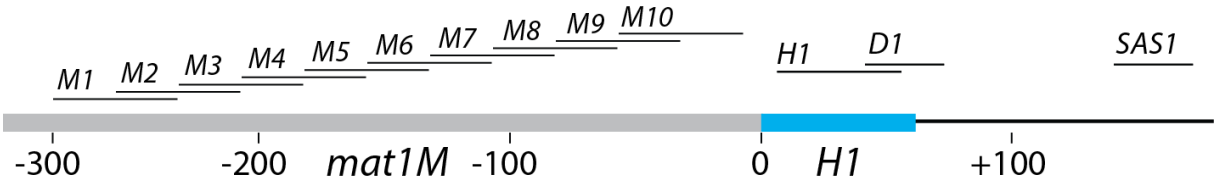


Figure 5 : Schéma de la position des oligonucléotides utilisés pour le gel shift.
La position des couples d'oligonucléotides est indiquée.

OL1436 M1+	**5'**	TATGCTTCTTAGAGTTACATTCACTGAAGATTATAATGTAATATTTTGTG	**3'**
OL1281 M1-	**5'**	CACAAAATATTACATTATAATCTTCAGTGAATGTAACTCTAAGAAGCATA	**3'**
OL1437 M2+	**5'**	GAAGATTATAATGTAATATTTTGTGTACCCCATTTGCGTTGAGTTATTCT	**3'**
OL1282 M2-	**5'**	AGAATAACTCAACGCAAATGGGGTACACAAAATATTACATTATAATCTTC	**3'**
OL1438 M3+	**5'**	TACCCCATTTGCGTTGAGTTATTCTATAGTAATTATTGTGTGTTCTATTA	**3'**
OL1283 M3-	**5'**	TAATAGAACACACAATAATTACTATAGAATAACTCAACGCAAATGGGGTA	**3'**
OL1439 M4+	**5'**	ATAGTAATTATTGTGTGTTCTATTAACGATGTATTGCGATTTATATCTGT	**3'**
OL1284 M4-	**5'**	ACAGATATAAATCGCAATACATCGTTAATAGAACACACAATAATTACTAT	**3'**
OL1440 M5+	**5'**	ACGATGTATTGCGATTTATATCTGTTATGCTAACATAACGTAGTTCTAAG	**3'**
OL1285 M5-	**5'**	CTTAGAACTACGTTATGTTAGCATAACAGATATAAATCGCAATACATCGT	**3'**
OL1441 M6+	**5'**	TATGCTAACATAACGTAGTTCTAAGCACTGTAATGCCATACTGTTTTAGA	**3'**
OL1286 M6-	**5'**	TCTAAAACAGTATGGCATTACAGTGCTTAGAACTACGTTATGTTAGCATA	**3'**
OL1442 M7+	**5'**	CACTGTAATGCCATACTGTTTTAGAGGGTGATGCTTCCTAAAATCTCCTT	**3'**
OL1287 M7-	**5'**	AAGGAGATTTTAGGAAGCATCACCCTCTAAAACAGTATGGCATTACAGTG	**3'**
OL1443 M8+	**5'**	GGGTGATGCTTCCTAAAATCTCCTTACATAAAGTAATACATGGATTTTAC	**3'**
OL1288 M8-	**5'**	GTAAAATCCATGTATTACTTTATGTAAGGAGATTTTAGGAAGCATCACCC	**3'**
OL1444 M9+	**5'**	ACATAAAGTAATACATGGATTTTACTGCCCTGATTCTATCGAAATATGCT	**3'**
OL1289 M9-	**5'**	AGCATATTTCGATAGAATCAGGGCAGTAAAATCCATGTATTACTTTATGT	**3'**
OL1445 M10+	**5'**	TGCCCTGATTCTATCGAAATATGCTGTTTTTTTTATTCGTTTTTATTTAT	**3'**
OL1290 M10-	**5'**	ATAAATAAAAACGAATAAAAAAAACAGCATATTTCGATAGAATCAGGGCA	**3'**
OL1446 H1+	**5'**	TTTGTAATATAAATGTATAGTCTTTCTCCTTTGTTTTCTCTCGTTCGTTT	**3'**
OL1291 H1-	**5'**	AAACGAACGAGAGAAAACAAAGGAGAAAGACTATACATTTATATTACAAA	**3'**
OL1447 SAS1+	**5'**	CTCTAACGAGATATTTGCTTCGCTACGCTACG	**3'**
OL1448 SAS1-	**5'**	CGTAGCGTAGCGAAGCAAATATCTCGTTAGAG	**3'**
OL1449 D1+	**5'**	TTCTCTCGTTCGTTTCCATGTTTCCAATTATG	**3'**
OL1292 D1-	**5'**	CATAATTGGAAACATGGAAACGAACGAGAGAA	**3'**
OL1492 Ter1+	**5'**	GGGATTTAACGCAGTGCAAGGAGCTATCTTGG	**3'**
OL1493 Ter1 -	**5'**	CCAAGATAGCTCCTTGCACTGCGTTAAATCCC	**3'**

Tableau 1 : Liste des oligonucléotides utilisés pour l'expérience de gel retard.

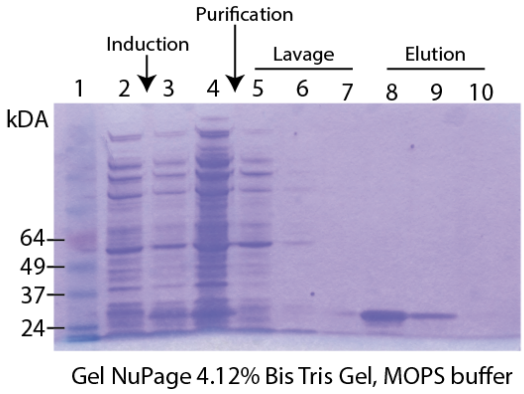


Figure 6 : Photographie du gel de vérification de la purification du domaine HMG marqué au bleu de Coomassie.
Dans la première piste, 20 µL de marqueur sont déposés (BenchMark Prestained N10748-010). Dans la piste 2, 20 µL d'extrait protéique total de la culture de *E. coli* non induite sont chargés. Dans les pistes 2 et 3, 20 µL d'extrait protéique total de la culture de *E. coli* induite sont chargés. Dans les pistes suivantes, 20 µL des fractions indiquées sont chargés.

2.2 Chromatine Immunoprécipitation chez *S. pombe*

La méthode de Chromatine ImmunoPrécipitation (ChIP) permet d'étudier le recrutement d'une protéine au niveau de la chromatine. J'ai utilisé cette technique couplée à une analyse quantitative par qPCR ou couplée à du séquençage à haut débit.

La technique utilise un anticorps dirigé contre la protéine d'intérêt afin de précipiter un complexe protéine/chromatine. Il est ensuite possible de purifier l'ADN présent dans ce complexe et de l'analyser. Cette technique permet de récupérer préférentiellement des fragments d'ADN liés par la protéine étudiée. L'ADN obtenu est analysé soit en qPCR pour quantifier l'enrichissement de la protéine immunoprécipitée à un locus spécifique soit par séquençage à haut débit pour analyser les sites de liaison préférentielle présents sur l'ensemble du génome. La différence technique entre la ChIP-qPCR et la ChIP-séq est la quantité d'ADN nécessaire. Typiquement, pour réaliser les librairies de séquençage, j'ai besoin de 5 ng d'ADN au minimum ce qui est 10 fois plus de ce que j'utilise pour une qPCR.

Pour la ChIP-qPCR, une culture de 3.10^8 cellules en croissance en milieu riche est fixée avec de la para-formaldéhyde. Les conditions de fixation pour chaque protéine sont indiquées dans le tableau 2. La méthode utilisée est décrite précédemment avec les modifications suivantes (Cam and Whitehall, 2016). L'étape de "preclearing" de la chromatine n'est pas réalisée et le temps d'incubation de l'extrait cellulaire avec les billes couplées à l'anticorps et de 3 h à 4 °C. La purification de l'ADN obtenu est réalisée par extraction au Phénol-Chloroforme. La phase aqueuse contenant l'ADN est ensuite concentrée en utilisant une précipitation à l'éthanol salin.

Les ADN ainsi obtenus sont analysés par qPCR (comme décrit dans la partie précédente) en utilisant deux couples d'oligonucléotides. L'un est la référence ou bruit de fond et l'autre hybride au niveau du locus d'intérêt. Le rapport ΔΔCq est analysé. Ce rapport tient compte d'une part, de la différence de quantité d'ADN au départ (rapport IP sur Whole Cell Extract (WCE) ou extrait cellulaire total de chaque expérience) et d'autre part de l'enrichissement en ADN du locus d'intérêt (par exemple la quantité d'ADN contenant *mat1* est comparée à celle contenant le locus situé à 3 kb en aval). Les différents couples d'oligonucléotides utilisés sont indiqués dans le tableau 3.

Chaque valeur d'enrichissement est une moyenne de trois répétitions techniques et d'au moins 3 ChIP-qPCR indépendantes. Un t-test non apparié est utilisé pour le calcul des *p values*.

	Type d'étiquette	Concentration	Temps	Température	Anticorps
Lsd1	Myc	1%	20 min	18 °C	1) 10 µL Bille couplée à un anticorps anti-Myc Tag
Lsd2	Myc	1%	20 min	18 °C	1) 10 µL Bille couplée à un anticorps anti-Myc Tag
Sap1	Aucune	3%	15 min	18 °C	2) 20 µL Bille couplée à 4 µL du sérum n°6
Abp1	Tap-Tag	3%	15 min	18 °C	3) 10 µL Bille couplée à 4 µL de l'anti-TAP Tag

Tableau 2 : Conditions techniques de la ChIP en fonction de chaque protéine.
Les conditions de fixation et l'anticorps utilisé pour les différentes ChIP sont indiqués.
1) Anti-Myc-Tag mAb-magnetic Beads (M047-11 MBL)
2) Dynabeads Protein G (10007D Invitrogen) couplées à 4 µL du sérum contenant l'anti-Sap1 (Lalonde et al., 1986).
3) Dynabeads Protein G (10007D Invitrogen) couplées à 4 µL de l'anticorps Tap Tag anticorps (CAB1001, Pierce Thermo)

	Noms	Séquences Forward	Noms	Séquences Reverse
mat1/2/3	**OL194**	CTCCTTTGTTTTCTCTCGT	**OL195**	TGGTTGATGGAGTGGTTG
mat1/2/3			**OL 196**	CTTCGTGGTATTCGGAAAT
mat1/2/3			**OL197**	CCTAGCGATTGATGTCAGT
matM	**OL1467**	TGGATTTTACTGCCCTGA	**OL1468**	GGGAAAATTGGAGGGAGA
matM	**OL1471**	GCTTCCTAAAATCTCCTTAC	**OL1472**	CTATCGGCCAAACAAACAC
matP	**OL1469**	CATCAGTTTTAATGGGCCAA	**OL1470**	CGACGGATATGATCTACAA
matP	**OL1473**	GGGCCAATTCTACGAAGT	**OL1474**	GGGAAGGGGAAGGTAGAAG
smt-0	**OL290**	CTTTCTCCCCATACATCT	**OL291**	GGAGTGTGATTGAAGGTG
Normalisation	**OL280**	CTCCCATCCTTGTCCTTT	**OL281**	GGCGCTCATGGTTATCTT
Normalisation	**OL477**	GAACGTTTCAATGAGACATGC	**OL478**	GTCATTCGGCGTTAGAAGGG
Control positif	**OL1431**	TGATCGCAGCATAGCATT	**OL1430**	GACGCAATGGAAGATGAA
Control positif	**OL1432**	TGATCGCAGCATAGCATT	**OL1433**	TGGTGGATGTTGATTTGTAG

Tableau 3 : Liste des oligonucléotides utilisés pour les expériences de ChIP-qPCR.

2.3 La ChIP-séquençage

2.3.1 Préparation de l'ADN en vue du séquençage

L'expérience de ChIP-séq est réalisée en suivant le même protocole décrit précédemment (ChIP-qPCR) sauf que la quantité de cellules utilisée est 10 fois supérieure (3.10^9 cellules) ainsi que la quantité de billes couplées à l'anticorps d'intérêt. La préparation des extraits cellulaires est donc réalisée en séparant le culot cellulaire en 10 tubes. Les différents extraits sont ensuite rassemblés et mis à incuber sur les billes couplées à l'anticorps.

L'ADN ainsi obtenu est ensuite confié à une plateforme d'analyse de l'Institut Pasteur (Plateforme Transcriptome et Epigénome). Les banques ont été préparées en utilisant le kit "TruSeq ChIP Library Preparation" en utilisant les instructions du fabricant (Illumina). J'ai

ensuite purifié sur gel d'agarose les banques obtenues (les tailles de fragments sélectionnés sont de 150 à 300 pb). Les banques purifiées ont été qualitativement vérifiées par Bioanalyser (Agilent). La quantité d'ADN dans chaque banque est mesurée avec un Qubit à l'aide du kit "dsDNA HS Assay Kit" (ThermoFisher).
Les séquences simples " Single Read" de 65 pb et de 130 pb ont été générées sur le séquenceur HiSeq 2500 selon les instructions du fabricant (Illumina).

2.3.2 Analyse des fichiers de séquence obtenus

L'analyse de ChIP-séq consiste à chercher des régions génomiques au niveau desquelles la protéine immunoprécipitée est enrichie. Cet enrichissement se traduit par une densité des lectures de séquençage plus importante dans ces régions.
Dans cette partie, je vais détailler les étapes de l'analyse jusqu'à l'obtention des pics qui sont les régions enrichies. Le code utilisé est décrit dans la section code 1.
Le fichier de sortie du séquenceur est un fichier contenant toutes les informations des séquences. Il est requis de vérifier la qualité du séquençage en utilisant un logiciel qui permet de donner la moyenne de la qualité pour chaque base des lectures (logiciel Fastqc). A la suite de cette vérification, si la qualité du séquençage est validée l'alignement sur le génome de référence est réalisé. Pour les différents séquençages, je n'ai pas traité les lectures avant leur alignement sur le génome de référence.
J'ai utilisé pour l'alignement l'algorithme Bowtie2 (Langmead et al., 2009) en utilisant les paramètres par défaut. Le génome "Schizosaccharomyces_pombe.ASM294v2.23" est utilisé pour l'alignement. Dans ce génome, la région *mat* n'est pas la région sauvage, c'est une région *mat* réarrangée. Elle ne permet donc pas d'étudier l'enrichissement des protéines d'intérêts au locus sexuel.
Les séquences sont donc alignées, en parallèle, sur les différentes séquences de la région *mat* détaillées dans la figure 7. Ces génomes sont les 45 kp contenant la région *mat* sauvage (h^{90}) avec l'allèle *M* à *mat1* (h^{90} *M*) ou avec l'allèle P (h^{90} *P*). Deux autres génomes utilisés sont composés d'environ 7 kb ne contenant que le locus *mat1* avec l'allèle *M* (*mat1M* Δ*2-3*) ou l'allèle *P* (*mat1P* Δ*2-3*). Ces séquences nous permettent d'analyser l'organisation de la région *mat*. La qualité des alignements est indiquée dans le tableau 4.
Les fichiers d'alignement ".sam" obtenues sont ensuite transformés en fichiers binaires ".bam" plus facilement manipulables. Ces fichiers sont ensuite ordonnés et indexés afin d'être visualisés sur le "Genome Browser IGV_2.3.69". La couverture pour chaque position du

génome est ensuite calculée puis normalisée en lecture par kb par la taille de la banque. Enfin la couverture de l'IP est divisée par celle du WCE en utilisant le Genome Browser IGV.

Code 1 : Alignement

```
### Les modules nécessaires
fastqc/0.10.1
bowtie2/2.1.0
samtools/0.1.19

### Vérification de la qualité
fastqc A.fastq

### Construction du génome pour l'alignement
bowtie2-build genome.fa genome

### Alignement
bowtie2 -x PREFIX.GENOME.BUILD.BOWTIE2 -U INPUT.fastq -S OUTPUT.sam

### Jusqu'au fichier binaire .bam et son index
samtools view -bS A.sam -o A.bam
samtools sort A.bam A.sort
samtools index A.sort.bam A.sort.bam.bai

### Calcul de la couverture
bamCoverage --normalizeUsingRPKM --binSize 1 -b A.sort.bam -o A.rpm.bw
bigWigToWig A.rpm.bw A.rpm.wig
igvtools toTDF A.rpm.wig  A.rpm.tdf genome.fa
```

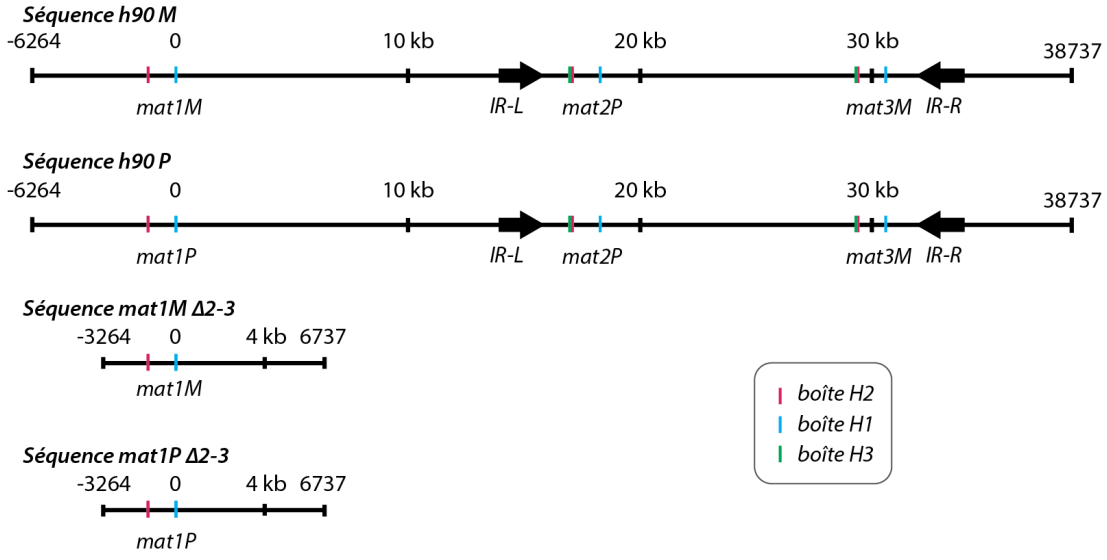


Figure 7 : Représentation schématique des séquences utilisées pour l'analyse.
Quatre séquences sont utilisées pour l'alignement. Les deux séquences du haut permettent l'alignement de la souche *h90*. La totalité de la région *mat* est représentée soit avec l'allèle *M* soit avec l'allèle *P* à *mat1*. Les deux séquences du dessous sont utilisées pour aligner les lectures des séquençages de l'ADN des souches portant l'allèle *M* ou *P* fixé à *mat1*. Dans ces souches les donneurs ne sont pas présents. Les loci *mat1*, *2* et *3* ainsi que les répétitions *IR-L* et *IR-R* sont indiqués. La boite d'homologie H2 est représentée en rouge et la boite H1 est représentée en bleue.

Génotype des souches	Banques	Type de séquençage	Nb tot de lectures	Nb de lectures alignées	% lectures alignées
h90 lsd1-Myc::KanMX6	A1IP	SR65	19681930	18482514	93.91%
h90 lsd1-Myc::KanMX6	A1WCE	SR65	21067533	19342359	91.81%
h90 lsd2-Myc::KanMX6	B1IP	SR65	8988245	8152554	90.70%
h90 lsd2-Myc::KanMX6	B1WCE	SR65	47617742	44206781	92.84%
h90	C1IP	SR65	12951987	12459329	96.20%
h90	C1WCE	SR65	62223348	55091901	88.54%
h90 abp1-CterTAP::KanMX6	D1IP	SR130	6462438	5690753	88.06%
h90 abp1-CterTAP::KanMX6	D1WCE	SR130	46506268	40369759	86.80%
h90 swi1Δ::KanMX6 abp1-CterTAP::KanMX6	D1swiIP	SR130	6868375	5599274	81.52%
h90 swi1Δ::KanMX6 abp1-CterTAP::KanMX6	D1swiWCE	SR130	28327346	22514091	79.48%
mat1M Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A2IP	SR65	78402964	70328896	89.70%
mat1M Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A2WCE	SR65	43201409	39627677	91.73%
mat1M Δ2-3::ScLEU2lsd2-Myc::KanMX6	B2IP	SR65	10184098	9714898	95.39%
mat1M Δ2-3::ScLEU2lsd2-Myc::KanMX6	B2WCE	SR65	9408560	8715332	92.63%
mat1M Δ2-3::ScLEU2	C2IP	SR65	12714852	12095278	95.13%
mat1M Δ2-3::ScLEU2	C2WCE	SR65	71445515	60200476	84.26%
mat1M Δ2-3::ScLEU2 abp1-CterTAP::KanMX6	D2IP	SR130	11015722	9764199	88.64%
mat1M Δ2-3::ScLEU2 abp1-CterTAP::KanMX6	D2WCE	SR130	42626692	37576928	88.15%
mat1P Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A3IP	SR65	36559765	33949628	92.86%
mat1P Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A3WCE	SR65	25466207	23759984	93.30%
mat1P Δ2-3::ScLEU2lsd2-Myc::KanMX6	B3IP	SR65	9443782	9170279	97.10%
mat1P Δ2-3::ScLEU2lsd2-Myc::KanMX6	B3WCE	SR65	30848920	28570601	92.61%
mat1P Δ2-3::ScLEU2	C3IP	SR65	9003795	8611021	95.64%
mat1P Δ2-3::ScLEU2	C3WCE	SR65	15378877	13342280	86.76%
mat1P Δ2-3::ScLEU2 abp1-CterTAP::KanMX6	D3IP	SR130	36287760	32457178	89.44%
mat1P Δ2-3::ScLEU2 abp1-CterTAP::KanMX6	D3WCE	SR130	78169993	35976808	46.02%
mat1MSS2 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A4IP	SR65	2405263	2137626	88.87%
mat1MSS2 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A4WCE	SR65	42445629	38607233	90.96%
mat1MSS2 Δ2-3::ScLEU2	C4IP	SR65	6003899	5124804	85.36%
mat1MSS2 Δ2-3::ScLEU2	C4WCE	SR65	52438070	46350961	88.39%
mat1MSS13 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A5IP	SR65	1322844	976586	73.82%
mat1MSS13 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A5WCE	SR65	37693504	34797905	92.32%
mat1MSS13 Δ2-3::ScLEU2	C5IP	SR65	9870989	9139481	92.59%
mat1MSS13 Δ2-3::ScLEU2	C5WCE	SR65	46317419	40703242	87.88%
mat1M smt-0 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A6IP	SR65	1102388	790954	71.75%
mat1M smt-0 Δ2-3::ScLEU2 lsd1-Myc::KanMX6	A6WCE	SR65	44260880	40696094	91.95%
mat1M smt-0 Δ2-3::ScLEU2	C6IP	SR65	5074834	4777575	94.14%
mat1M smt-0 Δ2-3::ScLEU2	C6WCE	SR65	35281724	31718229	89.90%
Séquençage issue de *Castel et al 2014*					
SRX736878	Rad52IPexp	PE 100	3554790	1872688	52.68%
SRX736879	Rad52WCEexp	PE 100	3443166	2219463	64.46%

Tableau 4 : Statistiques des banques de séquençage.

2.3.3 Comment obtenir des pics statistiquement valides ?

La détection des pics se fait en comparant la couverture de l'IP avec celle du WCE. Les bruits de fond local et global sont utilisés pour la détection. Le logiciel utilisé ici est le module MACS v2.1.0, qui permet de déterminer les coordonnés des pics liés par la protéine immunoprécipitée (Zhang et al., 2008).

J'ai réalisé 3 ChIP-séq de chacune des protéines d'intérêts au minimum. Seule la région *mat* est différente dans ces souches (*h90*, *mat1M* Δ*2-3*, *mat1P* Δ*2-3*). J'ai considéré ces souches comme des *triplicats* biologiques. En utilisant deux des séquençages comme *duplicat* j'ai réalisé l'analyse statistique "IDR" (Irreproducibility Discovery Rate) qui est utilisée pour la validation des données du ENCODE Consortium (Landt et al., 2012). Le détail de l'analyse utilisée est décrit ici :

https://sites.google.com/site/anshulkundaje/projects/idr#TOC-Intuitive-Explanation-of-IDR-and-IDR-plots

Le principe de l'approche consiste à créer une courbe, qui détermine quantitativement quand la découverte d'un pic n'est plus cohérente entre les *réplicats*. L'IDR détermine la probabilité attendue pour que le signal observé soit faux pour un seuil donné. Sachant que chaque pic déterminé par MACS2 a un *p value*, si cette dernière est au-dessus du seuil déterminé par l'analyse statistique, je ne le considère pas comme valide.

Le protocole de l'IDR est le suivant :

D'une part, les pics sont détectés dans chaque *réplicat*, puis les deux jeux de pics sont comparés en utilisant le script R (batch-consistency-analysis.r) décrit dans la méthode de l'IDR. D'autre part, les deux fichiers .bam des *réplicats* sont fusionnés, puis divisés aléatoirement en deux fichiers nommés : pseudo-*réplicats*. La détection des pics est faite sur chacun des pseudo-*réplicats* suivie de la comparaison des jeux de pics en utilisant le script R décrit (batch-consistency-analysis.r) dans la méthode.

A la suite de cette analyse, deux seuils d'IDR sont donnés. Ils permettent de sélectionner le nombre de pics partagé par les *réplicats* ou par les *pseudo-réplicats*. Les deux jeux de pics obtenus (celui des *réplicats* et celui des *pseudo-réplicats*) ne doivent pas différer de plus d'un facteur 2. Le nombre maximum de pics obtenus est utilisé. Dans la figure 8, nous pouvons observer les différentes probabilités que le signal soit faux en fonction de chaque signal (ici chaque pic).

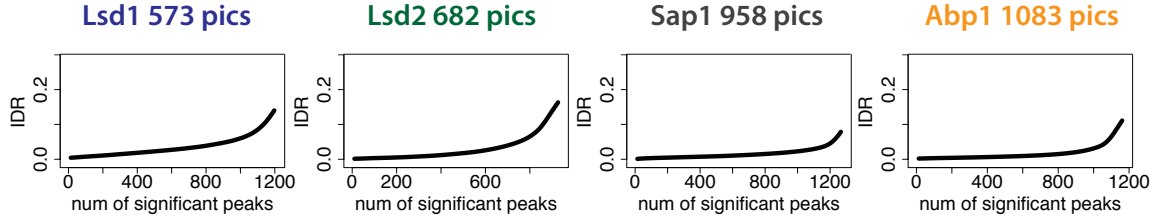


Figure 8 : Résultat de l'analyse de la cohérence statistique.
Les graphiques représentent l'IDR en fonction de la comparaison des pics des *réplicats*. Le nombre de pics significatif pour chaque protéine est indiqué au-dessus du graphique.

Cette analyse m'a permis de déterminer le jeu de pic statistiquement significatif pour chaque protéine. Cet ensemble de pics est ensuite utilisé pour déterminer les sites de liaisons potentiels. J'ai utilisé le logiciel RSAT "Regulatory Sequences Analysis Tools" (van Helden, J. 2003) afin de déterminer les motifs surreprésentés dans l'ensemble des pics. Nous pouvons supposer que de tels motifs sont de potentiels sites de liaison de la protéine immunoprécipitée. De plus, la position des motifs dans le pic est utilisée comme critère de sélection. Pour des protéines ayant des pics de forme "gaussienne", on suppose que le site de liaison est retrouvé au centre des pics. En outre, en ce qui concerne les protéines recrutées au niveau des séquences répétées, comme les LTRs, j'exclus de l'analyse la séquence des pics retrouvés dans ces régions. En effet, garder ces séquences répétées dans l'analyse biaise les résultats puisque n'importe quelle séquence d'homologie entre les séquences répétées seront décrite comme des potentiels sites de liaison.

2.3.4 Comparaison des pics et validation statistique

Dans cette partie, j'ai utilisé les ensembles de pics définis précédemment afin d'analyser les corrélations possibles entre différentes protéines immunoprécipitées.

Dans un premier temps, des intersections entre les coordonnées des pics et les coordonnées des gènes sont calculées. Pour ce faire, j'ai utilisé l'algorithme « bedtools » (bedtools/2.17.0). Ce dernier permet notamment de réaliser des intersections entre différents fichiers. Pour obtenir les régions biologiques qui correspondent aux coordonnées des pics, j'ai réalisé l'intersection du fichier des coordonnées de pics avec le fichier gff3 du génome utilisé. Cela donne la distribution de pics en fonctions des types de régions génomique.

```
### commande
intersectBed -a coordonées-pics.bed -b pombe.gff3 -wao |gawk '{print $13}' |sort |uniq -
c
```

Dans un second temps, j'ai comparé les différents ensembles de pics. Cette comparaison permet de déterminer les sites que deux ou plusieurs protéines partagent.

```
### commande
intersectBed -a coordonées-pics-Lsd1.bed -b coordonées-pics-Lsd2.bed - u >intersectLsd1-
Lsd2.bed
```

Dans un troisième temps, j'ai utilisé les données de transcription publiées par (Marguerat et al., 2012) dans le but d'analyser la corrélation entre le recrutement des protéines étudiées et le niveau de transcription. Le tableau S4 (colonne J) en données supplémentaire a été utilisé. Trois groupes transcriptionnels ont été réalisés

- fortement transcrit
- moyennement transcrit
- faiblement transcrit

Ensuite, les intersections de ces différents groupes de gènes et les ensembles de pics déterminés précédemment ont été calculées.

Enfin, j'ai réalisé une analyse statistique afin de déterminer la *p value* de chaque intersection. Pour ce faire, j'ai, à l'aide du HUB bio-informatique de l'Institut Pasteur, écrit un script qui permet de répartir de façon aléatoire les pics le long du génome (code 2).

Code 2 : Répartition aléatoire des pics sur le génome

```
#!/bin/bash

fileBed=$1
fileChrLen=$2
pathOut=$3

#bootstrap random
iter=0
while [ $iter -lt 100 ] ; do echo $iter
    cat $fileBed | while read chr start end rest ; do let lenPeak=$end-$start+1
       chr=`shuf -i 1-3 -n 1`
       len=`grep -w chromosome_$chr $fileChrLen | awk '{print $2}'`
       pos=`shuf -i 1-$len -n 1`
       let end=$pos+$lenPeak
       echo chromosome_$chr $pos $end
            done   |   sort   -k1,1   -k2n,2n   |   awk   '{print   $1"\t"$2"\t"$3}'   >>
result/${pathOut}_ranPeaks_$iter.bed
((iter++)) ; done #boostrap
```

J'ai donc distribué aléatoirement sur l'ensemble du génome les deux ensembles de pics à comparer, puis j'ai calculé leur intersection en utilisant les mêmes commandes que pour les pics réels. Ce calcul est réalisé sur 100 distributions aléatoires. La distribution des valeurs des intersections est ensuite représentée graphiquement (figure 9). Le calcul de la *p value* est ensuite réalisé de la façon suivante :

p value = (Nb intersection aléatoire > obs) / (Nb total d'intersections aléatoire)

Dans les différents cas de figure étudiés, j'observe que la *p value* est égale à zéro. Les corrélations que nous observons dévient fortement des répartitions aléatoires. Elles sont donc statistiquement significatives. Cependant, pour obtenir une estimation de la *p value*, il faudrait peut-être essayer en répétant plus de 100 fois le calcul.

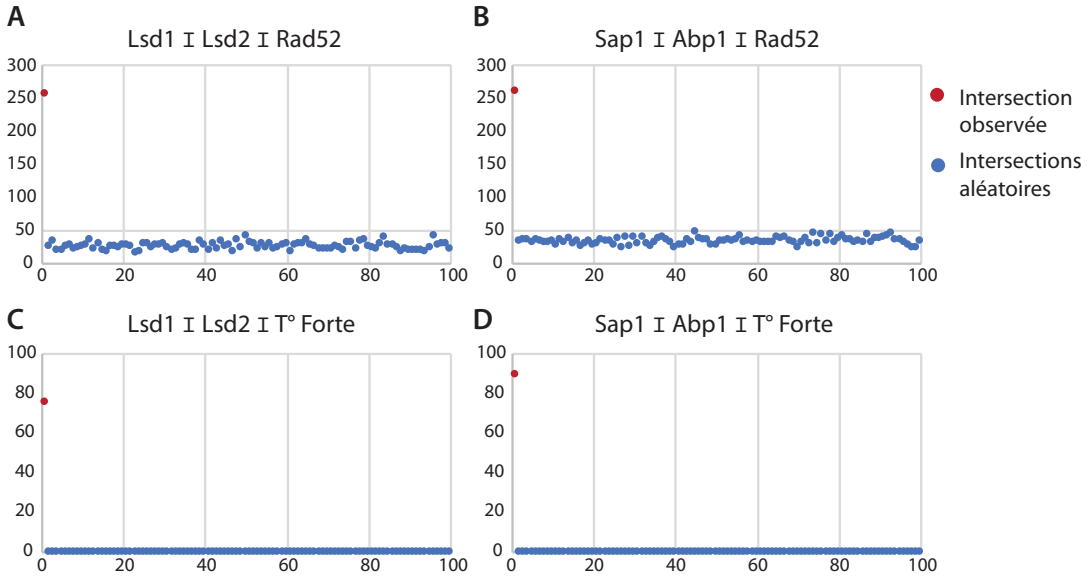


Figure 9 : Distribution des valeurs des intersections observées et aléatoires.
A) Distribution des intersections entre les pics de Lsd1, Lsd2 et Rad52. B) Distribution des intersections entre les pics de Sap1, Abp1 et Rad52. C) Distribution des intersections entre les pics de Lsd1, Lsd2 et du groupe de gènes fortement transcrits. D) Distribution des intersections entre les pics de Sap1, Abp1 et du groupe de gènes fortement transcrits. A-B-C-D) En rouge la valeur de l'intersection observée et en bleu les valeurs des intersections aléatoires.

3. Cartographie de l'empreinte par séquençage à haut débit

Afin d'analyser par séquençage le site de l'empreinte, j'ai utilisé le commencement des lectures d'ADN. Pour ce faire, j'ai utilisé un script développé par le Dr Jagla. Ce script permet d'extraire le commencement des lectures d'ADN en fonction de chaque brin. Ensuite, le script compte le nombre de début de lectures à chaque position génomique : j'appellerai ce nombre le **5' compte**. Théoriquement au site de l'empreinte les 5' comptes devraient être enrichis puisque l'ADN casse à ce niveau.

Afin de valider statistiquement la valeur présente au site de l'empreinte une distribution théorique est ajustée à la distribution des 5' comptes observés (Code 3. Cette distribution (binomial négative) est définie par sa taille et sa moyenne. Ces deux paramètres sont ensuite utilisés pour le calcul de la *p value* de la valeur des 5' comptes à l'empreinte. Les valeurs statistiques observées et théoriques des 8 séquençages étudiés sont listées dans le tableau 5. Le logiciel R est utilisé pour les calculs.

Code 3 : Calcule de la *p value* du 5' compte à *mat1* (R)

```
## Outils nécessaires

library(fitdistrplus)
library(vcd)

## Exemple pour un fichier d'alignement nommé : A.Plus-strand

A.Plus-strand <- read.delim(A.Plus-strand.bam, header=FALSE)

## Je détermine la distribution observée des valeurs des 5' comptes le long du génome

fitA.Plus-strand = fitdistrplus::fitdist(A.Plus-strand [,3], "nbinom")

########## Je calcule la p value de la valeur observée à l'empreinte en utilisant les paramètres du modèle de la distribution calculée

pValue = pnbinom("5' compte à l'empreinte", mu = fitA.Plus-strand["mu"] , size = fitA.Plus-strand$estimate["size"], lower.tail = F)
```

	allèle M	WCE1	WCE2	WCE3	WCE4		*allèle P*	WCE5	WCE6	WCE7	WCE8
Obs.	Min	0	0	0	0	Obs.	Min	0	0	0	0
	1er Qu	0	0	1	0		1er Qu	0	0	0	0
	Median	1	0	2	1		Median	1	1	0	1
	Moyenne	1,568	0,3452	2,381	1,404		Moyenne	1,404	1,131	0,5282	1,336
	3ème Qu	2	1	3	2		3ème Qu	2	2	1	2
	Max	250	40	387	216		Max	216	108	99	274
Théo.	size	2,737863	1,6878629	2,796443	1,856391	Théo.	size	2,666177	2,719106	2,43635	1,868058
	mu	1,568488	0,3452976	2,380682	1,403942		mu	0,940516	1,130586	0,5282618	1,335836

Tableau 5 : Valeurs observées et théoriques des distributions des 5' comptes du brin Watson.
Le tableau de gauche représente les valeurs statistiques observées et théoriques des 4 séquençages des souches *mat1MΔ2-3*. Le tableau de droite représente les valeurs statistiques observées et théoriques des 4 séquençages des souches *mat1PΔ2-3*.

4. Les fonds génétiques utilisées

Les différents fonds génétiques utilisés dans cette étude sont listés dans le tableau 6. Les souches sont obtenues après un croisement génétique et une méiose. Les asques sont disséqués sur milieu solide riche. Les génotypes de chaque souche sont vérifiés soit en suivant un marqueur phénotypique associé au génotype d'intérêt soit par PCR.

Nom	Génotype	Expérience	Origine
PB 70	*mat1M Δ2.3::ScLEU2 ade6-M216 ura- leu1-32*	ChIP	
PB 47	*h90 ade6-M210 leu1-32 uraΔ*	ChIP	
PB 2279	*mat1P Δ2.3::ScLEU2 ade6-M216 ura-*	ChIP	
PB 483	*mat1M SS2Δ Δ2.3::ScLEU2 ade6-M210 ura-*	ChIP	
PB 484	*mat1M SS13Δ Δ2.3::ScLEU2 ade6-M210 ura-*	ChIP	
PB 241	*mat1M smt-0 Δ2.3::ScLEU2 ade6-M210 hist2- leu1-32*	ChIP	
PB 2666	*mat1M Δ2.3::ScLEU2 cdc25Ts ade6-M210 ura-*	Cinétique	
PB 71	*mat1M smt-0 cdc25Ts ade6-M210 ura- leu1-32*	Cinétique	
PB 2668	*mat1M SS13Δ Δ2.3::ScLEU2 cdc25Ts ade6-M210 ura-*	Cinétique	
PB 2667	*mat1M SS2Δ Δ2.3::ScLEU2 cdc25Ts ade6-M210 ura-*	Cinétique	
PB 2678	*mat1M smt-0 Δ2.3::ScLEU2 cdc25Ts ade6-M210 hist2-*	Cinétique	
PB 337	*h90 swi1Δ::kan ade6-M210 leu1-32*	Fonction Abp1	
PB 2348	*h90 lsd1-Myc::KanMX6 ade6-M210 leu1-32 ura-*	ChIP	Modifié de Holmes et al 2012
PB 495	*mat1M Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M216 ura- leu1-32*	ChIP	
PB 2336	*mat1P Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M216*	ChIP	
PB 490	*mat1M SS2Δ Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M210 uraΔ leu1-32*	ChIP	
PB 492	*mat1M SS13Δ Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M210 ura- leu1-32*	ChIP	
PB 491	*mat1M smt-0 Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M210 hist2- leu1-32*	ChIP	
PB 431	*h90 lsd2-Myc::KanMX6 ade6-M210 leu1-32 ura-*	ChIP	
PB 489	*mat1M Δ2.3::ScLEU2 lsd2-Myc::KanMX6 ade6-M216*	ChIP	
PB 2350	*mat1P Δ2.3::ScLEU2 lsd2-Myc::KanMX6 ade6-M216 ura- leu1-32*	ChIP	
PB 306	*h90 lsd1-18 ade6-M216 leu1-32 ura4-D18*	Fonction Abp1	
PB 1997	*h90 abp1Δ::ScLEU2 cbh1Δ::ScLEU2 sap1-c ade6-M216 leu1-32*	Fonction Abp1	Modifié de Zaratiegui et al 2011
PB 1646	*h90 cbh1Δ::ScLEU2 ade6-M216 leu1-32*	Redondance	
PB 2000	*h90 cbh1Δ::ScLEU2 sap1-c ade6-M216 leu1-32*	Redondance	
PB 2210	*h90 cbh1Δ-CTAP::kanMX6 ade6-M210 leu1-32 ura-*	Redondance	
PB 2212	*h90 cbh1Δ-CTAP::kanMX6 abp1Δ::ScLEU2 ade6-M216 leu1-32*	Redondance	
PB 2264	*h90 cbh1Δ-CTAP::kanMX6 swi1Δ::kan ade6-M210 hist2-*	Redondance	
PB 1281	*h90 abp1Δ::ScLEU2 lsd1-18 ade6-M210 leu1-32*	Fonction Abp1	
PB 1240	*mat1M SS2Δ Δ2.3::ScLEU2 Abp1-CterTAP::KanMX6 ade6-M210*	ChIP	
PB 1264	*mat1M smt-0 Δ2.3::ScLEU2 Abp1-CterTAP::KanMX6 ade6-M210 hist2- leu1-32*	ChIP	
PB 1266	*h90 Abp1-CterTAP::KanMX6 ade6-M210 leu1-32 lsd1-18*	ChIP	
PB 1258	*h90 abp1Δ::ScLEU2 ade6-M210 leu1-32*	Fonction Abp1	
PB 1645	*h90 abp1Δ::ScLEU2 sap1-c ade6-M216 leu1-32*	Fonction Abp1	
PB 1268	*h90 Abp1-CterTAP::KanMX6 ura4-D18 leu1-32*	ChIP	
PB 1242	*mat1M Δ2.3::ScLEU2 Abp1-CterTAP::KanMX6 leu1-32 ade6-M216*	ChIP	
PB 1257	*mat1P Δ2:3::ScLEU2 Abp1-CterTAP::KanMX6 ade6-M216 ura4-DS/E*	ChIP	
PB 1269	*h90 Abp1-CterTAP::KanMX6 swi1Δ::KAN ade6-M210 leu1-32*	ChIP	
PB 2206	*h90 abp1Δ::ScLEU2 swi1Δ::kan ade6-M210*	Fonction Abp1	
PB 2288	*h90 abp1-DIMΔ-flag::KanMX6 ade6-M210 leu1-32 ura-*	Domaines Abp1	Modifié de Lorenz et al 2012
PB 2293	*h90 abp1-flag(3X)::KanMX ade6-M210 leu1-32 ura-*	Domaines Abp1	
PB 2304	*h90 abp1-DDEΔ-flag::KanMX6 ade6-M216 leu1-32 ura-*	Domaines Abp1	
PB 2302	*h90 abp1-DBDΔ-flag::KanMX6 ade6-M210 leu1-32 ura-*	Domaines Abp1	
PB 2116	*h90 nmt1::kanMX6 abp1-CterTAP::KanMX6 ade6-M210 leu1-32*	nmt1 système	Modifié de Holmes et al 2005

Tableau 6 : Liste des souches utilisées dans cette étude.

Résultats

1. Cartographie des protéines Lsd1/Lsd2, Sap1 et Abp1 dans la région *mat*

Différents complexes protéiques sont recrutés au niveau du locus *mat1*. Comme décrit précédemment, le complexe Lsd1/Lsd2 (*Lysine Specific Demethylase 1 et 2*) ainsi que la protéine Sap1 (*Switch Activating Protein 1*) sont présents à *mat1*. Des expériences, de gel retard, de footprint et de ChIP ont permis de déterminer plus ou moins précisément les sites de recrutement de ces protéines. Il a été montré que la protéine Sap1 est recrutée à *mat1* au niveau de la séquence *SAS1* (Arcangioli and Klar, 1991). Bien que la suppression ou la mutation de la séquence *SAS1* soit délétère pour le CTS, le rôle direct de Sap1 dans le mécanisme a été difficile à étudier puisqu'elle est essentielle et spécifique de *S. pombe* (Arcangioli et al., 1994a).

D'autre part, le complexe Lsd1/Lsd2, mis en évidence chez *S. pombe* en 2007 (Gordon et al., 2007), est décrit comme participant au contrôle de la transcription et des frontières chromatiniennes en partie via la déméthylation de la lysine 9 de l'histone H3 (H3K9). Les activités enzymatique et non enzymatique de Lsd1 et Lsd2 sont nécessaires pour la régulation de la propagation de l'hétérochromatine et pour le contrôle de la transcription de l'euchromatine (Gordon et al., 2007; Lan et al., 2007; Opel et al., 2007; Shi et al., 2004). Plus tardivement, le laboratoire a montré que ce complexe est également important pour le contrôle de la réplication et plus particulièrement au niveau du site *MPS1* (Holmes et al., 2012). Des données de ChIP suivie d'analyse à haut débit sont disponibles pour la protéine Lsd1, Lsd2 et Sap1. Cependant, l'influence de l'allèle présent à *mat1* ainsi que le recrutement de ces protéines dans l'intégralité de la région du CTS n'ont jamais été déterminées. Nous avons donc décidé d'utiliser la technique de ChIP couplée au séquençage à haut débit (ChIP-séq) pour mettre en évidence le site de recrutement de ces différentes protéines dans cette région. Afin de caractériser l'influence des allèles *M* et *P* sur le recrutement de ces protéines, nous avons utilisé plusieurs structures de la région *mat*.

Enfin, la protéine Abp1 (*ARS Binding Protein* 1) est importante pour le CTS car elle contrôle l'expression d'un transcrit alternatif du gène *swi2*, en collaboration avec la protéine *Mc*. Cette forme protéique courte de Swi2 permet de guider la recombinaison homologue vers le donneur qui porte l'information opposée à celle de *mat1*. Abp1 contrôle ainsi l'équilibre entre les deux types sexuels dans une population de cellule (Aguilar-Arnal et al., 2008; Matsuda et al., 2011; Yu et al., 2012b). De façon intéressante, les données de ChIP-séq de Abp1 (*ARS Binding Protein* 1) disponibles suggèrent qu'elle est recrutée dans la région *mat* (Zaratiegui et al.,

2011b). Une fois de plus, le fond génétique utilisé ainsi que la taille des fragments séquencés (35 pb) ne permettent pas de déterminer le site précis de recrutement d'Abp1, car des séquences de plus de 50 pb (H1 et H2) sont partagées par les trois loci *mat1*, *mat2* et *mat3*.

1.1 Les fonds génétiques et la technique utilisées

Comme précisé dans l'introduction, la région du CTS est localisée au niveau du bras droit du chromosome II. Elle est composée, dans le contexte sauvage *h90* de trois loci : *mat1*, *mat2P* et *mat3M* (figure 1). Afin d'éviter des problèmes d'alignement des séquences immunoprécipitées liés à la redondance de plusieurs séquences nous avons décidé d'utiliser également pour l'analyse des fonds génétiques ne contenant plus les donneurs *mat2P* et *mat3M* (figure 1). En outre, nous utilisons une taille de séquençage de 65 pb, une taille supérieure à la boite d'homologie *H1* (59 pb). Les séquences uniques situées au-delà des 59 pb nous permettent d'assigner sans ambiguïté les lectures aux loci *mat1*, *mat2* ou *mat3*.

La souche *mat1M* Δ*2-3*, contient l'allèle *M* au niveau de *mat1* et le gène *LEU2* de *S. cerevisiae* à la place d'une région de 18 kbp qui contient les donneurs *mat2P* et *mat3M* (Klar and Miglio, 1986). Dans ce contexte, l'empreinte est formée au même titre que dans la souche *h90* et est réparée en utilisant la chromatide sœur, conservant ainsi l'allèle *M* à *mat1* (Klar and Miglio, 1986).

De la même façon, la souche *mat1P* Δ*2-3* contient elle l'allèle *P* au niveau de *mat1*. Les souches *mat1M* Δ*2-3* et *mat1P* Δ*2-3* nous permettent d'aligner les lectures contenant l'information *M* ou *P* au niveau de *mat1*, contrairement à la souche *h90* dans laquelle ces lectures peuvent être alignées également à l'un des donneurs. D'autre part, ces souches nous permettent d'étudier le rôle des donneurs ainsi que le rôle des voies de réparation dans le recrutement des différentes protéines à *mat1* (Egel and Gutz, 1981; Klar and Miglio, 1986).

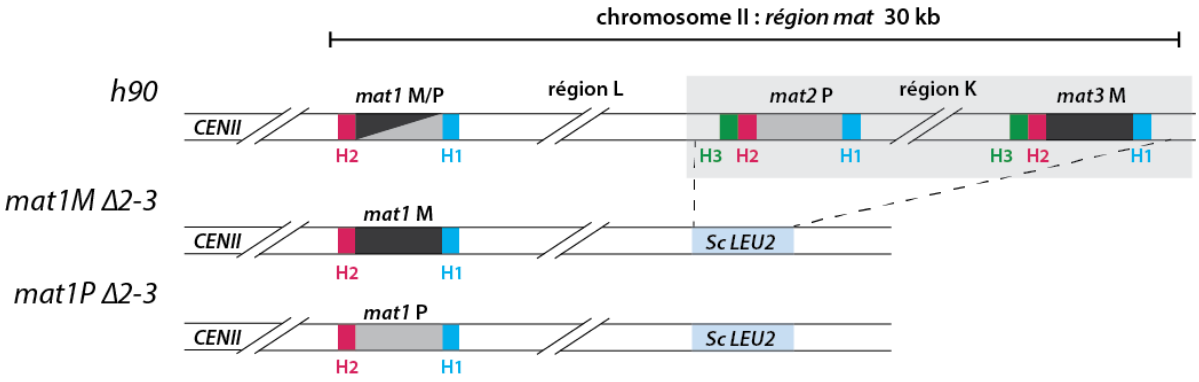


Figure 1 : Schéma de la région *mat* dans les souches utilisées.
La région du CTS est localisée au niveau du bras droit du chromosome 2. Le schéma du haut, représente la structure de la région *mat* d'une souche sauvage *h90* qui est capable de changer de type sexuel. Le locus *mat1,* exprimé, peut contenir l'allèle *M* en (noir) ou l'allèle *P* (en gris). Le locus *mat2P* est séparé de *mat1* par a région L (17 kb), qui contient au moins un gène essentiel. Le locus *mat3M* est séparé de *mat2P* par la région K (11 kb) qui contient le locus *CENH*. La boite grise entourant *mat2P* et *mat3M* représente l'hétérochromatine. En rouge, est représentée la boite d'homologie H2 (135 pb), en bleu, H1 (59 pb) et en vert, H3 (57 pb). Les souches *mat1M Δ2-3* et *mat1P Δ2-3* n'ont plus les loci *mat2P* et *mat3M.* Ils ont été remplacés par le gène *LEU2* de *S. cerevisiae* (bleu clair).

1.2 Cartographie des protéines Lsd1/2, Sap1 et d'Abp1 dans la région *mat*

Le complexe Lsd1/Lsd2 contrôle la formation de la pause et de l'empreinte à *mat1* (Holmes et al., 2012). Les données de ChIP-séq des protéines Lsd1 et Lsd2 dans une souche sauvage *h90* alignées sur toute la région *mat* montrent que ce complexe est recruté à l'intérieur des locus *mat1, 2* et *3* (Figure 2). Ce complexe est capable de reconnaitre l'allèle P et M mais nous ne pouvons pas déterminer si ces protéines sont recrutées seulement à *mat1* au vu de la redondance des allèles sexuels dans la région. Ce résultat suggère que les deux allèles partagent des homologies de séquence et/ou une structure chromatinienne.

La protéine Sap1 est présente à *mat1 in vitro*, du côté centromère distal de l'empreinte (Arcangioli et al., 1994a; Arcangioli and Klar, 1991). Les données de ChIP-séq montrent que Sap1 est recrutée *in vivo* à l'intérieur de l'allèle *M* et à l'extérieur du locus *mat3M* (Figure 2, *mat1* et *mat3*). A la différence du complexe Lsd1/2, la protéine Sap1 ne semble pas être recrutée au niveau de l'allèle *P* (Figure 2, *mat2*).

Ces résultats nous indiquent que les protéines Lsd1, Lsd2 et Sap1 ne sont pas recrutées par le même mécanisme. En effet, Sap1 n'étant pas recrutée au niveau de l'allèle *P* a certainement une spécificité de séquence de liaison plus stricte que le complexe Lsd1/Lsd2.

De façon intéressante, la protéine Abp1, décrite comme contrôlant une étape tardive du CTS (la directionnalité), est également recrutée dans la région *mat* et plus particulièrement à *mat1*

(figure 2, *mat1*). Cependant, contrairement au complexe Lsd1/2 et à la protéine Sap1, Abp1 est recrutée du côté centromère distal de *mat1* au niveau d'une séquence unique qui ne change pas en fonction de l'allèle présent à *mat1*.

Les donneurs sont dans une région transcriptionnellement inactive. En effet, l'hétérochromatine recouvre cette région comprise entre les deux répétitions *IR-R* et *IR-L* qui agissent comme frontières entre les deux types de chromatine (Grewal and Klar, 1997; Thon et al., 2002; Thon et al., 1994; Thon and Klar, 1992). Nous remarquons que Sap1 est recrutée au niveau des deux répétitions (figure 2 carrés rouges). Ces enrichissements de Sap1 au niveau de la fin des deux répétitions pourraient participer au blocage de la propagation de l'hétérochromatine. D'autre part, Abp1 est recrutée du côté centromère proximal de *CEN-H*. Au niveau de ce site, il a précédemment été montré qu'Abp1 permet le recrutement de Clr3, une histone acétylase (Cam et al., 2008). Cependant, en absence d'Abp1, la répression transcriptionnelle dans la région des donneurs est maintenue (Aguilar-Arnal et al., 2008; Nakagawa, 2002). Abp1 pourrait participer à la formation de l'hétérochromatine mais son rôle n'est pas indispensable.

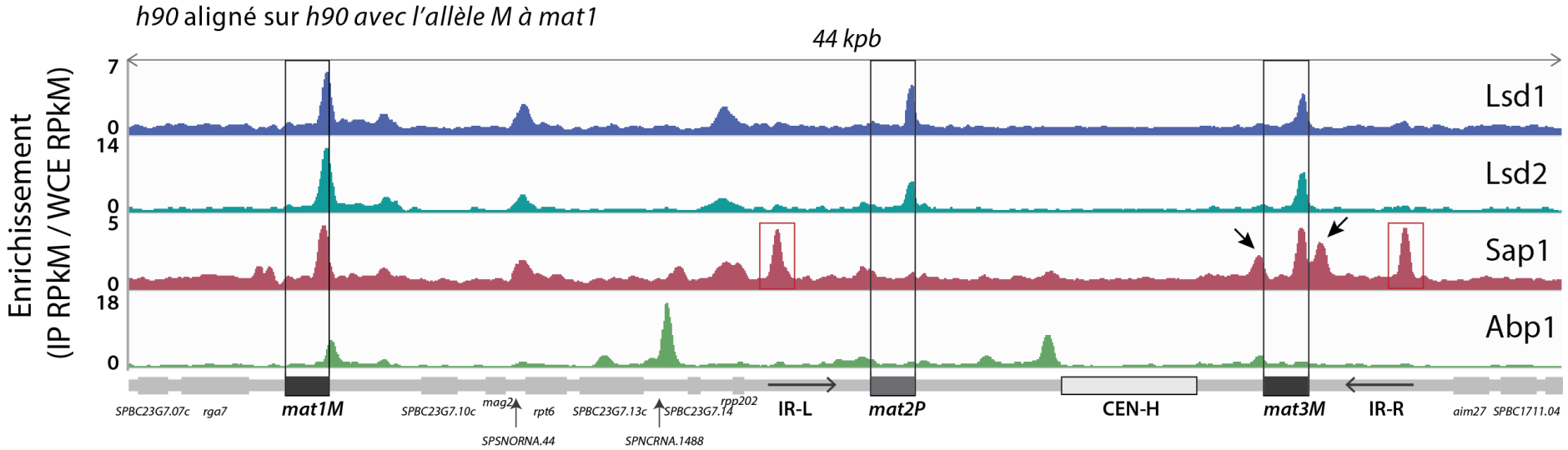


Figure 2 : Analyse de la région *mat.*

A) Distribution des enrichissements normalisés (IP RPkM *reads per kilobase million*/ WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *h90*. La séquence utilisée pour l'alignement est une région de 44 kb contenant l'allèle *M* à *mat1*. Les gènes, les séquences répétées *IR-R* et *IR-L* ainsi que le locus *CEN-H* sont notés en bas du panneau. Les rectangles noirs indiquent la position des trois loci *mat1M*, *mat2P* et *mat3M*. Les rectangles rouges mettent en évidence la liaison de Sap1 au niveau des limites de l'hétérochromatine. Les flèches noires indiquent les sites de recrutement de Sap1 à l'extérieur de *mat3M*.

Le problème que pose l'étude de la région *mat* est la présence de répétitions des allèles *M* et *P*. En effet, si on regarde précisément les signaux des IP de Lsd1 et de Lsd2 dans les allèles *M* et *P*, on remarque que les signaux à *mat2* et *mat3* se terminent abruptement à la fin de l'homologie de séquence avec *mat1*. Quant au signal à *mat1*, il se propage à l'extérieur de l'homologie. Cette observation suggère que ce complexe est recruté à *mat1M* et *mat1P* et non au niveau des donneurs. Afin d'assigner les signaux à l'intérieur de l'allèle *M* et *P* à un locus donné, j'ai étudié la distribution des lectures ne s'alignant qu'à un seul endroit sur la région *h90* ayant l'allèle *M* ou *P* à *mat1* (figure 3A et 3B). Les lectures ambiguës (pouvant s'aligner à plusieurs endroits) ne sont pas comptabilisées.

La figure 3A présente les deux types d'alignements (toutes les lectures, ligne 1 et les lectures ne s'alignant qu'à un endroit, ligne 2) pour chacune des IP sur la séquence *h90* ayant l'allèle *M* à *mat1*. La fin du signal d'enrichissement des protéines Lsd1 et Lsd2 est présent à *mat1M* (figure 3A, flèches noires). Nous remarquons que du côté centromère distal de *mat3M*, la couverture de ces IP est de l'ordre du bruit de fond. Cette comparaison nous permet d'assigner les lectures réparties aléatoirement entre *mat1M* et *mat3M* à *mat1M*.

D'autre part, nous remarquons qu'Abp1 n'est pas recrutée au niveau des donneurs.

De façon plus surprenante, Sap1 est recrutée au niveau de *mat1M* et *mat3M*. En effet, le signal de l'enrichissement se propage à la fois à l'extérieur de *mat1M* et à l'extérieur de *mat3M* (figure 3A, flèches noires). De plus, le signal de Sap1 au niveau de *IR-L* et *IR-R* n'est plus visible quand je retire les lectures ambiguës (figure 3A, carrés rouges). L'homologie de séquence de ces répétitions ne nous permet pas de déterminer si Sap1 fixe les deux répétitions ou une seule.

La même analyse a été réalisée en utilisant comme séquence de référence la séquence *h90* avec l'allèle *P* à *mat1*. Les signaux du complexe Lsd1-2 et d'Abp1 se répandent à l'extérieur de *mat1P*, comme pour *mat1M*. Ces résultats suggèrent que la nature de la chromatine plutôt que la séquence d'ADN contrôle le recrutement de Lsd1 et Lsd2. Cette chromatine serait ressemblante quand l'allèle est *M* à *mat1* ou bien quand il est *P* et permettrait la liaison de Lsd1-2, par contre elle serait différente au niveau des donneurs inhibant ainsi le recrutement du complexe.

En ce qui concerne Sap1, nous observons un faible signal à l'extérieur de *mat1P*. Ce signal peut être dû au recrutement de Sap1 à *SAS1* quand l'allèle à *mat1* est *P*. Alternativement ce signal peut être dû à des lectures provenant du signal à l'extérieur de *mat1M* aligné ici par défaut. En effet, l'utilisation de la souche h^{90} ne nous permet pas distinguer si les lectures à l'extérieur de *mat1* proviennent d'un chromosome *mat1M* ou *mat1P*. Comme déjà décrit, les allèles *P* et *M*

bien qu'ayant tous deux une empreinte ne sont pas équivalents. Nous pouvons supposer que la nature des protéines présentes au niveau de ces allèles pourra influencer les différentes étapes du CTS.

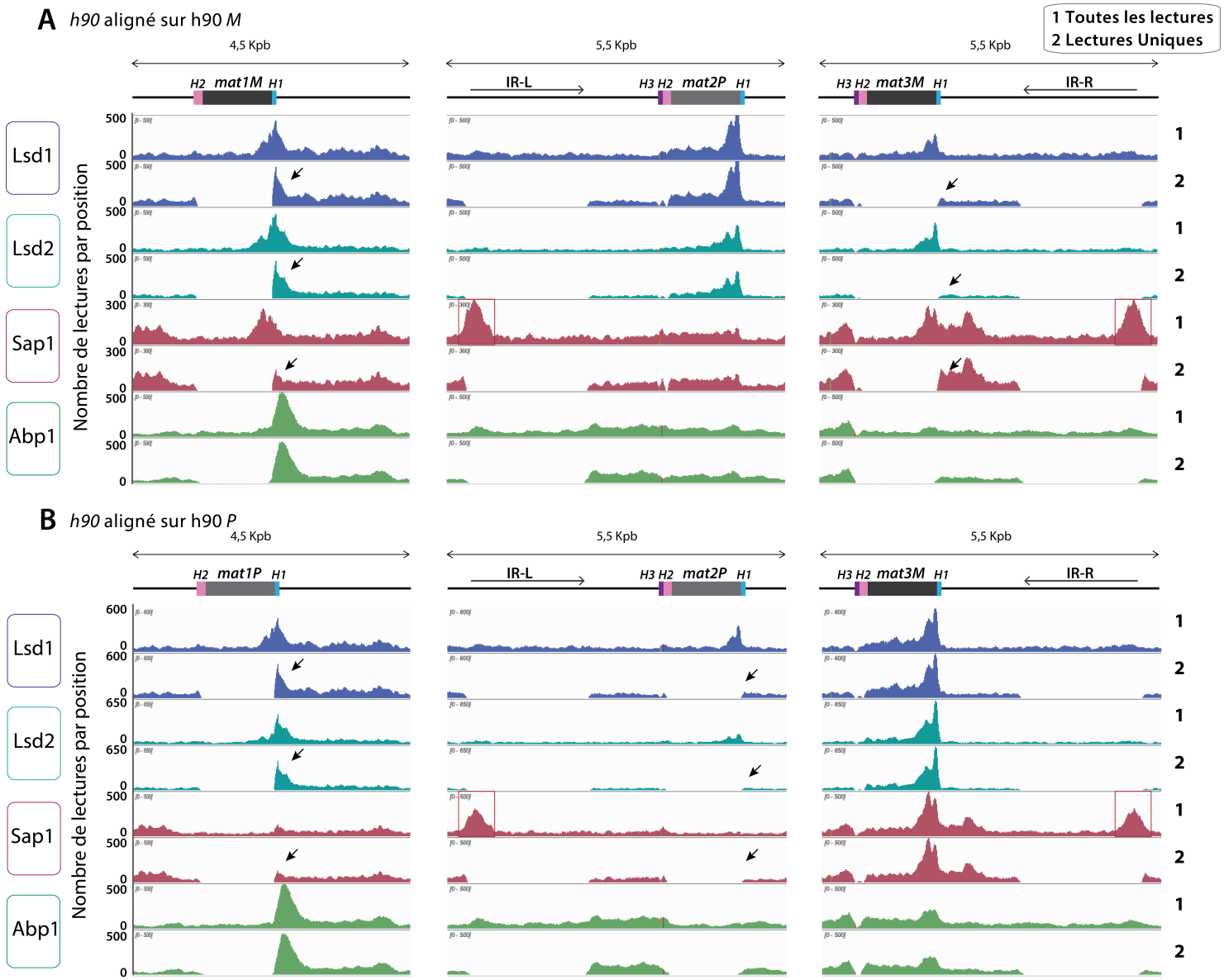


Figure 3 : Lsd1, Lsd2 et Abp1 ne sont recrutées qu'à *mat1*.

A) Distribution des lectures des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *h90*. L'alignement est fait en utilisant la séquence *h90* avec *M* à *mat1*. Les lignes 1 et 2 représentent la distribution de toutes les lectures et des lectures qui ne s'alignent qu'à un seul site, respectivement. Les flèches noires indiquent les sites uniques à l'extérieur de l'allèle *M*. Les rectangles rouges mettent en évidence la liaison de Sap1 au niveau des limites de l'hétérochromatine. B) Même que A en utilisant pour l'alignement la séquence *h90* avec l'allèle *P* à *mat1*.

1.3 Lsd1/2, Sap1 et d'Abp1 sont recrutées à *mat1* indépendamment des donneurs

Afin de valider l'analyse précédente et d'étudier le rôle des donneurs dans le recrutement de ces acteurs, nous avons déterminé les enrichissements de ces protéines dans des souches ne contenant plus les donneurs (*mat1M* Δ*2-3* et *mat1P* Δ*2-3*). Dans ces fonds génétiques seul le locus *mat1* est présent. Les données de ChIP-séq des protéines Lsd1 et Lsd2 montrent que ce complexe est recruté à l'intérieur du locus *mat1* indépendamment de l'allèle présent et des donneurs (figure 4A et 4B). Le locus *mat1* est donc suffisant au recrutement de ce complexe (figure 4). La protéine Sap1 est recrutée préférentiellement, *in vivo*, à l'intérieur de l'allèle *mat1M* et cela indépendamment des donneurs (Figure 4A, 4B, 4E). Ce résultat montre que le site de liaison caractérisé *in vitro* (du côté centromère distal de *mat1*) n'est pas celui préférentiellement lié *in vivo* par Sap1 dans les conditions utilisées.

Ainsi, les protéines Lsd1, Lsd2 et Sap1 ne sont pas recrutées par le même mécanisme. En effet, Sap1 n'étant pas recrutée au niveau de l'allèle *P* a certainement une spécificité de séquence de liaison plus stricte que le complexe Lsd1/Lsd2.

La protéine Abp1 est également recrutée indépendamment de l'allèle présent à *mat1* et des donneurs au niveau d'une séquence unique (figure 4A, 4B et 4E).

De façon globale, nous observons un enrichissement normalisé plus faible pour Sap1 que pour les trois autres protéines. Cela peut être expliqué par le fait que nous utilisons un sérum et non un anticorps monoclonal pour la ChIP de Sap1. En effet, Sap1 n'est pas étiquetée, à la différence de Lsd1, Lsd2 et d'Abp1. Dans le sérum, la quantité d'anticorps spécifiquement dirigé contre Sap1 n'est pas connue à contrario de l'utilisation d'anticorps monoclonaux contre une étiquette.

L'analyse des lectures uniques et l'utilisation des fonds génétiques ne comportant pas les donneurs nous a permis de valider le recrutement spécifique de ces protéines à *mat1*. J'ai donc aligné les séquençages des souche *h90* sur les séquences de 10 kb ne contenant que *mat1M* ou *mat1P* (cf. figure 4C et 4D). Les profils d'enrichissements sont globalement similaires à ceux observés avec les souches sans les donneurs (comparaison figure 4A et 4C, comparaison figure 4B et 4D). Le maximum du pic à *mat1* pour la protéine Abp1 ne change pas en fonction de l'allèle ou de la souche (figure 4E). Cela peut être dû au fait qu'elle soit à l'extérieur de *mat1*. Pour les protéines à l'intérieur de *mat1*, le maximum du pic est en général deux fois plus faible pour les souches h^{90} par rapport aux souches *mat1M* Δ*2-3* ou *mat1P* Δ*2-3* (figure 4E). Ce qui corrèle le fait que dans une souche h^{90} la moitié des cellules sont *M* et l'autre *P*.

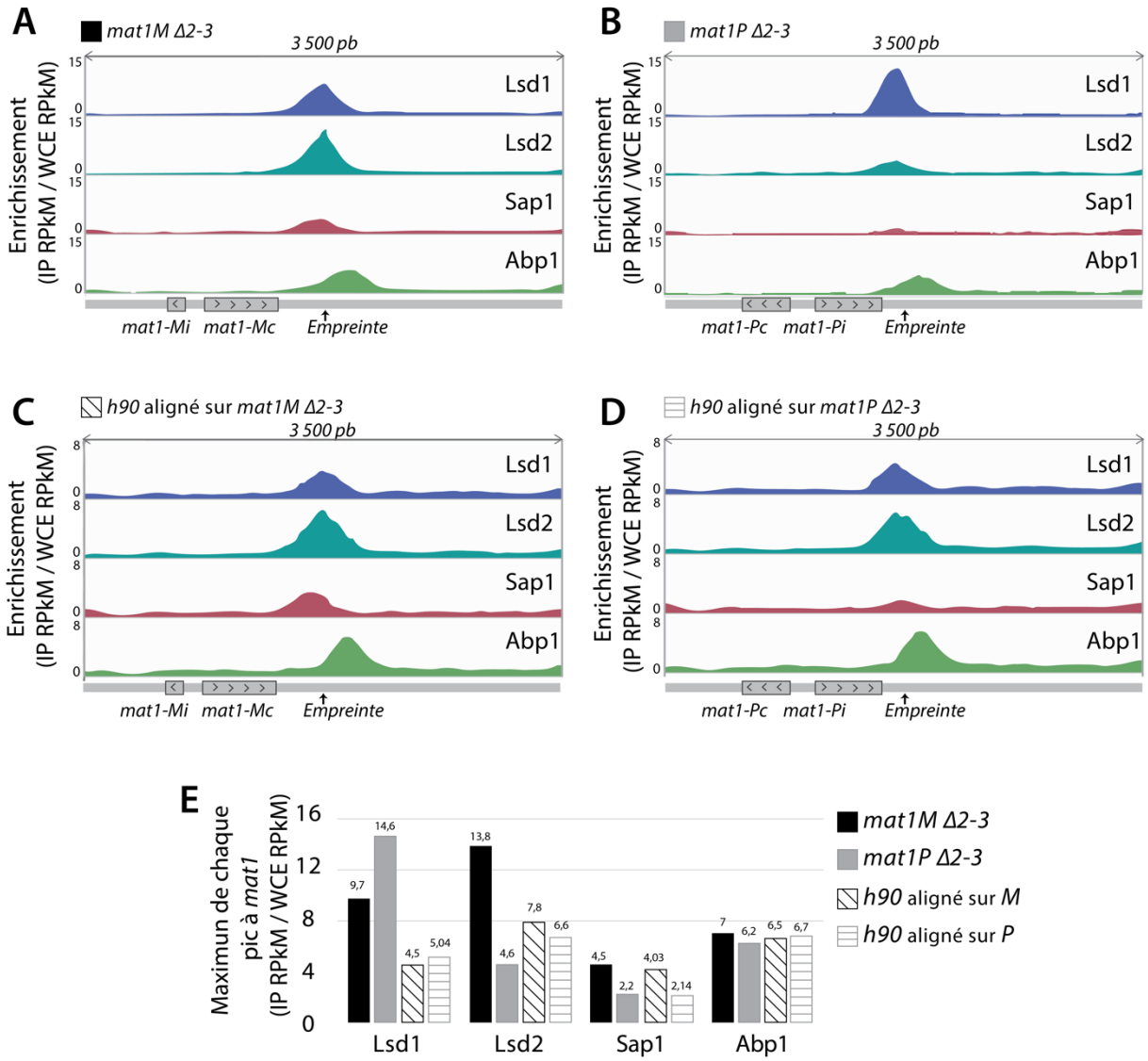


Figure 4 : Lsd1, Lsd2, Sap1 et Abp1 sont recrutées à *mat1* indépendamment des donneurs.

A) Distribution des enrichissements normalisés (IP RPkM *reads per kilobase million*/ WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *mat1MΔ2-3*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *M* à *mat1*. Le sens de la transcription de *Mc* et *Mi* est indiqué. La flèche marque le site de l'empreinte. B) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *mat1PΔ2-3*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *P* à *mat1*. C-D) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *h90*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *M* à *mat1* pour C et l'allèle *P* à *mat1* pour D. E) Enrichissement normalisé du maximum de chaque pic à *mat1* pour les ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 pour les séquençages des souches *mat1MΔ2-3* (■) et *mat1PΔ2-3* (■) et pour les séquençages des souches *h90* alignés sur la séquence *mat1M* (▨) ou sur la séquence *mat1P* (▤).

1.4 Sap1 est recrutée entre 211 et 111 pb du côté centromère proximal du site de l'empreinte

Afin d'identifier le site de fixation de la protéine Sap1, nous avons utilisé la technique de gel retard. Cette technique permet de caractériser une liaison protéine/ADN *in vitro*. Nous avons décidé d'analyser toute la région *mat1M* afin de déterminer si le site de recrutement observé en ChIP-séq pouvait être identifié. Pour ce faire, différents fragments d'ADN double brin marqués en 5' par un phosphate radioactif sont incubés 10 min avec un extrait protéique de *S. pombe* (figure 5A). Dans ces conditions, nous observons que les séquences *SAS1*, *Ter1*, M4, M5 et M6 sont retardées. Afin de confirmer que ce retard est dû à la liaison de Sap1, nous avons purifié la protéine Sap1 d'*Escherichia coli* et nous avons analysé par gel retard son recrutement au niveau de la région *mat1*. Nous observons le même profil de retard qu'avec l'extrait brut (figure 5B et 5C). Le fait que Sap1 lie *SAS1* et *TER1 in vitro* est en accord avec ce qui a déjà été décrit (Arcangioli and Klar, 1991; Krings and Bastia, 2005). En revanche, le recrutement de Sap1 au niveau d'une séquence à l'intérieur de l'allèle *M* n'a jamais été analysé. Les séquences liées par Sap1 à l'intérieur de *M* sont situées à 111 pb du site de l'empreinte (M4, M5 et M6) ce qui corrobore les données de ChIP-séq (figure 4). Le fait que trois séquences soient retardées suggère l'existence d'une coopération de différents sites *in vivo* qui permet le recrutement de Sap1 spécifiquement à l'intérieur de l'allèle *M*.

L'ensemble de ces résultats nous indique qu'au-delà de la séquence d'ADN, la structure de la région *mat1* influence le positionnement de la protéine Sap1 *in vivo*.

J'ai aussi produit et purifié le domaine HMG de la protéine Lsd1 dans *E. coli* afin de tenter de déterminer son site de fixation. Je n'ai pas été capable d'observer un retard sur gel avec ce domaine en utilisant les mêmes fragments d'ADN que précédemment. Ce résultat suggère, soit que le domaine HMG de Lsd1 produit chez *E. coli* ne fixe pas l'ADN, soit qu'une structure particulière de l'ADN et nécessaire pour son recrutement.

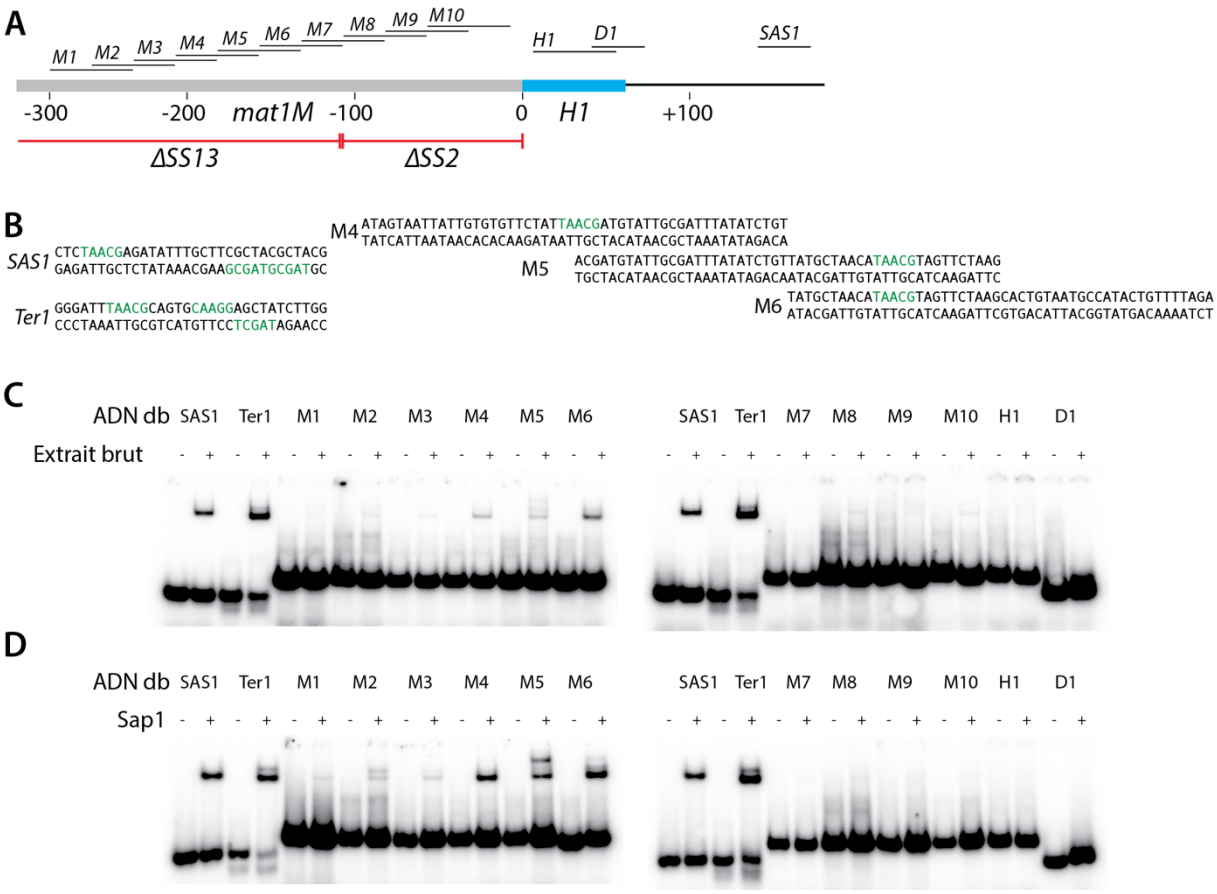


Figure 5 : Identification de la séquence liée par Sap1 au niveau de l'allèle *M*.
A) Représentation schématique du locus *mat1M* avec les ADN db (double brin) utilisés pour le gel-retard. Les ADN db de M1 à M10 sont à l'intérieur de l'allèle *M*. La boite H1 est représentée en bleu et le locus *mat1M* en gris. B) Les séquences des ADN db contrôles sont indiquées SAS1 et TER1. En vert, les sites de liaison précédemment caractérisés sont notés (Arcangioli et al., 1994b; Krings and Bastia, 2006). C) Le gel retard d'un extrait brut de protéines de *S. pombe*. Les oligonucléotides utilisés comme contrôles sont SAS1 et TER1. Dans les lignes (-), 1 ng de la sonde radio-marquée indiquée est déposé. Dans les lignes (+), un mélange contenant 1 ng de la sonde et 5 µg d'extrait total sont déposés D) Le gel retard de la protéine Sap1-6xHis purifiée chez *E. coli*. Les oligonucléotides utilisés comme contrôles sont SAS1 et TER1. Dans les ligne (-), 1 ng de la sonde radio-marquée indiquée est déposé. Dans les lignes (+), un mélange contenant 1 ng de la sonde et 40 ng d'extrait sont déposés.

1.5 Caractérisation du site de recrutement de Lsd1 et de Sap1 par ChIP-séq

Afin d'étudier les sites de recrutement de Lsd1 et Sap1 et leur rôle sur le CTS, nous avons analysé la liaison de ces protéines dans différents fonds génétiques contenant des délétions dans la région *mat*. Trois fonds génétiques sont utilisés : *mat1M smt-0*, *mat1M ΔSS13* et *mat1M ΔSS2* (figure 6). Au-delà de la suppression de différentes séquences, l'empreinte et/ou la pause sont également affectées dans ces souches (figure 6).

Dans le fond génétique *M smt-0* Δ*2-3* une séquence de 263 pb à droite de *mat1* est supprimée. Dans ce contexte, la pause *MPS1* est conservée mais l'empreinte n'est plus détectable (Styrkarsdottir et al,. 1993; Ulrike Engelke et al,. 1987; Dalgaard et al,. 2001) (figure 6A). Les délétions *SS13* et *SS2* sont dans l'allèle *mat1M* (figure 6B et 6C) (Sayrac et al., 2011). La délétion *SS13* est de 204 pb et la délétion *SS2* est de 110 pb. La délétion *SS2* commence au site de l'empreinte, alors que la délétion *SS13* commence à la fin de la délétion *SS2* (figure 6B et 6C).

Nous montrons par ChIP-séq que Lsd1 se lie à *mat1M* indépendamment de la délétion testée (figure 7). Ce résultat conforte l'hypothèse faite précédemment, à savoir que cette protéine n'est pas recrutée par une séquence d'ADN mais plutôt par une structure secondaire que forme la double hélice ou par une marque de la chromatine présente à *mat1*.

Par ailleurs, l'analyse de la position du pic de Lsd1 dans les différentes constructions confirme que cette protéine se lie du côté centromère proximal de l'empreinte. En effet, le pic d'enrichissement se rapproche de l'empreinte (ligne noire) dans les souches comportant une délétion à l'intérieur de l'allèle *mat1M* (*SS2* et *SS13*) et ne bouge pas dans la souche avec une délétion à l'extérieur (*smt-0*) (comparaison figure 7A et 7B avec 7C et 7D).

Il avait été montré que le recrutement de Lsd1 diminue dans une souche *mat1M* Δ*SS2* et que cette diminution affecte l'efficacité de la pause de la réplication à *MPS1* (Holmes et al., 2012). Mes résultats ne sont pas quantitatifs, mais ils suggèrent un déplacement du site de recrutement plutôt qu'une diminution de l'enrichissement. Les techniques utilisées sont différentes et peuvent jouer dans la variabilité (ChIP-séq vs ChIP-qPCR).

Plusieurs hypothèses peuvent expliquer que la pause ne soit plus détectable bien que Lsd1 soit recrutée :

- partenaire de Lsd1, la protéine Lsd2, n'est plus recrutée alors qu'elle est nécessaire à l'établissement de la pause ;
- le décalage du pic de Lsd1 ne permet plus la formation de la pause.

Cette dernière hypothèse étaye les observations faites précédemment qui montrent que la distance entre le site de pause et le gène *Mc* est déterminante pour l'efficacité de *MPS1* et du dépôt de l'empreinte (Sayrac et al., 2011).

La protéine Sap1 est toujours recrutée dans les souches *mat1M* Δ*SS2* et *mat1M smt-0* (figure 7A, 7B et 7D). Le site de recrutement se déplace de la même façon que pour Lsd1 dans la souche *mat1M* Δ*SS2* confirmant que Sap1 est à l'intérieur de l'allèle *M*. Par ailleurs, l'enrichissement de Sap1 est diminué de plus d'un facteur deux dans la souche *mat1M* Δ*SS13* (figure 7C). Ce résultat indique que le site préférentiel de recrutement de Sap1 est contenu dans

la séquence *SS13*. Cela corrobore les résultats de l'expérience de gel-retard qui montre que Sap1 se lie aux séquences M4-5-6. Ces résultats soutiennent le fait qu'*in vivo* le site préférentiel de recrutement de Sap1 est à l'intérieur de *mat1M*. Enfin, le fait que Sap1 ne soit plus enrichie dans une souche marquant encore la pause à *MPS1* mais n'ayant plus d'empreinte (*mat1M* Δ*SS13*) montre qu'elle n'est pas nécessaire à l'arrêt de fourche à *mat1* (Sayrac et al., 2011). L'efficacité de la pause dans cette souche est tout de même affectée par rapport à la souche sauvage (64% vs 100%). Il est donc possible que Sap1 optimise l'efficacité de la pause à *MPS1* ou/et qu'elle participe à la formation/maintenance de l'empreinte.

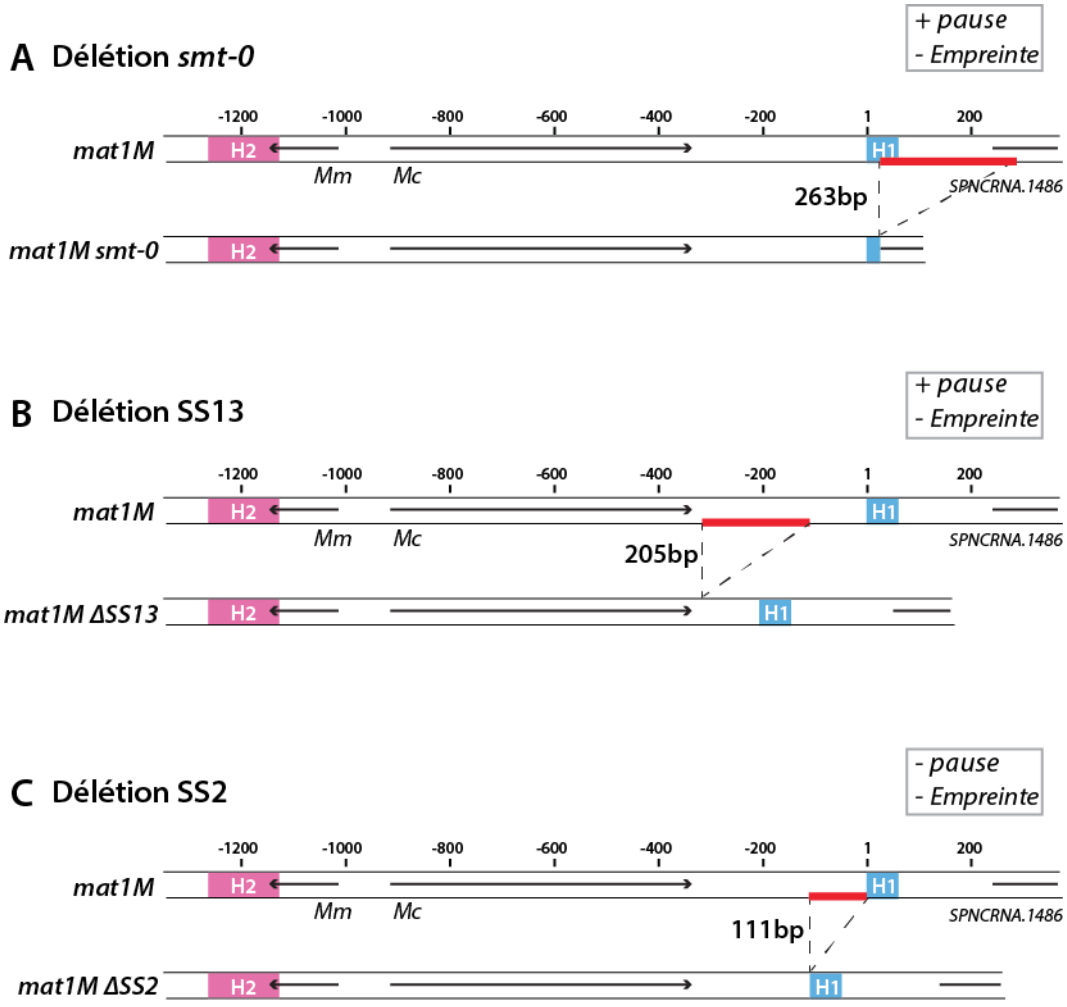


Figure 6 : Schéma de l'organisation de la région *mat* des souches utilisées.
A) Schéma représentant le locus *mat1*. Les boites d'homologie H2 et H1 sont indiquées en rose et bleu, respectivement. Les gènes présents dans la région et le sens de transcription sont indiqués par des flèches noires. La délétion *smt-0* est marquée en rouge. La structure du locus résultant de la délétion est représentée dans le panneau du bas. B) Même schéma, mais pour représenter la délétion *SS13*. C) Même schéma, mais pour représenter la délétion *SS2*.

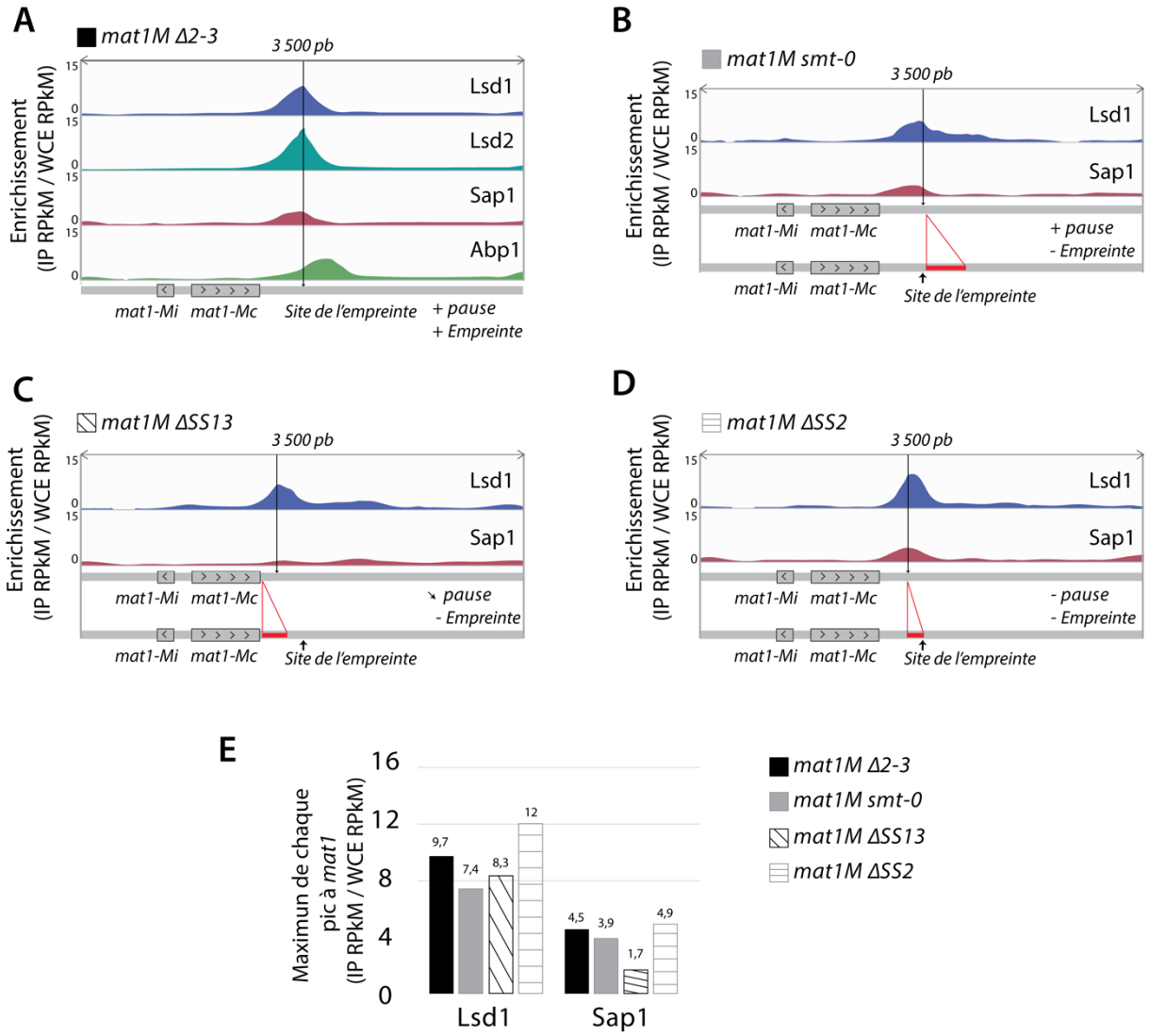


Figure 7 : Lsd1 et Sap1 sont recrutées dans *mat1*.

A) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *mat1MΔ2-3*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *M* à *mat1*. Le sens de la transcription de *Mc* et *Mi* est indiqué. La flèche marque le site de l'empreinte. B, C, D) En rouge sont indiquées les délétions utilisées. B) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, et Sap1 dans le fond génétique *mat1M smt-0*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *M smt-0* à *mat1*. C) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, et Sap1 dans le fond génétique *mat1MΔSS13*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *MΔSS13* à *mat1*. D) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, et Sap1 dans le fond génétique *mat1MΔSS2*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *MΔSS2* à *mat1*. E) Enrichissement normalisés du maximum de chaque pic à *mat1* pour les ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 pour les séquençages des souches *mat1MΔ2-3* (■), *mat1M smt-0* (▒), *mat1MΔSS13* (▨) et *mat1MΔSS2* (▤).

1.6 Sans empreinte, la fourche en pause casse spécifiquement

Le phénotype au niveau du CTS des souches *SS13*Δ et *smt-0* est similaire, la pause de la fourche à *MPS1* est conservée mais l'empreinte n'est plus détectable (Ulrike Engelke et al 1987; Dalgaard and Klar, 2001; Sayrac et al., 2011). Cependant, le recrutement de Lsd1 et de Sap1 diffère dans ces deux fonds génétiques. Dans la souche *mat1M SS13*Δ, la protéine Sap1 n'est plus recrutée à *mat1* et Lsd1 est recrutée plus proche de l'empreinte (figure 7), alors que dans la souche *mat1M smt–0*, le recrutement de Lsd1 et Sap1 est comparable à celui d'une souche sauvage. Afin d'analyser finement le phénotype au niveau du CTS dans ces contextes génétiques, j'ai analysé le niveau d'empreinte au cours du cycle cellulaire. L'hypothèse étant qu'en absence de ces séquences et des acteurs recrutés à ce niveau, l'empreinte n'est plus stabilisée ne permettant pas d'intervertir l'allèle à *mat1*. L'empreinte étant faite au redémarrage de la fourche bloquée à *MPS1* au cours de la phase S (Holmes et al., 2005), nous pouvons supposer que dans ces fonds génétiques elle sera formée en phase S puis réparée rapidement.

Afin de tester cette hypothèse, j'ai analysé le niveau d'empreinte au cours des différentes phases du cycle cellulaire en utilisant le mutant *cdc25*^{ts} pour synchroniser la population. Les cellules sont synchronisées en G2/M en étant incubées à température restrictive pendant 4 h à 37 °C. La culture est ensuite relâchée dans le cycle cellulaire en étant incubée à température permissive (25 °C). A ce moment, les cellules reprennent un cycle de division mitotique, entrent en mitose puis en phase S et enfin se divisent. Des fractions de culture sont récupérées à différents temps ; une partie du prélèvement servira au comptage des septa et une autre à l'analyse du niveau de l'empreinte. Chez la levure *S. pombe*, la phase S est concomitante à l'apparition des premiers septa, elle est donc facilement repérable (MacNeill and Fantes, 1997; Mitchison and Creanor, 1971).

En absence de la séquence *SS13* le niveau de l'empreinte est largement inférieur à celui d'une souche sauvage (figure 8D). Nous remarquons également que la taille de la bande supérieure (P) est plus faible que dans une souche sauvage en raison de la délétion d'une séquence de 205 pb dans cette souche. La détection de l'empreinte n'est cependant pas liée à la phase S, puisqu'au point -4 h qui correspond à une culture en croissance nous observons le même niveau d'empreinte que durant la phase S. De façon surprenante, nous remarquons aussi que le blocage à température restrictive entraine une diminution du niveau de l'empreinte (figure 8D point 0 à 80 min) qui n'est pas observé dans la souche sauvage (figure 8F). Cela suggère que l'empreinte est moins stable dans ce fond génétique. Pour résumer, en absence de *SS13*, la pause est faiblement diminuée et l'empreinte n'est pratiquement plus présente. Cela suggère que cette

souche conserve les éléments de séquences et protéiques qui déterminent la formation de l'empreinte mais pas ceux pour la maintenance de celle-ci. Nous pouvons suggérer que dans cette souche, l'absence de Sap1 ne permet plus à l'empreinte d'être stabilisée et/ou formée avec la même efficacité que dans un contexte sauvage.

De façon surprenante, dans la souche *mat1M smt–0*, un seul signal apparait durant la phase S (figure 8E point 80 min à 110 min). Afin de déterminer à quel fragment correspond ce signal, j'ai utilisé des sondes qui sont spécifiques du fragment distal (bleu) ou du fragment proximal (rouge). J'ai déshybridé puis réhybridé les membranes avec chacune des sondes (figure 8D et 8E panneau du milieu et du bas). Nous remarquons que la bande apparaissant en phase S n'est révélée qu'en hybridant avec la sonde distale (figure 8 panneau milieu et bas). Le fragment ainsi présent est le fragment distal de *mat1*, il est plus court que le sauvage car il ne contient plus la séquence *smt-0* (263 pb).

La détection d'une seule bande en phase S en absence de la séquence *smt–0* suggère que l'étape qui suit le blocage de la réplication à *MPS1* est différente de celle observée dans une souche sauvage. En effet, la séquence supprimée et probablement des acteurs protéiques recrutés à son niveau permettent le redémarrage correct de la fourche de réplication et la formation/maintenance de l'empreinte. En absence de *smt-0*, la pause de la fourche bloquée à *MPS1* doit être réparée en vue de la division cellulaire, sachant qu'elle ne peut pas être sauvée par une fourche arrivant de l'autre côté, puisque le site RTS1 bloque la réplication du côté centromère proximal (figure 9B). Cependant, un seul fragment est observé. Deux hypothèses pourraient expliquer la formation de ce fragment :

- La première est que la fourche bloquée à *MPS1* régresse et crée une structure de type "chicken foot" (Berti and Vindigni, 2016; Meng and Zhao, 2017; Neelsen and Lopes, 2015). Cette structure permettrait aux fragments d'ADN nouvellement synthétisés de s'hybrider et de créer un fragment double brin allant jusqu'au site de digestion HindIII (figure 9C). Ce fragment aurait la même taille que la bande que l'on observe en Southern-Blot durant la phase S. Cette hypothèse suggère que la pause au niveau du site *MPS1* dans ce contexte est instable et difficilement détectable. Or, la pause à *MPS1* dans la souche *mat1M smt–0* est comparable à celle observée dans la souche sauvage. Cette supposition est donc peu probable.
- La deuxième hypothèse est que l'empreinte est déposée sur le brin retardé au redémarrage de la pause à *MPS1*. Le brin avec l'empreinte casserait lors de la préparation de l'ADN et générerait deux fragments : un fragment "distal" de 5 kpb et un fragment engagé dans la réplication (figure 9D). Cependant, dans ce contexte génétique, l'empreinte n'est pas observée dans les autres phases du cycle. Elle n'est donc pas stabilisée et est réparée rapidement en utilisant la

chromatide sœur à la différence d'une souche sauvage. L'allèle à *mat1* est donc stable. Dans ce contexte la fourche de réplication ne passerait jamais à travers une empreinte mais toujours sur une chromatide "vierge".

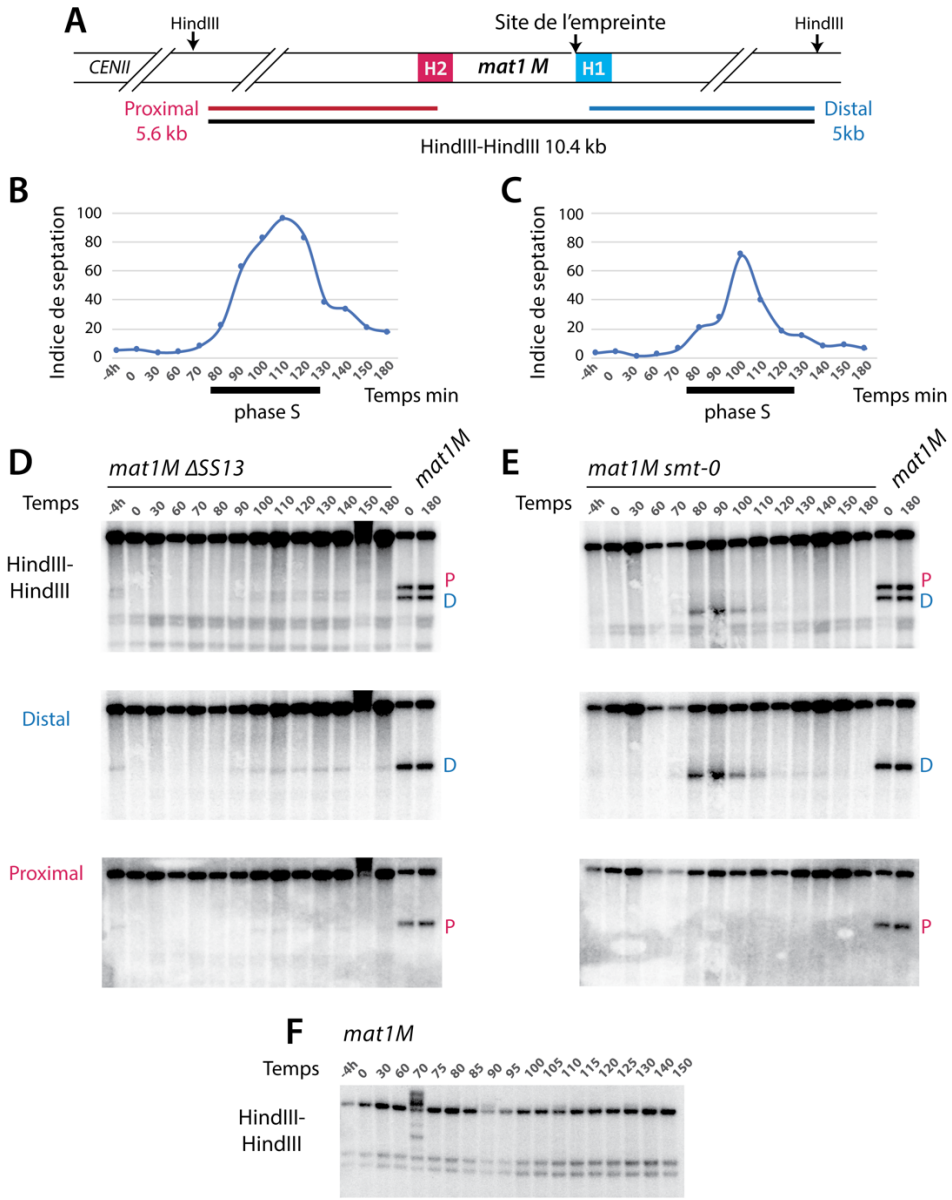


Figure 8 : La fourche casse durant la pause.

A) Schéma de la région *mat1*. Les tailles des fragments de restriction utilisés sont indiquées. Le panneau du bas indique les sondes utilisées. La sonde HindIII-HindIII hybride sur la totalité du fragment de 10,4 kb. La sonde Proximale hybride sur le fragment de 5,6 kb et la sonde Distale hybride sur le fragment de 5 kb. L'empreinte est présente dans environ 50 % des molécules *mat1*. Elle est transformée en cassure double brin lors de la préparation de l'ADN. Ainsi le fragment de 10.4 kb est cassé en deux sous-fragments (*mat1*-proximal 5.6 kb indiqué avec un "P", et *mat1*-distal 5 kb indiqué avec un "D"). B) Analyse du nombre de cellules *mat1M ΔSS13 Δ2-3 cdc25*^{ts} avec septum au cours du temps. C) Même qu'en B) pour la souche *mat1M smt-0 Δ2-3* cdc25^{ts}. D) Les ADN génomiques des souches *mat1M Δ2-3 cdc25*^{ts} et *mat1MSS13 Δ2-3 cdc25*^{ts} ont été préparés par une méthode classique, digérés par *Hind*III et analysés par Southern blot. Les ADN ont été extraits à différents temps après la synchronisation des cultures à 32 °C pendant 4 h. Le temps 0 indique le passage de 32 °C à 25 °C. La sonde utilisée pour le panneau du haut est un fragment radio-marqué HindIII préparé à partir de la souche *h90* (HindIII-HindIII). La sonde utilisée dans le panneau du milieu est un fragment radio-marqué *Hind*III-H1 qui s'hybride avec le fragment distal. La sonde utilisée dans le panneau du bas est un fragment radio-marqué *Hind*III-H2 qui s'hybride avec le fragment proximal. E) Même expérience qu'en D) avec le mutant *mat1M smt-0 Δ2-3* cdc25^{ts}. F) Même expérience qu'en D) avec la souche *mat1M Δ2-3* cdc25^{ts}.

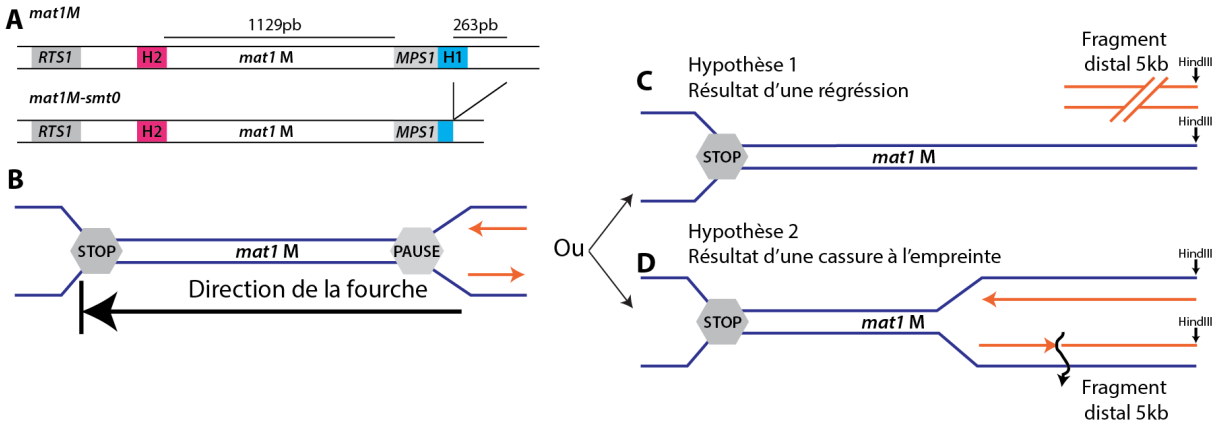


Figure 9 : Schéma du modèle de la réplication de *mat1* dans une souche portant la délétion *smt-0*.
A) Schéma de la région *mat1* dans un contexte sauvage et *smt-0*. B) Schéma de la réplication à *mat1*. C) Schéma du résultat de la régression de la fourche bloquée en utilisant une structure de type "chicken foot". D) Schéma du résultat de la cassure à l'empreinte.

1.7 Conclusions

Le complexe Lsd1/Lsd2 ainsi que la protéine Sap1 sont recrutés à l'intérieur de *mat1* indépendamment de la présence des donneurs. En absence des donneurs la CDB n'est pas réparée en utilisant les donneurs mais la chromatide sœur. Cette différence n'affecte pas le recrutement de ces protéines. Nous avons également déterminé le site de recrutement de Sap1 dans l'allèle *M* dans une séquence permettant la stabilisation de l'empreinte. Par ailleurs, Sap1 est présente aux limites de l'hétérochromatine entourant les donneurs. Lsd1 est toujours recrutée dans les différentes délétions utilisées. Pourtant, la délétion *SS2* est située au niveau du site de recrutement de Lsd1 dans une souche sauvage. Ce résultat ainsi que le fait que Lsd1-2 ne semblent être recrutées qu'au niveau de *mat1* suggèrent que la structure de la chromatine contrôle la liaison de ce complexe. Par ailleurs, le positionnement précis de Lsd1 par rapport au site de l'empreinte semble être un prérequis pour l'efficacité de *MPS1* et la formation/maintenance de l'empreinte.

La présence de la protéine Abp1 du côté centromère distal du locus sexuel nous a amené à étudier son rôle au niveau de ce locus.

2. Etude du rôle de la protéine Abp1 sur le changement de type sexuel chez *S. pombe*

Au cours de ma thèse j'ai étudié l'implication de la protéine Abp1 sur le mécanisme de division asymétrique chez *S. pombe*.

Une population de cellules de levures (*S. pombe*) est composée de 50 % de cellules de type sexuel *M* et de 50 % de cellules de type sexuel *P*. Il a été montré, qu'en absence d'Abp1, la population de cellules de levure n'est plus en équilibre (50 % *M* et 50 % *P*) mais qu'il y a 3 fois plus de cellules *M* que *P* (Aguilar-Arnal et al., 2008). Quand la cellule est *mat1M*, les protéines Abp1 et Mc permettent l'expression d'un transcrit alternatif de *swi2*. Ce transcrit est traduit en une protéine Swi2 courte qui permettrait que le donneur *mat2P* soit choisi lors de la réparation par conversion génique (Jia et al., 2004b; Matsuda et al., 2011; Yu et al., 2012b). En absence d'Abp1, le donneur *mat2P* ne serait plus choisi et donc la population tendrait vers une majorité de cellules de type *M*. Abp1 est donc impliquée dans la directionnalité du CTS. Son recrutement à *mat1* suggère qu'elle joue aussi un rôle dans les étapes précoces du CTS.

Tout comme à *mat1*, la fourche de réplication marque une pause au niveau des LTRs chez *S. pombe* (Dalgaard and Klar, 1999; Szilard et al., 2010). Des collaborateurs ont montré que cette pause de la fourche de réplication au niveau des LTRs est contrôlée par les protéines Sap1 et Abp1 (Zaratiegui et al., 2014). En effet, au niveau des LTRs, Sap1 permet de faire pauser la fourche de réplication et Abp1 permet de la faire redémarrer. En outre, la présence de Sap1 et d'une pause de la fourche guident l'insertion des rétrotransposons (Jacobs et al., 2015). Comme déjà décrit, Sap1 et Abp1 sont présentes au locus *mat1* (cette étude et Arcangioli and Klar, 1991).

Plusieurs questions ont donc été étudiées :

Le recrutement d'Abp1 diffère-t-il en fonction de l'allèle à *mat1* ?

Quels sont les éléments qui permettent le recrutement d'Abp1 à *mat1* ?

Abp1 contrôle-t-elle la pause de la fourche de réplication à *MPS1* ?

Peut-on récupérer un CTS efficace en mutant Sap1, la protéine antagoniste à Abp1 ?

Cbh1, un des homologue d'Abp1, joue t'il le même rôle qu'Abp1 dans le CTS ?

2.1 Le domaine de liaison à l'ADN est nécessaire au mécanisme de changement de type sexuel

La protéine Abp1 est composée de trois domaines : un domaine de liaison à l'ADN, un domaine de dimérisation et un domaine enzymatique (figure 10A). Pour déterminer quel domaine est important pour le mécanisme, j'ai utilisé des mutants de chacun d'eux (provenant du laboratoire du Dr Cam). Dans ces mutants, la séquence de chaque domaine est supprimée. J'ai strié chacune des souches sur Milieu Minimum puis j'ai coloré la boite avec des cristaux d'iode. Ce test permet d'analyser l'efficacité de CTS. En effet, l'iode colore l'amidon que l'on retrouve en grande quantité dans les spores, et témoigne de la capacité des cellules à se croiser et à sporuler. Ce test ne permet pas de déterminer quelles étapes du mécanisme sont défectives. Seul le mutant de délétion du domaine de liaison à l'ADN d'Abp1 montre une coloration réduite à l'iode. La mutation dans le domaine enzymatique de la protéine n'affecte pas le CTS, ce résultat nous permet d'exclure qu'Abp1 soit l'endonucléase responsable de la formation de l'empreinte.

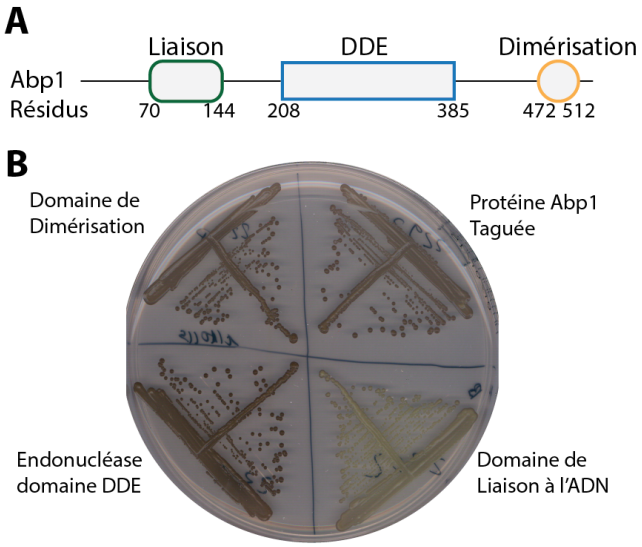


Figure 10 : Le domaine de liaison à l'ADN est nécessaire au CTS.
A) Représentation schématique des domaines de la protéine Abp1. En vert, le domaine de liaison à l'ADN du résidu 70 au résidu 144. En bleu, le domaine de la super famille des endonucléases DDE du résidu 208 au résidu 385 et en orange, le domaine de dimérisation du résidu 472 au résidu 512. Tiré de Pombase et de (Lorenz et al., 2012). B) Analyse de l'efficacité du CTS en utilisant la coloration à l'iode. Une coloration foncée est due à la fixation de l'iode sur l'amidon présent dans les spores. Une coloration claire signifie un défaut de CTS (ou de croisement). Les mutants des trois domaines sont testés ainsi que le fond génétique contenant la protéine Abp1 liée à une étique (TAP).

2.2 Comment Abp1 est-elle recrutée du côté centromère distal de *mat1* ?

2.2.1 Quantification du recrutement d'Abp1

Afin d'analyser quantitativement l'enrichissement de la protéine Abp1 au niveau du locus sexuel *mat1*, j'ai utilisé la technique de ChIP suivie d'une PCR quantitative (ChIP-qPCR). Cette technique permet de quantifier le recrutement d'une protéine au niveau d'un locus particulier en couplant la ChIP et la PCR quantitative. Les emplacements des oligonucléotides utilisés pour la PCR sont indiqués en figure 11A. Mes résultats montrent que la protéine Abp1 est recrutée à *mat1* et plus particulièrement du côté centromère distal (à droite) du locus (figure 11B). D'autre part, nous observons que Abp1 est 4 fois plus présente à *mat1* qu'au niveau des deux autres donneurs (figure 11B). De plus, j'ai testé l'influence de l'information à *mat1* sur le recrutement d'Abp1 (figure 11C). Pour ce faire, j'ai utilisé une souche sauvage *h90* dans laquelle 50 % des cellules sont M et 50 % sont P, une souche où l'allèle *M* est fixé à *mat1* (*mat1M* Δ*2-3*) et une souche où l'allèle *P* est fixé (*mat1P* Δ*2-3*). Dans ces trois fonds génétiques, mes résultats indiquent que la protéine Abp1 est recrutée à droite du locus *mat1M* et *mat1P* (figure 11C). Ces résultats corroborent ceux des ChIP-séq présentés dans la première partie. Le fait qu'Abp1 est présente au niveau du locus *mat1* suggère qu'elle participe au CTS, non seulement en jouant un rôle indirect en contrôlant *swi2,* mais également comme un acteur étant directement recrutée au niveau du site initiateur du mécanisme.

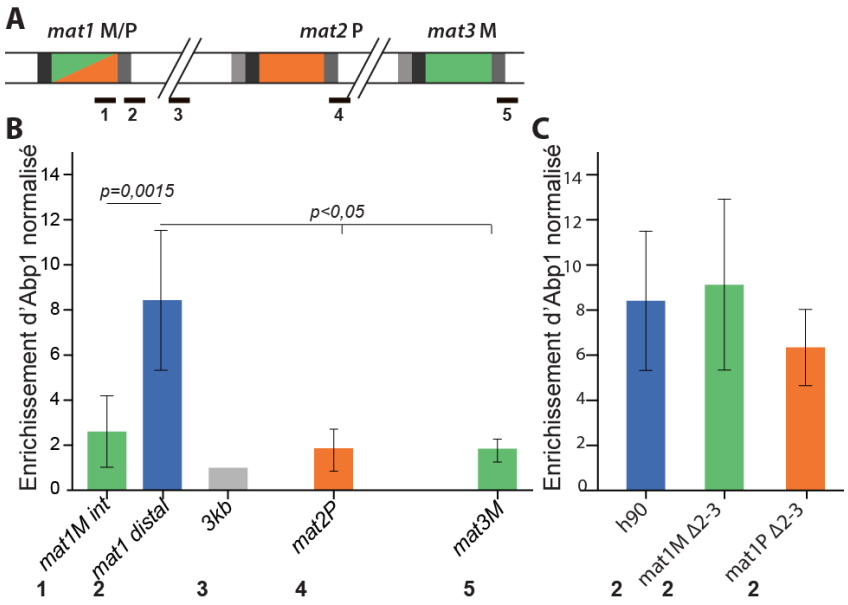


Figure 11 : Abp1 est recrutée à *mat1* indépendamment de l'allèle présent.
A) Représentation schématique du locus *mat1*, en noir, noté de 1 à 5, les couples d'amorces utilisées pour la qPCR. H1 et H2 sont les boites d'homologie qui permettent la recombinaison homologue avec les donneurs B et C) Résultat de ChIP-qPCR de la protéine Abp1. L'enrichissement est normalisé par la méthode des ΔCq en utilisant comme référence un couple d'amorces situé à 3 Kb du locus *mat1* (n°3). Les barres d'erreurs représentent les déviations standards de la moyenne (n>5). Un t-test non apparié est utilisé pour le calcul des *p values*.

2.2.2 Etude du recrutement d'Abp1 en fonction de mutations en Trans et en Cis

Dans un second temps, nous avons cherché à comprendre comment Abp1 était recrutée à *mat1*. Abp1 peut être recrutée à *mat1* grâce à une spécificité de séquence, grâce à une autre protéine ou bien par la structure que forme la pause à *MPS1* et / ou l'empreinte. Afin de déterminer lequel de ces éléments permet le recrutement d'Abp1, j'ai analysé par ChIP-qPCR son enrichissement à *mat1* dans différents contextes. J'ai donc analysé le recrutement d'Abp1 dans des mutants en *cis* (*SS2* et *M smt-0)* et dans des mutants en *trans* (*swi1*Δ et *lsd1-18*) (figure 12). Dans le mutant *SS2* une séquence de 111 pb est retirée à l'intérieur de *mat1M* et dans le mutant *M smt–0* une séquence de 263 pb est retirée du côté droit de *mat1M* (figure 12B). Ces deux mutants ne sont plus capables de changer de type sexuel (figure 12A). J'observe que l'enrichissement d'Abp1 est diminué d'un facteur trois dans la souche *mat1M smt–0* (figure 12B). Ces résultats indiquent qu'Abp1 est recrutée au niveau d'une séquence qui a été éliminée par la délétion de *mat1M smt–0* et non par celle de *SS2*.

Dans le mutant *swi1*Δ, la pause et l'empreinte sont fortement diminuées (figure 12A) (Dalgaard and Klar, 2000; Egel et al., 1984). D'autre part, le mutant *lsd1-18* présente une pause diminuée d'un facteur 2 par rapport au sauvage et une quasi absence d'empreinte (figure 12A) (Holmes

et al., 2012). J'observe une diminution du signal d'enrichissement de la protéine Abp1 d'un facteur 3 dans la souche *swi1*Δ mais pas dans une souche *lsd1-18* (figure 12C).

Afin d'exclure le fait que l'absence de Swi1 influence la quantité de la protéine Abp1, j'ai vérifié le niveau d'expression d'Abp1 par Western-blot. La quantité d'Abp1 est constante dans les différents fonds génétiques utilisés en ChIP (figure 13).

Pour conclure, au regard des résultats dans les souches *SS2* et *lsd1-18*, ni la pause ni l'empreinte ne sont nécessaires au recrutement d'Abp1 à *mat1*.

Le fait que le recrutement d'Abp1 soit affecté dans le mutant *swi1*Δ m'a amené à étudier le recrutement d'Abp1 au niveau de tout le génome en absence de Swi1.

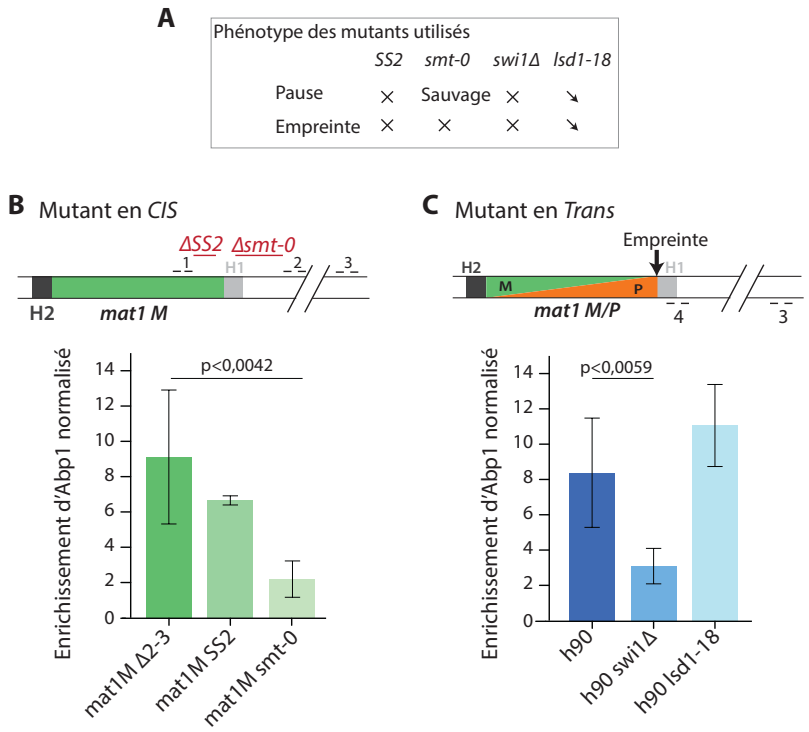


Figure 12 : Le site de liaison d'Abp1 est dans la délétion *smt–0*.
A) Tableau récapitulatif des phénotypes au niveau du CTS des mutants utilisés. B) Le panneau du haut est une représentation schématique du locus *mat1*. En noir de 1 à 3, les couples d'amorces utilisées pour la qPCR. En rouge, les délétions de séquence sont indiquées. La délétion *SS2* couvre 111 pb et la délétion *smt–0* 263 pb. H1 et H2 sont les boites d'homologie qui permettent la recombinaison homologue avec les donneurs. Le panneau du bas est le résultat de ChIP-qPCR de la protéine Abp1. Le couple 1 est utilisé pour les qPCR sur l'ADN des souches *mat1M Δ2-3* et *mat1M SS2* et le couple 2 pour l'ADN de la souche *mat1M smt-0*. C) Le panneau du haut est une représentation schématique du locus *mat1*. En noir de 4 et 3, les couples d'amorces utilisées pour la qPCR. Le panneau du bas est le résultat de ChIP-qPCR de la protéine Abp1 dans les mutants *swi1*Δ et *lsd1-18*. B et C) L'enrichissement est normalisé par la méthode des ΔCq en utilisant comme référence un couple d'amorces situé à 3 kb du locus *mat1* (n°3). Les barres d'erreurs représentent les déviations standards de la moyenne n>3. Un t-test non apparié est utilisé pour le calcul des *p values*.

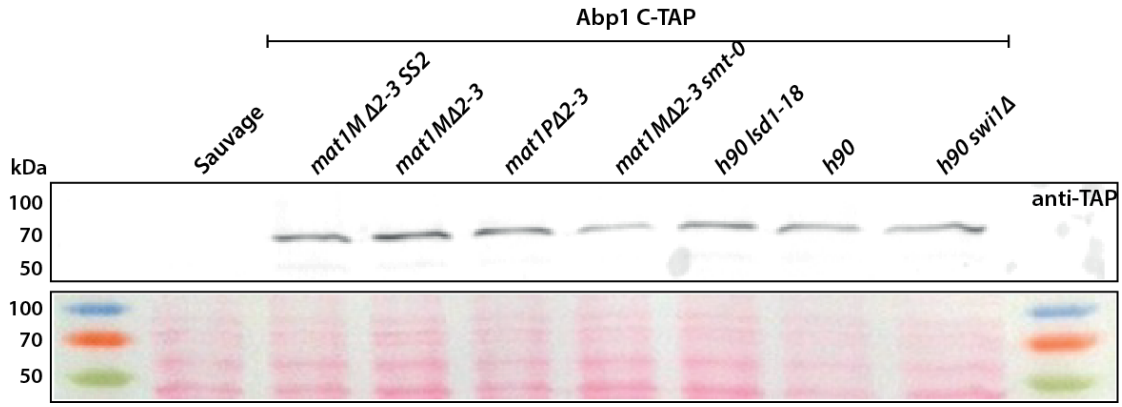


Figure 13 : Le niveau d'expression d'Abp1 est constant dans les différentes souches.
Western-blot d'Abp1-CTAP des différentes souches utilisées. Le génotype des souches utilisées pour la préparation de l'extrait protéique est annoté en haut de la membrane. Pour chaque piste, 20 µL de chaque extrait protéique sont déposés. La première piste contient 20 µL du marqueur de poids moléculaires. Les différentes tailles des bandes sont indiquées à gauche de la membrane. La deuxième piste, contient un extrait protéique isolé à partir de la souche sauvage ; les pistes 2 à 8 contiennent des extraits protéiques réalisés à partir des souches contenant la protéine Abp1-CTAP. Un anticorps reconnaissant l'étiquette de la protéine est utilisé. Le panneau du haut représente la numérisation de la membrane après la révélation, le panneau du bas correspond à une coloration au rouge Ponceau après transfert.

2.2.3 L'absence de Swi1 ne perturbe pas globalement le recrutement d'Abp1

En absence de la protéine Swi1, la perturbation de la réplication n'est pas restreinte au site *MPS1* (Noguchi et al., 2003; Noguchi et al., 2004b). Nous pouvons donc imaginer que le recrutement d'Abp1 au niveau d'autres sites dans le génome est perturbé dans ce contexte. Pour tester cette hypothèse, nous avons analysé le recrutement d'Abp1 sur l'ensemble du génome par ChIP-séq, dans le mutant *swi1*Δ.

Tout d'abord, il y a 1563 sites enrichis en Abp1 dans la souche *h90* (cf. Matériels et Méthodes) et 1780 en absence de *swi1*. Ces sites sont déterminés par le logiciel MACS2 en utilisant une *p value* de 10^{-2}. En comparant ces deux jeux de pics, j'observe que 1333 pics sont partagés. Ce qui est de l'ordre de ce qui est partagé par des réplicas biologiques (nombre de pics en commun chez les souches sauvages *h90*, *mat1M* Δ*2-3* et *mat1P* Δ*2-3*). Globalement, le recrutement d'Abp1 n'est pas affecté en absence de *swi1*. D'autre part, j'ai comparé l'enrichissement d'Abp1 dans les deux contextes au niveau des pics conservés (figure 14). Pour ce faire, j'ai utilisé les enrichissements calculés par le logiciel MACS2 (cf. Matériels et Méthodes). J'ai comparé l'enrichissement d'Abp1 dans un contexte sauvage *h90* à celui dans un contexte mutant *h90 swi1*Δ (figure 14). Nous observons une forte corrélation entre les deux enrichissements (Pearson test $R^2 = 0{,}7228$; *p value* <0.0001). Cela signifie que l'absence de Swi1 n'influence

pas le recrutement d'Abp1 au niveau de l'ensemble du génome. La diminution de l'enrichissement d'Abp1 en absence de Swi1 observée à *mat1* n'est pas un phénomène global.

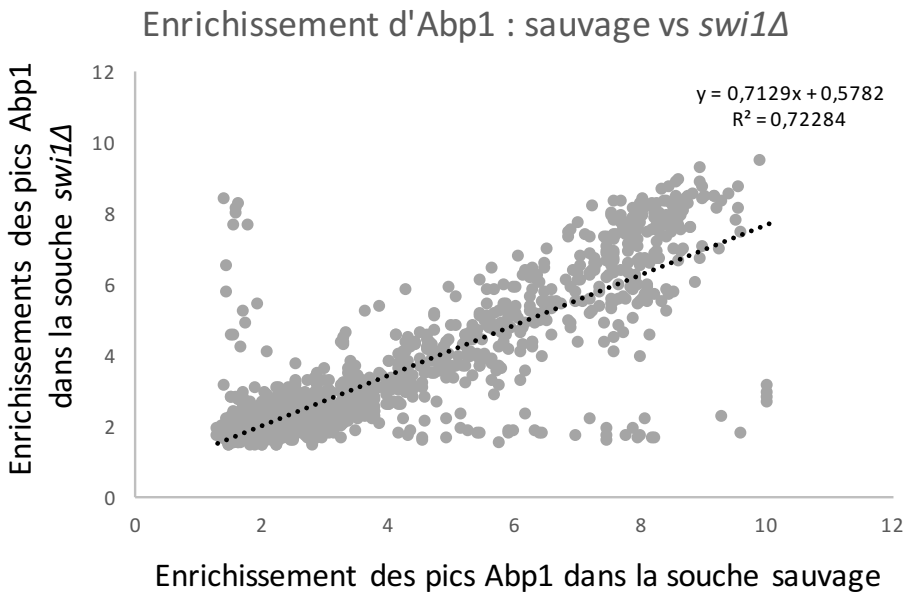


Figure 14 : Comparaison de l'enrichissement d'Abp1 dans un contexte sauvage ou en absence de Swi1.
Représentation graphique de l'enrichissement d'Abp1 dans une souche sauvage par rapport à une souche *swi1*Δ. Chaque jeu de pics et leur enrichissement ont été obtenus en utilisant le logiciel MACS2. Ensuite une intersection de ces deux jeux de données a été faite. Puis les enrichissements des pics ainsi obtenus ont été utilisés pour l'analyse. Un test de Pearson a été utilisé pour le calcul de la corrélation.

2.3 L'activité de *MPS1* et la formation de l'empreinte sont indépendantes d'Abp1

Dans le but de caractériser le rôle du recrutement d'Abp1 à *mat1*, nous avons étudié l'implication d'Abp1 dans deux étapes clefs du mécanisme :

- la pause de la fourche de réplication à *MPS1*
- la formation et le maintien de l'empreinte

Dans un premier temps, j'ai analysé l'accumulation des intermédiaires de réplication au niveau du site de pause de la fourche de réplication à *mat1* : *MPS1*. L'hypothèse était qu'Abp1 permettrait le redémarrage de la fourche par analogie à son rôle au niveau des LTRs (Zaratiegui et al., 2014). Mes résultats montrent qu'Abp1 n'est pas nécessaire au redémarrage de la fourche de réplication bloquée à *MPS1* (figure 15A), contrairement à ce qui est observé au niveau des rétrotransposons (Zaratiegui et al., 2014).

Dans un second temps, j'ai analysé le niveau de l'empreinte dans dix colonies indépendantes *h90 abp1*Δ. Nous observons que le niveau de l'empreinte n'est pas perturbé en absence d'Abp1

(figure 15C). Sur la figure 15B, les signaux de 5,4 kb et de 5 kb résultent de la cassure au niveau de l'empreinte du fragment génomique HindIII de 10,4 kb. De plus, j'ai observé que dans le mutant *abp1*Δ, *sap1-c* (mutation ponctuelle de la protéine Sap1 (Zaratiegui et al., 2014) l'empreinte est maintenue comme dans une souche sauvage (figure 15D). Le niveau d'empreinte dans le fond génétique *sap1-c* ne peut être testé car ce mutant est létal lorsqu'il n'est pas associé à la délétion d'*abp1*.

L'ensemble de ces résultats nous indique qu'Abp1 n'est pas nécessaire aux étapes précoces du mécanisme, à savoir l'activité du site *MPS1* et la formation/maintenance de l'empreinte.

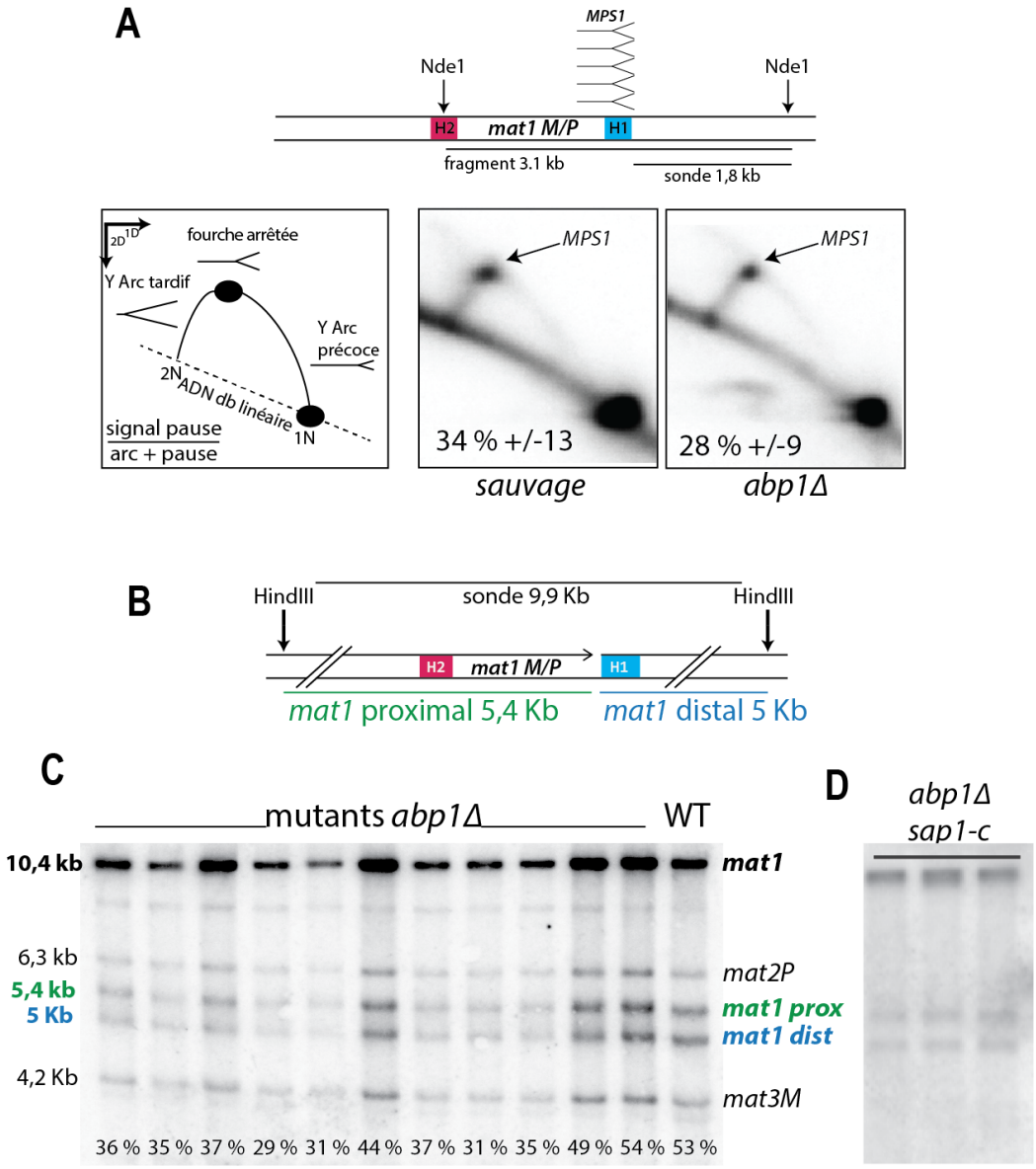


Figure 15 : Mesure de l'activité de *MPS1* et du niveau de l'empreinte.
A) Schéma de la région *mat1*, de la sonde utilisée et des sites de digestion par l'enzyme NdeI. Accumulation des intermédiaires de réplication au niveau du site *MPS1*. Schéma du résultat théorique d'un gel2D. Analyse des intermédiaires de réplication dans la souche sauvage et dans la souche *abp1*Δ. La quantification du signal d'hybridation est le rapport du signal arc total/pause. (n = 3) B) Schéma de la région *mat1,* de la sonde utilisée et des sites de digestion par l'enzyme HindIII. C) Southern-blot de l'ADN de onze colonies *abp1*Δ et d'une souche sauvage. Hybridation par une sonde de 9,9 kb produite dans une souche sauvage qui permet de révéler les trois loci *mat1*, *mat2P* et *mat3M*. En bas des pistes pourcentage qui représente la proportion de molécules qui portent l'empreinte. D) Southern-blot de l'ADN de trois colonies *abp1*Δ *sap1-c.*

2.4 Abp1 est cruciale pour l'équilibre entre les deux types sexuels

2.4.1 Lsd1 et Sap1 ne restaurent pas l'équilibre M/P en absence d'Abp1

Il a été montré qu'Abp1 et le gène Mc contrôlent l'expression de la protéine Swi2, qui en retour maintien l'équilibre entre les formes *M* et *P* en contrôlant l'accessibilité du donneur contenant l'allèle sexuel opposé (Aguilar-Arnal et al., 2008; Matsuda et al., 2011; Yu et al., 2012b). En vue d'établir le pourcentage de chaque type sexuel, j'ai utilisé la technique de qPCR en utilisant des amorces spécifiques à *mat1P* et à *mat1M*. Les résultats montrent qu'en absence d'Abp1, 90 % des cellules d'une colonie sont *M* (figure 16B). Au vu de la très faible proportion de cellules *mat1P abp1*Δ, j'ai vérifié que ce génotype n'est pas létal. Les cellules *mat1P abp1*Δ sont viables à 90% au même titre que les cellules *mat1M abp1*Δ (figure 16C). Nous remarquons également qu'une mutation dans la protéine Lsd1 engendre une faible augmentation de cellules *M* (>60%) par rapport au sauvage ; le simple mutant *sap1-c* n'est pas testé car il n'est pas viable. Enfin, les double-mutants *abp1*Δ *lsd1-18* et *abp1*Δ *sap1-c* ont le même phénotype que le simple mutant *abp1*Δ (>75% de cellules *M*) (figure 16B). Ces résultats confortent l'idée qu'Abp1 est nécessaire au changement de type sexuel en agissant sur le choix du donneur en fonction de l'information présente à *mat1*. D'autre part, la mutation *sap1-c* ne rétablit pas l'équilibre M/P an absence d'Abp1. Ce résultat montre qu'au niveau du CTS, Sap1 et Abp1 ne jouent pas des rôles antagonistes comme ce qui a pu être mis en évidence au niveau de la réplication des LTRs.

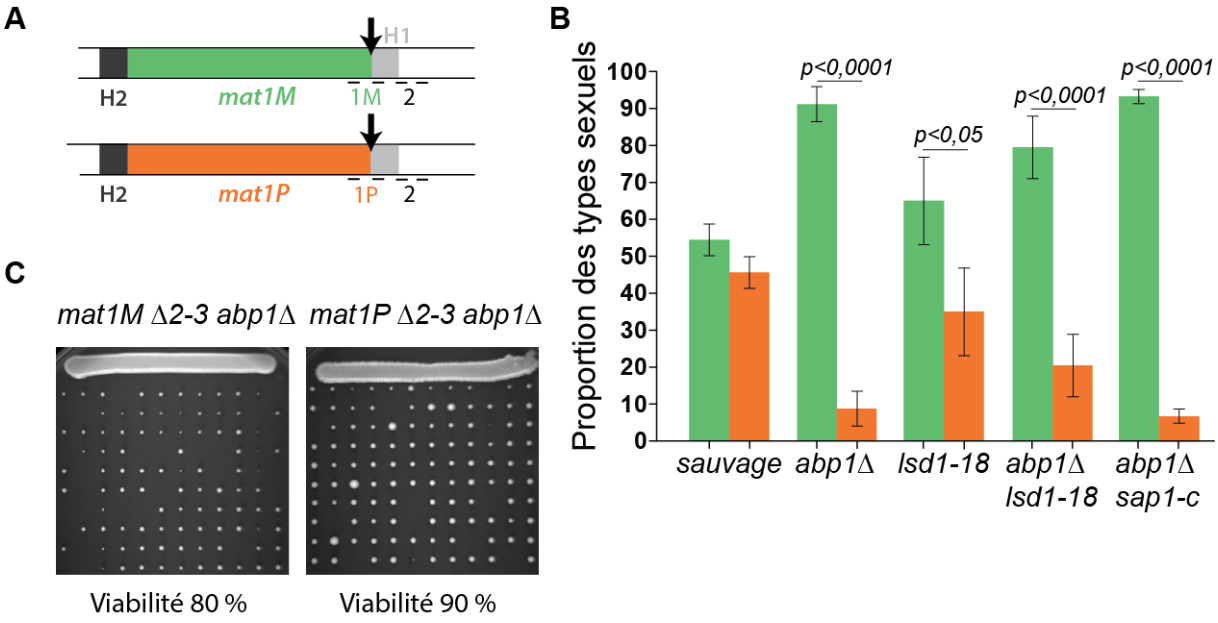


Figure 16 : La protéine Abp1 est nécessaire pour le maintien de l'équilibre entre les cellules M et les cellules P d'une population.
A) Représentation schématique du locus *mat1* avec en noir les couples d'amorces utilisés pour la qPCR. B) Représentation de la proportion des cellules M et P. La moyenne des valeurs obtenues pour les dix ADN issus de dix colonies est utilisée. La proportion de chaque type cellulaire est normalisée par la méthode des ΔCq en utilisant comme référence un couple d'amorces extérieur à *mat1*. Les barres d'erreurs représentent les déviations standards de la moyenne. Un t-test non apparié est utilisé pour le calcul des *p values*. C) Microdissection de 120 cellules de *abp1*Δ dans un fond génétique portant soit l'allèle M à *mat1* (gauche) soit l'allèle P (droite).

Bien que la mutation *sap1-c* ne rétablisse pas l'équilibre des deux types sexuels dans une population *abp1*Δ, nous remarquons qu'elle permet une croissance plus rapide de cette population en milieu riche (figure 17). Ce résultat suggère que les protéines Sap1 et Abp1 exercent des rôles antagonistes au regard du phénotype de croissance. De plus, on remarque que les simples mutants *lsd1-18* et *abp1*Δ ainsi que le double mutant *lsd1-18 abp1*Δ ont le même phénotype au regard de la croissance. Lsd1 ne semble pas participer au même mécanisme affectant la croissance que Sap1 et Abp1.

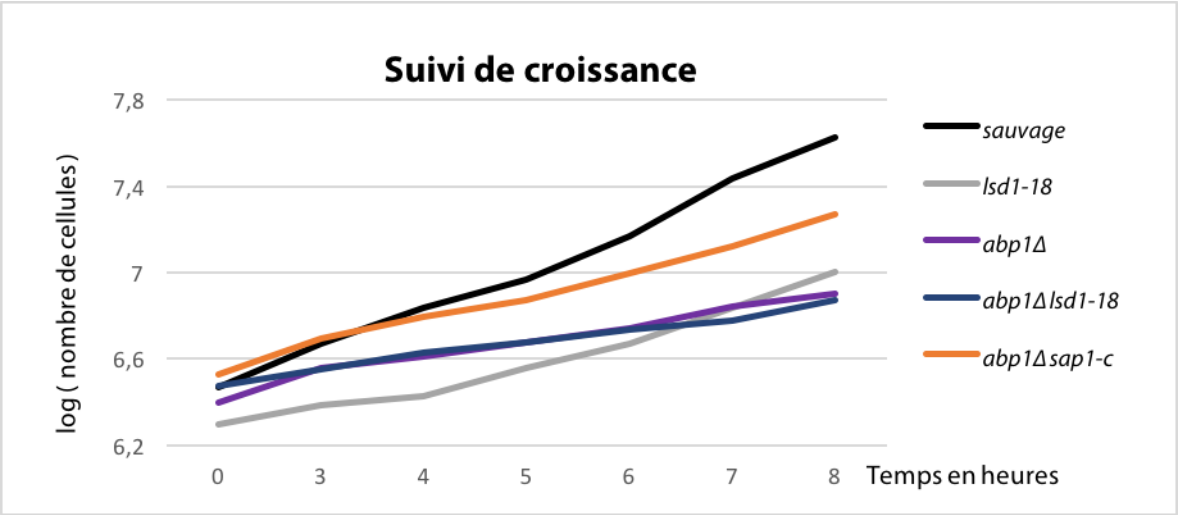


Figure 17 : Le mutant *sap1-c* favorise la croissance du mutant *abp1*Δ.
Analyse de la croissance des simples et doubles mutants *abp1*Δ, *sap1-c* et *lsd1-18*. En ordonnée, le $\log_{10}$ (nombre de cellules/ml de culture en milieu riche). En abscisse le temps en heures.

2.4.2 Swi1 agit en amont d'Abp1

En l'absence de la protéine Swi1, la pause et l'empreinte sont fortement diminuées (Dalgaard and Klar, 2000; Egel et al., 1984). Au vu des résultats précédents qui montrent qu'Abp1 n'est pas requise pour ces deux étapes, Swi1 jouerait un rôle en amont d'Abp1 dans le CTS. En absence de Swi1, l'allèle à *mat1* devrait être stable. Nous observons dans la figure 18 qu'en absence de Swi1, 6 colonies portent l'allèle *M* en majorité à *mat1* et 12 l'allèle *P*. En absence de Swi1, l'équilibre entre les deux types sexuels met probablement plus de temps à s'établir car le CTS est peu efficace. De la même façon, si Swi1 agit en amont d'Abp1, le simple mutant *swi1*Δ et le double mutant *swi1*Δ *abp1*Δ devraient avoir le même biais M/P avant d'être dans un état stable vis à vis des deux types sexuels. J'ai analysé quatre colonies, non saturées, *h90 swi1*Δ *abp1*Δ et j'ai trouvé que 3/4 des colonies sont composées en majorité de cellules de type sexuel P (figure 18). Le gène *swi1* est épistatique sur le gène *abp1* indiquant ainsi que Swi1 agit en amont d'Abp1. Cependant, si l'expérience était réalisée sur une colonie étant dans un état stable vis à vis de la proportion des deux sexuels, le phénotype d'Abp1 devrait reprendre le dessus, puisqu'en majorité les cellules réparant l'empreinte deviendront M.

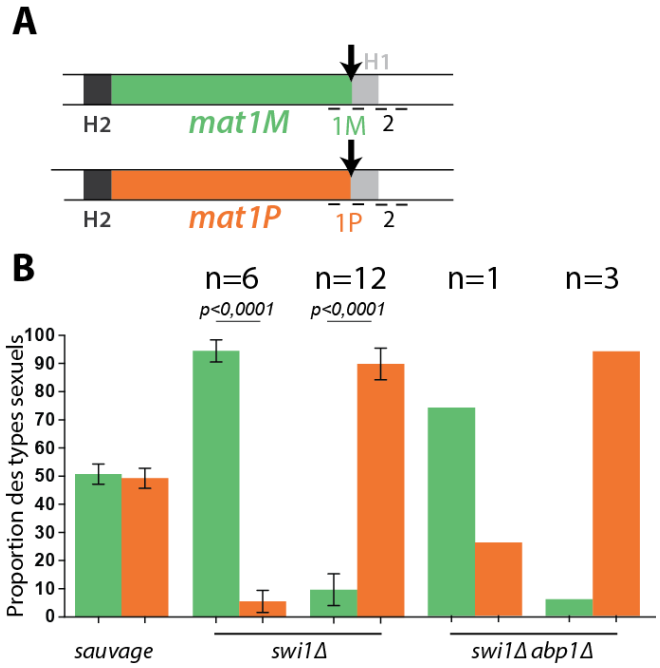


Figure 18 : Swi1 agit en amont d'Abp1.
A) Représentation schématique du locus *mat1*, en noir les couples d'amorces utilisés pour la qPCR. B) Représentation de la proportion des cellules M et P. La moyenne des valeurs obtenues est utilisée. La proportion de chaque type cellulaire est normalisée par la méthode des ΔCq en utilisant comme référence un couple d'amorces extérieur à *mat1*. Les barres d'erreurs représentent les déviations standards de la moyenne. En orange la proportion de cellules P est représentée et en vert celle de cellules M. Un t-test non apparié est utilisé pour le calcul des p values.

2.4.3 Abp1 se lie à *swi2* seulement dans les cellules M

Il a été proposé précédemment qu'Abp1 est recrutée au niveau de *swi2* et contrôle sont l'expression (Matsuda et al., 2011; Yu et al., 2012a). Nous savions que le recrutement de la protéine Mc au locus *swi2* dépend d'Abp1 et nous montrons ici que le recrutement d'Abp1 disparaît dans la souche *mat1P* Δ*2-3*, alors que la quantité de la protéine Abp1 est inchangée dans cette souche (figure 19B et 13), (Matsuda et al., 2011). De plus, seul le pic présent au niveau de *swi2* est modifié en fonction de l'allèle présent à *mat1*. En effet, le pic observé au niveau de *SPRRNA.16* par exemple est inchangé (figure 19B). Nous montrons également que ni Lsd1, ni Lsd2, ni Sap1 sont recrutées au niveau de *swi2* (figure 19A). Dans une cellule de type sexuel M, Abp1 en collaboration avec Mc est recrutée au niveau de *swi2* et engendrerait la formation d'une protéine courte de Swi2. Cette protéine se propagerait jusqu'au locus *mat2P* permettant qu'il soit choisi dans 90 % des cas pour la réparation de la cassure à *mat1* (figure 19C), (Aguilar-Arnal et al., 2008; Matsuda et al., 2011; Yu et al., 2012b).

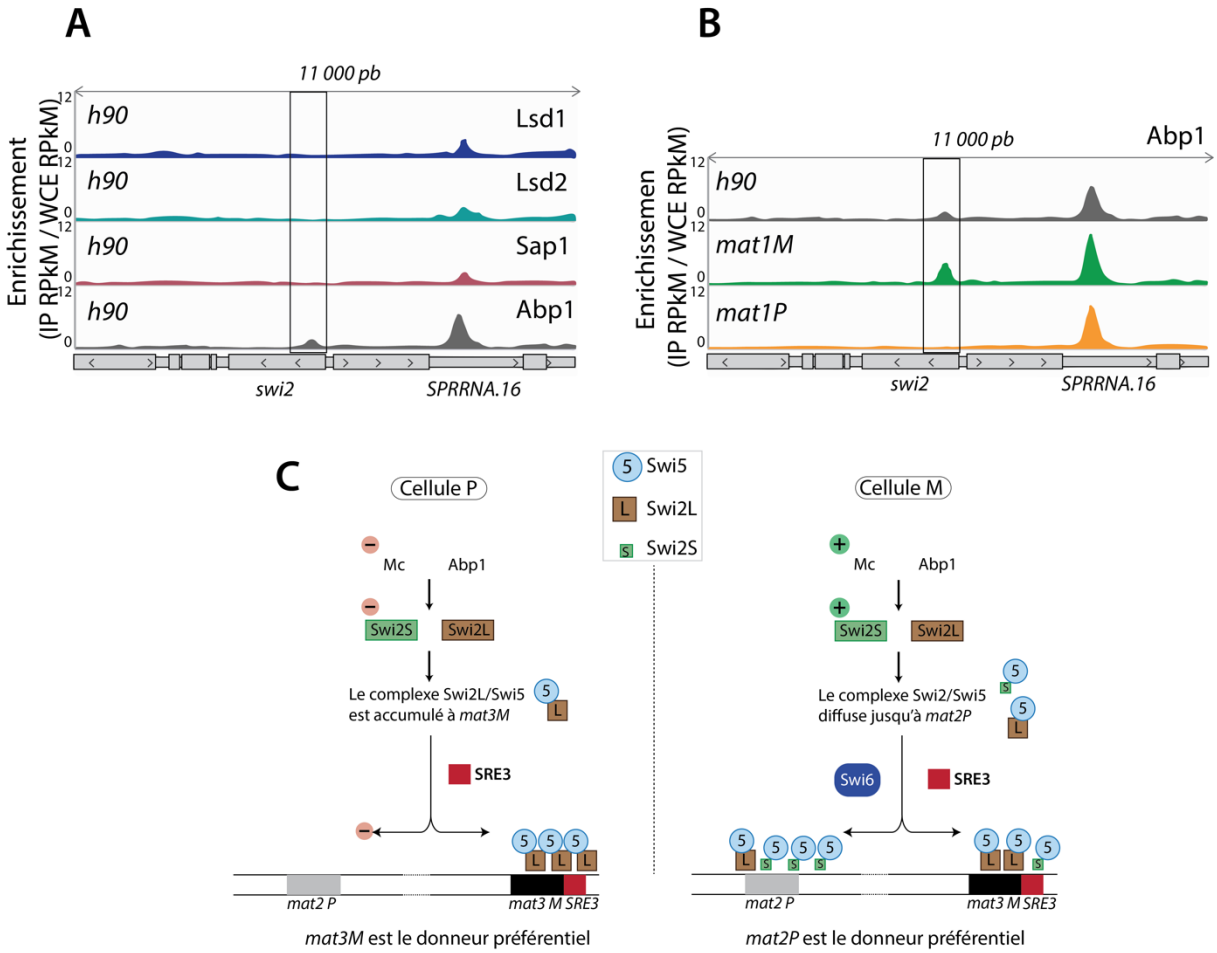


Figure 19 : Abp1 est recrutée à *swi2* en fonction de l'allèle *M*.
A) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et de Abp1 dans le fond génétique *h90* au locus *swi2*. B) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Abp1 dans le fond génétique *h90* (gris), *mat1MΔ2-3* (vert) et *mat1PΔ2-3* (orange) au locus *swi2*. C) Modèle d'action de Abp1 et Mc modifié de (Yu *et al*, 2012).

2.5 Abp1 et Cbh1 ne sont pas redondant au niveau du contrôle du CTS

La famille des Cenp-b chez *S. pombe* est composée de trois paralogues : Abp1, Cbh1 et Cbh2. J'ai analysé la proportion de cellules *M* et *P* en absence de la protéine Cbh1 car il a été montré que Cbh1 et Abp1 possèdent des fonctions redondantes (Baum and Clarke, 2000; Irelan et al., 2001; Lee et al., 1997). A la différence d'Abp1, Cbh1 n'est pas nécessaire au maintien de l'équilibre M et P (figure 20C). Le double mutant *abp1Δ cbh1Δ* n'a pas été testé car il pousse très lentement. Le triple mutant *abp1Δ cbh1Δ sap1-c* a le même phénotype que le simple mutant *abp1Δ*. De plus, par ChIP-qPCR, on ne détecte pas d'enrichissement de Cbh1 à *mat1* et cela même en absence d'Abp1 ou de Swi1 (figure 20D). Cbh1 et Abp1 n'ont probablement pas les

mêmes sites de liaison et ils n'exercent pas la même fonction dans le mécanisme de CTS. Ces résultats suggèrent que la protéine Abp1 agit en tant qu'homodimère.

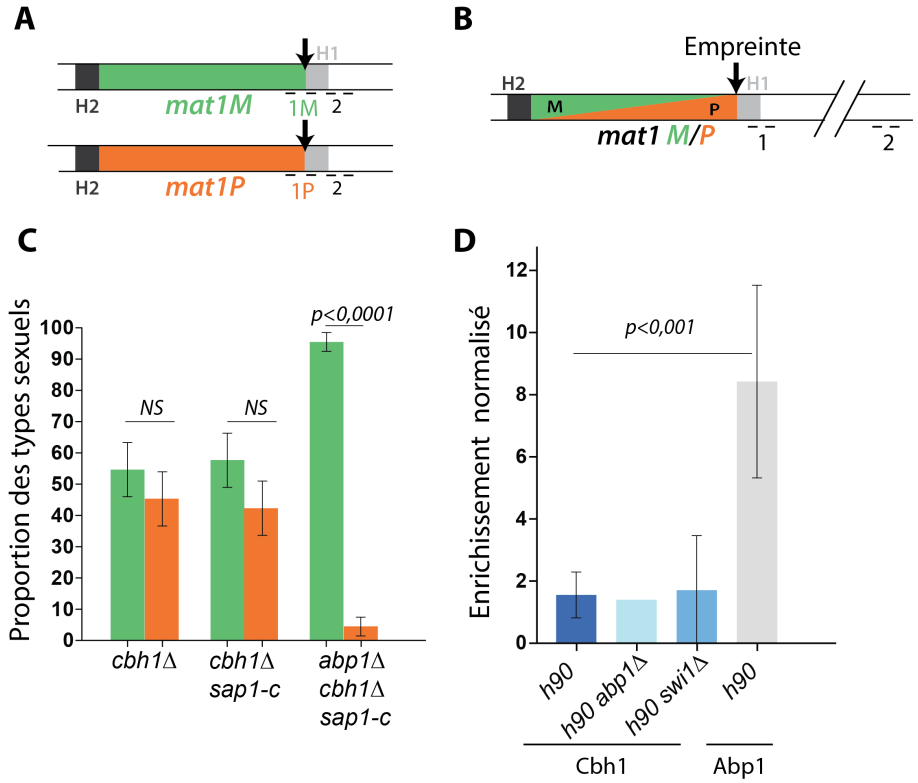


Figure 20 : La protéine Cbh1 ne joue pas le même rôle qu'Abp1 au niveau du CTS.
A) Représentation schématique du locus *mat1* avec en noir les couples d'amorces utilisés pour la qPCR pour tester les proportions des deux types sexuels. B) Représentation schématique du locus *mat1* avec en noir les couples d'amorces utilisés pour la ChIP-qPCR d'Abp1 et de Cbh1. C) Représentation de la proportion des cellules M et P. La moyenne des valeurs obtenues pour les dix ADN issus de dix colonies est utilisée. La proportion de chaque type cellulaire est normalisée par la méthode des ΔCq en utilisant comme référence un couple d'amorces extérieur à *mat1*. Les barres d'erreurs représentent les déviations standards de la moyenne. Un t-test non apparié est utilisé pour le calcul des p values. D) Résultat de ChIP-qPCR de la protéine Cbh1 en bleu et Abp1 en gris dans différents fonds génétiques. L'enrichissement est normalisé par la méthode des ΔCq en utilisant comme référence un couple d'amorces situé à 3Kb du locus *mat1* (B, couple 2). Les barres d'erreurs représentent les déviations standards de la moyenne. n=4 sauf pour le fond génétique *h90 abp1*Δ n=1. Un t-test non apparié est utilisé pour le calcul des *p values*.

2.6 L'insertion du promoteur *nmt1* à droite de *mat1* diminue le recrutement d'Abp1

Afin d'établir à quel moment du CTS Abp1 est recrutée à *mat1,* j'ai utilisé un système inductible développé au laboratoire, qui permet de disséquer les étapes du mécanisme (figure 21A, (Holmes et al., 2005). Le promoteur inductible *nmt1* est introduit à 196 pb de *mat1*. La transcription via *nmt1* élimine l'empreinte, et les cellules deviennent stable *M* ou *P* (en fonction

de l'allèle présent à *mat1* lors de la transcription). Cela permet de synchroniser toutes les cellules de la population pour le CTS. Après répression de *nmt1,* les cellules M (ou P) sont de nouveau capables de changer de type sexuel. Au cours du premier cycle de réplication, toutes les cellules vont former une empreinte au niveau du brin retardé. Durant la deuxième réplication, elles répareront cette empreinte en changeant l'allèle présent à *mat1*. J'ai analysé par ChIP-qPCR à quel moment Abp1 est recrutée à *mat1* dans ce système. Mes résultats n'ont pas été concluants vis-à-vis de la période de recrutement d'Abp1. En effet, comme le montre la figure 19B, Abp1 n'est pratiquement plus recrutée à *mat1* dans ce système. L'analyse du potentiel site de fixation grâce aux résultats de ChIP-séq à haut débit a permis de montrer qu'il est disrupté dans ce fond génétique tout comme dans le fond génétique *M smt–0* (figure 12).
L'utilisation du logiciel MACS a permis la détection de 1083 pics. En utilisant le logiciel RSAT j'ai déterminé les motifs d'ADN fixés par Abp1. Dans le pic présent à *mat1*, plusieurs motifs suggérés par RSAT sont retrouvés : les motifs proposés sont riches en A/T (figure 21D). Ces sites sont proches de l'insertion du promoteur *nmt1* (196 pb de l'empreinte). Cela peut expliquer pourquoi Abp1 n'est pas recrutée à *mat1* dans le système inductible précédemment utilisé. Le site préférentiel de fixation d'Abp1 à *mat1* serait donc "TAATATTATA" (figure 21C et 21D).
En présence de *nmt1* réprimé, l'efficacité du CTS n'est pas affectée (Holmes et al., 2005). La cellule est capable de produire la pause, l'empreinte et de choisir le donneur de façon comparable à la souche sauvage. Cependant, en présence de *nmt1*, Abp1 n'est plus liée du côté centromère distal de *mat1*. Cela indique que la présence d'Abp1 à *mat1* n'est pas requise pour le CTS. Nous pouvons imaginer qu'à *mat1*, Abp1 joue un rôle d'optimisation. En son absence, la variation de l'efficacité du CTS est faible et donc difficilement analysable.

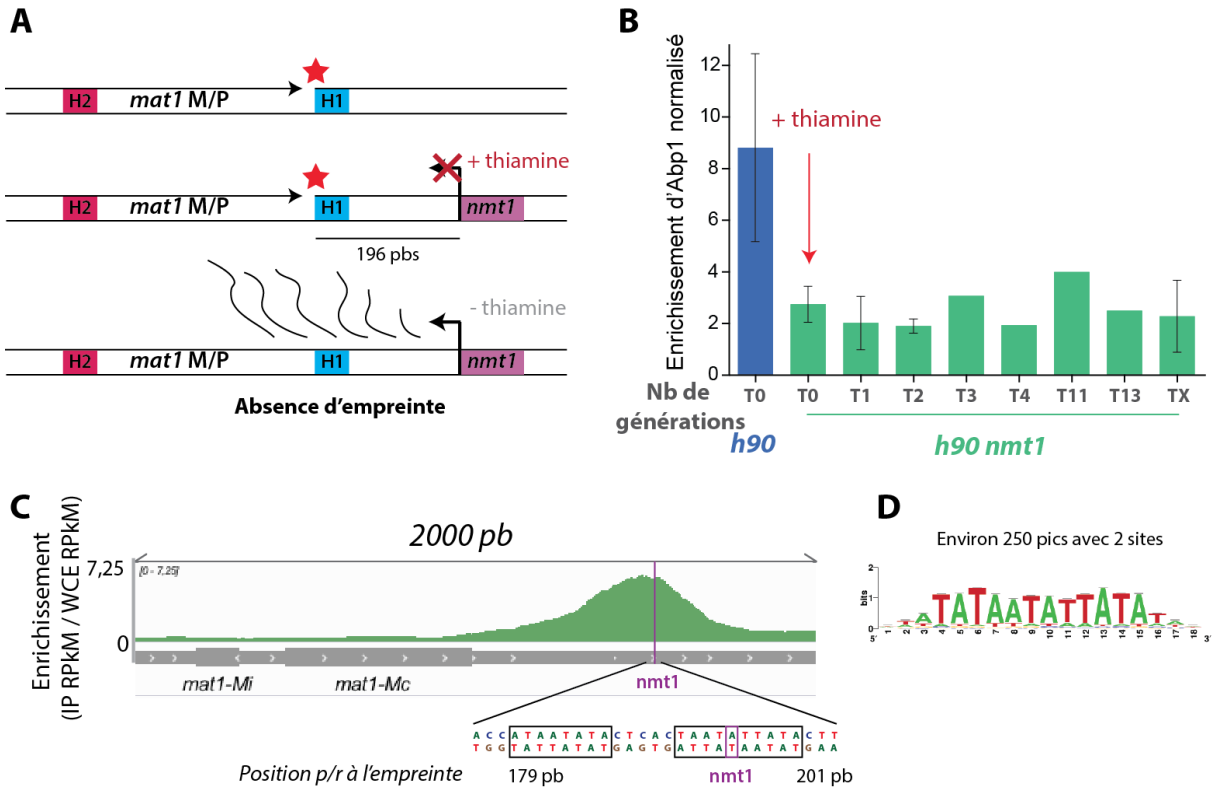


Figure 21 : Le système *nmt1* ne permet pas d'étudier le recrutement d'Abp1.
A) Représentation schématique du locus *mat1*. H1 et H2 sont les boites d'homologie qui permettent la recombinaison homologue avec les donneurs. Le promoteur inductible *nmt1*est représenté en violet. En absence de thiamine la transcription est activée, et l'empreinte n'est pas formée. B) Résultat de ChIP-qPCR de la protéine Abp1 en fonction du temps après arrêt de la transcription. Le temps X représente une culture qui n'a été incubée que dans un milieu avec de la thiamine (de la pré-culture à l'expérience). L'enrichissement est normalisé par la méthode des ΔCq en utilisant comme référence un couple d'amorces situé à 3 Kb du locus *mat1*. Les barres d'erreurs représentent les déviations standards de la moyenne. C) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) de la ChIP-séq de Abp1 dans le fond génétique *mat1MΔ2-3*. Le sens de la transcription de *Mc* et *Mi* est indiqué. La ligne violette indique le site d'insertion du promoteur *nmt1*. Les sites de liaison potentiels d'Abp1 déterminés par le logiciel RSAT sont indiqués (carrés noirs). D) Représentation d'un des sites potentiels déterminés par RSAT et présent au centre de 250 pics.

2.7 Conclusions

En résumé, Abp1 est recrutée du côté centromère-distal de *mat1* au niveau d'une séquence absente dans les souches *M smt–0* et *nmt1* (figure 11, 12 et 21). Tout comme au niveau des LTRs, une pause de la fourche de réplication est observée à *mat1* au niveau du site *MPS1* (Dalgaard and Klar, 1999; Zaratiegui et al., 2014). Cependant, en absence d'Abp1 l'accumulation des intermédiaires de réplication à *mat1* est la même que dans une souche sauvage (figure 15). De plus, l'insertion du promoteur *nmt1* au niveau du site de recrutement d'Abp1 à *mat1* n'abolit pas le CTS (figure 21). Enfin, il a été montré qu'en absence d'Abp1 la majorité des cellules sont de type sexuel M, un phénotype non sauvé par la mutation *sap1-c* (figure 16). Ces résultats suggèrent qu'à contrario du contrôle des fourches de réplication au niveau des LTRs et de l'ADNr (Zaratiegui et al., 2014), Abp1 et Sap1 ne jouent pas des rôles antagonistes dans le CTS et qu'Abp1 joue un rôle mineur au niveau de *mat1*. Malgré le recrutement d'Abp1 au niveau du site initiateur du mécanisme, son rôle au niveau du CTS n'intervient que dans le choix du donneur, une étape tardive du mécanisme (figure 16, 18 et 19). Son rôle dans le choix des donneurs est expliqué essentiellement par son action sur *swi2* (Yu *et al*, 2012 et Matsuda *et al*, 2011).

3. Analyse du recrutement de Lsd1/2, Sap1 et Abp1 sur l'ensemble du génome

La colocalisation de Sap1, Abp1, Lsd1 et de Lsd2 au niveau de *mat1* ainsi que leurs rôles transcriptionnels et réplicatifs ont motivé une analyse de la répartition de ces différents acteurs sur l'ensemble du génome.

3.1 Analyse des sites de recrutements de Lsd1/Lsd2, Sap1 et Abp1 sur l'ensemble du génome

Grâce à la technique de ChIP-séq, j'ai identifié des régions préférentiellement liées par chacune des protéines étudiées. La méthode d'analyse décrite dans le Matériels et Méthodes m'a permis de valider statistiquement ces ensembles de pics. Nous constatons que le nombre de pics varie de 573 à 1083 en fonction de la protéine immunoprécipitée (tableau 1). Selon la méthode décrite par le " ENCODE Consortium" le nombre de pics statistiquement significatif est celui obtenu à la suite de l'analyse IDR (Irreproducibility Discovery Rate). Cette analyse utilise les réplicas biologiques et les pseudo réplicas dans le but de déterminer un seuil statistique. Ce seuil sera ensuite utilisé pour choisir un ensemble de pics "statistiquement significatif" (cf. Matériels et Méthodes). Ici les seuils utilisés sont indiqués dans le tableau 1. La valeur utilisée comme seuil pour la ChIP-séq de Lsd1 est 10 fois inférieure aux seuils des autres ChIP-séq. Ceci peut être expliqué par le fait que la ChIP-séq de Lsd1 est moins efficace. Le tableau 2 présente la comparaison de nos données avec celles déjà publiées. Nous remarquons une forte variabilité entre les différentes études. Dans un premier temps, nous pouvons dire que les données de ChIP-ChIP sont difficilement comparables à celles de ChIP-séq puisque la méthode de détection des pics et la résolution n'est pas la même. Cependant, les conclusions biologiques ne devraient pas être affectées (cf. figure 22, 23 et 24). En outre, le nombre de lectures dans les ChIP-séqs publiées par Zaratiegui, M. est inférieure à celui dans notre étude (Communication personnelle). Cela influence largement la détection des pics par le logiciel MACS2. En ce qui concerne le nombre de pics d'Abp1 publié par Daulny, A. nous remarquons qu'il est comparable à celui observé dans chaque réplica biologique avant l'IDR (cf. Matériels et Méthodes). L'analyse statistique que j'ai réalisée entraine une réduction de ce nombre.

	Nombre de pics après IDR	Seuil = -Log_{10}(*p value*)
Lsd1	573	10,22646
Lsd2	682	19,19591
Sap1	958	26,07609
Abp1	1083	25,11338

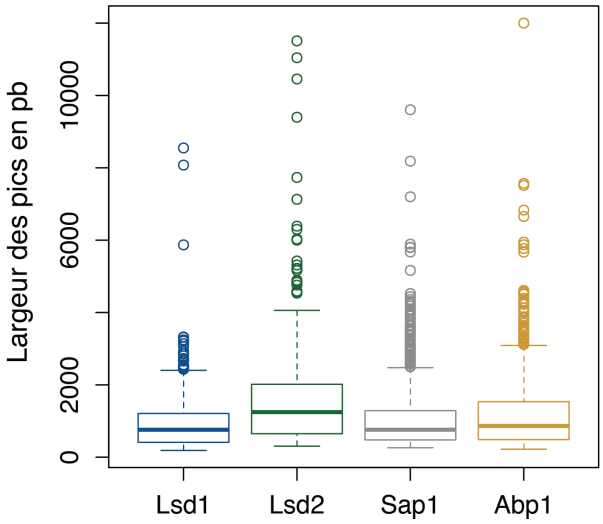
Tableau 1 : Analyse du nombre de pics.
Les différents nombres de pics, statistiquement validés, obtenus pour Lsd1, Lsd2, Sap1 et Abp1 sont indiqués dans la deuxième colonne. La troisième colonne indique le "–log_{10}(*p value*)", déterminée par l'IDR, utilisé comme seuil pour déterminer le nombre de pics statistiquement significatifs.

Etudes	Années	Techniques	Lsd1	Lsd2	Sap1	Abp1
Nicolas *et al*	2006	ChIP-ChIP	105	164		
Gordon *et al*	2007	ChIP-ChIP		426		
Lan *et al*	2007	ChIP-ChIP	370	283		
Cam *et al*	2008	ChIP-ChIP				NA
Zaratiegui *et al*	2011	ChIP-séq			151	136
Daulny *et al*	2016	ChIP-séq				1623
Cette étude	2016	ChIP-séq	573	682	958	1083

Tableau 2 : Comparaison avec les données publiées.
Les différentes études à haut débit sont répertoriées, ainsi que le nombre de pics trouvés dans chacune. La technique utilisée dans chaque étude est indiquée ainsi que l'année de publication.

Dans un second temps, la distribution de la largeur des pics a été analysée (figure 22). Nous remarquons que la médiane de la largeur des pics est comprise entre 761 pb et 1247 pb en fonction de la protéine immunoprécipitée. La protéine Lsd2 se lie à des régions en moyenne plus larges comparées aux trois autres protéines. En ce qui concerne Lsd1, Sap1 et Abp1, les régions enrichies sont de taille similaire (environ 800 bp). Nous avons exclu de cette analyse les extrémités du chromosome 3 qui donnent deux pics très larges (supérieurs à 15 kb). En effet, ces pics sont dus aux répétitions de l'ADNr et télomèriques. Dans l'ensemble, nous remarquons que la majorité des pics pour chaque protéine a une taille inférieure à 2 kb ce qui nous permet de conclure que ces protéines ne recouvrent pas de larges régions mais sont plutôt réparties de façon ponctuelle sur l'ensemble du génome.

Dans un troisième temps, nous constatons que les pics sont répartis uniformément sur l'ensemble du génome (figure 23). Nous ne pouvons pas déterminer de région génomique ayant une densité de pics plus importante. Ce résultat nous indique qu'il n'y a pas de biais de répartition des pics sur l'ensemble du génome.



Analyse de la distribution des pics

	Lsd1	Lsd2	Sap1	Abp1
Min	188	305	264	222
1er Qu	410	650.2	480	488
Mediane	**761**	**1247**	**759.5**	**864**
Moyenne	**967.3**	**1559.5**	**1094.8**	**1188**
3ème Qu	1209	2016.2	1286.8	1534
Max	8549	11508	9605	12000

Figure 22 : Analyse de la largeur des pics.
Le panneau de gauche représente la distribution de la largeur des pics en fonction de la protéine immunoprécipitée : Lsd1 (bleu), Lsd2 (vert), Sap1 (gris) et Abp1 (jaune). Le panneau de droite indique les valeurs statistiques de ces distributions.

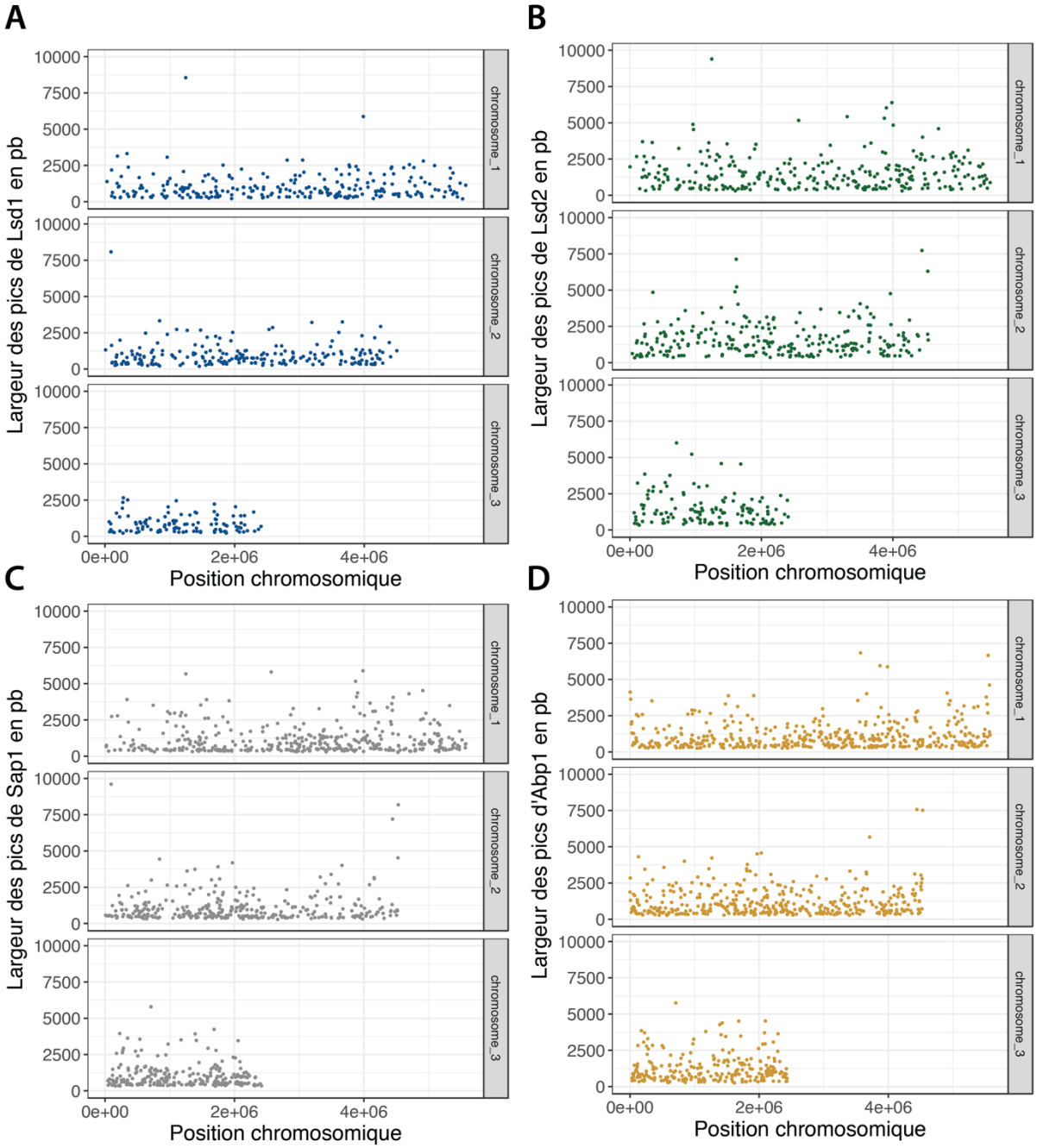


Figure 23 : Analyse de la répartition des pics sur l'ensemble du génome.
La largeur (en bp) de chaque pic est représentée en fonction de la position du pic dans le génome (en bp). A) Répartition génomique de la distribution de la largeur des pics de Lsd1. B) Répartition génomique de la distribution de la largeur des pics de Lsd2. C) Répartition génomique de la distribution de la largeur des pics de Sap1. D) Répartition génomique de la distribution de la largeur des pics de Abp1.

Dans un quatrième temps, nous avons déterminé quel type de régions biologiques étaient particulièrement enrichies pour chacune des protéines étudiées. Pour ce faire, nous avons utilisé les annotations du génome de *S. pombe* disponibles sur la base de données "pombase" (http://www.pombase.org/) auxquelles nous avons ajouté les coordonnées des régions répétées. Ces 238 régions répétées sont les rétrotransposons complets et les solo LTRs. Pour les quatre protéines, la majorité des pics se répartit au niveau des régions codantes (exons, miRNA, gènes, CDS, 3'/5' UTR, transcrits) (figure 24). Plus particulièrement, les ChIP-ChIP de Lsd1 et Lsd2 précédemment publiées montrent également que ces protéines sont enrichies au niveau de régions activement transcrites (Lan et al., 2007; Nicolas et al., 2006). De façon intéressante, nous observons, qu'à la différence des protéines Lsd1/2, les protéines Sap1 et Abp1 sont également enrichies dans des régions silencieuses appelées "repeat region". Ces régions rassemblent les rétrotransposons et les LTRs. Ces résultats corroborent le fait que Sap1 et Abp1 contrôlent la réplication des LTRs (Zaratiegui et al., 2011b), mais également le fait que Abp1 participe à la répression transcriptionnelle des rétrotransposons (Cam et al., 2008; Daulny et al., 2016).

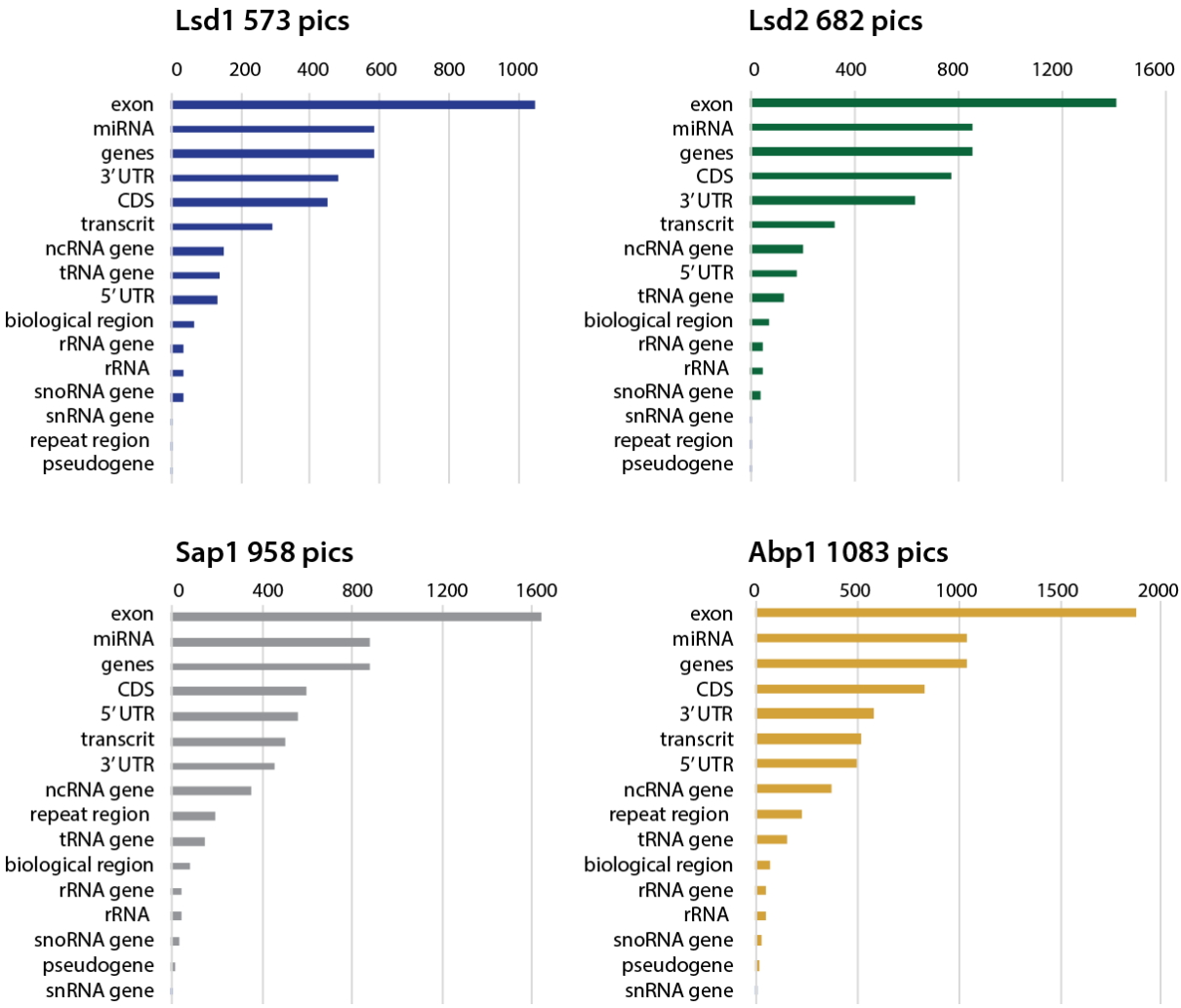


Figure 24 : Analyse des régions enrichies.
Représentation du nombre de régions chromosomiques, classées par type, chevauchées par les pics d'enrichissement déterminés précédemment. Lsd1 (bleu), Lsd2 (vert), Sap1 (gris) et Abp1 (jaune). Le nombre de pics est également indiqué.

3.2 Lsd1/Lsd2, Sap1 et Abp1 partagent des sites avec Rad52

Les déméthylases Lsd1 et Lsd2 font partie d'un complexe et exercent des fonctions partiellement redondantes (Gordon et al., 2007; Holmes et al., 2012). Sans surprise, Lsd1 et Lsd2 partagent la majorité de leurs pics comme le montre le diagramme de Venn de la figure 25A. La validation statistique de toutes les analyses d'intersection est détaillée dans la partie Matériels et Méthodes.

Nous avons également analysé le chevauchement des pics des protéines Abp1 et Sap1. En effet, ces deux acteurs cellulaires agissent sur des régions similaires : les LTRs (Zaratiegui et al., 2011b). Nous constatons que les protéines Abp1 et Sap1 partagent la moitié de leurs pics (figure 25B), mais que seule une partie de ces derniers sont retrouvés au niveau des régions

répétées (figure 25C). Ce résultat montre qu'Abp1 et Sap1 ont d'autres cibles communes. J'ai également répertorié le nombre de pics partagés par Lsd1-2 et par Sap1 Abp1 déjà publié (tableau 3). Une fois de plus, nous pouvons remarquer que le nombre de pics partagés diverge en fonction des techniques utilisées. L'utilisation de trois expériences de ChIP–séq indépendantes pour chaque protéine me permet de valider statistiquement les pics et les intersections.

Les quatre protéines partagent également 314 sites d'enrichissement. Sachant que ces protéines ont comme point commun le contrôle de fourches de réplications en pause à différents endroits dans le génome, nous avons décidé d'étudier de façon globale leur implication au niveau de zones "difficiles à répliquer" (Holmes et al., 2012; Krings and Bastia, 2005; Mejia-Ramirez et al., 2005; Zaratiegui et al., 2011b) . Plusieurs études ont cartographié ces zones qui sont susceptibles d'engendrer des arrêts de fourche (Magdalou et al., 2014; Sabouri et al., 2012). Il a été montré que la recombinaison homologue peut être utilisée pour le redémarrage de la réplication dans ces zones (Ahn et al., 2005; Irmisch et al., 2009; Sabouri et al., 2012).

Nous avons donc analysé le lien entre ces protéines et la recombinaison homologue (RH). A cette fin, j'ai utilisé les données publiées de ChIP-séq de Rad52 afin de les comparer avec les protéines d'intérêts (Castel et al., 2014). Dans un premier temps, j'ai analysé la ChIP-séq de Rad52 en utilisant les mêmes paramètres que pour les autres ChIP-séq. Pour obtenir un jeu de pics statistiquement significatif, j'ai utilisé un seuil ($-\mathrm{Log}_{10}(p\ value)=10$) qui a été choisi en fonction des seuils déterminés précédemment par les analyses IDR (tableau 1). J'ai obtenu 382 pics significativement enrichis en Rad52 sur l'ensemble du génome. L'analyse de ces pics nous montre qu'ils ne sont pas présents dans les régions répétées, à contrario des pics partagés par Abp1 et Sap1 (figure 25C). Enfin, nous remarquons que 227 pics sont partagés par les protéines Lsd1/2, Sap1, Abp1 et Rad52, soit environ 60 % des pics de Rad52. Ces résultats suggèrent un potentiel rôle des protéines étudiées sur le contrôle de ces zones "difficilement répliquées".

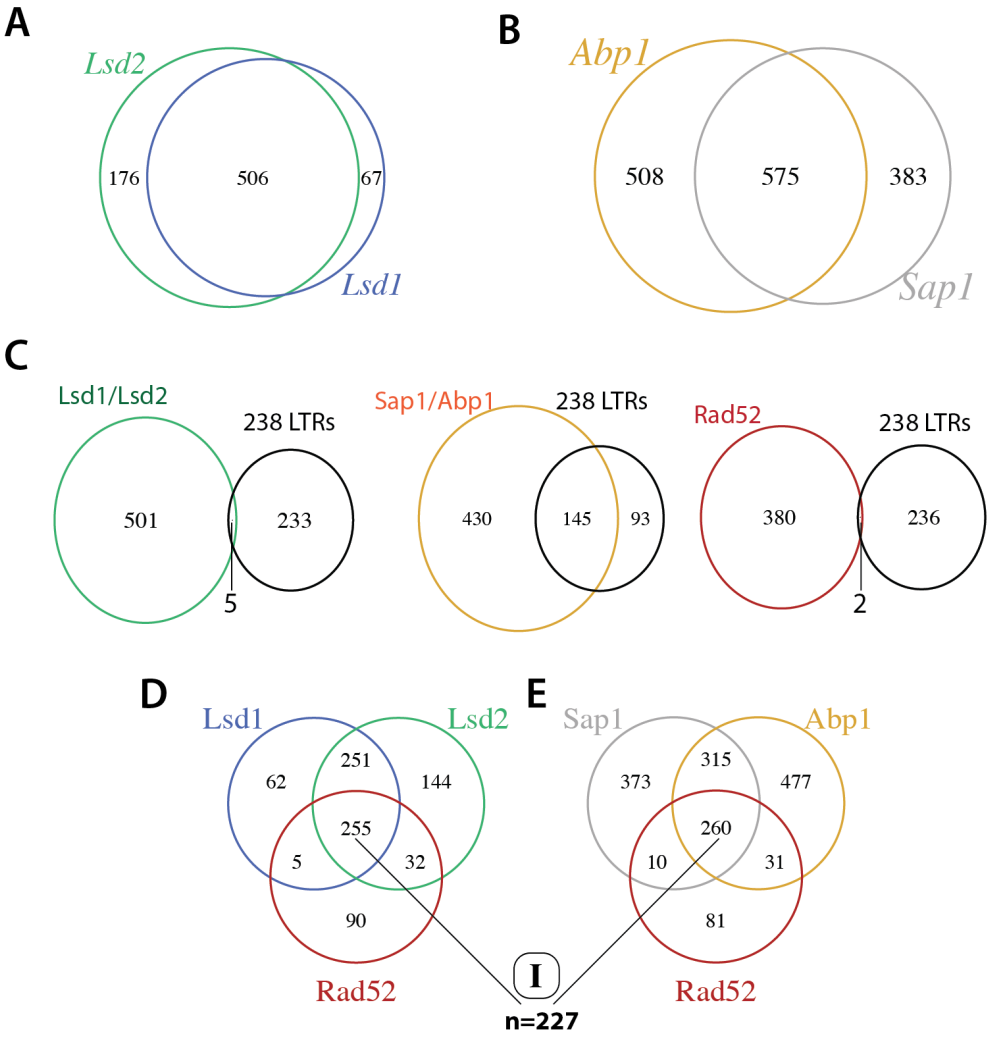


Figure 25 : Analyse des intersections des régions enrichies.
A) Diagramme de Venn représentant l'intersection des pics Lsd1 et Lsd2. B) Diagramme de Venn représentant l'intersection des pics Sap1 et Abp12. C) Le panneau de gauche représente un diagramme de Venn qui montre l'intersection des pics Lsd1/ Lsd2 et les LTRs. Le panneau du milieu représente un diagramme de Venn qui montre l'intersection des pics Sap1/Abp1 et les LTRs. Le panneau de droite représente un diagramme de Venn qui montre l'intersection des pics Rad52 et les LTRs. D) Diagramme de Venn représentant l'intersection des pics Lsd1, Lsd2 et Rad52. E) Diagramme de Venn représentant l'intersection des pics Abp1, Sap1 et Rad52. Le nombre de pics partagés par Lsd1, Lsd2, Sap1, Abp1 et Rad52 est indiqué (n=227).

Etudes	Années	Techniques	Lsd1	Lsd2	Sap1	Abp1	Lsd1 I Lsd2	Sap1 I Abp1
Nicolas *et al*	2006	ChIP-ChIP	105	164			94	
Lan *et al*	2007	ChIP-ChIP	370	283			93	
Zaratiegui *et al*	2011	ChIP-séq			151	136		105
Cette étude	2016	ChIP-séq	573	682	958	1083	506	575

Tableau 3 : Comparaison des intersections avec les données publiées
Les différentes études à haut débit sont répertoriées. Les deux dernières colonnes indiquent le nombre de pics partagés par les protéines Lsd1 et Lsd2 (noté Lsd1 I Lsd2), ou par Sap1 et Abp1 (noté Sap1 I Abp1).

3.3 Ces protéines sont recrutées au niveau de gènes fortement transcrits

Récemment, la participation de Rad52, une protéine importante pour la réparation de l'ADN, au contrôle de la réplication de zones hautement transcrites sujettes à des arrêts de fourche, a été établie (Castel et al., 2014). J'ai donc analysé plus finement les régions enrichies en Lsd1, Lsd2, Sap1, Abp1 et Rad52. Le postulat principal repose sur l'idée que ces protéines devraient se lier au niveau de gènes hautement transcrits, car ils sont décrits comme pouvant engendrer des collisions entre la fourche de réplication et la machinerie transcriptionnelle (Magdalou et al., 2014; Sabouri et al., 2012).

L'analyse des régions enrichies en Lsd1/2, Sap1, Abp1 et Rad52 révèlent qu'elles se situent en majorité au niveau d'exons, comme l'indique la figure 26A. Nous avons ensuite décidé d'analyser plus en détail ces régions et de déterminer si elles recouvrent des gènes fortement exprimés ou non. Pour cela, je me suis basée sur l'analyse de la transcription réalisée précédemment chez *S. pombe* (Marguerat et al., 2012). Trois groupes de transcription ont été déterminés :

- Un groupe de 108 gènes fortement transcrits ; c'est-à-dire présentant plus de 200 copies d'ARN par cellule (cpc)
- Un groupe de 108 gènes moyennement transcrits ; c'est-à-dire présentant de 5.5 à 6 cpc
- Un groupe de 108 gènes faiblement transcrits ; c'est-à-dire présentant moins de 1 cpc

Nous avons ensuite réalisé l'intersection de l'ensemble des 227 pics partagés par les protéines étudiées avec chacun des groupes de gènes définis. Le résultat indique que 30 % des gènes fortement transcrits sont enrichis en Lsd1/2, Sap1, Abp1 et Rad52 (figure 26B). Cette analyse suggère que les régions difficiles à répliquer nécessitent l'action de régulateurs de la réplication et de modulateurs de la chromatine.

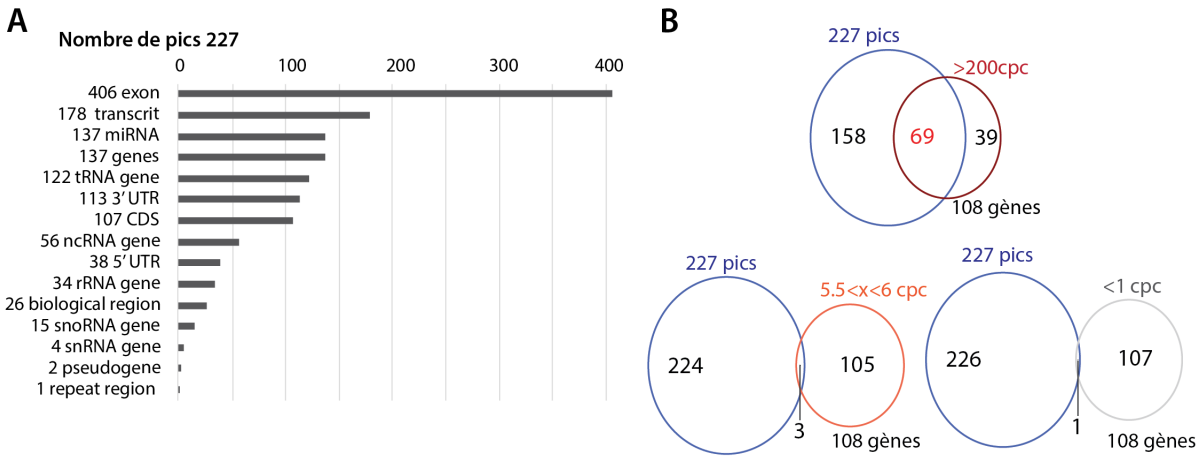


Figure 26 : Analyse des pics en fonction de la transcription.
A) Représentation du nombre de régions chromosomiques, classées par type, chevauchées par les pics d'enrichissement déterminés précédemment. Le nombre de pics est également indiqué. B) Diagramme de Venn représentant l'intersection des pics analysés avec les trois groupes de gènes définis. En rouge les gènes fortement exprimés (>200cpc), en orange les gènes moyennement exprimés (5.5<cpc<6) et en gris les gènes faiblement exprimés (<1cpc).

3.4 Conclusions

Le rôle du complexe Lsd1 et Lsd2 au niveau de la régulation de la réplication est peu documenté. Comme montré précédemment, ce complexe est retrouvé au niveau de gènes fortement exprimés. Ce résultat peut être corrélé au fait que ces zones sont difficilement répliquées mais aussi au fait que ce complexe participe à la régulation de la transcription (Gordon et al., 2007; Lan et al., 2007; Nicolas et al., 2006). Par ailleurs, le fait qu'une partie des cibles du complexe, en dehors de ces gènes fortement exprimés, est aussi reconnue par Rad52 suggère que Lsd1 et Lsd2 ont un rôle dans le contrôle de la réplication/recombinaison sur l'ensemble du génome. De façon intrigante, la protéine Abp1, connue comme étant d'une part un répresseur transcriptionnel et d'autre part une protéine contrôlant la réplication des LTRs, est retrouvée au niveau de gènes fortement exprimés (Cam et al., 2008; Daulny et al., 2016; Zaratiegui et al., 2011b). Ces résultats suggèrent que Sap1 et Abp1 participent au contrôle de la réplication dans ces régions "difficiles à répliquer" par analogie à ce qui est observé au niveau des LTRs et de l'ADNr (Mejia-Ramirez et al., 2005; Zaratiegui et al., 2011b).

De plus, il a été montré d'une part, que la mutation *sap1-1* engendre une augmentation du nombre de foci Rad52 et que d'autre part, que le double mutant *rad52*Δ *sap1-1* présente un phénotype synthétique de retard de croissance (Mizuguchi et al., 2017; Noguchi and Noguchi, 2007). Ces éléments, ainsi que le fait que Sap1 et Rad52 partagent des sites de liaison le long du génome, suggèrent que Sap1 est un facteur préventif des lésions d'ADN pris en charge par Rad52.

4. Signature de l'empreinte par séquençage haut débit

L'empreinte, située à la jonction du locus *mat1* et de la boite d'homologie *H1*, crée un site fragile qui casse durant la préparation de l'ADN (Arcangioli, 1998). Un fragment d'ADN contenant l'empreinte sera donc cassé au niveau du site de cette dernière. Bien que la nature de l'empreinte soit toujours sujette à débat, les spécialistes du domaine sont d'accord sur la conséquence de cette dernière (Kaykov and Arcangioli, 2004; Vengrova and Dalgaard, 2004). Nous avons décidé d'analyser l'empreinte par séquençage haut débit. En utilisant des génomes portant une empreinte, nous devrions être capables d'observer un enrichissement en lectures commençant au niveau du site fragile.

4.1 L'utilisation du séquençage haut débit pour l'analyse d'un site fragile

Afin de réaliser un séquençage haut débit Illumina, l'ADN doit être découpé en fragments de 150 pb à 350 pb. Une digestion enzymatique ou une sonication de l'ADN peuvent être effectuées pour la fragmentation. Dans notre étude, nous avons utilisé la sonication qui permet de mettre en évidence le site fragile. Cette technique fragmente l'ADN aléatoirement. Les conditions de fragmentation sont mises au point en fonction de la taille désirée des fragments. Une chromatide ne comportant pas de site fragile sera découpée aléatoirement (figure 27A). Par contre, une chromatide portant un site fragile cassera préférentiellement au niveau de ce site (figure 27B). Partant de cette logique, les lectures d'ADN résultant du séquençage seront réparties différemment en fonction de la présence ou non de cette empreinte. Si l'ADN a été fragmenté aléatoirement, les lectures seront également réparties aléatoirement le long du génome. Par contre si l'ADN porte une empreinte, les lectures commenceront/finiront préférentiellement à l'empreinte (figure 27). Si l'on s'intéresse au commencement des lectures (ce que nous appellerons le " **5'** "), nous devrions observer un enrichissement en 5' commençant au site de l'empreinte et cela sur le brin Watson et sur le brin Crick (figure 27B). Ainsi, en utilisant ce raisonnement et avec l'aide de Bernd Jagla j'ai compté le commencement des lectures sur l'ensemble du génome et j'ai analysé ce que nous appellerons les "5' comptes".

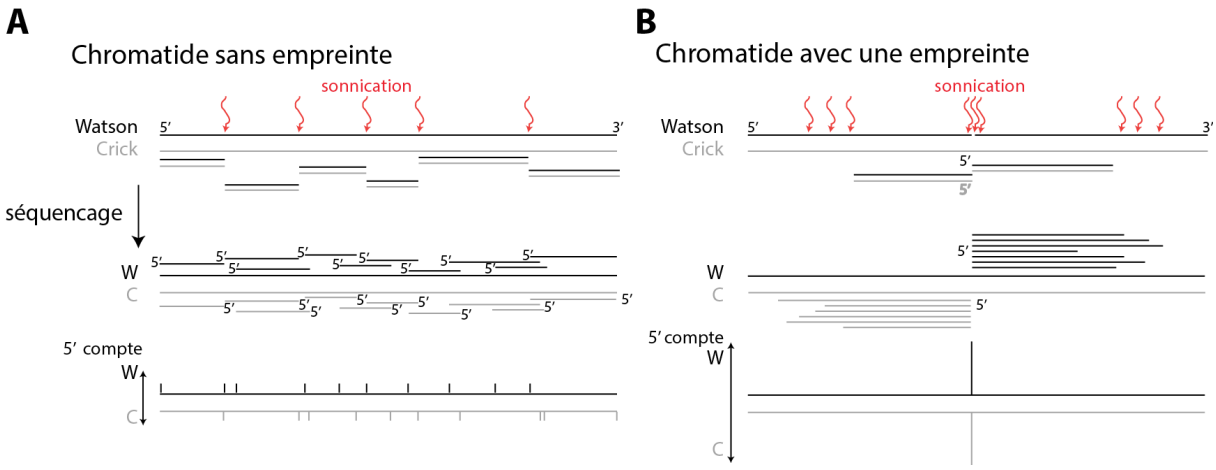


Figure 27 : Schéma de l'analyse des 5' comptes.
A) Une chromatide qui ne porte pas d'empreinte sera fragmentée aléatoirement et donnera donc des 5' comptes aléatoires et uniformes sur l'ensemble du génome. B) Une chromatide portant un site fragile cassera préférentiellement à ce niveau. On observera donc un enrichissement du 5' compte au site de l'empreinte.

4.2 Observation du commencement des lectures sur tout le génome

J'ai analysé huit séquençages différents. Comme décrit précédemment, le site de l'empreinte quand l'allèle *M* est présent à *mat1* est le même, que les cellules soient ou non capable de changer de type sexuel (c.-à-d. que les cellules soient *h90* (avec *M* à *mat1*) ou *mat1M* Δ*2-3*). Cela est également vrai quand l'allèle *P* est présent (Nielsen and Egel, 1989). Cette observation m'a permis d'utiliser les génomes des différentes souches précédemment étudiées :

- *h90*
- *mat1M* Δ*2-3*
- *mat1P* Δ*2-3*

Pour cette analyse, j'ai utilisé un script (cf. Matériels et Méthodes) qui permet de ne regarder que le commencement des lectures d'ADN (les 5'). Grâce à ce script, j'ai extrait du fichier d'alignement le début des lectures et j'ai déterminé leur nombre pour chaque nucléotide du génome au niveau du brin Watson ou Crick (figure 27). La méthode de préparation des banques d'ADN permet d'analyser précisément le début des lectures des deux brins d'ADN. En effet, les extrémités non franches seront réparées durant la préparation des banques de séquençage par l'action d'une exonucléase 3' ou par celle d'une polymérase en fonction du type d'extrémité présent (figure 28). L'action de ces deux enzymes préserve donc les extrémités 5' des brins Watson et Crick.

Dans un premier temps, j'ai analysé la distribution des 5' comptes dans une souche *mat1M* Δ*2-3* (figure 29). Pour un 5' compte observé, je détermine le nombre d'occurrence dans le génome

(par exemple : plus de 3. 10^6 bases ont un 5' compte de 1, alors qu'une seule position a un 5' compte de 211 sur le brin Watson) (figure 29A). Nous constatons que pour la majorité des positions génomiques, le 5' compte est inférieur à 10. Par ailleurs, les distributions des 5' comptes du brin Watson et Crick sont comparables (figure 29A et 29B). Ces distributions sont utilisées pour calculer la *p value* du 5' compte au site de l'empreinte (cf. Matériels et Méthodes). Nous remarquons que le 5' compte du brin Watson au site de l'empreinte est de 44 et que seulement 164 bases dans le génome ont le même compte alors que le 5' compte du brin Crick au site de l'empreinte est de 5 et plus de 250 000 bases ont le même compte (figure 29C et 29D). La valeur du 5' compte au niveau de l'empreinte est donc enrichie sur le brin Watson et non sur le brin Crick (figure 29). Il a été montré que l'empreinte est faite sur le brin Watson uniquement, et que c'est durant la préparation de l'ADN que le brin Crick casse (Arcangioli, 1998). Le fait que nous n'observons pas d'enrichissement en 5' compte du brin Crick au niveau du site de l'empreinte suggère que la cassure de ce brin se fait à des positions variables. Par exemple, si les 44 5' observés sur le brin Watson au site de l'empreinte se retrouvent répartis à +/- 10 nucléotides sur le brin Crick, nous ne pourrons plus observer d'enrichissement. La cartographie du brin Crick cassé avait déjà été questionnée par les Dr Nielsen et Egel (Nielsen and Egel, 1989). En effet, ces auteurs n'avaient pas pu observer la molécule d'ADN Crick cassée au niveau de l'empreinte en utilisant la procédure de séquençage décrite par les Dr Church et Gilbert (Church and Gilbert, 1985). Ces résultats indiquent qu'au cours de la sonication, le brin Crick ne casse pas précisément en face du site de l'empreinte.

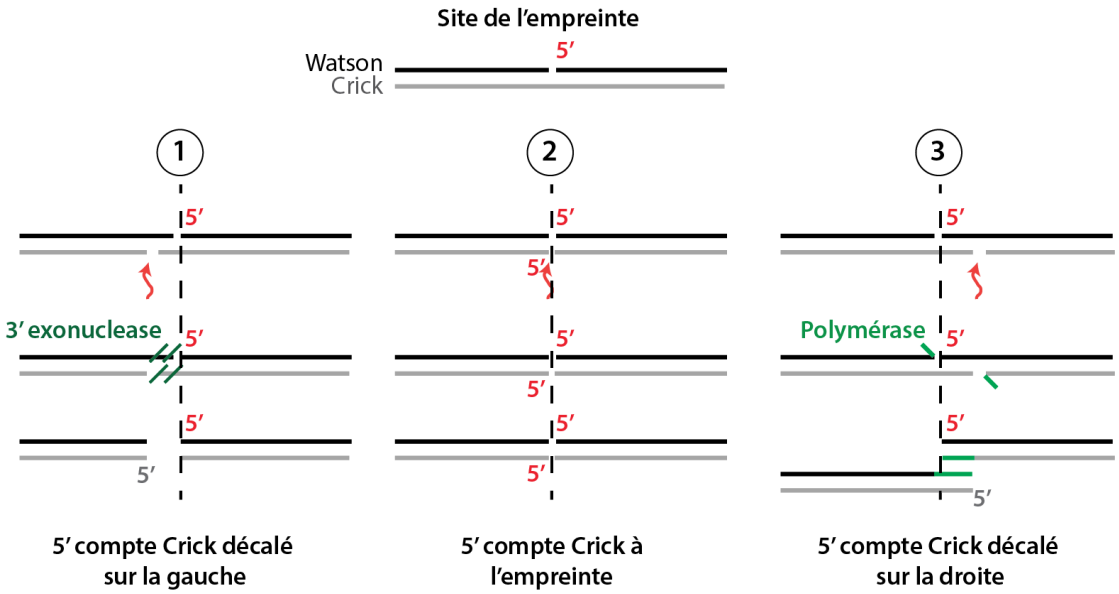


Figure 28 : Représentation des différents 5' du brin Crick autour du site de l'empreinte.
Lors de la préparation de l'ADN, le brin Crick ne casse pas exactement en face de l'empreinte. Les trois cas présentent la répartition des 5' comptes des deux brins en fonction de la position de la cassure du brin Crick. Le cas n°1 schématise le résultat d'une cassure du brin Crick du côté gauche de l'empreinte. Le cas n°2 représente une cassure en face de l'empreinte et le cas n°3 une cassure du côté droit de l'empreinte.

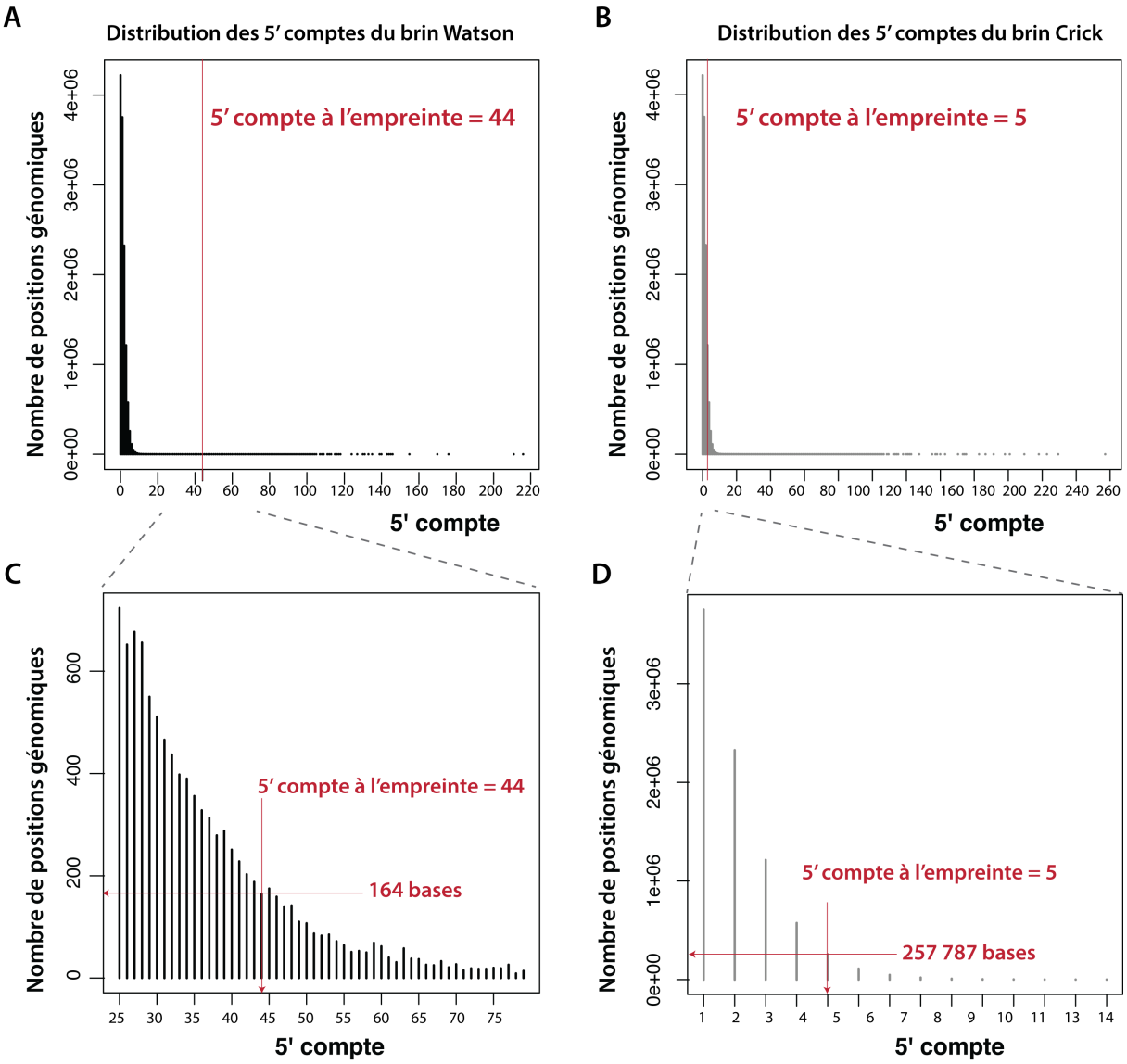


Figure 29 : Distribution du nombre de 5' comptes par positions génomiques.
A) Nombre de positions du génome en fonction du 5' compte du brin Watson. La ligne rouge indique le 5' compte au niveau de l'empreinte. B) Même que A) pour le brin Crick. C) Agrandissement du graphique A) autour du 5' compte du site de l'empreinte. Le 5' compte à l'empreinte est indiqué ainsi que le nombre de bases partageant ce 5' compte. D) Même qu'en C) pour le brin Crick.

Dans un second temps, nous avons analysé la répartition des 5' comptes sur l'ensemble du génome (figure 30). En plus du site de l'empreinte, six régions présentent un enrichissement en 5' compte. A la différence du locus sexuel, l'enrichissement est observable sur le brin Watson (en noir) et sur le brin Crick (en gris), et cela dans les deux fonds génétiques représentés (figure 30A et 30B représentent le séquençage de la souche *mat1M* Δ*2-3* et 30C et 30D représentent le séquençage de la souche *mat1P* Δ*2-3*). La question sous-jacente est de

déterminer si l'ADN présent au niveau de ces sites est plus fragile et casse préférentiellement. L'analyse de ces sites montre qu'ils se situent dans des régions génomiques répétées (TelII, ADNr et CENIII) ou bien qu'ils sont dans des gènes qui présentent des homologies avec l'ADN mitochondrial (SPNUMT). Ces éléments suggèrent que ces enrichissements ne sont pas dûs à la présence de sites fragiles, mais plutôt à des biais d'alignement des lectures dans les régions répétées. L'empreinte à *mat1* semble donc constituer le seul site fragile récurrent dans le génome de *S. pombe*. Il est probable que d'autre loci portent des cassures simples brins mais que la fréquence de ces lésions est trop faible pour être détectée par cette méthode.

L'empreinte est donc détectable par séquençage génomique, quelle que soit l'allèle présent à *mat1*.

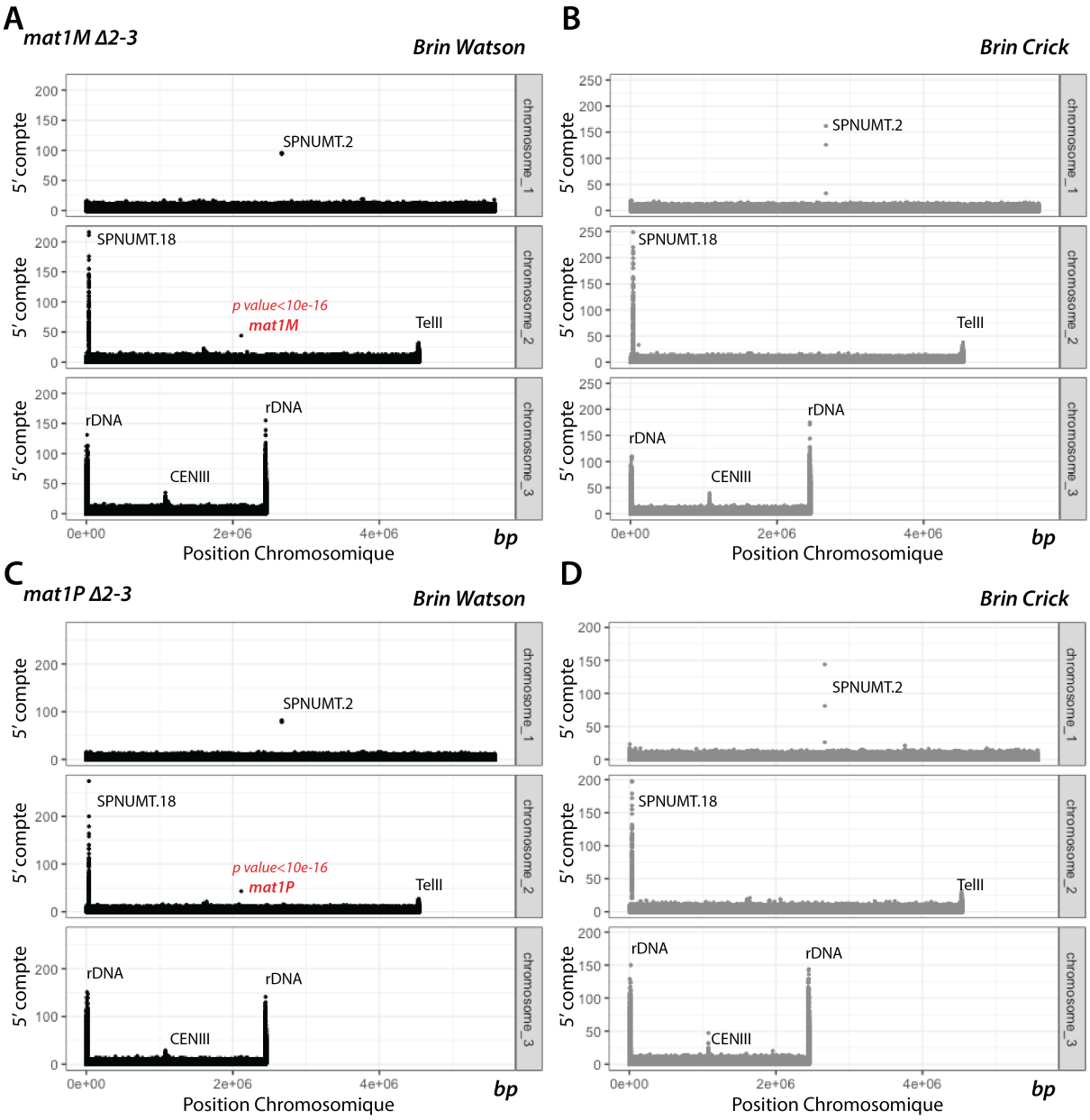


Figure 30 : Distribution des 5' comptes sur l'ensemble du génome.
A et B) Analyse des 5' comptes d'une souche *mat1M* Δ*2-3*. A) Distribution des 5' comptes du brin Watson sur l'ensemble du génome. B) Distribution des 5' comptes du brin Crick sur l'ensemble du génome. C et D) Même qu'en A et B) pour une souche *mat1P* Δ*2-3*. Les *p values* sont calculées en utilisant comme modèle une distribution binomiale négative choisie après l'analyse de la distribution des 5' comptes (figure 29) (cf. Matériels et Méthodes).

L'utilisation d'une taille de séquençage de 130 pb nous permet de répartir la plupart des lectures correctement sur le génome et de réduire les biais au niveau des séquences répétées. En effet, avec un séquençage de plus petite taille (65 pb) 11 régions présentent des 5' comptes élevés comparé à 7 régions avec une séquençage de 130 pb (figure 31A, 31B et 31C comparer à 31D). À l'exception du site de l'empreinte, toutes les autres régions enrichies se situent dans des séquences répétées ou homologues à l'ADN mitochondriale (tableau 4).

Numéro du pic	Chromosome	Position	Nom externe	Nom logique	Séquençage de 130 bp
1	1	71982	SPNUMT.8	nuclear mt pseudogne	
2	1	2667776	SPNUMT.2	nuclear mt pseudogne	✗
3	1	3810366	SPNUMT.17	nuclear mt pseudogne	
4	2	36899	SPNUMT.18	nuclear mt pseudogne	✗
5	2	2115364	empreinte	empreinte	✗
6	2	3360976	SPNUMT.7	nuclear mt pseudogne	
7	2	4532473	TELII	telomeric region	✗
8	3		SPRRNA	rDNA	✗
9	3	814835	SPRRNA	rDNA	
10	3	1070904	CENIII	centromeric region	✗
11	3		SPRRNA	rDNA	✗

Tableau 4 : Sites qui présentent un enrichissement en 5' compte.
Les numéros des pics sont indiqués, ainsi que toutes les annotations disponibles. La dernière colonne indique les pics qui sont conservés avec un séquençage de 130 pb.

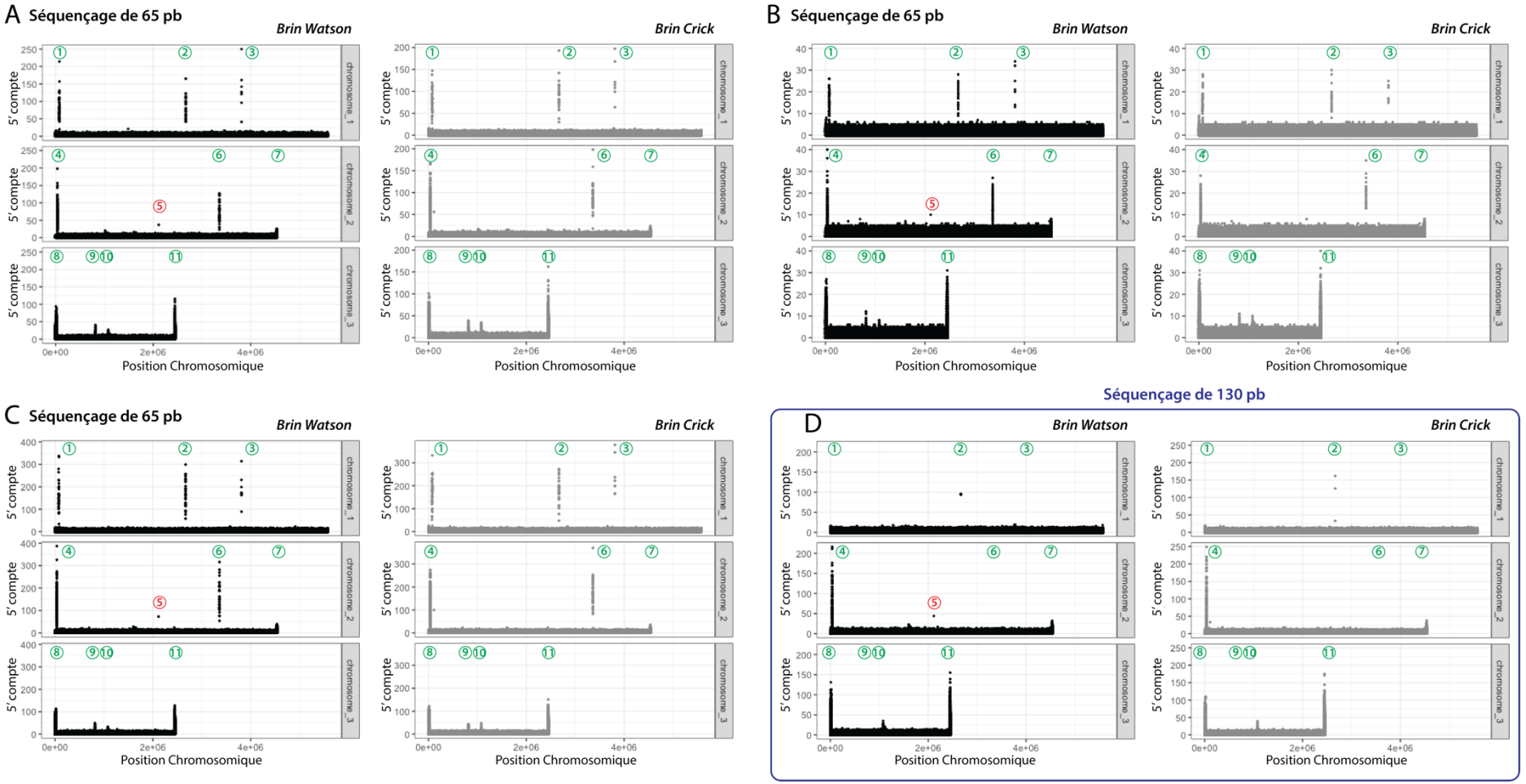


Figure 31 : Distribution des 5' comptes sur l'ensemble du génome dans quatre souches *mat1M* Δ*2-3*.
A, B et C) Analyse des 5' comptes de trois souches *mat1M* Δ*2-3* en utilisant une taille de séquençage de 65 pb. En noir est représentée la distribution des 5' comptes du brin Watson sur l'ensemble du génome. En gris est représentée la distribution des 5' comptes du brin Crick sur l'ensemble du génome. Les régions enrichies sont numérotées de 1 à 11. D) Même qu'en A) avec une taille de séquençage de 130 pb.

4.3 Analyse de l'empreinte à *mat1*

J'ai ensuite analysé les 5' comptes dans quatre séquençages dont l'ADN provient des souches *mat1M* Δ*2-3*, *mat1P* Δ*2-3*, *h90* et *h90 swi*Δ (figure 32). L'agrandissement de la région *mat1* nous permet d'analyser précisément la position des 5' comptes. Nous remarquons tout d'abord que seul un site est biaisé en 5' compte dans la région autour de *mat1*. Par ailleurs, la probabilité d'avoir le compte observé quand l'allèle à *mat1* est fixé (*M* ou *P*) est faible (*p value* $<10^{-15}$), ce qui indique que cette position est significativement enrichie en 5' compte (figure 32 A et B). Comme exposé précédemment, il a été montré que le site de l'empreinte est décalé de 3 nucléotides en fonction de l'allèle présent à *mat1* (Nielsen and Egel, 1989). Ce décalage est également observable lorsque nous comparons la position du 5' compte enrichi dans la souche *mat1M* Δ*2-3* avec celui dans la souche *mat1P* Δ*2-3* (figure 32A et 32B). Dans la souche *h90* représentée, les deux positions présentent un 5' compte enrichi, ce qui est attendu puisque dans une population de cellules de fond génétique *h90*, 50 % des cellules contiennent l'allèle *P* à *mat1* et 50 % l'allèle *M*. Enfin, la souche *h90 swi1*Δ qui a un niveau très faible d'empreinte ne présente pas d'enrichissement en 5' comptes au niveau de la région du CTS (figure 32D et (Egel et al., 1984)).

Afin de confirmer l'observation faite sur un génomes *mat1M* Δ*2-3*, j'ai analysé 3 séquençages supplémentaires (figures 33A et 34A). Le site de l'empreinte possède un 5' compte sur le brin Watson significativement plus élevé que la moyenne dans le génome. Cette analyse a également été réalisée sur 3 génomes supplémentaires portant l'information *P* à *mat1* (figures 33B et 34B). Dans 3 de ces génomes, le 5' compte est enrichi au site de l'empreinte sur le brin Watson. Dans le quatrième, nous n'avons pas observé d'enrichissement et ce sera discuté ci-dessous. Enfin, j'ai aussi utilisé 3 séquençages supplémentaires de la souche *h90* (figure 34C). Dans ce contexte, les 5' comptes sont enrichis en majorité seulement au niveau du site de l'empreinte quand l'allèle *M* est à *mat1* (figure 34C). Dans une souche *h90*, un enrichissement du site de l'empreinte *mat1P* n'est pas décelé quand la taille du séquençage est de 65 nt. Un séquençage de 130 nt est donc préférable pour cette analyse car il permet de meilleurs alignements. Par ailleurs, le site de l'empreinte est décalé de 3 nucléotides dans une souche *P*, ce qui rajoute 3 T à l'extrémité 5' du brin Watson. Cette différence pourrait influencer la stabilité de l'extrémité, et/ou l'efficacité de séquençage.
Enfin, indépendamment du séquençage analysé, les 5' comptes du brin Crick ne montrent jamais d'enrichissement à l'empreinte (figures 33 et 34).

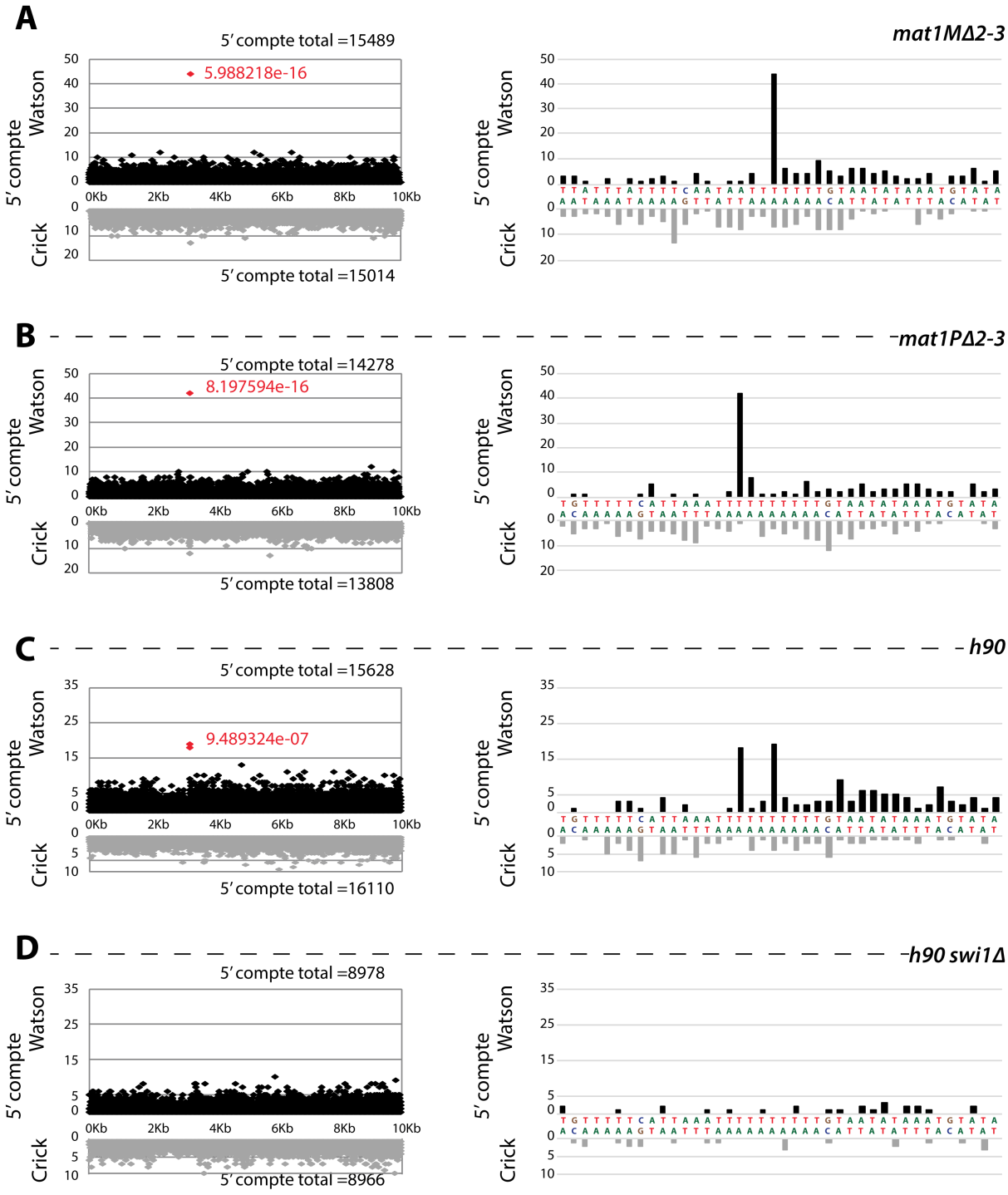


Figure 32 : Analyse des 5' comptes dans la région *mat*.
A) Distribution des 5' comptes en fonction de la position dans le génome, sur 10 kb autour de *mat1* (panneau de gauche), ou sur 40 bp (panneau de droite) dans un souche *mat1M* Δ*2-3*. B-C-D) Même qu'en A) pour une souche *mat1P* Δ*2-3* (B), *h90* (C) et pour une souche *h90 swi1*Δ (D). En noir sont représentés les 5' comptes du brin Watson et en gris ceux du brin Crick.

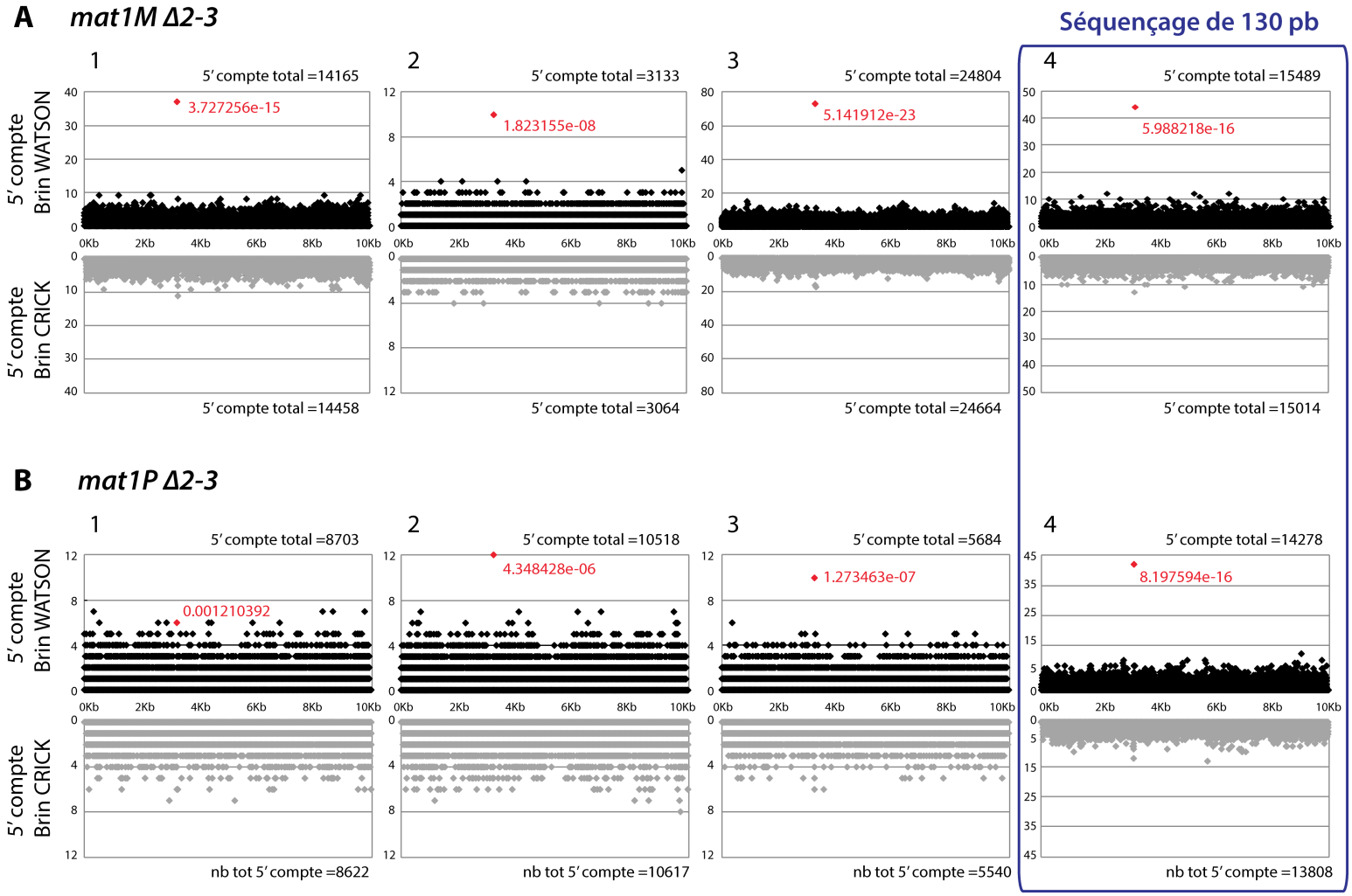


Figure 33 : Analyse des 5' comptes dans les 10 kbp autour de *mat1*.
A) Distribution des 5' comptes en fonction de la position dans le génome, sur 10 kb autour de *mat1* dans quatre souches *mat1M* Δ*2-3*. B) Même qu'en A) dans quatre souches *mat1P* Δ*2-3*. En noir sont représentés les 5' comptes du brin Watson et en gris ceux du brin Crick. La taille de séquençage est de 65 pb excepté pour les panneaux entourés en bleu.

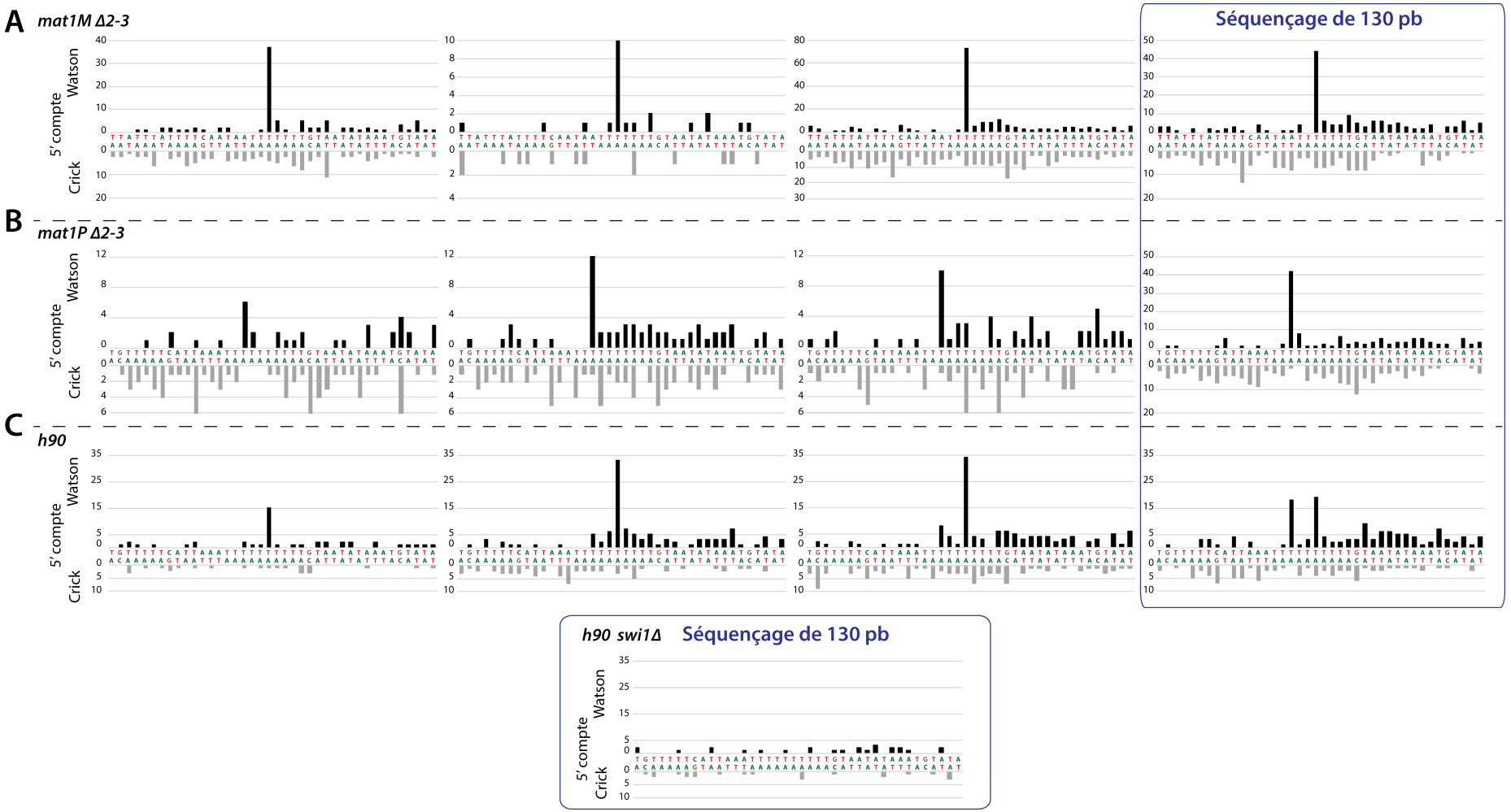


Figure 34 : Analyse des 5' comptes au niveau des 40 pb autour de l'empreinte.
A) Distribution des 5' comptes en fonction de la position génomique, sur 40 pb autour de *mat1* dans quatre souches *mat1M* Δ*2-3*. B-C) Même qu'en A) dans quatre souches *mat1P* Δ*2-3* (B), dans quatre souches *h90* (C) et dans la souche *h90 swi1*Δ (C panneau du bas). En noir sont représentés les 5' comptes du brin Watson et en gris ceux du brin Crick. La taille de séquençage est de 65 pb excepté pour les panneaux entourés en bleu.

4.5 Le 5’ compte à l'empreinte est enrichi dans les IP

4.5.1 Abp1 immunoprécipite la molécule qui porte l'empreinte

L'analyse du commencement des lectures permet donc de cartographier l'empreinte. J'ai utilisé cette analyse pour déterminer la liaison de la protéine Abp1 à la chromatide qui porte l'empreinte.

Nous avons constaté précédemment que la protéine Abp1 est recrutée à environ 150 pb du côté centromère distal de *mat1*, indépendamment de l'allèle présent.

Théoriquement, pour la ChIP-séq d'Abp1, les 5' comptes de chacun des deux brins d'ADN devraient être "enrichis" de part et d'autre du site de recrutement de cette protéine (figure 35A). Nous supposons ici que la probabilité d'avoir un fragment avec la protéine en son centre est plus élevée. En absence d'empreinte, les 5’ comptes enrichis du brin Watson devraient se répartir en une gaussienne du côté gauche du site de recrutement d'Abp1 et les 5’ comptes enrichis du brin Crick devraient se répartir en une gaussienne du côté droit du site de recrutement. Ces distributions devraient être à une distance du site de liaison d'Abp1 égale à la moitié de taille de la fragmentation des ADN (de 150 à 350 pb) (figure 35A). Les deux distributions devraient être à une distance égale à celle de la sonication. En présence de l'empreinte, la sonication casse l'ADN au niveau de l'empreinte. Les 5 'comptes du brin Watson devraient donc être enrichis au niveau du site de l'empreinte et ceux du brin Crick devraient former une gaussienne restreinte du côté droit du site de liaison d'Abp1 (figure 35B). Nous supposons que le fragment casse dans un premier temps au niveau de l'empreinte et dans un deuxième temps à une distance de 150 à 300 pb de celui-ci. Le site de la deuxième cassure est donc influencé par le premier. Dans les différents fonds génétiques capables de changer de type sexuel, environ la moitié des cellules auront une empreinte. Le résultat de l'analyse de l'IP d'Abp1 devrait donc être un mélange des deux types de distributions. Dans un fond génétique sans empreinte, comme *h90 swi1*Δ, seule la distribution de la figure 35A devrait être observée.

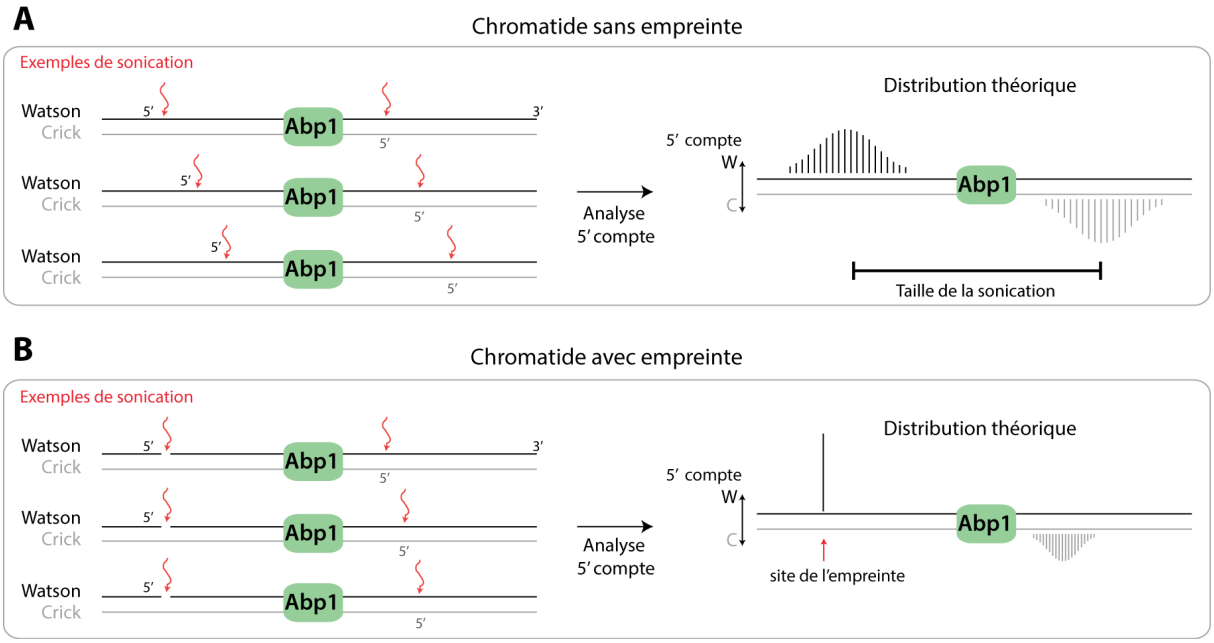


Figure 35 : Distributions théoriques des 5' comptes de l'IP Abp1 en fonction de la présence de l'empreinte.
En noir, les 5' comptes du brin Watson et en gris ceux du brin Crick. La protéine Abp1 est représentée en vert. Les sites cassés au cours de la sonication sont marqués par une flèche rouge. A) Résultat théorique de l'analyse des 5' comptes de l'IP d'Abp1 sur une chromatide sans empreinte et en B) sur une chromatide avec empreinte.

Pour la représentation des données observées, j'ai utilisé les données d'alignement non normalisées des lectures de l'IP Abp1 (couverture). Cela me permet de les comparer aux 5' comptes.

J'ai dénombré les 5' comptes de la ChIP-séq de la protéine Abp1, puis j'ai normalisé le 5' compte au site de l'empreinte en utilisant la taille de la banque et les 5' comptes du WCE (figure 36B). La distribution des 5' comptes des IP d'Abp1 dans les souches *h90* et *h90 swi1*Δ est représentée figure 36. Les pistes vertes (claire et foncée) sont les couvertures des IP au locus sexuel. La distribution des 5' comptes de l'IP Abp1 est proche de la distribution théorique. Tout d'abord, nous observons que le 5' compte du brin Watson est enrichi au niveau du site de l'empreinte *M* et *P* (figures 36A et 36C). D'autre part quand je change l'échelle des 5' comptes afin qu'ils ne soient pas écrasés par le compte au site de l'empreinte, nous constatons que les 5' comptes enrichis sont distribués sur environ 400 pb sur le brin Watson et sur le brin Crick (figure 36B). Dans la souche sauvage, les deux distributions théoriques sont donc observées (figure 36D, panneau du haut). Dans la souche *h90 swi1*Δ, la distribution des 5' comptes est similaire à celle de la souche *h90* à l'exception du site de l'empreinte. Au niveau de ce site, nous n'observons pas d'enrichissement. Ceci corrèle avec l'absence d'empreinte détectable dans ce mutant. Dans

cette souche, seule la distribution théorique en absence d'empreinte est observée (figure 36D, panneau du bas). Ces résultats indiquent que la protéine Abp1 est recrutée du côté centromère distal de *mat1M* et *mat1P*, indépendamment de la présence de l'empreinte (figure 36B et 36C).

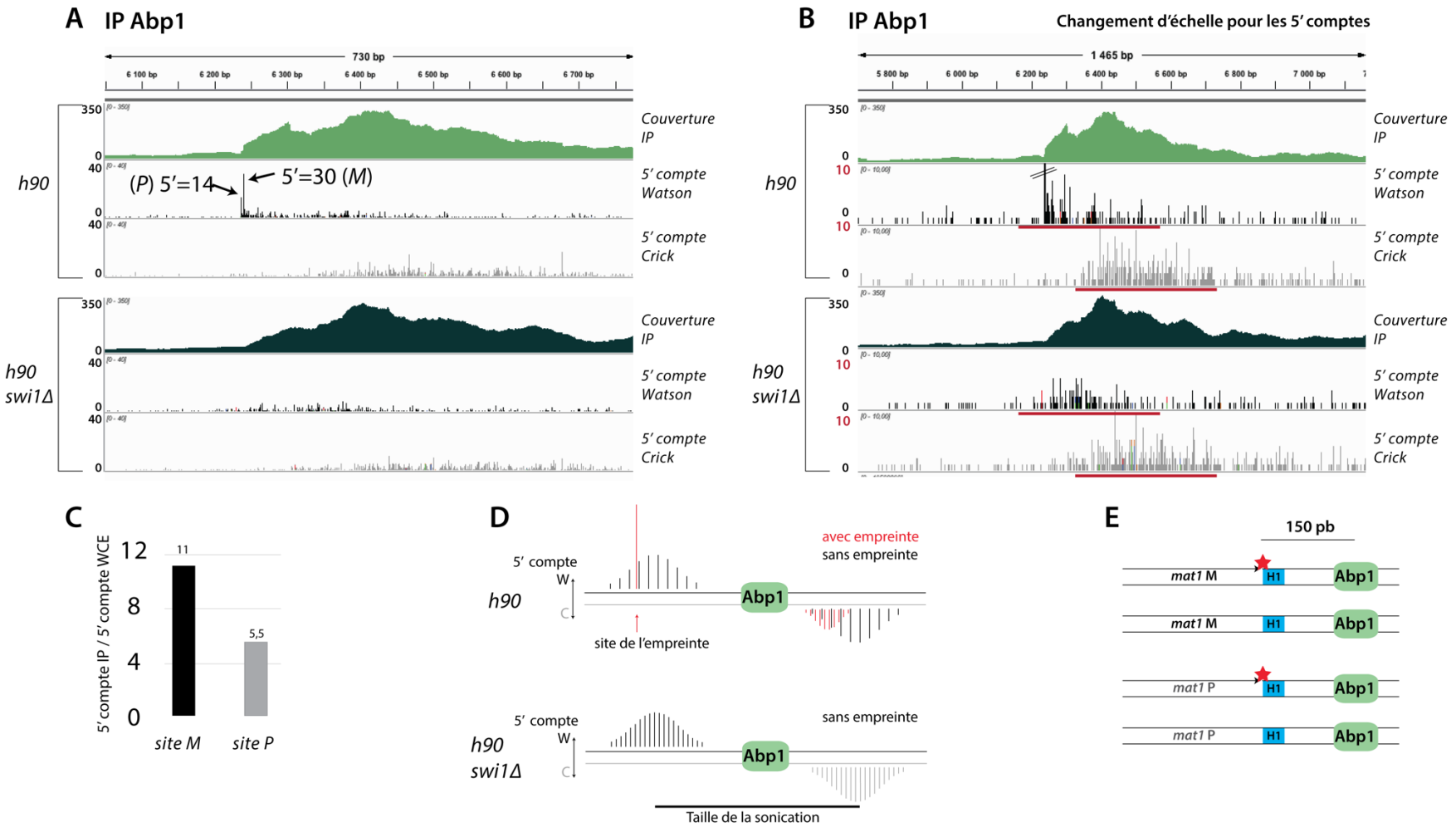


Figure 36 : Analyse des 5’ comptes au niveau de l'empreinte dans l'IP d'Abp1.

A) Les lignes Couverture IP représentent la couverture des lectures d'ADN des ChIP-séq d'Abp1 dans le fond génétique *h90* en vert clair et *h90 swi1*Δ en vert foncé. La séquence utilisée pour l'alignement est une région de 35 kb contenant l'allèle *P* à *mat1*. Les lignes nommées "Watson" sont les distributions des 5’ comptes du brin Watson et celle nommées "Crick" des 5’ comptes du brin Crick et cela pour chacune des ChIP-séq indiquées. Les flèches indiquent la position du 5’ compte du brin Watson enrichi au site de l'empreinte quand l'allèle à *mat1* est P (P) ou M (M). B) Même qu'en A) avec une échelle différente pour les 5' comptes. Les lignes rouges indiquent les 5' comptes enrichis. C) Enrichissements normalisés du 5’ compte à l'empreinte de la ChIP-séq d'Abp1 dans une souche *h90*, quand l'allèle est *M*, en noir et quand l'allèle est *P*, en gris. La taille de la banque ainsi que le 5’ compte des WCE sont pris en compte dans la normalisation. D) Schéma représentant le résultat de l'analyse des 5' comptes de l'IP Abp1 dans les contextes *h90* et *h90 swi1*Δ. E) Schéma représentant le recrutement d'Abp1 en fonction de l'allèle présent à *mat1* ainsi qu'en fonction de l'empreinte.

4.5.2 Lsd1, Lsd2 et Sap1 lient la chromatide qui porte l'empreinte

Afin de réaliser la même analyse sur les IP de Lsd1, Lsd2 et Sap1, j'ai utilisé les ChIP-séq qui ont été réalisées dans les fonds génétiques *mat1M* Δ*2-3* et *mat1P* Δ*2-3*. À la différence d'Abp1, ces protéines sont recrutées à l'intérieur de *mat1*, et l'utilisation de ces souches au lieu de la souche *h90* permet d'éviter la répartition aléatoire des lectures au niveau des donneurs *mat2P* et *mat3M*.

En absence de l''empreinte, la distribution théorique des 5' comptes d'une protéine se fixant du côté centromère proximal du site de l'empreinte est présentée figure 37A. Deux gaussiennes de 5' comptes enrichis seront présentes à une distance égale la taille des fragments générés à la sonication.

En présence de l'empreinte, l'ADN casse dans un premier temps au niveau de ce site comme présenté sur la figure 37B. La distribution des 5' comptes enrichis du brin Crick sera donc restreinte autour du site de l'empreinte. Celle des 5' comptes du brin Watson sera également restreinte. Elle sera séparée de la distribution des 5' comptes enrichis du brin Crick par la distance égale à la taille moyenne des fragments produits par la sonication.

Dans les différents fonds génétiques capables de changer de type sexuel, environ la moitié des cellules aura une empreinte. Le résultat de l'analyse de l'IP de Lsd1, Lsd2 et Sap1 dans une souche *mat1M* Δ*2-3* devrait donc être un mélange des deux types de distributions. Dans un fond génétique *mat1P* Δ*2-3*, nous devrions observer le même résultat pour les protéines enrichies à savoir Lsd1 et Lsd2. Sap1 ne se liant pas au niveau de cet allèle, nous ne devrions pas observer de 5' comptes enrichis autour du locus *mat1P* dans cette IP.

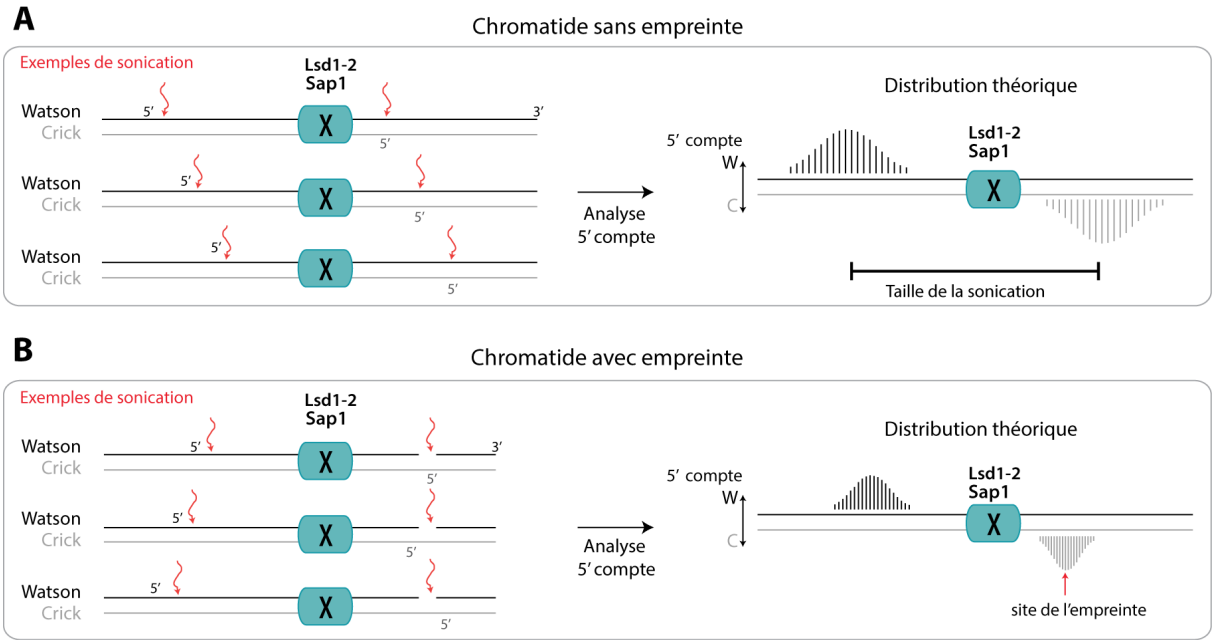


Figure 37 : Distributions théoriques des 5' comptes des IP Lsd1-2 et Sap1 en fonction de la présence de l'empreinte.
En noir les 5' comptes du brin Watson et en gris ceux du brin Crick. Les protéines Lsd1-2 et Sap1 sont représentées par le rectangle bleu. Les sites cassés au cours de la sonication sont marqués par une flèche rouge. A) Résultat théorique de l'analyse des 5' comptes de l'IP de Lsd1-2 et Sap1 sur une chromatide sans empreinte et en B) sur une chromatide avec empreinte.

Quand l'allèle *M* est à *mat1,* les distributions des 5' comptes des IP de Lsd1, Lsd2 et Sap1 sont similaires, mais sont différentes de la distribution théorique (figure 38A). En effet, de façon surprenante le 5' compte du brin Watson au site de l'empreinte est enrichi par rapport au WCE dans les IP de Lsd1/2 et Sap1 dans le fond génétique *mat1M Δ2-3* (figures 37A et 37C). Cela indique que ces protéines sont capables d'immunoprécipiter le fragment distal portant l'empreinte.

En revanche, les distributions des 5' comptes enrichis du brin Watson et Crick sont effectivement un mélange de celles attendues avec et sans empreinte (figure 37 et figure 38B). À l'intérieur de *mat1M*, à environ 300 pb du site de l'empreinte, les 5' comptes du brin Watson sont plus élevés. Autour du site de l'empreinte, une partie des 5' comptes du brin Crick sont plus élevés, et à 300 pb de ce site, une autre partie de ces 5' comptes est enrichie.

Ainsi, dans les IP des protéines Lsd1, Lsd2 et Sap1, il semble que trois types de distributions soient présentent : (figure 38D)

- la distribution d'une protéine se fixant à l'intérieur de *mat1M* sans empreinte ;
- la distribution d'une protéine se fixant à l'intérieur de *mat1M* avec empreinte ;

- la distribution d'une protéine se fixant à l'extérieur (centromère distal) de *mat1M* avec empreinte.

Enfin, nous ne pouvons pas savoir si la distribution d'une protéine se fixant à l'extérieur (centromère distal) de *mat1M* sans empreinte est aussi présente, puisque les enrichissements en 5' comptes de cette distribution seront confondus avec les autres distributions.

Dans l'ensemble, ces résultats suggèrent qu'un repliement spécifique de la chromatine est formé et permet aux protéines qui se fixent à l'intérieur de l'allèle *M*, d'immunoprécipiter le fragment distal portant l'empreinte (figure 38E).

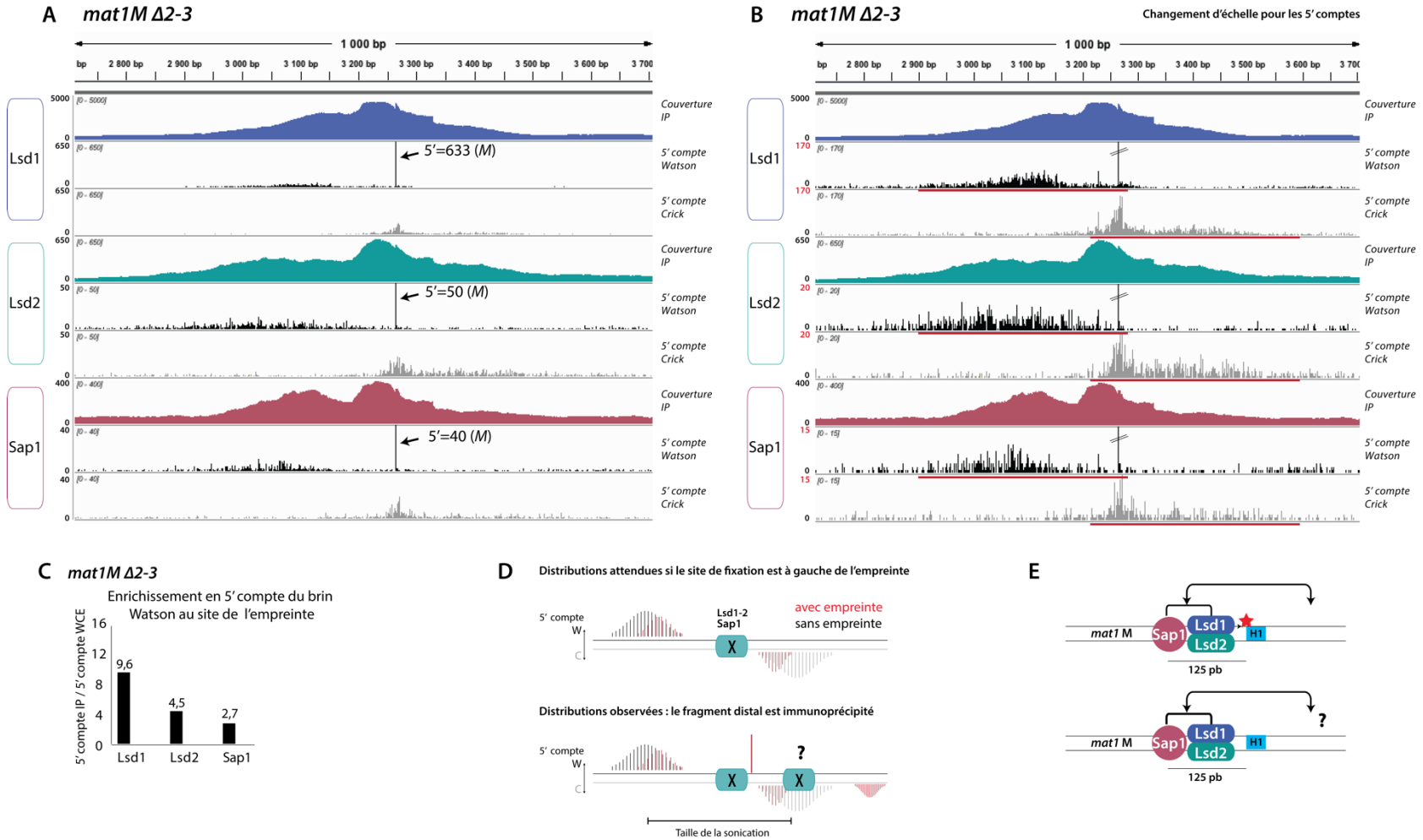


Figure 38 : Analyse des 5' comptes au niveau de l'empreinte dans les IP de Lsd1/2 et Sap1.

A) Les lignes nommées, Couverture IP, représentent la couverture des lectures d'ADN des ChIP-séq de Lsd1, en bleu, de Lsd2, en vert, et de Sap1, en rouge, dans le fond génétique *mat1M Δ2-3*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *M* à *mat1*. Les lignes nommées "Watson" sont les 5' comptes du brin Watson et celle nommées "Crick" des 5' comptes du brin Crick et cela pour chacune des ChIP-séq indiquées. Les flèches indiquent la position du 5' compte du brin Watson enrichi au site de l'empreinte (*M*). B) Même qu'en A) avec une échelle différente pour les 5' comptes. Les lignes rouges indiquent les 5' comptes enrichis. C) Enrichissements normalisés du 5' compte à l'empreinte des ChIP-séq de Lsd1, de Lsd2 et de Sap1 dans une souche *mat1M Δ2-3*. La taille de la banque ainsi que le 5' compte des WCE sont pris en compte dans la normalisation. D) Schéma représentant le résultat de l'analyse des 5' comptes de l'IP des protéines étudiées se fixant à l'intérieur de *mat1M* et à l'extérieur dans les contextes *mat1M Δ2-3*. E) Schéma représentant le recrutement de Lsd1/2 et Sap1 à *mat1M* en fonction l'empreinte suggérant un recrutement de ces protéines du côté centromère distal de l'empreinte.

Pour la protéine Lsd1, les distributions des 5’ comptes sont globalement comparables quand les allèles *P* ou *M* sont présents à *mat1* de quand l'allèle *M* (figures 38A et 38B). Lsd1 est capable de lier le fragment distal qui porte l'empreinte quand l'allèle est *P* à *mat1* (figure 39B). Lsd2 semble moins enrichie à *mat1P* par rapport à *mat1M*. En effet, nous remarquons que la molécule qui porte l'empreinte est moins enrichie dans l'IP de Lsd2 quand l'allèle à *mat1* est *P* (figure 39B comparée à 38C). La molécule qui porte l'empreinte n'est pas enrichie dans l'IP de Sap1, ce qui corrobore le fait que Sap1 n'est pratiquement pas recrutée à l'intérieur de *mat1P* (figures 39A et 39B). Ces résultats montrent que Lsd1 et Lsd2 sont recrutées à *mat1P* et notamment en présence de l'empreinte.

L'ensemble de ces résultats suggère que les allèles *M* et *P* permettent tous deux la formation d'un repliement de la chromatine reconnu par Lsd1 et Lsd2. Seule la protéine Sap1 est recrutée préférentiellement au niveau de l'allèle *M* (figures 38E et 39C).

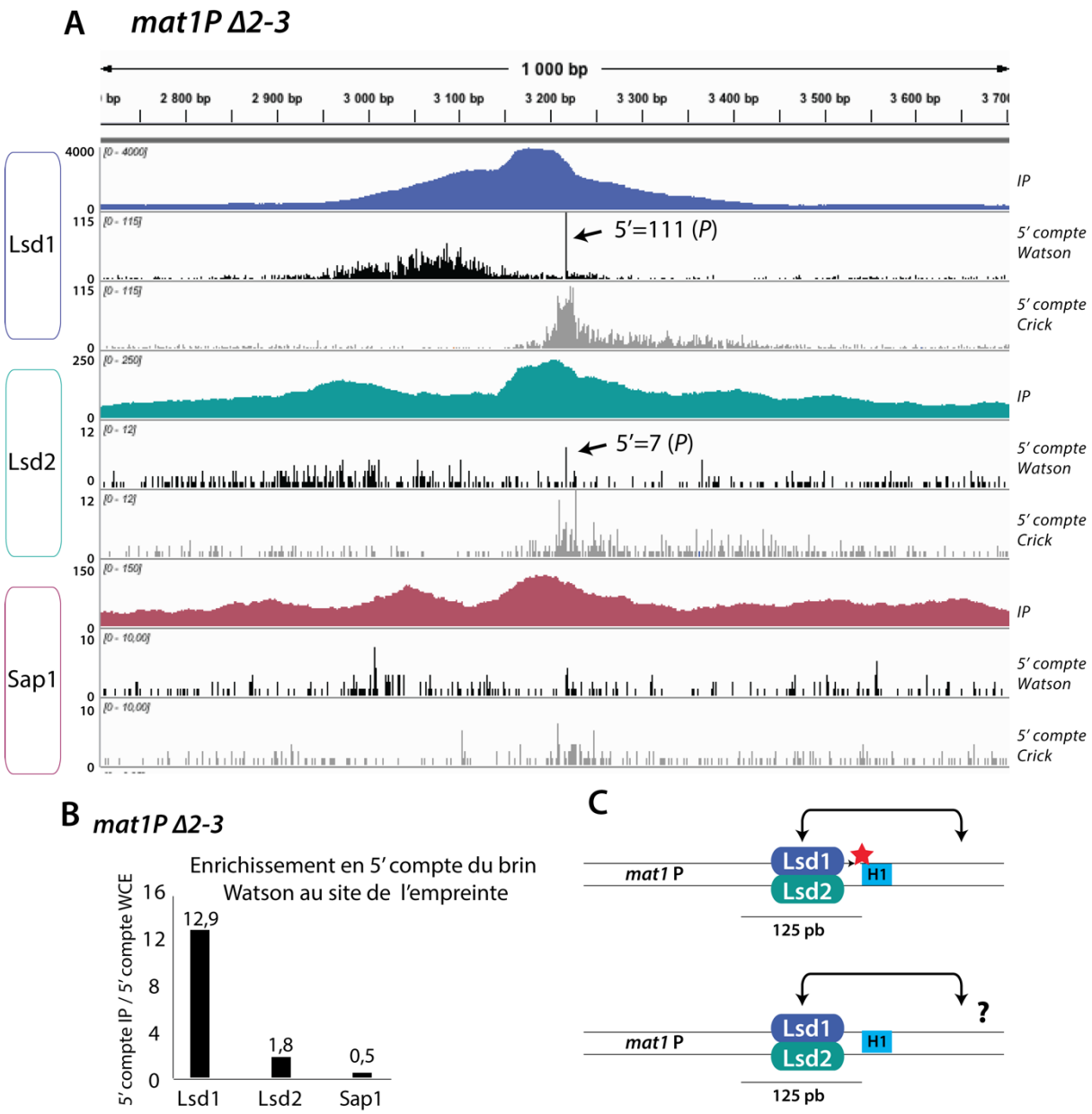


Figure 39 : Analyse des 5' comptes au niveau de l'empreinte dans les IP de Lsd1, Lsd2 et Sap1.

A) Les lignes nommées, Couverture IP, représentent la couverture des lectures d'ADN des ChIP-séq de Lsd1, en bleu, de Lsd2, en vert, et de Sap1, en rouge, dans le fond génétique *mat1P Δ2-3*. La séquence utilisée pour l'alignement est une région de 10 kb contenant l'allèle *P* à *mat1*. Les lignes nommées "Watson" sont les 5' comptes du brin Watson et celle nommées "Crick" des 5' comptes du brin Crick et cela pour chacune des ChIP-séq indiquées. Les flèches indiquent la position du 5' compte du brin Watson enrichi au site de l'empreinte (M). B) Enrichissements normalisés du 5' compte à l'empreinte des ChIP-séq de Lsd1, de Lsd2 et de Sap1 dans une souche *mat1P Δ2-3*. La taille de la banque ainsi que le 5' compte des WCE sont pris en compte dans la normalisation. C) Schéma représentant le recrutement de Lsd1/2 à *mat1P* en fonction l'empreinte suggérant un recrutement de ces protéines du côté centromère distal de l'empreinte.

4.4 Conclusions

En utilisant des séquençages génomiques ne résultant pas d'une immunoprécipitation, nous constatons qu'il est possible de cartographier un site fragile grâce à la détection d'un nombre accru d'extrémités 5' de lectures qui démarrent au site de la discontinuité. L'étude de l'empreinte au locus sexuel nous a permis de montrer qu'un ADN avec une discontinuité simple chaine ne casse pas exactement en face de cette lésion lors de la sonication. La caractéristique d'une telle lésion sera donc, un enrichissement en 5' comptes sur l'un des deux brins à un site précis et un enrichissement en 5' compte aux alentours de ce site sur le brin opposé (ici Watson et Crick, respectivement).

D'autre part, la séquence d'ADN est importante dans cette analyse. En effet, nous observons des différences notables en fonction de l'allèle présent à *mat1*. Par exemple, quand l'information *M* est présente à *mat1*, le compte des 5' est enrichi au niveau du site de l'empreinte sur le brin Watson alors que quand l'information à *mat1* est *P*, le 5' compte à ce site n'est enrichi que dans 3 des 4 génomes testés. Ces disparités peuvent être liées à la profondeur ainsi que la taille du séquençage.

Les résultats de cette analyse faite sur les ChIP-séq suggèrent que les protéines présentes à l'intérieur du locus *mat1* sont également capables de lier le fragment distal portant l'empreinte. Nous pourrions imaginer que la structure de la chromatine autour de l'empreinte permet un contact entre les protéines à l'intérieur de *mat1* et la chromatine du côté centromère distal. Cette conformation pourrait notamment jouer un rôle dans la stabilisation de l'empreinte.

5. Contrôle de la réplication de *wtf9*

Un des projets du laboratoire consiste à déterminer l'implication de la protéine Lsd1 au niveau de la réplication. Un complexe contenant Lsd1 a été isolé par différents laboratoires, et il est composé des protéines Lsd1, Lsd2, Phf1 et Phf2 (Gordon et al., 2007; Lan et al., 2007; Nicolas et al., 2006). Lsd1 est décrit comme contrôlant la transcription de certains gènes, mais également comme contrôlant la propagation de l'hétérochromatine (Gordon et al., 2007; Lan et al., 2007; Nicolas et al., 2006). De plus, le laboratoire a montré que Lsd1 contrôle également la réplication à *mat1* (Holmes et al., 2005). Cependant, la cible de cette protéine n'a pas encore été identifiée.

En absence de Lsd1, la cellule est très malade et pousse très lentement. En utilisant ce phénotype, un crible a pu être réalisé. Des cellules *lsd1*Δ ont été étalées sur boite, et les colonies formées ont été repiquées jusqu'à l'obtention de colonies à croissance plus rapide. L'ADN de ces colonies a ensuite été préparé puis séquencé afin de déterminer les mutations responsables de la suppression du défaut de croissance du mutant *lsd1*Δ. Vingt-quatre suppresseurs ont ainsi été isolés. Une analyse de la présence de duplications réalisée par Béatrice Regnault a montré que quatorze des vingt-quatre suppresseurs contiennent une duplication au niveau du chromosome 3. L'analyse fine de ce réarrangement par le Dr Singh, un post-doctorant du laboratoire, a montré une duplication d'une région d'au moins 60 kb portant le gène *phf1*. Le produit de ce gène (Phf1) est retrouvé dans le complexe formé par Lsd1, Lsd2 et Phf2. Le fait qu'une majorité des mutants isolés contiennent cette duplication et que Lsd1 participe au contrôle de la réplication nous a amené à penser que Lsd1 pourrait influencer la réplication au niveau des bordures de la duplication sur le chromosome 3. En l'absence de Lsd1, la formation de duplications serait facilitée. Mon implication dans ce projet a été de caractériser l'influence de Lsd1 sur la réplication au niveau de la bordure de la duplication isolée.

5.1 Caractérisation de la duplication du bras gauche du chromosome 3

La duplication présente dans les différents mutants est comprise entre deux gènes appartenant à la famille des "*wtf*" (*With Tf*) : *wtf8* et *wtf9*. La famille de gènes *wtf* a été découverte chez *S. pombe* (Wood et al., 2002). Au total, vingt-cinq séquences *wtf* sont présentes dans le génome de *S. pombe*. Ces séquences ont été nommées ainsi car des *LTRs* sont très souvent retrouvés à leurs côtés (Bowen et al., 2003). La région dupliquée s'étend sur 60 kp sur le bras gauche du chromosome 3 (figure 40A). L'analyse des bordures de cette duplication a montré qu'elle consiste en une fusion de *wtf8* et de *wtf9* (figure 40B et données non montrées). La cartographie

par PCR suivie de séquençage Sanger des bordures montre que le dernier site utilisé pour la recombinaison est situé dans le *LTR* n°3 (figure 40C). Ces données suggèrent que ce dernier site serait l'endroit où la fourche de réplication casserait pour initier la RH. J'ai donc analysé la réplication au niveau du LTR n°3.

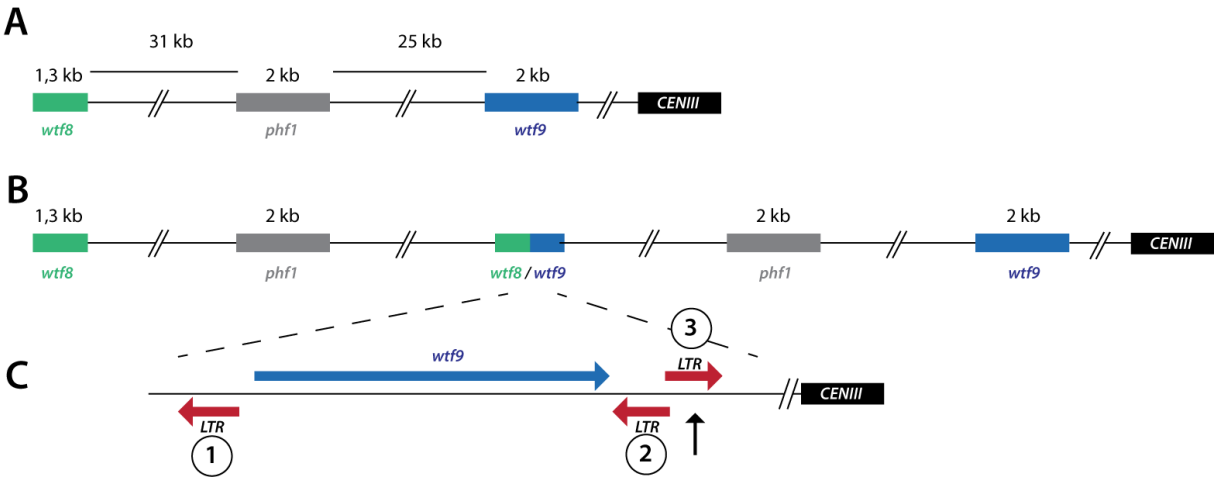


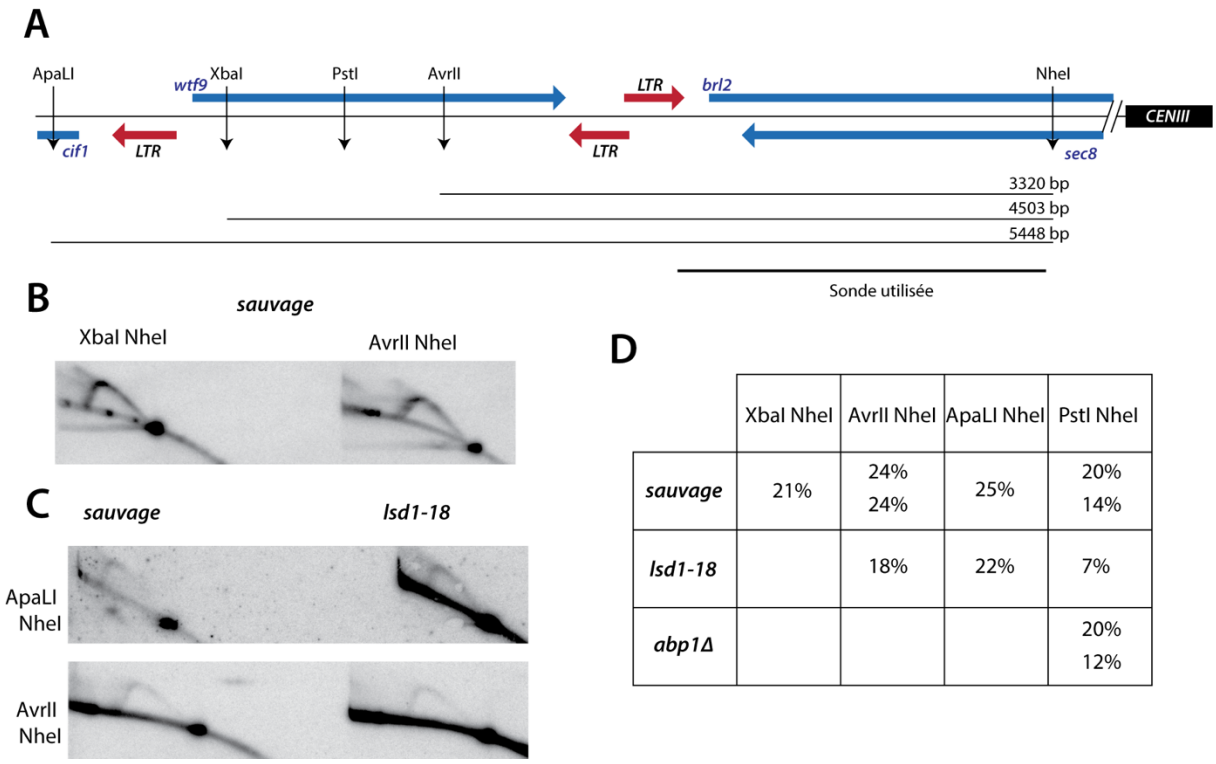
Figure 40 : Schéma de la région dupliquée.
A) Schéma de la région du bras gauche du chromosome 3 contenant *wtf8*, en vert, *phf1*, en gris et *wtf9*, en bleu. La distance séparant chaque gène est indiquée en kb. B) Représentation de la duplication observée chez les mutants suppresseurs. C) Zoom sur le locus *wtf9* qui contient la bordure de la duplication. Les *LTRs* sont numérotés. Le site de la bordure la plus extrême est indiqué par une flèche.

5.2 Lsd1 contrôle-t-elle la réplication de *wtf9* ?

L'analyse de la propagation de la fourche de réplication au niveau des *LTRs* a montré que ce sont des régions difficiles à répliquer, et que des pauses de la fourche de réplication peuvent être observées. J'ai, dans un premier temps, déterminé si une pause de la fourche de réplication était observable au niveau de la bordure de cette duplication. Pour ce faire, j'ai analysé l'accumulation d'intermédiaires de réplication par gel bidimensionnel d'agarose (figure 41). Les résultats suggèrent que la fourche de réplication fait une pause au niveau de *wtf9* (figure 41B). D'autre part, la mutation *lsd1-18* (mutation ponctuelle dans le domaine HMG de *lsd1*) ne semble pas affecter la progression de la fourche de réplication (figure 41C). Le résumé des différentes analyses nous suggère que la mutation *lsd1-18* n'affecte pas la réplication de *wtf9* à la différence de ce qui est observé à *mat1* (figure 41D).

Ces résultats sont en accord avec les résultats de ChIP-séq qui ont été à notre disposition plus tardivement. En effet, Lsd1 n'est pas enrichie au niveau de *wtf9* ou de *wtf8* (figure 42). Ces résultats excluent le fait que Lsd1 contrôle directement la fourche de réplication à ces loci.

Comme déjà souligné précédemment, Abp1 et Sap1 sont présentes au niveau des LTRs adjacents à *wtf9* (figure 42). Cependant, des résultats préliminaires d'analyse de la réplication suggèrent qu'Abp1 ne contrôle pas la fourche de réplication au niveau de *wtf9* (figure 41D).



	XbaI NheI	AvrII NheI	ApaLI NheI	PstI NheI
sauvage	21%	24% 24%	25%	20% 14%
lsd1-18		18%	22%	7%
abp1Δ				20% 12%

Figure 41 : Lsd1 ne contrôle pas la réplication au niveau de *wtf9*.
A) Schéma de la région du bras gauche du chromosome 3 contenant *wtf9*. Les sites des enzymes de restriction utilisés sont indiqués. La taille des fragments obtenus ainsi que la sonde utilisée sont notés. B) Analyse des intermédiaires de réplication dans la souche sauvage au niveau de *wtf9*. Les enzymes de restriction utilisées sont indiquées. C) Analyse des intermédiaires de réplication dans la souche sauvage et *lsd1-18* au niveau de *wtf9*. Deux digestions enzymatiques ont été testées. D) Résumé des différentes analyses. La quantification du signal d'hybridation est le rapport du signal (pause / arc total).

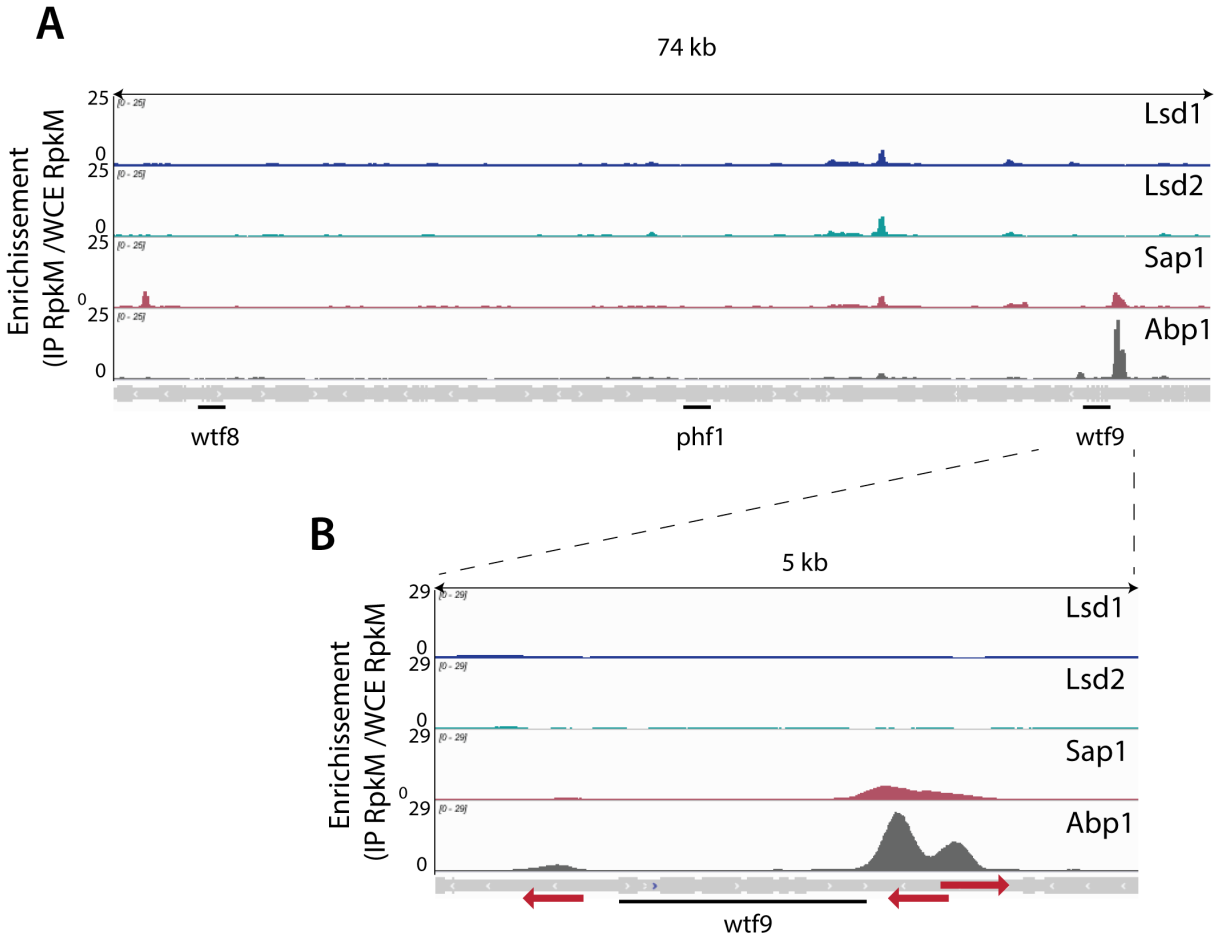


Figure 42 : Lsd1 ne contrôle pas directement la réplication au niveau de *wtf9*.
A) Distribution des enrichissements normalisés (IP RPkM / WCE RPkM) des ChIP-séq de Lsd1, Lsd2, Sap1 et Abp1 dans le fond génétique h^{90}. Les loci *wtf8*, *phf1* et *wtf9* sont indiqué. B) Agrandissement du locus *wtf9*. Sap1 et Abp1 sont enrichies du côté droit de *wtf9*.

5.3 Conclusions

Les données d'analyses des intermédiaires de la réplication et de ChIP-séq au niveau de la région comportant la duplication de *phf1* montrent que Lsd1 ne contrôle pas directement la réplication de ce locus. Le fait que dans le crible effectué quatorze des vingt-quatre mutants isolés comportent cette duplication indique un fort avantage sélectif. De plus, d'autres duplications n'impliquant pas *wtf8* et *wtf9* ont été détectées (données non montrées). Ces duplications, contrairement à celle de *phf1*, n'ont pas d'effet suppresseur sur la croissance très ralentie des mutants *lsd1*Δ. Le fait que des mutations "passagères" soient mises en évidence suggère qu'en absence de *lsd1* le contrôle du nombre de copies de différents loci est affecté. Le mécanisme sous-jacent n'est pour le moment pas connu.

6. Computational signature of the imprintosome complex at the *mat1* locus in fission yeast

Raimondi[1] C, Jagla B[2], Waxin H[3], Gangloff[1] S, Arcangioli[1] B.

[1] *Genome Dynamics Unit, UMR 3525 CNRS, Institut Pasteur, 25-28 rue du Docteur Roux, Paris , France.*

[2] *Plate-forme Transcriptome et Epigenome, BioMics, Centre d'innovation et recherche technologique, Institut Pasteur, 25-28 rue du Docteur Roux, Paris, France.*

[3] *Directeur délégué à l'enseignement, Institut Pasteur, 25-28 rue du Docteur Roux, Paris, France.*

Cet article est en cours de préparation et reprend les résultats majeurs détaillés précédemment. Dans cet article les protéines Abp1, Lsd1/2 et Sap1 sont cartographiées au niveau de la région sexuelle. Plus particulièrement, les résultats montrant que Lsd1/2 sont localisés à *mat1* indépendamment de l'allèle sont présentés. D'autre part, la signature de l'empreinte par séquençage à haut débit est également détaillée. Cette signature se caractérise sur le brin Watson, par un 5' compte biaisé au niveau du site de l'empreinte et sur le brin Crick par une plus large région (20 nucléotides) enrichie en 5' compte. Cette signature est ensuite utilisée afin de déterminer la structure des protéines à *mat1*. L'analyse suggère la présence d'un complexe nucléo-protéique : l'imprintosome, qui permettrait la protection de l'empreinte.

Abstract

Genetic and molecular studies have indicated that an epigenetic imprint at *mat1*, the sexual locus of fission yeast, initiates mating type switching. The polar DNA replication of *mat1* generates an imprint on the Watson strand. The process by which the imprint is formed and maintained through the cell cycle remains unclear. To understand better the mechanism of imprint formation and stability, we characterized the recruitment of early players of mating type switching at the *mat1* region. We found that the switch activating protein 1 (Sap1) is preferentially recruited inside the *mat1M* allele on a sequence (*SS13*) that enhances the imprint. The lysine specific demethylases, Lsd1/2, that control the replication fork pause at *MPS1* and the formation of the imprint are specifically drafted inside of *mat1*, regardless of the allele. The CENP-B homolog, Abp1, is highly enriched next to *mat1* but it is not required in the process. Additionally, we established the computational signature of the imprint. Using this signature, we show that both sides of the imprinted molecule are bound by Lsd1/2 and Sap1, suggesting a nucleoprotein protective structure defined as imprintosome.

Introduction

Haploid *Schizosaccharomyces pombe* cells exist as two mating types (MTs), *P* (for plus) and *M* (for minus), that switch during cell divisions. The MT of the cell is determined by the allele (*M* or *P*) expressed at the *mat1* locus on the right arm of chromosome II, 500 kb away from the centromere: *mat1P* in *P* cells and *mat1M* in *M* cells (Beach et al., 1982; Kelly et al., 1988). The *mat1* allele can be replaced efficiently by genetic information contained in one of the two silent donor cassettes *mat2P* and *mat3M* (Beach, 1983; Egel and Gutz, 1981; Kelly et al., 1988). *mat2P* and *mat3M* are located 17 kb and 29 kb centromere-distal to *mat1*, respectively (Beach, 1983; Egel, 1984), and are maintained in a silent chromatin state, preventing their transcription or recombination (Egel, 1984; Grewal, 2000; Thon and Verhein-Hansen, 2000). Two inverted repeats flank the donor region (*IR-L* and *IR-R*) and act as boundary elements (Noma K et al., 2001; Thon et al., 2002). Additionally, a centromeric repeated sequences located between *mat2P* and *mat3M*, *CEN-H*, promotes the formation of heterochromatin {Grewal:1997vm}, see Figure 1). A salient feature of the MT loci is that they are flanked by homologous sequences. The H1 homology box (59 bp) is located on the centromere distal (right) side of the cassettes and the H2 homology box (135 bp) on the centromere proximal (left) side (Kelly et al., 1988). Both sequences are thought to be essential for base pairing during the initiation and resolution steps of the gene conversion process required for MT switching (Arcangioli and de Lahondès,

2000; Kaykov et al., 2004). The H3 box (57 bp) is located to the left of H2 at *mat2* and *mat3* (Kelly et al., 1988), and is not required for MT switching (Fleck et al., 1994).

Extensive pedigree analysis at the single-cell level demonstrated that two consecutive divisions are required to produce one switched cell among four related cousins. Fission yeast possesses a remarkable genetically programmed system initiated by a site and DNA strand-specific imprint at the *mat1* locus for changing its MT (reviewed in (Arcangioli and Thon, 2004; Dalgaard and Klar, 2001; Klar et al., 2014)). The first division produces the imprint on one of the sister chromatids at *mat1*, while the second division triggers a double-strand break (DSB) on the imprinted chromatid that initiates MT switching (Egel and Eie, 1987; Miyata and Miyata, 1981).

It was shown that the polarity of DNA replication of *mat1* is instrumental in the establishment and the strand specificity of the imprint {Dalgaard:1999ee}. Two sites control the replication polarity at *mat1*, *RTS1* that blocks the fork coming in from the centromere and *MPS1* located near the site of the imprint that pauses the fork coming in from the right. The pausing of the replication fork at *MPS1* is a prerequisite for the formation of the imprint (Dalgaard and Klar, 2001). Thus, the imprint is made on the newly synthesized lagging strand during the resumption of DNA synthesis at *MPS1* (Holmes et al., 2005; Sayrac et al., 2011).

Molecular studies have suggested that the imprint at *mat1* is either a single-strand DNA break (Arcangioli, 1998; Kaykov and Arcangioli, 2004) or one or two ribonucleotides (Vengrova and Dalgaard, 2006). The imprint was mapped to the Watson strand corresponding to the neo-synthesized lagging strand, at the junction of the *mat1* allele and the H1 homology box (Nielsen and Egel, 1989). Furthermore, the position of the imprint is site-specific but sequence-independent (Kaykov and Arcangioli, 2004; Sayrac et al., 2011). Interestingly, the position of the break on the Watson strand at *mat1* differs by three nucleotides between the *mat1P* and *mat1M* alleles (Nielsen and Egel, 1989), indicating that the position is dictated from within the *mat1M* and *mat1P* sequences. The imprint is protected against DNA repair and remains stable throughout the entire length of the cell cycle to be transiently converted to a polar DSB during the following S-phase (Arcangioli, 1998; Holmes et al., 2005; Kaykov and Arcangioli, 2004). Repair of this break is not random; *mat1P* prefers *mat3M*, and *mat1M* prefers the nearby *mat2P* in 90 % of switches. This preference is called directionality of switching {Thon:1993tn}. In strains in which the donor sequences are deleted {Klar:1986uv}, the position and level of the imprint are identical to that in wild type, but the DSB is repaired off the sister (Nielsen and Egel, 1989; Roseaulin et al., 2008).

Several *cis*-acting sequences are required for the formation and maintenance of the imprint. The SS2 and SS13 deletions are located within *mat1M*, SS2 is important for the pause and the imprint, whereas SS13 acts as a spacer element that allows efficient imprinting (Sayrac et al., 2011). To the right of *mat1M*, the *Msmt-0* deletion maintains efficient pausing but abolishes the imprint {Styrkarsdottir:1993wp Dalgaard:1999ee}. The Lsd1/2 (Lysine Specific Demethylases) and the Swi1/Swi3 (Switch gene) complexes are necessary for the pause and the subsequent imprint (Dalgaard and Klar, 2000; Holmes et al., 2005; 2012). Lsd1 was found into a complex including Lsd2 together with Phf1/Phf2, two plant homeodomain fingers. The complex also associates with, topoisomerase 2 (top2), replication factor A (Rfa1) and Sap1 (Lan et al., 2007b), essential proteins with a role in DNA replication. The implication of Lsd1 in DNA replication and DNA damage responses, beside transcription regulation is not new (Nicholson and Chen, 2009). The Lsd1 and Lsd2 act redundantly to promote the pause through both their HMG domains and amine oxidase activities and are recruited to *mat1* through SS2, (Holmes et al., 2012). Swi1/3 proteins are associated to the replication fork and stabilize the fork during the pause (Dalgaard and Klar, 2000; Noguchi et al., 2004). Sap1, the Switch Activating Protein 1 is an essential gene that interacts with SAS1, a sequence covered by the *Msmt-0* deletion (Arcangioli et al., 1994). Sap1 was also referenced as having a role into DNA replication (Guan et al., 2017; Krings and Bastia, 2005; Mejia-Ramirez et al., 2005; Zaratiegui et al., 2011) and chromosomal organization (Mizuguchi et al., 2017). Finally, Abp1 (ARS Binding Protein 1), a CENP-B (Centromeric Protein B) homolog that antagonizes Sap1 at LTRs, is enriched next to *SAS1*, to the right of *mat1M* (Zaratiegui et al., 2011). Abp1 is proposed to control the directionality of switching by regulating the alternative expression of the *swi2* gene (Switch gene 2) (Aguilar-Arnal et al., 2008; Matsuda et al., 2011; Yu et al., 2012) important for the distribution of the Swi2/Swi5 recombination-promoting complex in the *mat2/3* region (Yu et al., 2012).

To advance our understanding of the process of imprint formation and stability, we analyzed by chromatin-immunoprecipitation and whole genome sequencing (ChIP-seq) the recruitment of Abp1, Lsd1, Lsd2, and Sap1 proteins at the *mat1* region. Because Lsd1/2 complex interacts with a *cis*-acting element within the MT cassettes and promotes replication pause at *mat1* but not at *mat2P* and *mat3M*, we compare the interactions of the four proteins in the wild-type strain and in two strains containing a deletion of the *mat2P* and *mat3M* cassettes. Doing so, we established the computational signature of the imprint and proposed a model whereby Lsd1 and Lsd2 are exclusively recruited to the *mat1* locus.

Result

ChIP-sequencing mapping of Abp1, Lsd1/2 and Sap1 to the h^{90} mating type region

It has been previously reported that Abp1, Lsd1/2 and Sap1 bind to the *mat1* region (Arcangioli and Klar, 1991; Holmes et al., 2012; Zaratiegui et al., 2011). However, several limitations arise from those studies: The strains that were initially used to map Abp1 and Sap1 by ChIP-seq were rearranged in the MT region (h^{-S}), and the length of the reads (30 nucleotides) prevented their unambiguous assignment to *mat1, 2 or 3* cassettes (Zaratiegui et al., 2011). Concerning Lsd1/2, whole genome analyses have been performed by ChIP-chip (Gordon et al., 2007), but no detailed analysis of the MT region was performed. More recently, we mapped Lsd1/2 to *mat1* by ChIP-qPCR (Holmes et al., 2012).

To extend the mapping of these proteins to the entire MT region, we performed ChIP-seq of Abp1, Lsd1/2 and Sap1 on the wild-type and switching proficient h^{90} strain (Figure 1). To achieve proper mapping, we constructed two reference genomes carrying either the allele *M* at *mat1* or *P* (Figure 1). We found that both Lsd1 and Lsd2 are enriched at the *mat1*, *2* and *3* cassettes, whereas Sap1 was more enriched at the *mat1M* and *mat3M* cassettes than *mat1P* and *mat2P* (Figure 1). Additionally, Sap1 was found to be enriched at the IR-L and IR-R boundary elements (Noma et al., 2006; Thon et al., 2002). Concerning Abp1, we confirmed that it is indeed enriched at the left border of *CEN-H* (Aguilar-Arnal et al., 2008). Interestingly, we found that Abp1 is enriched at the right border of *mat1* in an h^{90} background. This observation prompted us to investigate the potential involvement of Abp1 in the early steps of MT switching.

Abp1 is recruited to *swi2* when *M* is present at *mat1*

Using 2D gels analysis we address the role of Abp1 for *MPS1* activity. Unlike the situation observed at LTRs, Abp1 is not required for the pausing activity of *MPS1* (Figure 2A) consistent with Abp1 being dispensable for the formation and maintenance of the imprint (Aguilar-Arnal et al., 2008). Earlier work has shown that Mc, encoded from the *mat1M* cassette, and Abp1 control the expression of Swi2, a key player in the directionality process (Matsuda et al., 2011; Yu et al., 2012). Moreover, the recruitment of Mc to *swi2* depends on Abp1 (Matsuda et al., 2011). Reciprocally, using the two stable *M* and *P* strains with deleted donor sequences, we found that the enrichment of Abp1 at *swi2* is restricted to *M* cells, while the enrichment at the neighboring peak is unaffected by the MT of the cells (Figure 2B). The genome wide analysis showed that Abp1 M-specific interaction is unique for *swi2* gene.

Lsd1/2, Sap1 and Abp1 map at *mat1*

The enrichment locations of the four proteins and the repeated nature of the MT cassettes introduce an ambiguity that cannot be resolved during alignment. To improve the location of the Lsd1/2 and Sap1 binding sites to *mat1*, we used the *mat1M* Δ*2-3* and *mat1P* Δ*2-3* strains deleted for the donor sequences, {Klar:1986uv} (Figure 3A). Abp1 is enriched outside of the *mat1* locus (to the right of the H1 cassette), and its enrichment is independent of the allele present at *mat1* (Figure 3B, 3C). Lsd1 and Lsd2 are enriched both in *mat1M* and *mat1P* with Lsd2 and Sap1 more enriched in *mat1M* (Figure 3B, 3C).

We then analyzed the reads that can only map to a unique position on the reference genome (unique mapper) in a h^{90}, wild type context (Figure 4A), where the donor sequences are present (Supplementary Figure 4). As expected, in the two h^{90} genomes (h^{90} *M* and h^{90} *P*), the highly repetitive sequences of IR-L and IR-R are no longer covered (Supplementary Figure 4A and 4B). When the *M* allele is at *mat1*, we found no unique mapper reads at *mat1* and *mat3* (Figure 4B). Similarly, when the *P* allele is at *mat1*, no unique mapper reads at *mat1* and *mat2* were found (Figure 4C). Abp1 is located outside of the *mat1* cassette in a unique region and can thus be used for alignment. In the case of Lsd1/2, the enriched unique mappers are well maintained at *mat1* (both in the *M* and *P* context) and strongly diminished, almost to background level, at *mat2P* and *mat3M* loci, indicating that Lsd1/2 are preferentially recruited at *mat1* and not at the two silent cassettes. The situation for Sap1 is less clear, its profile suggests that it is recruited to *M* but poorly to *P*, suggesting that in addition to SAS1, Sap1 is recruited within *mat1M*.

Characterization of the Sap1 binding in the *M* allele

We used *in vitro* and *in vivo* complementary approaches to further determine Sap1 binding sequences in the *M* allele. The gel shift experiments were achieved using 14 double-stranded DNA probes labeled in 5'. Two positive controls were included (*SAS1* and *Ter1*) in addition to 10 regions mapping inside the *M* allele and 2 outside (Figure 5A). Using crude protein extracts, we observed the expected mobility shift for *SAS1* and *Ter1* as well as for 3 probes inside the M allele (Figure 5B). We used purified Sap1 protein to confirm that it is responsible for the shift (Figure 5C). Since M4, M5 and M6 are part of a region involved in the stability of the imprint (SS13) (Sayrac et al., 2011), we analyzed by ChIP-seq the enrichment of Sap1 in several strains with no donor sequences containing deletions spanning the *mat1M* locus. The *Msmt-0* deletion does not change the enrichment signal within *mat1M*, whereas the SS13 deletion strongly decreases the Sap1 signal, supporting the *in vitro* interaction result (Figure 5D *Msmt-0* and SS13). The slight decrease on the right of the *mat1Msmt-0*-containing deletion together with

the remaining weak Sap1 signal in SS13 is attributed to SAS1 interaction, located 140 bp distal to the imprint. The SS2 deletion, that abolishes DNA replication pausing and imprinting (Sayrac et al., 2011) does not affect Sap1 enrichment but brings the signal on the top of the imprint position (Figure 5D ΔSS2). These results indicate that Sap1 is preferentially enriched *in vivo* on two sequences (SS13 and SAS1) that are not required for replication fork pausing but required for imprinting formation of maintenance, at least for the *mat1M* locus.

Detection of an imprinted site by high throughput sequencing

The imprint was described as either a single stranded break (Arcangioli, 1998; Kaykov and Arcangioli, 2004) or a one/two ribonucleotide(s) (Vengrova and Dalgaard, 2004; 2006) that behave(s) as a mechanical fragile site (Arcangioli, 1998). For the sake of clarity, we portray the imprint as a single strand break in the remainder of the text. In Figure 6A, we reasoned that upon chromatin sonication, the strand opposite to the nick will preferentially break at its vicinity. Library preparation using the Illumina kit (TruSeq Chip) uses an exonuclease and a DNA polymerase to process the ends in order to achieve efficient ligation prior to sequencing (Figure 6B). This method conserves the first nucleotide of the 5' end of all sonicated fragments. Consequently, reads starting at the imprint are expected to be more frequent in the whole chromatin extract even prior to any chromatin immunoprecipitation. Furthermore, it is possible to count for each DNA strand (Watson and Crick) the number of 5' ends at each position in the genome, and we will refer to this count as "5' count" in the rest of the manuscript (see 5' count analysis in Materials and Methods). A whole genome distribution of the 5' count was established on the *mat1M* Δ*2-3* (Figure 6C) and *mat1P* Δ*2-3* and h^{90} strains (Figure 7B-C) using 130 nucleotide-long reads. 7 enriched regions were detected on the Watson strand (Figure 6C). 6 such regions are also detected on the Crick strand (Figure 6C). They delineate either repeated regions or regions exhibiting homology with the mitochondrial genome (Supplementary table 1). The last enriched 5'count is only found on one strand, the Watson, and is not a region but corresponds to the imprint site at *mat1* (Figure 6C, in red).

We next focused on the 10 kb region surrounding the *mat1* locus (Figure 7). The overall distribution of 5' count on the Watson and Crick strands is similar, except for one nucleotide (in red) significantly over-represented on the Watson strand (Figure 7). The position in *mat1M* Δ*2-3* and *mat1P* Δ*2-3* corresponds to the imprinted position that is 3 nucleotides apart in *M* and *P* (Nielsen and Egel, 1989) (Figure 7A and 7B). As expected, both positions are enriched in 5' count in the h^{90} background that contains a mixture of *mat1M* and *mat1P* alleles (Figure 7C). To confirm that the signal results from the imprint, we used the *swi1*Δ mutant

strain that exhibits a low level of imprint. In this genetic background, no nucleotide shows any bias in 5' count (Figure 7D).

Because the bias in 5’ count is only visible on the imprinted strand, we conclude that sonication does not preferentially break the phosphate bond facing the nick. Therefore, we searched on the Crick strand for a region with a strong 5’ count statistical bias. Using a sliding window with different window sizes we calculated the p-value in the 10 kb region surrounding *mat1* (Figure 8). As expected, on the Watson strand the window of one nucleotide at the site of the imprint has the lowest p-value (Figure 8A and 8B). On the Crick strand, the most significant p-value is contained in a window of 17 to 23 nucleotides facing the imprint (Figure 8C and 8D). This correspond to the molecular signature upon chromatin sonication of a confined and persistent single-strand break, as a fixed nucleotide on one strand with larger region on the opposite strand (Figure 8E and 8F).

Immunoprecipitation of the imprinted chromatid

To determine whether Lsd1, Lsd2, and Sap1 interact with the imprinted chromatid and on the right or left of the imprint we used the molecular signature defined above. Figure 9A shows the expected distributions of enriched 5’ count upon immunoprecipitation. Since Lsd1/2 and Sap1 bind the repeated *mat* cassettes, we circumvented the alignment problem by performing the experiment in the donor less strains. For Lsd1, when *M* is at *mat1*, the enrichment of the 5' count at the imprint on the Watson and Crick strands (Figure 9B) exhibits the molecular signature defined above (Figure 9C). This result shows that Lsd1 interacts on both sides or directly with the imprinted chromatid. The additional spreading of the 5’ count on both DNA strands indicates that Lsd1 also interacts with the chromatid in the absence of imprint (Figure 9B). A similar 5’ count distribution is observed when *P* is at *mat1* (Figure 9D and 9E). For Lsd2 and Sap1 a similar result is observed when *M* is at *mat1*. When *P* is at *mat1*, Lsd2 and Sap1 interaction is weaker (Figure 9D and 9E). Altogether, the results indicate that Lsd1, Lsd2 and Sap1 interact with *mat1M* and *mat1P*, though with a weaker contribution of Lsd2 and Sap1 with *mat1P,* regardless of the imprint and on both sides of the imprint.

Discussion

In this work, we proposed a molecular definition of the imprint upon sonication and genomic sequencing. Our computational analysis demonstrates that a site- and strand-specific persistent imprint/nick can be unambiguously mapped. The computation of the 5’ count reveals a unique

accumulation at the site of the imprint on the Watson strand, whereas the opposite Crick strand is mechanically broken by sonication within a window of 20 nucleotides, thus yielding a more diffuse accumulation of 5' ends. This molecular signature is consistent with the free rotation of the Crick strand that generates a population of molecules with weakened A-T rich Watson-Crick pairing on both sides of the nick and makes the resulting single-strand in this window highly prone to breakage.

We also provide data on the chromosomal enrichment at the genome wide level of four proteins, Lsd1, Lsd2, Sap1, and Abp1. Abp1 is strongly interacting with the sequences near *mat1*, and a function in replication fork pausing and imprinting formation was ruled out. The M-specific enrichment of Abp1 next to the Mc binding site located in the *swi2* gene reinforces the role of Swi2 protein in directionality (Matsuda et al., 2011; Yu et al., 2012).

In mammals, Lsd1 plays an important role in gene expression by establishing and maintaining proper epigenetic marks. Several layers control the chromosomal targeting of Lsd1 for positive or negative transcription regulation. Globally, CoREST targets Lsd1 complex to histone H3K4 (Shi et al., 2004), whereas human androgen receptor targets Lsd1 to histone H3K9 (Metzger et al., 2005). The plant homeodomain finger protein BHC80 and the HMG protein Braf35 are part of the core complex Lsd1-CoREST (reviewed in (Meier and Brehm, 2014)). BHC80 participates to histone mark reading (Lan et al., 2007a) and Braf35 (Hakimi et al., 2002) is a structural DNA-binding protein. In fission yeast, several pieces of evidence support the specific demethylation of H3K9 at many promoters and boundary elements but its role on H3K4 was not established (Lan et al., 2007b; Opel et al., 2007). Lsd1 is found in a complex including Lsd2 and Phf1/Phf2, two plant homeodomain finger proteins (Lan et al., 2007a). More generally, Lsd1 and Lsd2 are fused to an HMG domain in fungi (Nicolas et al., 2006). Taken together, the protein composition of the Lsd1/2 core complex is quite similar in fission yeast and mammals and the main functional difference resides in the absence of H3K4 demethylation. Because Sap1 interacts with the Lsd1/2 complex (Lan et al., 2007b) and that Sap1 and Lsd1/2 complex are found at active promoters (see Supplemental Figure 9), we propose that Sap1 participates to the targeting of the Lsd1/2 complex to prevent H3K9 methylation at key promoters and defines heterochromatin boundaries (Lan et al., 2007b; Li et al., 2008).

The situation is more complicated at the mating type loci. Although *mat1P* and *mat1M* exhibit no homology, a remarkable finding is that Lsd1 and Lsd2 are enriched in both alleles on sequences that have been shown to be necessary on the chromosome to promote replication fork pauses at *MPS1* and sufficient in a plasmid system (Sayrac et al., 2011). Intriguingly, when *P* and *M* sequences are present at the silent *mat2P* and *mat3M* loci, Lsd1 and Lsd2 are absent

consistent with the absence of replication fork pause at these loci (Sayrac et al., 2011). This result suggested that the heterochromatin at the two silent cassettes is preventing the binding of Lsd1/2 and pointed the Phf1/Phf2 histone code readers as targeting partners. Limiting the access of H3K9 demethylase enzymes is also relevant to maintain the two cassettes silent.

By using the molecular signature of the imprint, we provide strong evidence that the Lsd1/2 complex is interacting at the imprint site, prior and after imprinting formation. We had reached the same conclusion earlier by showing that Lsd1/2 are working upstream of Swi1/3 and are still enriched at *mat1* in the absence of imprint (Holmes et al., 2012). The two Sap1 binding sites on both sides of the imprint (SS13 and SAS1), overlaps with two regions that are dispensable for replication fork pause and Lsd1/2 initial recruitment but are essential for imprinting formation or maintenance. The simplest model will be that Lsd1/2 complex is recruited to *mat1*, independently of Sap1, to promote the replication fork pause at *MPS1* (Figure 10). Following replication fork restart and imprinting formation, Sap1 together with Lsd1/2 are forming a chromatin structure that is protecting the imprinting from being repair. Perhaps, the molecular signature of the imprint on the Crick strand is influenced by the imprintosome. We now propose that Lsd1/2 also control the pausing of the replication fork at *mat1* on both the imprinted and the non-imprinted chromatids, therefore allowing to setup the stage for the fork to restart in the same context at each round of DNA replication and generate recurrently and efficiently the imprint.

Materials and Methods

Fission Yeast Strains and Genetic Procedures:

The *S. pombe* strains used in this study are listed in Table 1. Standard media (YES) and genetic protocols for fission yeast were used (Moreno et al., 1991).

2D Gel Analysis:

2D gel analysis of replication intermediates was carried out as described previously (Brewer and Fangman, 1988). DNA was prepared and digested in agarose plugs with the indicated restriction enzymes (Kaykov et al., 2004). Enriched fractions for replication intermediates were obtained using BND cellulose columns (SIGMA). Gels were blotted onto Hybond-N+ nylon membrane. The probes were labeled with alpha-^{32}P dCTP and the blot analyzed on a PhosphoImager and quantified with the ImageQuant software (pause/Y arc + pause in percentage) (GE Healthcare Life Sciences). Sequences of primers used to produce the probe are indicated in Table2.

Gel shift assay:

Production of Sap1-6His by *E. coli* was induced with 1 mM IPTG as previously described {Arcangioli:1994un}. Sap1-6His was purified using column of Ni-NTA His-tag from Qiagen according to the manufacturer's instructions. Sequences used are listed in Table 2. 1.50 ng of a given oligonucleotide was labeled with gamma-^{32}P with T4 polynucleotide Kinase (NEB) and incubated with its cold complementary sequence. 1 ng of the resulting double stranded probe was incubated with 5 µg of total protein extract or with 40 ng of purified Sap1-6His (Arcangioli and Klar, 1991). Gel shift was performed as previously described {Arcangioli:1985to}.

Chromatin ImmunoPrecipitation (ChIP):

ChIP-sequencing was performed as previously described (Zaratiegui et al., 2011) with the following modifications: fixation for ChIP of Lsd1 and Lsd2 were carried out at 25 °C for 20 min with 1 % of paraformaldehyde (PFA) and fixation for ChIP of Sap1 and Abp1 were performed at 18 °C for 15 min with 3 % of PFA. For Lsd1 and Lsd2, an Anti-Myc-Tag mAb-magnetic Beads (M047-11 MBL) was used. For Abp1 and Sap1, Dynabeads Protein G 10007D (Invitrogen) coupled with Tap Tag antibody CAB1001 (Pierce Thermo) or with polyclonal serum against the native Sap1(de Lahondès et al., 2003), respectively, were used.

Illumina Sequencing of DNA:

Libraries were prepared using TruSeq ChIP Library Preparation kit following the manufacturer's instructions (Illumina). Libraries are checked for concentration and quality on DNA chips with a Bioanalyser (Agilent). The libraries were quantified by fluorimetric measurements with the Qubit® dsDNA HS Assay Kit (ThermoFisher). 65-bp and 130-bp Single Read sequences were generated on the Hiseq2500 sequencer according to manufacturer's instructions (Illumina). Library statistics are provided in Table 3.

Illumina Read Processing and Alignment of DNA Libraries:

Illumina reads were aligned to the *S. pombe* genome assembly ASM294v2.23 using Bowtie v2.1.0 (Langmead et al., 2009) with default parameters and multi-mapper reads randomly distributed. Using BamTools (Barnett et al., 2011), read counts were normalized to reads per million per kilo base (RPKM), using the total library size and a bin size of 1 Kb. For the analysis of multi-mapper reads, we filtered aligned read files with samtools (Li and Durbin, 2009) using the parameter -q 10.

ChIP-Sequencing Analysis:

ChIP peaks versus appropriate inputs (whole-cell extracts) were called using MACS v2.1.0 (Zhang et al., 2008). Statistic validation of peaks was performed using the IDR method with ENCODE recommendations (Landt et al., 2012). Genome browser tracks and meta-analysis were created using enrichment (IP reads per million (RPM) / input RPM) of indicated sequencing. Intersection of lists of peaks were calculated with BEDTools v.2.17.02 (Quinlan and Hall, 2010) (parameters: -u). For the transcriptional analysis 2 groups of 108 genes were defined and intersected with the list of peaks shared by Lsd1/2 and Sap1. The first group contains the 108 most present transcripts and the 108 less present transcripts using the normalized data from (Marguerat et al., 2012).

5' count analysis:

We used a script to extract the beginning of reads that map on the Watson or Crick strands. Enrichments of 5' count were obtained using the total library size. p-values were obtained using the negative binomial distribution fitting function in R.

All scripts used for the analysis and outputs are available upon request.

Acknowledgments

We thank the Pasteur Institute PF2 platform to provide the sequencing data. We thank Caroline Proux for the library preparation and the sequencing result. This project was support by the ARC foundation and the FRM.

References

Aguilar-Arnal, L., Marsellach, F.-X., and Azorín, F. (2008). The fission yeast homologue of CENP-B, Abp1, regulates directionality of mating-type switching. Embo J. *27*, 1029–1038.

Arcangioli, B. (1998). A site- and strand-specific DNA break confers asymmetric switching potential in fission yeast. Embo J. *17*, 4503–4510.

Arcangioli, B., and de Lahondès, R. (2000). Fission yeast switches mating type by a replication-recombination coupled process. Embo J. *19*, 1389–1396.

Arcangioli, B., and Klar, A.J. (1991). A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient *mat1* switching in *Schizosaccharomyces pombe*. Embo J. *10*, 3025–3032.

Arcangioli, B., and Thon, G. (2004). Mating-Type Cassettes: Structure, Switching and Silencing. In The Molecular Biology of *Schizosaccharomyces Pombe*, (Berlin, Heidelberg: Springer, Berlin, Heidelberg), pp. 129–147.

Arcangioli, B., Copeland, T.D., and Klar, A.J. (1994). Sap1, a protein that binds to sequences required for mating-type switching, is essential for viability in *Schizosaccharomyces pombe*. Mol. Cell. Biol. *14*, 2058–2065.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stroemberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics *27*, 1691–1692.

Beach, D.H. (1983). Cell type switching by DNA transposition in fission yeast. Nature *305*, 682–688.

Beach, D., Nurse, P., and Egel, R. (1982). Molecular rearrangement of mating-type genes in fission yeast. Nature *296*, 682–683.

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell *55*, 637–643.

Dalgaard, J.Z., and Klar, A.J. (2000). *swi1* and *swi3* perform imprinting, pausing, and termination of DNA replication in *S. pombe*. Cell *102*, 745–751.

Dalgaard, J.Z., and Klar, A.J. (2001). Does *S. pombe* exploit the intrinsic asymmetry of DNA synthesis to imprint daughter cells for mating-type switching? Trends Genet. *17*, 153–157.

de Lahondès, R., Ribes, V., and Arcangioli, B. (2003). Fission yeast Sap1 protein is essential for chromosome stability. Eukaryotic Cell *2*, 910–921.

Egel, R. (1984). Two tightly linked silent cassettes in the mating-type region of *Schizosaccharomyces pombe*. Curr. Genet. *8*, 199–203.

Egel, R., and Eie, B. (1987). Cell lineage asymmetry in *Schizosaccharomyces pombe*: unilateral transmission of a

high-frequency state for mating-type switching in diploid pedigrees. Curr. Genet. *12*, 429–433.

Egel, R., and Gutz, H. (1981). Gene activation by copy transposition in mating-type switching of a homothallic fission yeast. Curr. Genet. *3*, 5–12.

Fleck, O., Rudolph, C., Albrecht, A., Lorentz, A., Schar, P., and Schmidt, H. (1994). The mutator gene *swi8* effects specific mutations in the mating-type region of *Schizosaccharomyces pombe*. Genetics *138*, 621–632.

Gordon, M., Holt, D.G., Panigrahi, A., Wilhelm, B.T., Erdjument-Bromage, H., Tempst, P., Bähler, J., and Cairns, B.R. (2007). Genome-wide dynamics of SAPHIRE, an essential complex for gene activation and chromatin boundaries. Mol. Cell. Biol. *27*, 4058–4069.

Grewal, S.I.S. (2000). Transcriptional silencing in fission yeast. Journal of Cellular Physiology *184*, 311–318.

Guan, L., He, P., Yang, F., Zhang, Y., Hu, Y., Ding, J., Hua, Y., Zhang, Y., Ye, Q., Hu, J., et al. (2017). Sap1 is a replication-initiation factor essential for the assembly of pre-replicative complex in the fission yeast *Schizosaccharomyces pombe*. J. Biol. Chem. *292*, 6056–6075.

Hakimi, M.-A., Bochar, D.A., Chenoweth, J., Lane, W.S., Mandel, G., and Shiekhattar, R. (2002). A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. Proc. Natl. Acad. Sci. U.S.a. *99*, 7420–7425.

Holmes, A.M., Kaykov, A., and Arcangioli, B. (2005). Molecular and cellular dissection of mating-type switching steps in *Schizosaccharomyces pombe*. Mol. Cell. Biol. *25*, 303–311.

Holmes, A.M., Roseaulin, L., Schurra, C., Waxin, H., Lambert, S., Zaratiegui, M., Martienssen, R.A., and Arcangioli, B. (2012). Lsd1 and lsd2 control programmed replication fork pauses and imprinting in fission yeast. Cell Rep *2*, 1513–1520.

Kaykov, A., and Arcangioli, B. (2004). A programmed strand-specific and modified nick in *S. pombe* constitutes a novel type of chromosomal imprint. Curr. Biol. *14*, 1924–1928.

Kaykov, A., Holmes, A.M., and Arcangioli, B. (2004). Formation, maintenance and consequences of the imprint at the mating-type locus in fission yeast. Embo J. *23*, 930–938.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. Embo J. *7*, 1537–1547.

Klar, A.J.S., Ishikawa, K., and Moore, S. (2014). A Unique DNA Recombination Mechanism of the Mating/Cell-type Switching of Fission Yeasts: a Review. Microbiol Spectr *2*.

Krings, G., and Bastia, D. (2005). Sap1p binds to Ter1 at the ribosomal DNA of *Schizosaccharomyces pombe* and causes polar replication fork arrest. J. Biol. Chem. *280*, 39135–39142.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007a). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718–722.

Lan, F., Zaratiegui, M., Villén, J., Vaughn, M.W., Verdel, A., Huarte, M., Shi, Y., Gygi, S.P., Moazed, D., Martienssen, R.A., et al. (2007b). *S. pombe* LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. Mol. Cell *26*, 89–101.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, F., Huarte, M., Zaratiegui, M., Vaughn, M.W., Shi, Y., Martienssen, R., and Cande, W.Z. (2008). Lid2 is

required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. Cell *135*, 272–283.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell *151*, 671–683.

Matsuda, E., Sugioka-Sugiyama, R., Mizuguchi, T., Mehta, S., Cui, B., and Grewal, S.I.S. (2011). A homolog of male sex-determining factor SRY cooperates with a transposon-derived CENP-B protein to control sex-specific directed recombination. Proc. Natl. Acad. Sci. U.S.a. *108*, 18754–18759.

Meier, K., and Brehm, A. (2014). Chromatin regulation: how complex does it get? Epigenetics *9*, 1485–1495.

Mejia-Ramirez, E., Sanchez-Gorostiaga, A., Krimer, D.B., Schvartzman, J.B., and Hernandez, P. (2005). The mating type switch-activating protein Sap1 is required for replication fork arrest at the rRNA genes of fission yeast. Mol. Cell. Biol. *25*, 8755–8761.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436–439.

Miyata, H., and Miyata, M. (1981). Mode of conjugation in homothallic cells of *Schizosaccharomyces pombe*. J. Gen. Appl. Microbiol. *27*, 365–371.

Mizuguchi, T., Taneja, N., Matsuda, E., Belton, J.-M., FitzGerald, P., Dekker, J., and Grewal, S.I.S. (2017). Shelterin components mediate genome reorganization in response to replication stress. Proc. Natl. Acad. Sci. U.S.a. *114*, 5479–5484.

Moreno, S., Klar, A., and Nurse, P. (1991). Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. In Molecular Biology of the Fission Yeast *S. Pombe*, (Elsevier), pp. 795–823.

Nicholson, T.B., and Chen, T. (2009). LSD1 demethylates histone and non-histone proteins. Epigenetics *4*, 129–132.

Nicolas, E., Lee, M.G., Hakimi, M.-A., Cam, H.P., Grewal, S.I.S., and Shiekhattar, R. (2006). Fission yeast homologs of human histone H3 lysine 4 demethylase regulate a common set of genes with diverse functions. J. Biol. Chem. *281*, 35983–35988.

Nielsen, O., and Egel, R. (1989). Mapping the double-strand breaks at the mating-type locus in fission yeast by genomic sequencing. Embo J. *8*, 269–276.

Noguchi, E., Noguchi, C., McDonald, W.H., Yates, J.R., and Russell, P. (2004). Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. Mol. Cell. Biol. *24*, 8342–8355.

Noma K, Allis, C.D., and Grewal, S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science *293*, 1150–1155.

Noma, K.-I., Cam, H.P., Maraia, R.J., and Grewal, S.I.S. (2006). A role for TFIIIC transcription factor complex in genome organization. Cell *125*, 859–872.

Opel, M., Lando, D., Bonilla, C., Trewick, S.C., Boukaba, A., Walfridsson, J., Cauwood, J., Werler, P.J.H., Carr, A.M., Kouzarides, T., et al. (2007). Genome-wide studies of histone demethylation catalysed by the fission yeast homologues of mammalian LSD1. PLoS ONE *2*, e386.

Pohl, A., and Beato, M. (2014). bwtool: a tool for bigWig files. Bioinformatics *30*, 1618–1619.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Roseaulin, L., Yamada, Y., Tsutsui, Y., Russell, P., Iwasaki, H., and Arcangioli, B. (2008). Mus81 is essential for sister chromatid recombination at broken replication forks. Embo J. *27*, 1378–1387.

Sayrac, S., Vengrova, S., Godfrey, E.L., and Dalgaard, J.Z. (2011). Identification of a novel type of spacer element required for imprinting in fission yeast. PLoS Genet *7*, e1001328.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Thon, G., and Verhein-Hansen, J. (2000). Four chromo-domain proteins of *Schizosaccharomyces pombe* differentially repress transcription at various chromosomal locations. Genetics *155*, 551–568.

Thon, G., Bjerling, P., Bünner, C.M., and Verhein-Hansen, J. (2002). Expression-state boundaries in the mating-type region of fission yeast. Genetics *161*, 611–622.

Vengrova, S., and Dalgaard, J.Z. (2004). RNase-sensitive DNA modification(s) initiates *S. pombe* mating-type switching. Genes Dev. *18*, 794–804.

Vengrova, S., and Dalgaard, J.Z. (2006). The wild-type *Schizosaccharomyces pombe mat1* imprint consists of two ribonucleotides. EMBO Rep. *7*, 59–65.

Yu, C., Bonaduce, M.J., and Klar, A.J.S. (2012). Going in the right direction: mating-type switching of *Schizosaccharomyces pombe* is controlled by judicious expression of two different *swi2* transcripts. Genetics *190*, 977–987.

Zaratiegui, M., Vaughn, M.W., Irvine, D.V., Goto, D., Watt, S., Bähler, J., Arcangioli, B., and Martienssen, R.A. (2011). CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. Nature *469*, 112–115.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Figure and Table Legends

Figure 1: Lsd1, Lsd2, Sap1 and Abp1 are recruited to the mating type region
A) Distribution of normalized enrichments of the Abp1, Lsd1, Lsd2, and Sap1 ChIPs in a h^{90} strain (IP RPkM *reads per kilobase million*/ WCE RPkM). The sequence used for the alignment is a 44 kb region that contains the MT region with the *M* allele at *mat1*: h^{90} *M*. The 3 mat loci are indicated as well as the CEN-H region and the inverted repeats IR-L and IR-R. Light gray boxes show genes present in this region. B) Same than in A) except that the sequence used for the alignment is a 44 kb region that contains the MT region with the P allele at *mat1*: h^{90} *P*.

Figure 2: Abp1 is recruited at *swi2* as a function of the mating type
A) The upper panel indicates the probe and restriction enzyme used for the analysis of the replication intermediates. The lower panel represents the 2D gel analysis of the MPS1 site at *mat1* in h^{90} strain and in h^{90} *abp1*Δ strain (n=3). B) Distribution of normalized enrichments of the Abp1 ChIPs at the *swi2* locus (IP RPkM / WCE RPkM) in h^{90} (gray), *mat1M* stable (green) and *mat1P* stable (orange). Gray boxes indicate the genes in the region and arrows the transcription direction. Loci where Abp1 is enriched are annotated.

Figure 3: Abp1, Lsd1/2 and Sap1 are recruited at *mat1*
A) Schematic view of the strains used. The stable *mat1M* Δ*2-3* and *mat1P* Δ*2-3* strains as well as the wild type h^{90} strain are drawn. B) On the left: distribution of normalized enrichments of the Abp1, Lsd1, Lsd2, and Sap1 ChIPs on the h^{90} background (IP RPkM / WCE RPkM). The sequence used for the alignment is a 44 kb region that contains the MT region with the *M* allele at *mat1*: h^{90} *M*. Only *mat1M* is shown. On the right: same as at the left, with the h^{90} P sequence used for the alignment. Only *mat1P* is shown. C) Same as in B) with the *mat1M* Δ*2-3* background on the left and the *mat1P* Δ*2-3* background on the right. The sequence used for the alignments is a 10 kb region containing *mat1M* (for the left panel) or *mat1P* (for the right panel). B-C) the imprint site is indicated as well as the H2 (pink) and H1 (blue) homology boxes.

Figure 4: Lsd1/2 and Abp1 are recruited only at *mat1*
A) Schematic representation of the analysis used. Green boxes represent homologous sequences. Reads that can be unambiguously mapped are represented by black lines and multi-mapper reads are represented by green lines. Upper drawing represents the result of an alignment without filter (1). Lower drawing represents the alignment after that multi-mapper reads are removed using "samtools" algorithm (2). B) Distribution of raw coverage of the Abp1, Lsd1, Lsd2, and Sap1 ChIPs. The sequence used for the alignment is a 44 kb region that contains the MT region with the *M* allele at *mat1*: h^{90} *M*. The lines numbered 1 are the coverage without filter and the lines numbered 2 are the coverage obtained after removal of multi-mapper reads. The arrows indicate the unique-mapper reads at the junction of the *mat* loci. B) Same as in A with the h^{90} *P* sequence used for the alignment.

Figure S4: Lsd1/2 and Abp1 are recruited only at *mat1*
A) Distribution of raw coverage of the Abp1, Lsd1, Lsd2, and Sap1 ChIPs. The sequence used for the alignment is a 44 kb region that contains the MT region with the *M* allele at *mat1*: h^{90} *M*. The lines numbered 1 are the coverage without filter and the lines numbered 2 are the coverage obtained after removal of multi-mapper reads. The arrows indicate the unique-mapper reads at the junction of the *mat* loci. Red rectangles indicate the Sap1 enrichment at the inverted repeats IR-L and IR-R. B) Same as in A with the h^{90} *P* sequence used for the alignment.

Figure 5: Sap1 is recruited in *mat1M*
A) Schematic view of right region of the *mat1M* locus. The probes used for the shift assay as well as the sites of the deletion strains used for the ChIP-seq of Sap1 are indicated. The sequence of the probes that trigger a shift in B and C are indicated. B) Electrophoretic mobility shift analysis (EMSA) of crude extracts from *S. pombe*. *SAS1* and *Ter1* are used as controls. In lanes (-), 1 ng of radio labelled dsDNA is used with no proteins. In lanes (+), 5 µg of total proteins and 1 ng of radio labelled ds DNA are used. C) EMSA of Sap1-6xHis protein purified from *E. coli*. In lanes (-), 1 ng of radio labelled ds DNA is used with no proteins. In lanes (+), 40 ng of Sap1-6xHis protein and 1 ng of radio labelled ds DNA are used. D) Distribution of normalized enrichments of Sap1 ChIPs (IP RPkM / WCE RPkM) in the *mat1M* Δ*2-3, mat1M smt-0* Δ*2-3, mat1M* Δ*SS2* Δ*2-3* and *mat1M* Δ*SS13* Δ*2-3* backgrounds. The sequence used for the alignment is a 10 kb region that contains the *mat1M*, *mat1M smt-0, mat1M* Δ*SS2* or *mat1M* Δ*SS13*, respectively. Only *mat1M* is shown. The MT switching phenotype of the deletion strains is indicated.

Figure 6: The imprint site is detectable by high throughput sequencing on the Watson strand
A) Schematic view of the result of the sonication of both chromatid types. Red arrows indicate sites of sonication breakage. The black strand is the Watson strand (W) and the gray the Crick (C). 5' of Watson are shown in black and those of Crick in gray. Lower panel indicates the different types of reads obtained after sequencing. B) Schematic representation of the blunting of the end that is used during libraries preparation. The Crick and Watson 5' ends are preserved after treatment. The Crick strand can break at 3 positions compared to the imprint site: to the left, across or to the right. C) Distribution of the 5' count per position in the genome on both strands. *mat1M* Δ*2-3* stain was used for the analysis. p-value is calculated using a theoretical t-student distribution that fits on the distribution of the whole-genome observed 5' count. Regions that have a high 5' count are numerated.

Supplementary table 1: Region enriched in 5' count are indicated.

Figure 7: 5' count on the Watson strand is biased precisely at the site of the imprint
On the left: distribution of the 5' count of the Watson strand (black) and the Crick strand (gray) on the 10 kb region containing *mat1*. On the right: zoom on the imprint site. Red bars represent the 5' count enriched at the imprint sites. p-value is calculated using a theoretical t-student distribution that fits on the distribution of the whole-genome observed 5' count. A) The sequencing of *mat1M* Δ*2-3* strain is represented, B) *mat1P* Δ*2-3* strain, C) h^{90} strain and D) h^{90} *swi1*Δ strain. ABCD) The *mat1M* Δ*2-3*, *mat1P* Δ*2-3* and h^{90} *P* genomes are used, respectively.

Figure 8: Signature of sonication across the imprint
Statistical analysis of the 5' count in the *mat1* region. -log(p-value) of 5' count as a function of the window size and genome position is represented. The -log(p-value) scale is indicated on the right of the graphics. The p-value is calculated using theoretical t-student distributions that fit with the distribution of 5' count observe in the 10 kb region that contains *mat1*. A) A sliding window (+/- 1 Kb around the imprint site) with step size one and varying window size (1 to 40 nt) is applied when calculating the -log(p-value) of the 5' count of the Watson strand shown. The sequencing of a *mat1M* Δ*2-3* strain is used. B) Same as in A for the Crick strand. C) Window sizes and window positions compared to imprint site for the top 10 -log(p-value) are indicated. D, E and F) Same as in A, B and C, respectively for the sequencing of a *mat1P* Δ*2-3* strain.

Figure 9: Lsd1/2 and Sap1 immunoprecipitate chromatids on both sides of the imprint
Result of the 5' count analysis of the ChIP data presented in Figure 3C. A) Schematic representation of the theoretical 5' count distribution of Lsd1/2 and Sap1. In black, the 5' count of a non-imprinted chromatid and in red that of an imprinted chromatid. B) Distribution of the 5' count of the Lsd1/2 and Sap1 ChIPs for the Watson strand (black) and Crick strand (gray) in the *mat1M* Δ*2-3* background. The 5' count at the imprint site is indicated. The p-value is calculated using theoretical t-student distributions that fit with the distribution of 5' count observed in the 10 kb region that contains *mat1*. C) Analysis of the 5' count bias on the crick strand using a 20 nt window in a *mat1M* Δ*2-3* strains for the IPs of Lsd1-2 and Sap1. A sliding window of 20 nt with set size one is used in a 10 Kb region around *mat1* is applied. Gray line indicates the bias of the crick across the imprint. D) Same as in B) in a *mat1P* Δ*2-3* strain. E) Same as in C) in a *mat1P* Δ*2-3* strain.

Figure S9: Lsd1/2 and Sap1 are recruited at the same regions
A) Overlap of significant enriched peaks of Lsd1, Lsd2 and Sap1. B) Characterization of the 371 peaks enriched in Lsd1/2 and Sap1. Over the 371 peaks 78 are present on highly transcribed gene and 6 on poorly transcribed genes. Transcriptional data from (Marguerat et al., 2012) was used.

Figure 10: Model of recruitment of Lsd1/2 and Sap1
Schematic view of the model of binding of Lsd1/2 and Sap1 at *mat1*. Lsd1/2 complex is represented in blue with Phf1/2 in green. Sap1 is represented in pink. The imprint is represented by a red star. Lsd1/2 complex is recruited on the imprint site and Sap1 stabilizes the complex when the imprint is present.

Figures and Tables

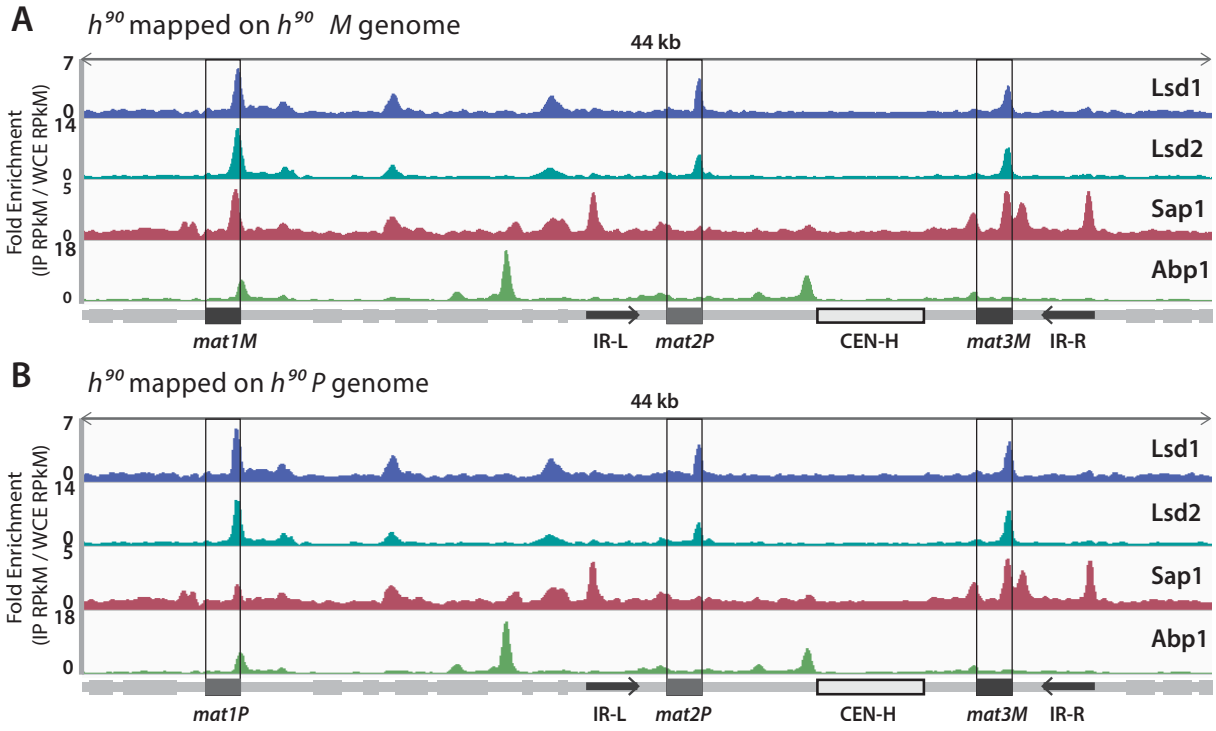


Figure 1: Lsd1, Lsd2, Sap1 and Abp1 are recruited to the mating type region

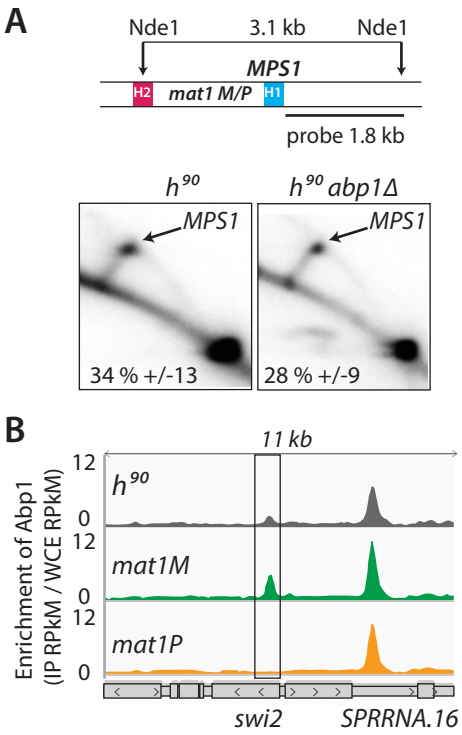


Figure 2: Abp1 is recruited at *swi2* as a function of the mating type

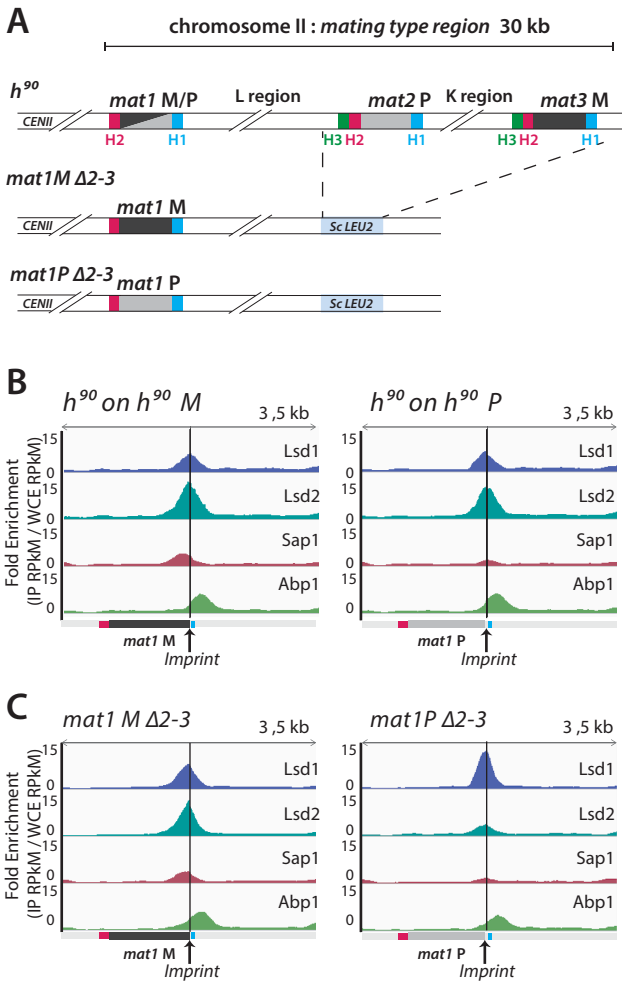


Figure 3: Abp1, Lsd1/2 and Sap1 are recruited at *mat1*

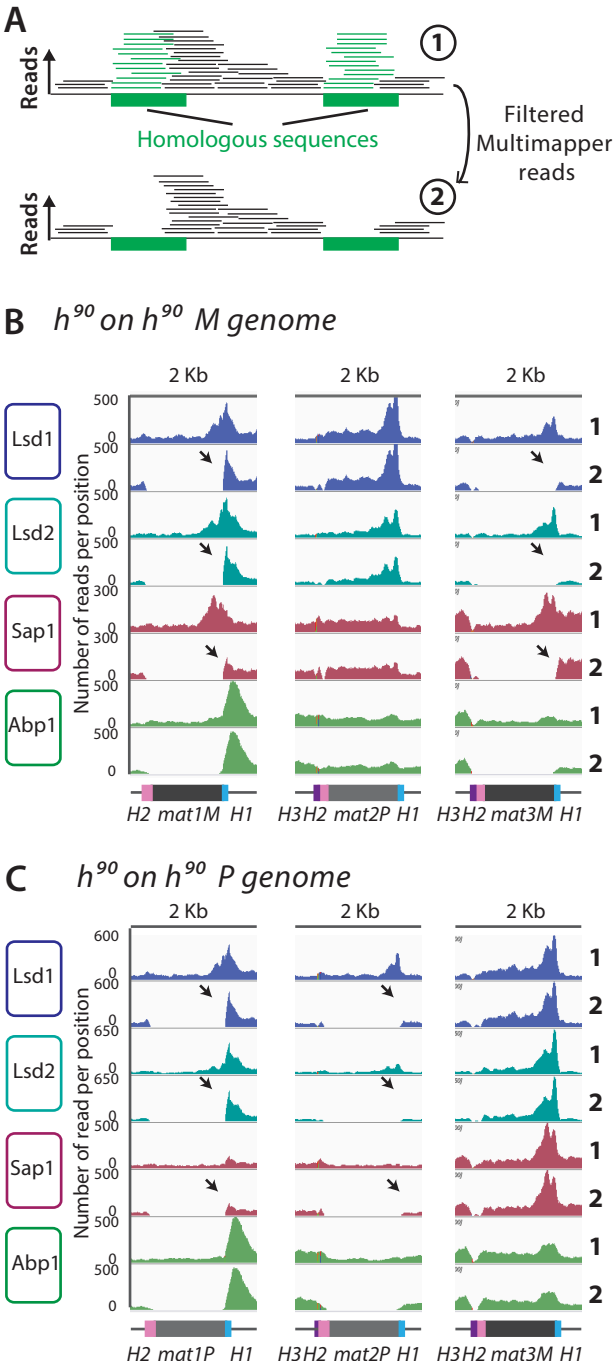


Figure 4: Lsd1/2 and Abp1 are recruited only at *mat1*

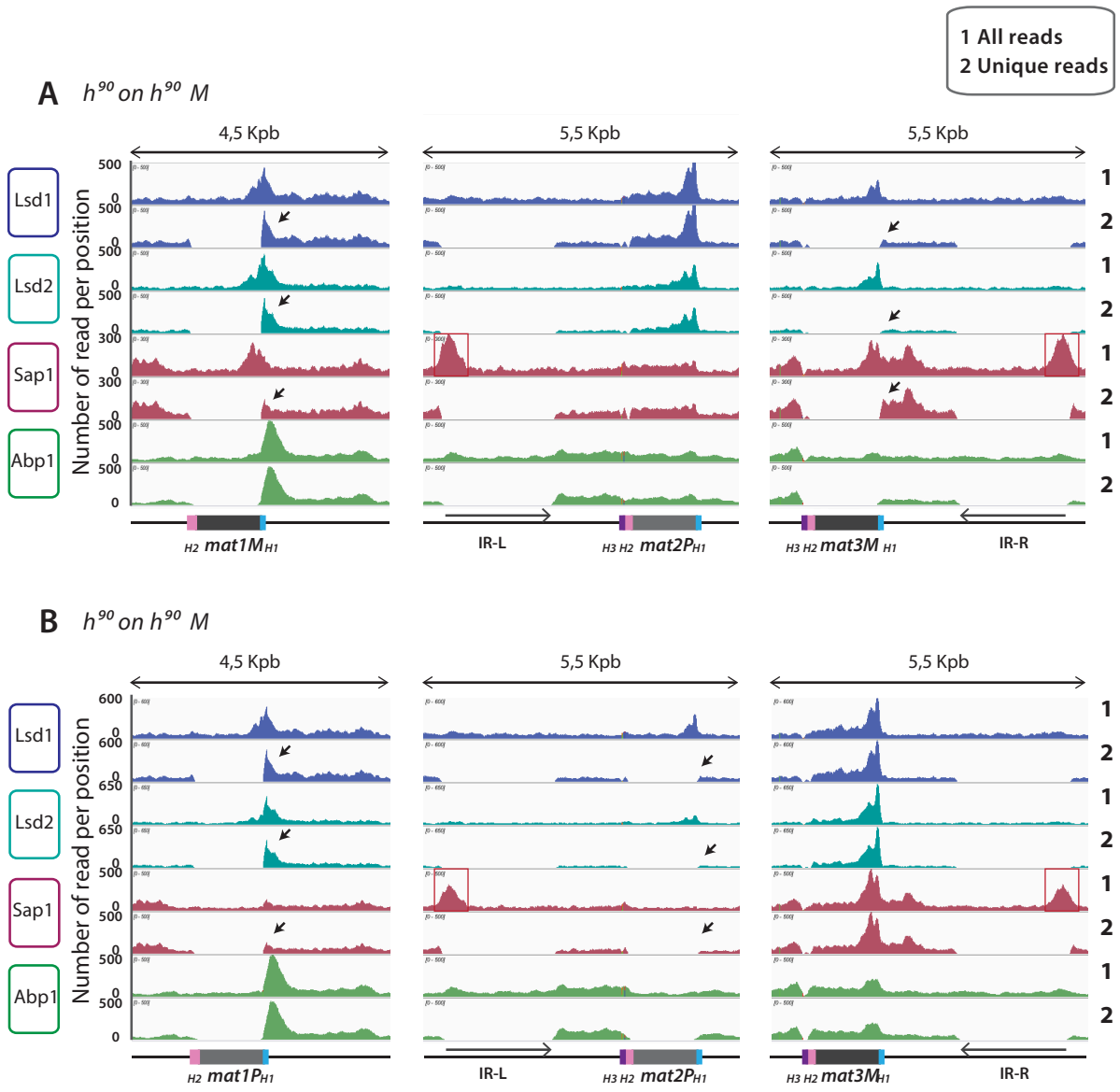


Figure S4: Lsd1/2 and Abp1 are recruited only at *mat1*

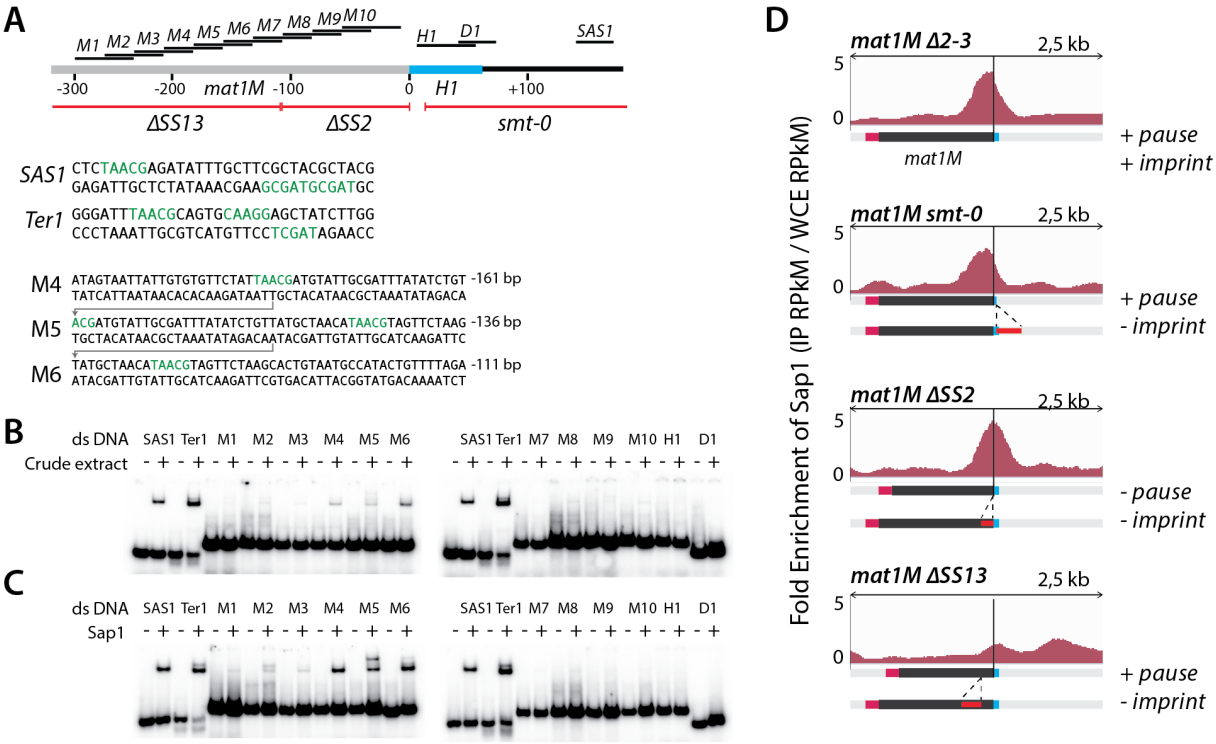


Figure 5: Sap1 is recruited in *mat1M*

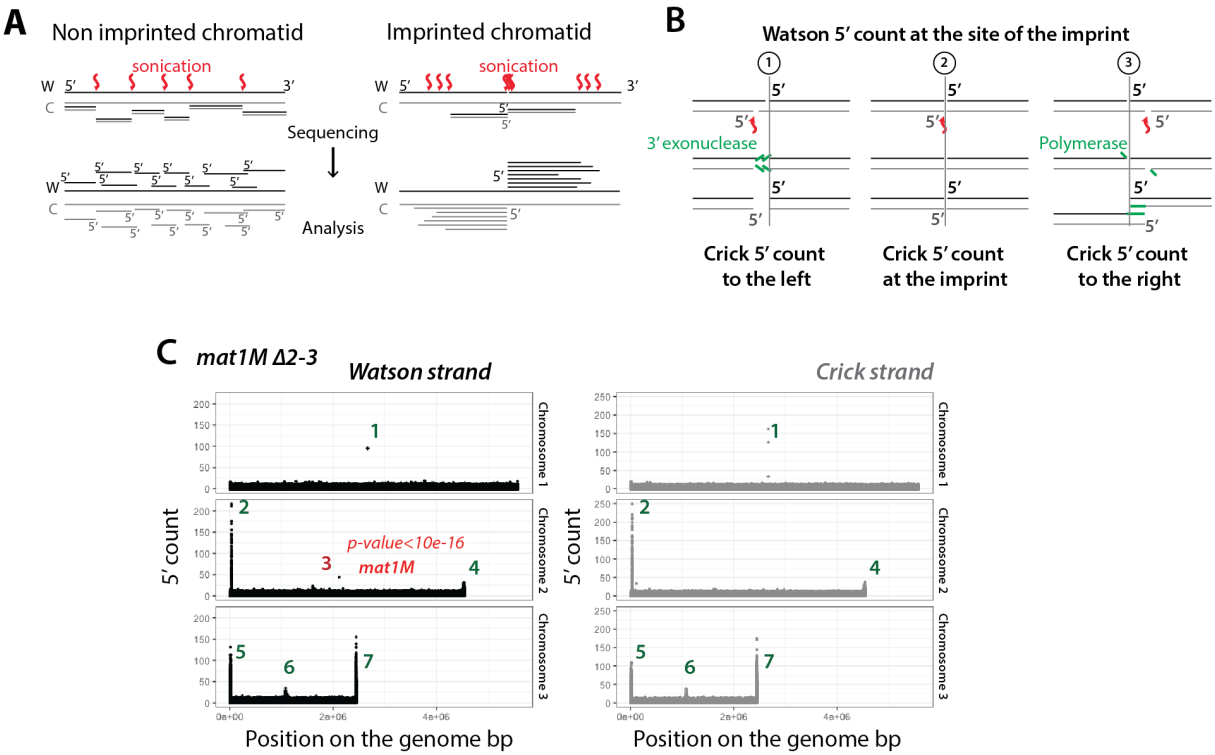


Figure 6: The imprint site is detectable by high throughput sequencing on the Watson strand

Supplementary Table 1: Region enriched in 5' count are indicated.

Peak number	Chr	First enriched 5'	Last enriched 5'	External Name	Logical Name
1	1	71982	72380	SPNUMT.8	nuclear_mt_pseudogne
2	1	3810366	3810458	SPNUMT.17	nuclear_mt_pseudogne
3	2	2115364		Imprint	Imprint
4	2	4532473	4540005	TELII	telomeric region
5	3	0	24683	SPRRNA	rDNA
6	3	1070904	1137003	CENIII	centromeric region
7	3	2439435	2452883	SPRRNA	rDNA

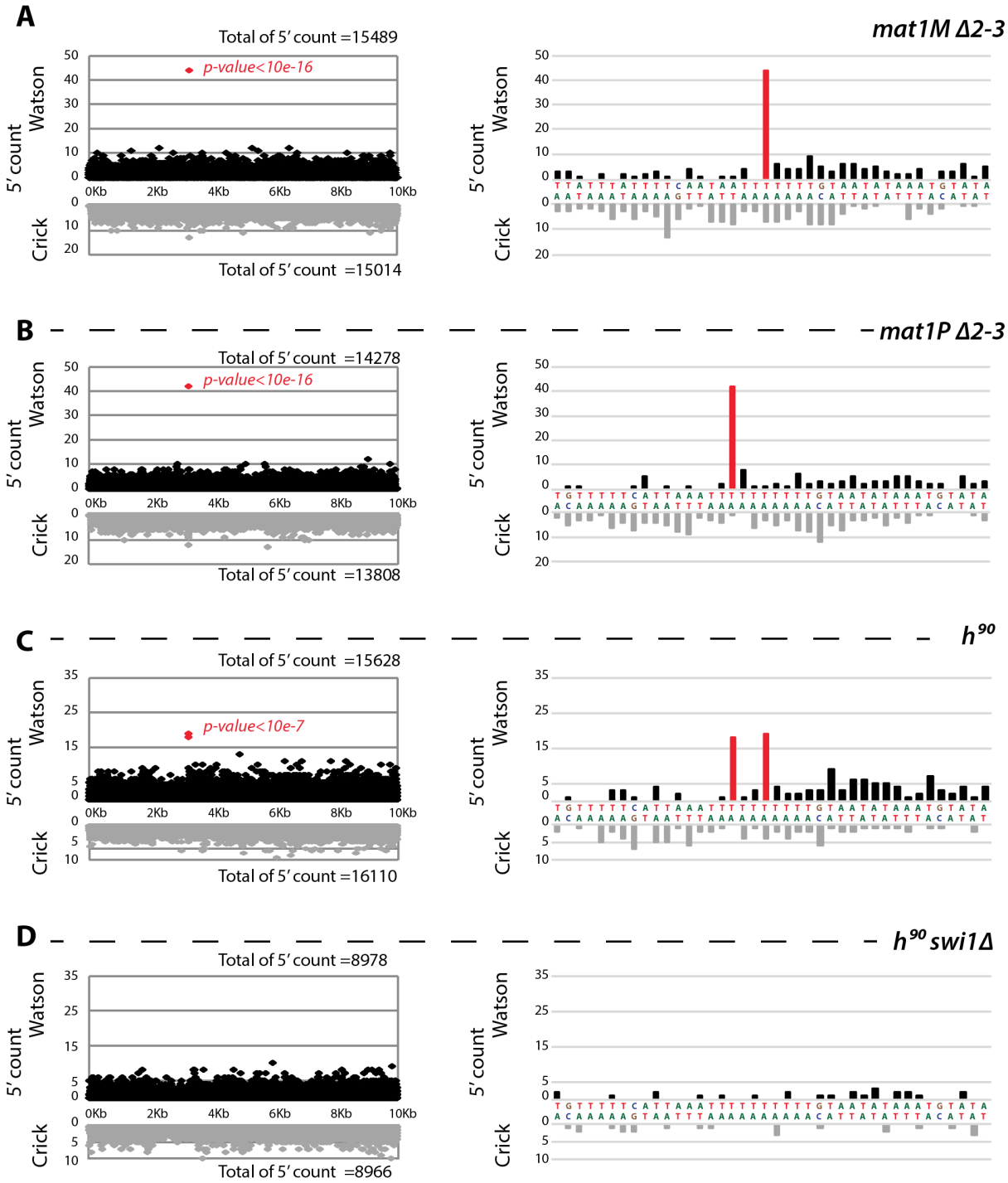


Figure 7: 5' count on the Watson strand is biased precisely at the site of the imprint

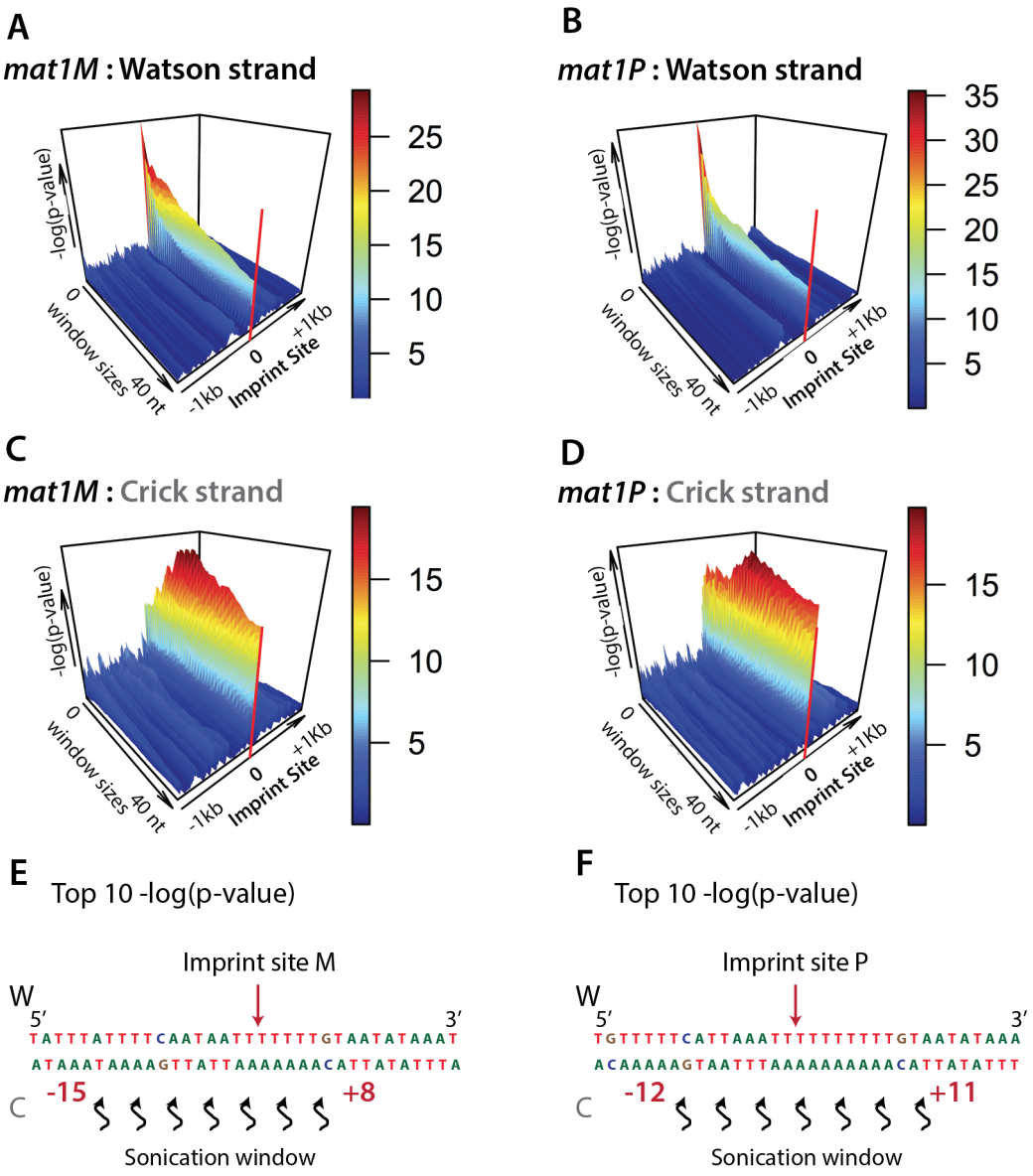


Figure 8: Signature of sonication across the imprint

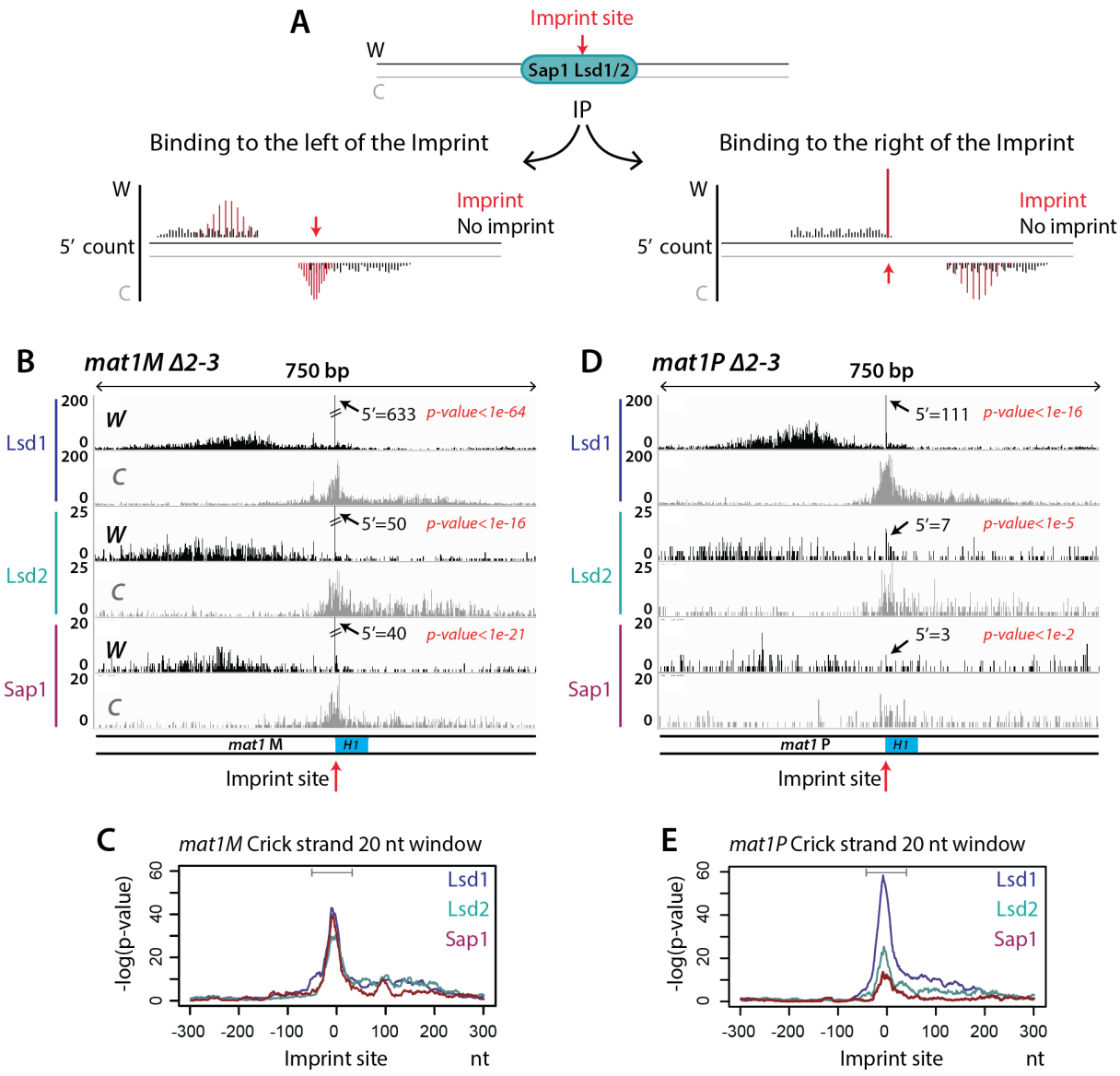


Figure 9: Lsd1/2 and Sap1 immunoprecipitate chromatids on both sides of the imprint

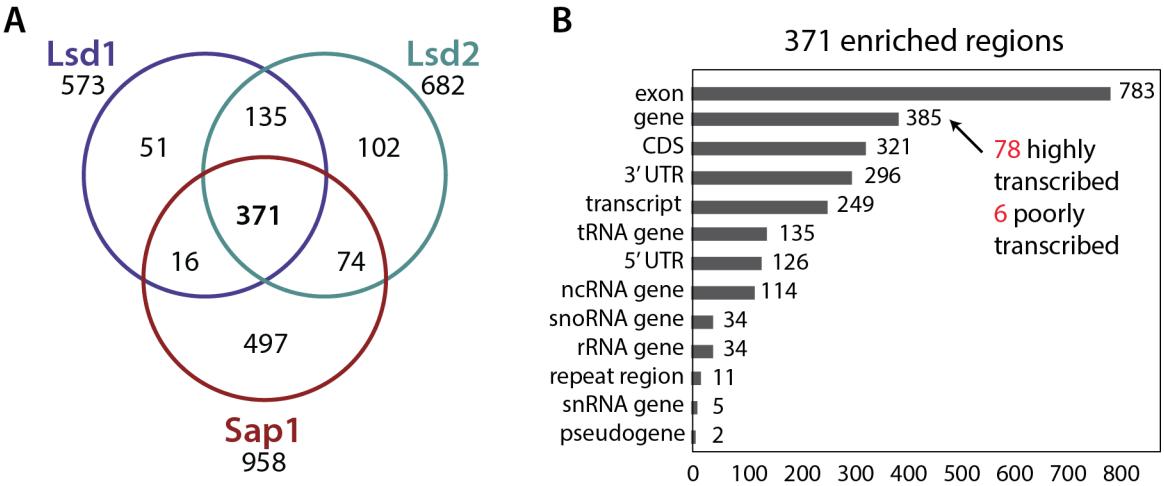


Figure S9: Lsd1/2 and Sap1 are recruited at the same regions

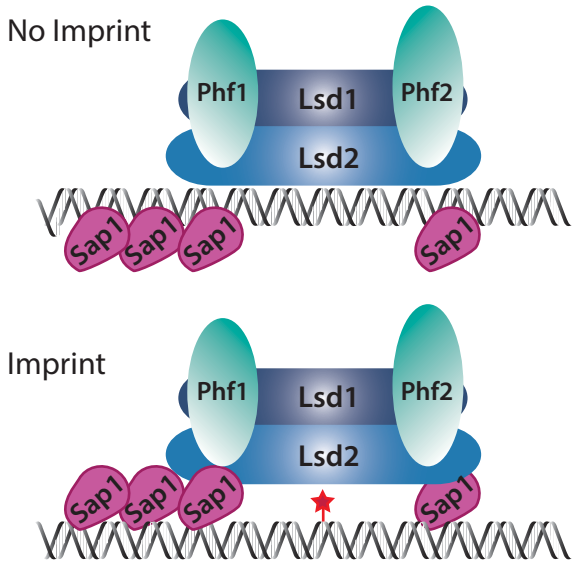


Figure 10: Model of recruitment of Lsd1/2 and Sap1

Supplementary Tables Materials and Methods

Table 1: List of strains

Name	Genotype	Experiment	Origin
PB 70	*mat1M Δ2.3::ScLEU2 ade6-M216 ura- leu1-32*	ChIP	
PB 47	*h90 ade6-M210 leu1-32 uraΔ*	ChIP/ 2D-gel	
PB 2279	*mat1P Δ2.3::ScLEU2 ade6-M216 ura-*	ChIP	
PB 483	*mat1M SS2Δ Δ2.3::ScLEU2 ade6-M210 ura-*	ChIP	
PB 484	*mat1M SS13Δ Δ2.3::ScLEU2 ade6-M210 ura-*	ChIP	
PB 241	*mat1M smt-0 Δ2.3::ScLEU2 ade6-M210 hist2- leu1-32*	ChIP	
PB 2348	*h90 lsd1-Myc::KanMX6 ade6-M210 leu1-32 ura-*	ChIP	Modify from Holmes et al 2012
PB 495	*mat1M Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M216 ura- leu1-32*	ChIP	
PB 2336	*mat1P Δ2.3::ScLEU2 lsd1-Myc::KanMX6 ade6-M216*	ChIP	
PB 431	*h90 lsd2-Myc::KanMX6 ade6-M210 leu1-32 ura-*	ChIP	
PB 489	*mat1M Δ2.3::ScLEU2 lsd2-Myc::KanMX6 ade6-M216*	ChIP	
PB 2350	*mat1P Δ2.3::ScLEU2 lsd2-Myc::KanMX6 ade6-M216 ura- leu1-32*	ChIP	
PB 1268	*h90 Abp1-CterTAP::KanMX6 ura4-D18 leu1-32*	ChIP	Modify from Zaratiegui et al 2011
PB 1242	*mat1M Δ2.3::ScLEU2 Abp1-CterTAP::KanMX6 leu1-32 ade6-M216*	ChIP	
PB 1257	*mat1P Δ2:3::ScLEU2 Abp1-CterTAP::KanMX6 ade6-M216 ura4-DS/E*	ChIP	
PB 1269	*h90 Abp1-CterTAP::KanMX6 swi1Δ::KAN ade6-M210 leu1-32*	ChIP	
PB 1258	*h90 abp1Δ::ScLEU2 ade6-M210 leu1-32*	2D-gel	

Table 2: List of sequences of primers used in this study

Name		Sequence		Experience
OL1436 M1+	**5'**	TATGCTTCTTAGAGTTACATTCACTGAAGATTATAATGTAATATTTTGTG	**3'**	gel-shift
OL1281 M1-	**5'**	CACAAAATATTACATTATAATCTTCAGTGAATGTAACTCTAAGAAGCATA	**3'**	
OL1437 M2+	**5'**	GAAGATTATAATGTAATATTTTGTGTACCCCATTTGCGTTGAGTTATTCT	**3'**	
OL1282 M2-	**5'**	AGAATAACTCAACGCAAATGGGGTACACAAAATATTACATTATAATCTTC	**3'**	
OL1438 M3+	**5'**	TACCCCATTTGCGTTGAGTTATTCTATAGTAATTATTGTGTGTTCTATTA	**3'**	
OL1283 M3-	**5'**	TAATAGAACACACAATAATTACTATAGAATAACTCAACGCAAATGGGGTA	**3'**	
OL1439 M4+	**5'**	ATAGTAATTATTGTGTGTTCTATTAACGATGTATTGCGATTTATATCTGT	**3'**	
OL1284 M4-	**5'**	ACAGATATAAATCGCAATACATCGTTAATAGAACACACAATAATTACTAT	**3'**	
OL1440 M5+	**5'**	ACGATGTATTGCGATTTATATCTGTTATGCTAACATAACGTAGTTCTAAG	**3'**	
OL1285 M5-	**5'**	CTTAGAACTACGTTATGTTAGCATAACAGATATAAATCGCAATACATCGT	**3'**	
OL1441 M6+	**5'**	TATGCTAACATAACGTAGTTCTAAGCACTGTAATGCCATACTGTTTTAGA	**3'**	
OL1286 M6-	**5'**	TCTAAAACAGTATGGCATTACAGTGCTTAGAACTACGTTATGTTAGCATA	**3'**	
OL1442 M7+	**5'**	CACTGTAATGCCATACTGTTTTAGAGGGTGATGCTTCCTAAAATCTCCTT	**3'**	
OL1287 M7-	**5'**	AAGGAGATTTTAGGAAGCATCACCCTCTAAAACAGTATGGCATTACAGTG	**3'**	
OL1443 M8+	**5'**	GGGTGATGCTTCCTAAAATCTCCTTACATAAAGTAATACATGGATTTTAC	**3'**	
OL1288 M8-	**5'**	GTAAAATCCATGTATTACTTTATGTAAGGAGATTTTAGGAAGCATCACCC	**3'**	
OL1444 M9+	**5'**	ACATAAAGTAATACATGGATTTTACTGCCCTGATTCTATCGAAATATGCT	**3'**	
OL1289 M9-	**5'**	AGCATATTTCGATAGAATCAGGGCAGTAAAATCCATGTATTACTTTATGT	**3'**	
OL1445 M10+	**5'**	TGCCCTGATTCTATCGAAATATGCTGTTTTTTTTATTCGTTTTTATTTAT	**3'**	
OL1290 M10-	**5'**	ATAAATAAAAACGAATAAAAAAAAACAGCATATTTCGATAGAATCAGGGCA	**3'**	
OL1446 H1+	**5'**	TTTGTAATATAAATGTATAGTCTTTCTCCTTTGTTTTCTCTCGTTCGTTT	**3'**	
OL1291 H1-	**5'**	AAACGAACGAGAGAAAACAAAGGAGAAAGACTATACATTTATATTACAAA	**3'**	
OL1447 SAS1+	**5'**	CTCTAACGAGATATTTGCTTCGCTACGCTACG	**3'**	
OL1448 SAS1-	**5'**	CGTAGCGTAGCGAAGCAAATATCTCGTTAGAG	**3'**	
OL1449 D1+	**5'**	TTCTCTCGTTCGTTTCCATGTTTCCAATTATG	**3'**	
OL1292 D1-	**5'**	CATAATTGGAAACATGGAAACGAACGAGAGAA	**3'**	
OL1492 Ter1+	**5'**	GGGATTTAACGCAGTGCAAGGAGCTATCTTGG	**3'**	
OL1493 Ter1 -	**5'**	CCAAGATAGCTCCTTGCACTGCGTTAAATCCC	**3'**	
probe mat1	**5'**	TTCGGTATTTAAGTCTGGCG	**3'**	2D gel
probe mat1	**5'**	CCAATTATGCTGTTCGTGTC	**3'**	

Table 3: Libraries statistics

Name	Library	Run Type	Total Reads	Aligned Reads	% Aligned Reads
PB 2348	h90lsd1IP	SR65	19681930	18482514	93.91%
PB 2348	h90lsd1WCE	SR65	21067533	19342359	91.81%
PB 431	h90lsd2IP	SR65	8988245	8152554	90.70%
PB 431	h90lsd2WCE	SR65	47617742	44206781	92.84%
PB 47	h90sap1IP	SR65	12951987	12459329	96.20%
PB 47	h90sap1WCE	SR65	62223348	55091901	88.54%
PB 1268	h90abp1IP	SR130	6462438	5690753	88.06%
PB 1268	h90abp1WCE	SR130	46506268	40369759	86.80%
PB 495	mat1Mlsd1IP	SR65	78402964	70328896	89.70%
PB 495	mat1Mlsd1WCE	SR65	43201409	39627677	91.73%
PB 489	mat1Mlsd2IP	SR65	10184098	9714898	95.39%
PB 489	mat1Mlsd2WCE	SR65	9408560	8715332	92.63%
PB 70	mat1Msap1IP	SR65	12714852	12095278	95.13%
PB 70	mat1Msap1WCE	SR65	71445515	60200476	84.26%
PB 1242	mat1Mabp1IP	SR130	11015722	9764199	88.64%
PB 1242	mat1Mabp1WCE	SR130	42626692	37576928	88.15%
PB 2336	mat1Plsd1IP	SR65	36559765	33949628	92.86%
PB 2336	mat1Plsd1WCE	SR65	25466207	23759984	93.30%
PB 2350	mat1Plsd2IP	SR65	9443782	9170279	97.10%
PB 2350	mat1Plsd2WCE	SR65	30848920	28570601	92.61%
PB 2279	mat1Psap1IP	SR65	9003795	8611021	95.64%
PB 2279	mat1Psap1WCE	SR65	15378877	13342280	86.76%
PB 1257	mat1Pabp1IP	SR130	36287760	32457178	89.44%
PB 1257	mat1Pabp1WCE	SR130	78169993	35976808	46.02%
PB 483	mat1MSS2sap1IP	SR65	6003899	5124804	85.36%
PB 483	mat1MSS2sap1WCE	SR65	52438070	46350961	88.39%
PB 484	mat1MSS13sap1IP	SR65	9870989	9139481	92.59%
PB 484	mat1MSS13sap1WCE	SR65	46317419	40703242	87.88%
PB 241	mat1Msmtsap1IP	SR65	5074834	4777575	94.14%
PB 241	mat1Msmtsap1WCE	SR65	35281724	31718229	89.90%
PB 1269	h90abp1swi1IP	SR130	6868375	5599274	81.52%
PB 1269	h90abp1swi1WCE	SR130	28327346	22514091	79.48%

Supplementary Materials and Methods

Processing of sequencing data

Reads obtained using the illumina technology (see MM) were aligned using the bowtie2 algorithm (Langmead et al., 2009) with the genome of *S. pombe* (ASM294v2.23) and with several contigs depending on the genotype of the strain. The strains h^{90} were aligned with

contigs that contain 45 kb of the wild type MT region with the *M* allele at *mat1* (file named h90M-45.fa) or with the *P* allele at *mat1* (file named h90P-45.fa). Strains *mat1M* Δ*2-3* were aligned with a contig that contains the 10 kb region with *mat1M* (file named mat1M.fa) and strains *mat1P* Δ*2-3* were aligned with a contig that contains the 10 kb region with *mat1P* (file named mat1P.fa). Strains containing deletions at *mat1M* were first aligned with mat1M.fa contigs and next with contigs deleted for the appropriated sequence (file named mat1MSS2.fa, mat1MSS13.fa and mat1Msmt0.fa).
Next, the normalized coverage was calculated using bamtools package version 2.2.3 (Barnett et al., 2011) (parameters: --normalizeUsingRPKM --binSize 1). The bigWig file was converted to Wig files with bedtools v.2.17.0 and BigWig_tools.v4 (Pohl and Beato, 2014; Quinlan and Hall, 2010). IGV genome browser was used to normalize IP coverage with WCE coverage, using combined data track we divide the IP normalized coverage by the WCE normalized coverage.

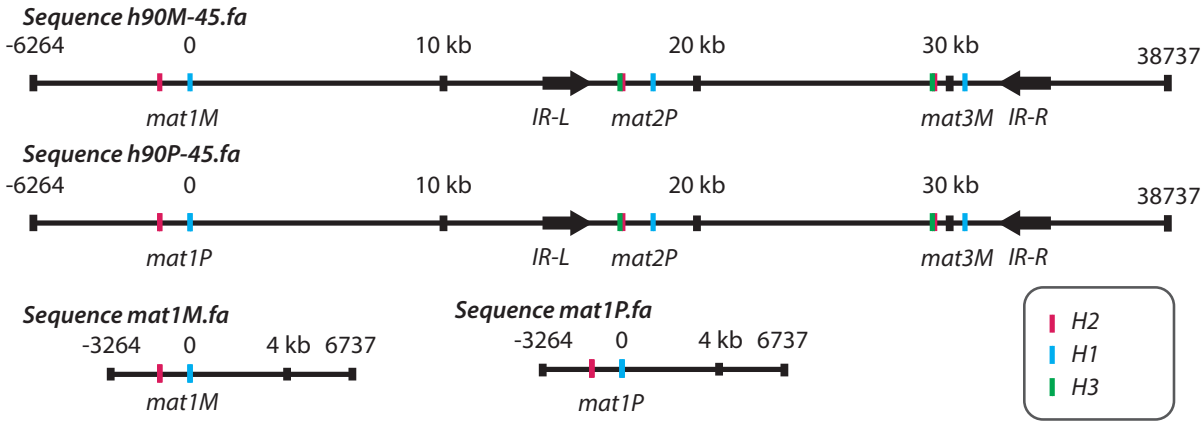


Figure 1: Contigs used for the alignment
Contigs used for the different alignments are represented. The sequences are available one the server with the accession number XXXX.

Statistical validation of ChIP peaks

After alignment, peaks were called using MACS v2.1.0 (Zhang et al., 2008). Statistic validation of peaks was performed using the IDR method with ENCODE recommendations (Landt et al., 2012). ChIP in a strain h^{90}, *mat1M* Δ*2-3* and *mat1P* Δ*2-3* were considered biological replicates for the whole genome since only 10 kb of the silenced donor regions differ. Peaks were called on individual duplicates (h^{90} and *mat1M* Δ*2-3* sequencing were used) and were called on pseudo-replicates. Pseudo-replicates consist of 2 files that result from merging of biological replicates and their random separation. Irreproducibility Discovery Rate (IDR) was calculated as described by the ENCODE Consortium (Landt et al., 2012). The lists of peaks of the biological replicates were challenged with the lists of the pseudo-replicates. This comparison permits to obtain a low IDR threshold that permits to extract the significant peaks (Figure 2 and Table 1).

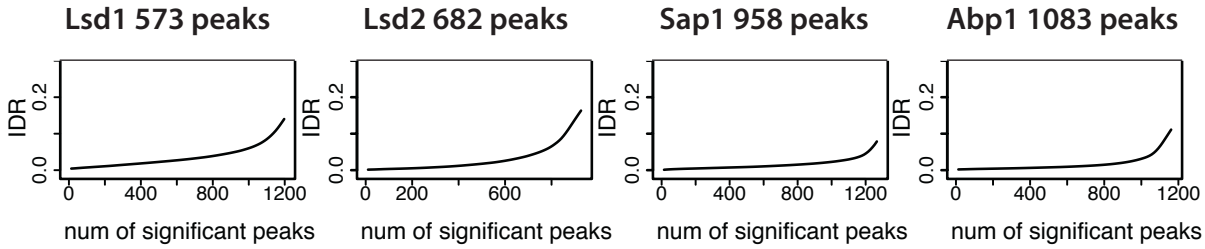


Figure 2: Result of the IDR analysis

The IDR is presented as a function of the number of significant peaks for Lsd1, Lsd2, Sap1 and Abp1.

Table 1: Thresholds obtained after IDR analysis are presented for each immunoprecipitated protein.

	Number of peaks IDR	Threshold = -log10(p-value)
Lsd1	573	10,22646
Lsd2	682	19,19591
Sap1	958	26,07609
Abp1	1083	25,11338

5' count analysis and statistical validation

5’ nucleotides were extracted from alignments using a custom script (available upon request). Statistical analysis was performed on a coverage file non-normalized of the 5' count in R. The p-value was calculated using a negative binomial distribution in R (fdist lowtail = FALSE (fitdistrplus)). The theoretical t-student distributions were calculated using the whole genome data (Figure 5 and 6) or on the 10 Kb containing *mat1* data (Figure 7 and 8).

Data and R scripts are available upon request.

Conclusions et Discussion

La division asymétrique chez la levure *S. pombe* initie le changement de type sexuel d'une des deux cellules sœurs. La dissection de ce processus a permis de mettre en évidence plusieurs acteurs protéiques impliqués dans des mécanismes divers comme la réplication, la réparation ou la formation de l'hétérochromatine. L'empreinte présente à *mat1* est formée asymétriquement sur l'une des deux chromatides sœur lors de la réplication. Le génome de la cellule répliquée contient donc deux chromatides différentes à l'instar de ce qui est observé au niveau des gènes soumis à empreinte parentale chez les métazoaires. Par exemple, chez la souris, l'ADN de loci particuliers est asymétriquement méthylé. Ainsi, un seul des deux génomes parentaux est marqué par cette méthylation qui réprime son expression. Bien que la nature des empreintes soit différente entre les métazoaires et la levure *S. pombe*, elles sont maintenues au cours des divisions cellulaires et participent, chez les eucaryotes pluricellulaires, au développement de l'organisme. Ainsi, l'utilisation du modèle de division asymétrique de *S. pombe* permet d'étudier les processus fondamentaux contrôlant l'établissement et le devenir d'une empreinte chromosomique.

Un repliement chromosomique autour du site de l'empreinte ?

Dans cette étude, nous avons mis en évidence que les protéines présentes du côté centromère proximal de l'empreinte (Lsd1, Lsd2 et Sap1) sont capables d'immunoprécipiter le fragment contenant l'empreinte du côté centromère distal. Ces observations suggèrent qu'une structure particulière est formée autour de l'empreinte qui permet le rapprochement de l'ADN proximal et distal (figure 1). Il est probable que cette structure participe à la stabilisation de l'empreinte au cours du cycle cellulaire. Plus particulièrement, le complexe Lsd1-Lsd2 qui se lie à l'intérieur de *mat1* est capable d'immunoprécipiter le fragment distal de *mat1* portant l'empreinte. Par ailleurs, ce complexe est nécessaire à l'établissement de la pause de la fourche de réplication, l'étape précoce dans la formation de l'empreinte (Holmes et al., 2012). En étudiant le recrutement de ce complexe en détail, j'ai pu mettre en évidence qu'il est présent dans *mat1M* et *mat1P*. Plus particulièrement, Lsd1 se lie au niveau des séquences minimales décrites comme permettant la pause de la fourche de réplication à *MPS1* (Sayrac et al., 2011). Cependant, Lsd1 est encore capable de se lier à *mat1M* dans une souche ne contenant plus la séquence minimale établissant la pause (*SS2*) ainsi que dans des souches contenant la pause et non l'empreinte (*SS13* et *smt-0*). Cela suggère que Lsd1 et Lsd2 ne sont ni recrutées par une séquence d'ADN spécifique, ni par la pause, ni par l'empreinte. Une des hypothèses est que ce complexe est recruté par un élément chromatinien au niveau de *mat1*, tel qu'une marque d'histone par exemple. Cette hypothèse est soutenue par l'absence de Lsd1 et Lsd2 à l'intérieur des allèles

donneurs qui sont dans une région hétérochromatique, à contrario de *mat1*. D'autre part, au niveau des donneurs, l'absence de Lsd1 et Lsd2 corrèle avec l'absence de pause de la fourche de réplication (Sayrac et al., 2011; Vengrova and Dalgaard, 2004). Il serait intéressant d'analyser le recrutement de Lsd1 et de Lsd2 et le comportement de la fourche de réplication au niveau des donneurs quand la nature de l'hétérochromatine est perturbée, en utilisant par exemple des mutants *swi6*Δ ou *clr4*Δ.

Bien que le recrutement de Swi1 (membre du complexe de la protection des fourche) est diminué dans un mutant affectant le domaine HMG de Lsd1 (*lsd1-18*) (Holmes et al., 2012), en absence de pause, Lsd1 est toujours présent à *mat1* (*SS2*). Dans leur ensemble, ces résultats suggèrent que le complexe Lsd1 et Lsd2 est nécessaire et non suffisant pour que la pause de la fourche soit réalisée et soit utilisée efficacement pour la formation de l'empreinte. Nous ne pouvons pas déterminer si Lsd1 et Lsd2 sont requis pour la formation de la pause ou pour son maintien. En effet, il est impossible de distinguer en absence de Swi1 ou dans la délétion *SS2* si la pause est formée et non stabilisée, donc non détectable, ou si elle n'est plus formée.

Plus particulièrement, des éléments suggèrent que Lsd1 contrôle l'établissement de la pause en absence d'empreinte. En effet, il est surprenant que la pause à *MPS1* dans une souche mutante *lsd1-18* varie en fonction de la structure de *mat1*. Par exemple, dans un contexte *h90 lsd1–18*, la pause est réduite d'un tiers alors que dans une souche *mat1M smt-0 lsd1-18* elle n'est plus détectable (Holmes et al., 2012). De plus, Lsd1 est toujours présente dans la souche *mat1M smt-0*. Ce résultat indique qu'en absence d'empreinte (*smt-0*) ou de la séquence *smt-0*, l'efficacité de la pause est totalement dépendante de la protéine Lsd1. Il est concevable que la structure de la pause de la fourche à *MPS1* soit influencée par l'empreinte et que la fonction du complexe Lsd1-Lsd2 soit différente avec ou sans empreinte. Il serait intéressant d'analyser le niveau de pause dans la souche *mat1M* Δ*SS13 lsd1-18*, puisqu'il permettrait de confirmer ce qui est observé dans une souche *mat1M smt-0 lsd1-18*, à savoir qu'en absence d'empreinte la pause est totalement dépendante de Lsd1. L'analyse du niveau de pause en utilisant les mutants catalytiques des protéines Lsd1 et Lsd2 (*lsd1-ao lsd2-ao*) dans les fonds génétiques *mat1M* Δ*SS13* et *mat1M smt-0* permettrait de déterminer si l'activité amine oxydase est impliquée dans cette pause sur une chromatide vierge.

Nous savons que deux réplications consécutives sont nécessaires pour qu'une cellule change de type sexuel. Cette observation reflète une des difficultés que présente l'étude de la fourche de réplication à *mat1* à savoir que la fourche parcours à la première réplication une chromatide sans empreinte et à la seconde une chromatide avec empreinte. Le type de chromatide utilisée comme matrice pour la réplication détermine le devenir de l'allèle à *mat1*. Soit l'empreinte sera

formée soit l'allèle présent sera remplacé. L'utilisation de souche ne présentant que le premier type de pause (comme probablement *mat1M smt*-0) nous permettra peut-être d'isoler des acteurs important dans l'initiation du mécanisme et donc pour la formation de l'empreinte.

La question sous-jacente concerne l'élément qui permet le recrutement du complexe Lsd1-Lsd2. Il est possible qu'une marque d'histone permette le recrutement de ce complexe. Cette marque est inconnue à ce jour, et serait, via le recrutement de Lsd1 et Lsd2, l'élément précurseur de la pause. De plus, le positionnement précis de ce complexe par rapport au site de l'empreinte semble être requis pour son activité sur la pause de la fourche. En effet, dans la souche *mat1M* Δ*SS2*, qui ne présente pas de pause, le site de recrutement de la protéine Lsd1 est décalé par rapport à une souche sauvage. Le positionnement particulier de Lsd1 serait à l'origine de la pause et du maintien de l'empreinte à ce locus. Par analogie, il a été montré par exemple que le positionnement des nucléosomes le long du génome participe à la spécification des origines de réplication ou encore à la formation de cassure double brin par la protéine Rec12 (SPO11) en méiose (de Castro et al., 2012). Dans d'autres organismes, l'homologue de Lsd1 a également été montré comme étant important pour le maintien des empreintes parentales au travers du contrôle de la méthylation de H3K4 ou indirectement via le contrôle de la méthylation de l'ADN. Chez *Caenorhabditis elegans* (*C. elegans*) et chez *Drosophila melanogaster* (Drosophile), une mutation affectant les homologues de Lsd1 (SRP-5 et (var)3-3, respectivement) entraine une stérilité due à un défaut dans la formation des gamètes (ovogenèse et gamétogénèse). Cela corrèle avec le fait que dans les précurseurs des cellules germinales, H3K4me2 persiste au niveau de gènes exprimés pour la spermatogénèse. Cette méthylation est donc transmise de génération en génération et inhibe la formation correcte des gamètes (Katz et al., 2009; Rudolph et al., 2007; Szabad et al., 1988). Par ailleurs, chez la souris femelle déficiente pour LSD1 (KDM1), les ovocytes ont un taux de méthylation de H3K4 plus élevé et ils échouent à établir la méthylation de l'ADN au niveau de gènes soumis à empreinte maternelle. La conséquence est une activation ou une répression bi-allélique de ces loci qui entraine la mort de ces embryons avant le stade "mid-gestation" (Ciccone et al., 2009). D'autre part, durant la différentiation hématopoïétique, LSD1 réprime le programme transcriptionnel des progéniteurs et des cellules souches hématopoïétiques. Si ce programme n'est pas réprimé correctement, la différenciation des cellules souches hématopoïétiques ainsi que la formation d'une lignée mature de cellules sanguines sont compromises (Kerenyi et al., 2013). Finalement, chez *C. elegans*, LSD1/SRP-5 a été décrit comme participant à la réponse aux lésions doubles brins de l'ADN (Nottke et al., 2011). De la même façon, dans des lignées cellulaires humaines, LSD1 participe à la réparation des lésions d'ADN, et plus particulièrement en fin de phase S

(Mallette et al., 2012; Mosammaparast et al., 2013). Par ailleurs, la perte de LSD1 augmente la recombinaison homologue (Mosammaparast et al., 2013).
Au niveau du CTS chez *S. pombe*, il a été montré que les activités catalytiques de Lsd1 et de Lsd2 sont requises pour la formation de la pause et de l'empreinte (Holmes et al., 2012). Nous pouvons imaginer qu'au travers de la modification des histones H3 autour de l'empreinte, Lsd1 et Lsd2 permettent de stabiliser l'empreinte. De façon intéressante, la protéine Sap1, également recrutée à *mat1*, est copurifiée avec le complexe Lsd1 Lsd2 (Lan et al., 2007).
Le rôle de Sap1 à *mat1* est encore mal compris. Sap1 est la seule protéine décrite comme liant préférentiellement l'allèle *M* plutôt que *P*. Par ailleurs, sa liaison à la cassette *M* ne semble pas être dictée par la nature de la chromatine. En effet, les analyses suggèrent que Sap1 est recrutée dans *mat1M* et dans *mat3M*.

Au niveau de *mat1*, Sap1 est présente du côté centromère proximal de Lsd1-Lsd2 dans une séquence importante pour la stabilisation de l'empreinte (*SS13*) (Sayrac et al., 2011). De façon intéressante, quand la séquence recrutant Sap1 est substituée par une autre, l'empreinte est rétablie (Sayrac et al., 2011). A ce jour, nous ne savons pas si Sap1 est de nouveau recrutée dans la substitution. Cet élément de séquence (*SS13*) permettrait un espacement correct entre le site de l'empreinte et la fin du transcrit *Mc* (Sayrac et al., 2011). Il contribuerait à la formation d'une organisation particulière de la chromatine, qui serait nécessaire pour la stabilisation de l'empreinte. Sachant que Sap1 a été décrite comme participant au positionnement des nucléosomes (Tsankov et al., 2011), il est probable qu'elle participe également au positionnement d'un nucléosome à *mat1*. Ce nucléosome serait important pour le recrutement de Lsd1 et Lsd2 et la stabilisation de l'empreinte. La question qui reste en suspens est de déterminer quel(s) acteur(s) pourrai(en)t jouer ce rôle dans une cellule *mat1P*. Aucune protéine recrutée spécifiquement au niveau de *mat1P* n'a été mise en évidence pour le moment.
Du côté centromère distal de l'empreinte, j'ai pu mettre en évidence le recrutement de la protéine Abp1. La protéine CENP-B des mammifères possède trois homologues chez *S. pombe* : Abp1, Cbh1 et Cbh2. Abp1 semble jouer un rôle d'optimisation dans le CTS à *mat1*. En effet, les différentes expériences réalisées ont permis de conclure que le site de recrutement d'Abp1 à *mat1* n'est pas nécessaire à l'activité du CTS. Il est possible que la protéine Abp1 joue un rôle dans la formation du repliement chromatinien à *mat1* en permettant la stabilisation du nucléosome présent du côté centromère distal de l'empreinte. Par analogie, au niveau des centromères, il a été suggéré que CENP-B participe à la stabilisation et au positionnement du variant d'histone CENP-A (Fachinetti et al., 2013; Fachinetti et al., 2015).

Enfin, l'hélicase Chd1 (ou Hrp1, orthologues de CHD1) a aussi été copurifiée avec le complexe Lsd1, Lsd2, Phf1 et Phf2 (Lan et al., 2007). Cette protéine participe à la formation de l'hétérochromatine et au positionnement des nucléosomes (Hennig et al., 2012; Pointner et al., 2012). Il serait intéressant de tester le recrutement de Hrp1 au niveau du locus sexuel et d'autre part, son effet sur la structure formée à *mat1*.

Ainsi, le modèle proposé en figure 1 permet d'expliquer le fait que les protéine Sap1, Lsd1 et Lsd2 présentes préférentiellement du côté centromère proximal de *mat1* sont capables de lier le fragment centromère distal portant l'empreinte. Dans une cellule portant l'allèle *M* à *mat1*, la protéine Sap1 est cartographiée entre 211 et 111 pb du côté centromère proximal de *mat1*. Elle permettrait la stabilisation du nucléosome lié par Lsd1-Lsd2. De l'autre côté, à 180 pb du site de l'empreinte, la liaison d'Abp1 permettrait la stabilisation du nucléosome centromère distal. Une marque chromatinienne permettrait le recrutement du complexe Lsd1-Lsd2 sur le nucléosome centromère proximal, stabilisé par Sap1. Lsd1 et Lsd2 au travers de leur activité catalytique pourraient modifier les histones autour de *mat1*. Cette structure permettrait à la fourche de réplication de faire une pause à *MPS1* et de former l'empreinte. L'empreinte serait ainsi stabilisée durant la division cellulaire permettant le CTS à la réplication suivante. Dans une cellule portant l'allèle *P* à *mat1*, Sap1 n'est pas présente, mais il est possible qu'un autre acteur soit présent (gris) permettant également de stabiliser le recrutement de Lsd1 et Lsd2 entrainant la formation de la pause et de l'empreinte.

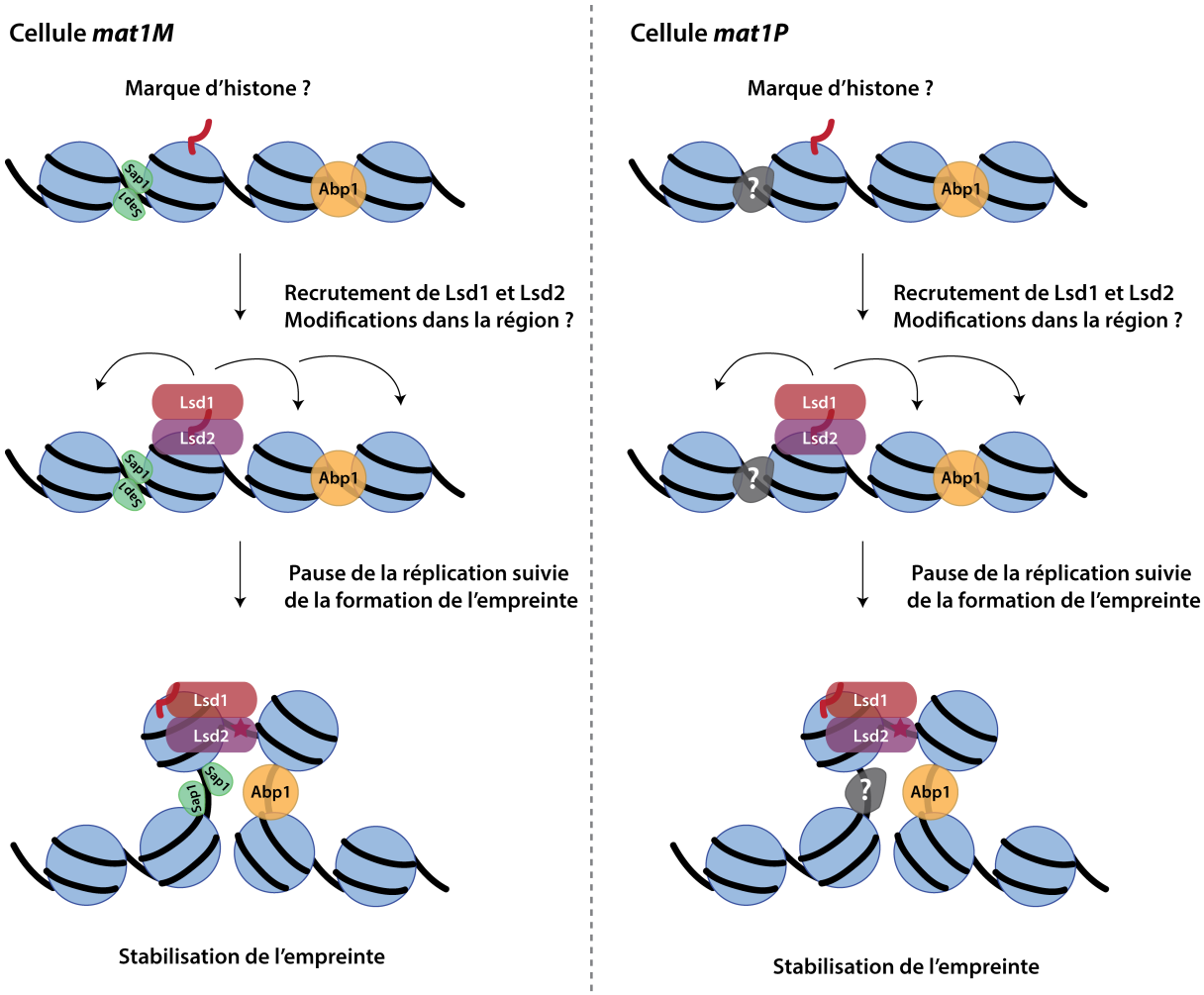


Figure 1 : Modèle du repliement autour de l'empreinte.
Schéma représentant le modèle d'étude de la formation et de la stabilisation de l'empreinte à *mat1*. A gauche quand l'allèle est *M*, et à droite quand l'allèle est *P*. Les protéines Sap1 (vert), Abp1 (orange) et le complexe Lsd1-Lsd2 (rouge et violet respectivement) sont représentées. Les nucléosomes sont indiqués en bleu et l'ADN en noir.

La directionnalité et l'hétérochromatine à *mat2P* et à *mat3M*

Mes résultats montrent que la protéine Abp1 est recrutée à *swi2* dans une cellule exprimant l'allèle *M* et non dans une cellule exprimant l'allèle *P*. Par ailleurs, la localisation de la protéine Mc (produite par la cassette *mat1M*) au niveau de *swi2* dépend de Abp1 (Matsuda et al., 2011). La présence de *Mc* à *swi2* contrôle l'expression d'un transcrit alternatif de *swi2* (Yu et al., 2012a). De plus, l'expression de la protéine Mc permet l'étalement du complexe Swi2-Swi5 jusqu'à *mat2P* (Aguilar-Arnal et al., 2008; Matsuda et al., 2011). Ces observations suggèrent qu'il y a un effet coopératif entre Abp1 et Mc. La présence des deux protéines est nécessaire pour leur recrutement respectif au niveau du gène *swi2*. De façon intéressante, il a été montré que généralement, Abp1 joue un rôle répressif sur la transcription (Cam et al., 2008; Daulny et

al., 2016; Lorenz et al., 2012; Mallet et al., 2017). Par exemple, en absence d'Abp1 il y a une accumulation de transcrits cryptiques instables (CUTs) au niveau des éléments répétés *Tf2* et des répétitions de l'ADNr (Daulny et al., 2016). Par ailleurs, au niveau des rétroéléments, il a été montré qu'en condition normale de croissance un transcrit instable est produit alors qu'en condition de stress le site d'initiation de la transcription est décalé pour permettre l'expression du rétroélément (Persson et al., 2016). Il est possible qu'au niveau de *swi2*, Abp1 en collaboration avec Mc permette l'activation de la transcription au niveau d'un promoteur cryptique, par analogie a ce qui est observé aux rétroéléments. Ce transcrit serait stable et permettrait l'expression de la protéine Swi2 courte. Les preuves moléculaires démontrant l'action de Mc et d'Abp1 sur l'activation d'un transcrit alternatif au niveau de *swi2* n'ont pas été apportées. Les modèles proposés expliquant la directionnalité reposent sur la répartition différentielle au niveau de la région des donneurs de facteurs protéiques en fonction de l'allèle présent à *mat1* (Aguilar-Arnal et al., 2008; Jakociunas et al., 2013b; Jia et al., 2004b; Matsuda et al., 2011; Yu et al., 2012a). Cependant, des analyses du CTS sur des cellules diploïdes ont révélé que l'allèle à *mat1* alterne de la "même" façon que dans des cellules haploïdes (Egel and Eie, 1987; Klar, 1990). Comment, dans un même noyau, des facteurs peuvent être répartis de façon préférentielle sur l'une ou l'autre des régions des donneurs ? Il est probable que des éléments dont la présence à *mat1* est dépendante de l'allèle, participent à la directionnalité. Il est possible par exemple que la protéine Sap1, préférentiellement recrutée au niveau de *mat1M,* prenne part au choix du donneur. Par ailleurs, Sap1 est également présente au niveau de l'élément *SRE3* à *mat3M*, un élément décrit comme favorisant la RH (Jakociunas et al., 2013b; Jia et al., 2004b). Sap1 est décrite comme guidant l'intégration des rétrotransposons et également comme participant à la conformation tridimensionnelle du noyau (Hickey et al., 2015; Jacobs et al., 2015; Mizuguchi et al., 2017). Au niveau de la région *mat*, Sap1 pourrait favoriser le recrutement de *mat1* au niveau de *mat3M* et la RH. Cependant, le fait que Sap1 soit préférentiellement à *mat1M* et à *mat3M* suppose que la recherche du donneur se fasse d'abord par le rapprochement de *mat1M* au niveau de *mat3M*, puis qu'un autre élément permette de ne pas initier la RH dans ce contexte mais de glisser jusqu'à *mat2P*. L'élément inhibant la recombinaison avec *mat3M* quand *mat1* porte l'allèle *M* n'est à ce jour toujours pas connu. En outre, la protéine qui permettrait le rapprochement spécifique de *mat1P* au niveau du donneur manque à ce modèle.

Les protéines Sap1 et Abp1 sont présentes dans la région des donneurs. Tout d'abord, Abp1 se lie au niveau d'une séquence participant à la nucléation de l'hétérochromatine, du côté

centromère proximal de *CENH* (cette étude, (Aguilar-Arnal et al., 2008; Cam et al., 2008)). Au niveau de ce site, il a été montré qu'Abp1 permet le recrutement de Clr3, une histone déacétylase (Cam et al., 2008). Cependant, en absence d'Abp1, à la différence de ce qui est observé au niveau des centromères, la répression transcriptionnelle dans la région des donneurs est maintenue (Aguilar-Arnal et al., 2008; Nakagawa, 2002). Dans la région des donneurs, il est possible qu'Abp1 ne joue qu'un rôle d'optimisation dans la formation de l'hétérochromatine. D'autre part, Sap1 est recrutée au niveau des régions répétées *IR-L* et *IR-R* ainsi qu'au niveau de *REM3* du côté centromère proximal de *mat3M*. De façon intéressante, *REM3* est décrit comme participant à la formation de l'hétérochromatine (Thon et al., 1999). Sap1 en se liant à cet élément pourrait participer à la nucléation de l'hétérochromatine.

Les études sur l'organisation tridimensionnelle du noyau ont montré que la région *mat* est localisée en périphérie nucléaire, proche du SPB et des centromères (Noma et al., 2006). De plus, cette localisation est dépendante de la nature hétérochromatinniene au niveau des donneurs (Alfredsson-Timmins et al., 2007). Plus particulièrement, la méthyle transférase Clr4, ainsi que les éléments de séquences *CENH*, *IR-L* et *IR-R* participent à cette localisation (Alfredsson-Timmins et al., 2007). Deux éléments suggèrent que Sap1 en se liant aux éléments frontières *IR-L* et *IR-R* participe à la localisation de la région *mat* au niveau du SPB via un des composants de la "*shelterin*", la protéine Taz1. Premièrement, Sap1 permet les interactions à longue distance dans le noyau via la protéine Taz1 (Mizuguchi et al., 2017). Deuxièmement, les protéines Taz1 et Raf1 participent à la formation de la frontière hétérochromatique au niveau de *IR-L* et *IR-R* (Toteva et al., 2017).

Par ailleurs, la protéine Abp1 pourrait elle aussi participer à cette localisation spécifique, puisqu'il a été montré qu'elle est importante pour l'organisation nucléaire et en particulier pour le regroupement des rétroéléments au niveau des centromères (Cam et al., 2008; Tanaka et al., 2012).

La colocalisation de Lsd1-Lsd2, Sap1 et Abp1 avec Rad52

L'analyse des sites de recrutement de Lsd1-Lsd2, Sap1 et Abp1 sur l'ensemble du génome a révélé une association avec les sites de liaison de la protéine Rad52. Cette corrélation suggère que ces protéines présentes au locus sexuel participent au maintien de l'intégrité génomique. Plus particulièrement, une partie des pics partagés par Lsd1-Lsd2, Sap1, Abp1 et Rad52 est retrouvée au niveau des gènes hautement transcrits. Comme décrit dans l'introduction, les gènes hautement transcrits sont des lieux sujets à des conflits entre la machinerie transcriptionnelle et réplicative. Il est possible que les protéines Lsd1-Lsd2, Sap1 et Abp1 participent au travers de

mécanismes distincts au maintien de l'intégrité de ces régions. En particulier, il a été montré que les protéines Sap1 et Abp1 contrôlent de façon antagoniste les fourches bloquées au niveau des rétrotransposons (Zaratiegui et al., 2014). Par ailleurs, Sap1 promeut le recrutement du complexe Swi1-Swi3 qui stabilise les fourches de réplication en pause (Noguchi and Noguchi, 2007; Noguchi et al., 2004a). De plus, une mutation ponctuelle dans le gène *sap1* ou une délétion du gène *abp1* entraine une accumulation de foci Rad52, suggérant que des problèmes surviennent lors de la réplication (Mizuguchi et al., 2017; Noguchi and Noguchi, 2007; Zaratiegui et al., 2011b). Enfin, il a été aussi montré qu'en absence d'Abp1, la Pol II phosphorylée sur la sérine 5 est présente le long des rétroéléments et de l'ADNr. Cette forme de la Pol II reflète le fait qu'elle est dans une étape d'élongation (Daulny et al., 2016). En absence d'Abp1, une transcription non désirée de ces éléments est observée. Dans leur ensemble, ces éléments suggèrent que les protéines Sap1 et Abp1 participent au contrôle de la réplication des zones hautement transcrites sujettes à la RH. Enfin, deux phénomènes peuvent expliquer le rôle du complexe Lsd1 et Lsd2 au niveau des gènes hautement transcrits. Premièrement, Lsd1 et Lsd2 participent au contrôle du programme transcriptionnel de la cellule et pourraient être requis pour l'activité transcriptionnelle à ces loci. Deuxièmement, ce complexe contrôle la fourche de réplication à *mat1* et pourrait être nécessaire à la réplication de ces zones difficiles. De plus, en absence de LSD1 dans une lignée cellulaire humaine, il a été observé une augmentation de la RH (Mosammaparast et al., 2013). Ces deux activités pourraient être requise au niveau de ces gènes hautement exprimés. Afin de favoriser l'un ou l'autre de ces phénomènes, j'ai utilisé des données de transcriptomique publiées précédemment (Opel et al., 2007). J'ai comparé l'ensemble des gènes dérégulés en absence de *lsd1* avec l'ensemble des gènes hautement transcrits lié par Lsd1-Lsd2, Rad52. Premièrement, 438 gènes sont dérégulés en absence de Lsd1. Deuxièmement, 74 gènes hautement transcrits sont liés par Lsd1-Lsd2 et Rad52. Parmi ces 74 gènes hautement transcrits seuls 7 sont dérégulés en absence de Lsd1. Ce résultat supporte l'hypothèse que la protéine Lsd1 participe au contrôle de la réplication au niveau de ces zones plutôt qu'à une action directe sur la transcription. Il est possible que cela passe par la distribution correcte des nucléosomes qui permettent le maintien des marques épigénétiques dans ces régions à l'instar du rôle que joue LSD1 au cours du développement chez d'autres eucaryotes.

Perspectives

Mon étude a permis de mettre en évidence une structure particulière à *mat1*. Nous ne savons pas si cette structure participe à la formation/maintenance de l'empreinte, puisque l'une et l'autre sont interdépendantes. Au vu de la complexité de mettre en évidence l'acteur formant l'empreinte, il est possible que plusieurs éléments coopèrent dans ce processus.

L'utilisation de la réplication pour générer de l'asymétrie ne semble pas être restreinte à *S. pombe*. Plusieurs modèles expliquant la division asymétrique des cellules souches proposent que la réplication soit au cœur du processus.

Modèle de John Cairns 1975 : L'immortalité du brin d'ADN

John Cairns formula l'hypothèse de "l'immortalité" du brin d'ADN qui permettrait aux cellules souches de protéger leur ADN des erreurs créées lors de la réplication. Ce modèle propose que les chromatides contenant les brins d'ADN "anciens" soient ségrégées asymétriquement dans l'une des deux cellules filles. Les cellules souches hériteraient des brins d'ADN immortels et cela participerait au maintien de l'intégrité de l'ADN de ces cellules. Plus tardivement, des éléments ont été ajoutés à cette hypothèse, suggérant que les statuts épigénétiques des deux chromatides sœurs sont différents et que ces statuts contrôleraient dans certains cas la répartition asymétrique des chromatides (Lansdorp, 2007). Ce mécanisme nécessite plusieurs postulats, et notamment que l'un des deux brins d'ADN, à savoir l'ancien ou le néo-synthétisé, soit marqué durant la réplication afin qu'ils soient distinctement reconnus et répartis lors de la division cellulaire. Ce modèle soutient donc le fait que la réplication promeut la division asymétrique des cellules souches. Cependant, les études dans ce domaine sont variées et parfois contradictoires ne permettant pas d'avoir une vision claire sur le mécanisme utilisé (pour revue se référer à (Yennek and Tajbakhsh, 2013)). L'avènement de techniques permettant de suivre une cellule unique participera probablement à la compréhension de ce processus.

Modèle de la répartition asymétrique des histones

Lors de la réplication, la chromatine est désassemblée afin de d'ouvrir l'ADN double brin et permettre de synthétiser l'ADN complémentaire. Derrière la fourche, les deux doubles hélices sont rapidement réenroulées autour des nucléosomes. Une question importante qui reste en suspens est la détermination du devenir des nucléosomes parentaux présents devant la fourche. Sont-ils déposés préférentiellement sur l'une des deux chromatides derrière la fourche, ou distribuées de façon aléatoire ? Une répartition biaisée permettrait de créer de l'asymétrie au cours de la réplication. Un modèle propose que le tétramère d'histone $(H3\text{-}H4)_2$ parentales qui porte les marques épigénétiques est réparti sur l'une des chromatides. Ainsi, une seule des chromatides porterait toutes des marques d'histones précédemment établies. À la suite de la réplication, une ségrégation asymétrique en fonction de "l'âge des histones" permettrait de

donner lieu à une cellule similaire à la cellule mère. Ce modèle expliquerait en partie la division des cellules souches et leur capacité à maintenir au cours des divisions un ensemble de cellules souches. Plusieurs études valident le modèle conservatif de la répartition des histones derrière la fourche de réplication, c'est-à-dire que les histones parentales se retrouvent préférentiellement associées à une seule des chromatides (Riley and Weintraub, 1979; Seale, 1976; Seidman et al., 1979; Weintraub, 1976). Cependant, d'autres études montrent que les nucléosomes parentaux sont répartis aléatoirement entre les deux chromatides ; c'est le modèle semi-conservatif (Jackson and Chalkley, 1981, 1985; Russev and Hancock, 1982). Finalement, il est possible qu'à une échelle globale, les histones soient réparties de façon aléatoire, mais que localement un groupe d'histone puisse se coordonner pour se retrouver sur l'une des chromatides (pour revue se référer à (Xie et al., 2017)). Lors de la division cellulaire, la cellule devra être capable de reconnaitre les chromatides contenant les nucléosomes parentaux pour les ségréger préférentiellement dans la cellule mère. Une étude a montré par exemple, que la phosphorylation de la thréonine 3 de l'histone H3 participe à la ségrégation asymétrique des chromatides (Xie et al., 2015). Cette phosphorylation est présente sur l'histone H3 parentale, et son abolition engendre une répartition aléatoire des chromatides.

Bien que ces modèles soient encore controversés, il est concevable que comme chez *S. pombe* la réplication soit au cœur de la division asymétrique chez les métazoaires. L'étude des différents types d'asymétrie permettra une compréhension globale de ces mécanismes qui sont à l'origine de la diversité cellulaire.

Bibliographie

Aguilar-Arnal, L., Marsellach, F.X., and Azorin, F. (2008). The fission yeast homologue of CENP-B, Abp1, regulates directionality of mating-type switching. EMBO J *27*, 1029-1038.

Ahn, J.S., Osman, F., and Whitby, M.C. (2005). Replication fork blockage by RTS1 at an ectopic site promotes recombination in fission yeast. EMBO J *24*, 2011-2023.

Akamatsu, Y., Tsutsui, Y., Morishita, T., Siddique, M.S., Kurokawa, Y., Ikeguchi, M., Yamao, F., Arcangioli, B., and Iwasaki, H. (2007). Fission yeast Swi5/Sfr1 and Rhp55/Rhp57 differentially regulate Rhp51-dependent recombination outcomes. EMBO J *26*, 1352-1362.

Alfredsson-Timmins, J., Henningson, F., and Bjerling, P. (2007). The Clr4 methyltransferase determines the subnuclear localization of the mating-type region in fission yeast. J Cell Sci *120*, 1935-1943.

Allshire, R.C., and Ekwall, K. (2015). Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. Cold Spring Harb Perspect Biol *7*, a018770.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M.*, et al.* (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835-846.

Arcangioli, B. (1998). A site- and strand-specific DNA break confers asymmetric switching potential in fission yeast. EMBO J *17*, 4503-4510.

Arcangioli, B. (2000). Fate of mat1 DNA strands during mating-type switching in fission yeast. EMBO Rep *1*, 145-150.

Arcangioli, B., Copeland, T.D., and Klar, A.J. (1994a). Sap1, a protein that binds to sequences required for mating-type switching, is essential for viability in Schizosaccharomyces pombe. Mol Cell Biol *14*, 2058-2065.

Arcangioli, B., and de Lahondes, R. (2000). Fission yeast switches mating type by a replication-recombination coupled process. EMBO J *19*, 1389-1396.

Arcangioli, B., Ghazvini, M., and Ribes, V. (1994b). Identification of the DNA-binding domains of the switch-activating-protein Sap1 from S.pombe by random point mutations screening in E.coli. Nucleic Acids Res *22*, 2930-2937.

Arcangioli, B., and Klar, A.J. (1991). A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in Schizosaccharomyces pombe. EMBO J *10*, 3025-3032.

Arcangioli, B., and Lescure, B. (1985). Identification of proteins involved in the regulation of yeast iso- 1-cytochrome C expression by oxygen. EMBO J *4*, 2627-2633.

Ayoub, N., Goldshmidt, I., Lyakhovetsky, R., and Cohen, A. (2000). A fission yeast repression element cooperates with centromere-like sequences and defines a mat silent domain boundary. Genetics *156*, 983-994.

Azvolinsky, A., Giresi, P.G., Lieb, J.D., and Zakian, V.A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. Mol Cell *34*, 722-734.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Baum, M., and Clarke, L. (2000). Fission yeast homologs of human CENP-B have redundant functions affecting cell growth and chromosome segregation. Molecular and Cellular Biology *20*, 2852-2864.

Beach, D. (1983). Cell type switching by DNA transposition in fission yeast. Nature *305*, 682-688.

Beach, D., Nurse, P., and Egel, R. (1982). Molecular rearrangement of mating-type genes in fission yeast. Nature *296*, 682-683.

Beach, D.H., and Klar, A.J. (1984). Rearrangements of the transposable mating-type cassettes of fission yeast. EMBO J *3*, 603-610.

Bermejo, R., Lai, M.S., and Foiani, M. (2012). Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. Mol Cell *45*, 710-718.

Berti, M., and Vindigni, A. (2016). Replication stress: getting back on track. Nat Struct Mol Biol *23*, 103-109.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

Birse, C.E., Lee, B.A., Hansen, K., and Proudfoot, N.J. (1997). Transcriptional termination signals for RNA polymerase II in fission yeast. EMBO J *16*, 3633-3643.

Bjerling, P., Silverstein, R.A., Thon, G., Caudy, A., Grewal, S., and Ekwall, K. (2002). Functional divergence between histone deacetylases in fission yeast by distinct cellular localization and in vivo specificity. Mol Cell Biol *22*, 2170-2181.

Bowen, N.J., Jordan, I.K., Epstein, J.A., Wood, V., and Levin, H.L. (2003). Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of Schizosaccharomyces pombe. Genome Res *13*, 1984-1997.

Brewer, B.J. (1988). When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell *53*, 679-686.

Brewer, B.J., and Fangman, W.L. (1987). The localization of replication origins on ARS plasmids in S. cerevisiae. Cell *51*, 463-471.

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. Cell *55*, 637-643.

Cam, H.P., Noma, K., Ebina, H., Levin, H.L., and Grewal, S.I. (2008). Host genome surveillance for retrotransposons by transposon-derived proteins. Nature *451*, 431-436.

Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., and Grewal, S.I. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet *37*, 809-819.

Cam, H.P., and Whitehall, S. (2016). Chromatin Immunoprecipitation (ChIP) in Schizosaccharomyces pombe. Cold Spring Harb Protoc *2016*, pdb prot091546.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.Y., Sanchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell *159*, 572-583.

Church, G.M., and Gilbert, W. (1985). The genomic sequencing technique. Prog Clin Biol Res *177*, 17-21.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature *461*, 415-418.

Cooper, J.P., Nimmo, E.R., Allshire, R.C., and Cech, T.R. (1997). Regulation of telomere length and function by a Myb-domain protein in fission yeast. Nature *385*, 744-747.

Dalgaard, J.Z., and Klar, A.J. (1999). Orientation of DNA replication establishes mating-type switching pattern in S. pombe. Nature *400*, 181-184.

Dalgaard, J.Z., and Klar, A.J. (2000). swi1 and swi3 perform imprinting, pausing, and termination of DNA replication in S. pombe. Cell *102*, 745-751.

Dalgaard, J.Z., and Klar, A.J. (2001). A DNA replication-arrest site RTS1 regulates imprinting by determining the direction of replication at mat1 in S. pombe. Genes Dev *15*, 2060-2068.

Daulny, A., Mejia-Ramirez, E., Reina, O., Rosado-Lugo, J., Aguilar-Arnal, L., Auer, H., Zaratiegui, M., and Azorin, F. (2016). The fission yeast CENP-B protein Abp1 prevents pervasive transcription of repetitive DNA elements. Biochim Biophys Acta *1859*, 1314-1321.

de Castro, E., Soriano, I., Marin, L., Serrano, R., Quintales, L., and Antequera, F. (2012). Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. EMBO J *31*, 124-137.

Deshpande, A.M., and Newlon, C.S. (1996). DNA replication fork pause sites dependent on transcription. Science *272*, 1030-1033.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. Cell *146*, 533-543.

Ebrahimi, A., Habibi-Khorassani, M., Akher, F.B., Farrokhzadeh, A., and Karimi, P. (2013). Caffeine as base analogue of adenine or guanine: a theoretical study. J Mol Graph Model *42*, 81-91.

Egel, R. (1984a). The pedigree pattern of mating-type switching in Schizosaccharomyces pombe. Curr Genet *8*, 205-210.

Egel, R. (1984b). Two tightly linked silent cassettes in the mating-type region of Schizosaccharomyces pombe. Curr Genet *8*, 199-203.

Egel, R., Beach, D.H., and Klar, A.J. (1984). Genes required for initiation and resolution steps of mating-type switching in fission yeast. Proc Natl Acad Sci U S A *81*, 3481-3485.

Egel, R., and Eie, B. (1987). Cell lineage asymmetry in Schizosaccharomyces pombe: unilateral transmission of a high-frequency state for mating-type switching in diploid pedigrees Current Genetics *12*.

Egel, R., and Gutz, H. (1981). Gene activation by copy transposition in mating-type switching of a homothallic fission yeast. Curr Genet *3*, 5-12.

Eissenberg, J.C., and Elgin, S.C. (2000). The HP1 protein family: getting a grip on chromatin. Curr Opin Genet Dev *10*, 204-210.

Ekwall, K., Nielsen, O., and Ruusala, T. (1991). Repression of a mating type cassette in the fission yeast by four DNA elements. Yeast *7*, 745-755.

Ekwall, K., Nimmo, E.R., Javerzat, J.P., Borgstrom, B., Egel, R., Cranston, G., and Allshire, R. (1996). Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. J Cell Sci *109 (Pt 11)*, 2637-2648.

Ekwall, K., and Ruusala, T. (1994). Mutations in rik1, clr2, clr3 and clr4 genes asymmetrically derepress the silent mating-type loci in fission yeast. Genetics *136*, 53-64.

Engelke, U., Grabowski, L., Gutz, H., Heim, H., Schmidt, H. (1987). Molecular characterization of h- mutants of Schizosaccharomyces pombe.

Eydmann, T., Sommariva, E., Inagawa, T., Mian, S., Klar, A.J., and Dalgaard, J.Z. (2008). Rtf1-mediated eukaryotic site-specific replication termination. Genetics *180*, 27-39.

Fachinetti, D., Folco, H.D., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.*, et al.* (2013). A two-step mechanism for epigenetic specification of centromere identity and function. Nat Cell Biol *15*, 1056-1066.

Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev Cell *33*, 314-327.

Fleck, O., Rudolph, C., Albrecht, A., Lorentz, A., Schar, P., and Schmidt, H. (1994). The mutator gene swi8 effects specific mutations in the mating-type region of Schizosaccharomyces pombe. Genetics *138*, 621-632.

Funabiki, H., Hagan, I., Uzawa, S., and Yanagida, M. (1993). Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. J Cell Biol *121*, 961-976.

Gordon, M., Holt, D.G., Panigrahi, A., Wilhelm, B.T., Erdjument-Bromage, H., Tempst, P., Bahler, J., and Cairns, B.R. (2007). Genome-wide dynamics of SAPHIRE, an essential complex for gene activation and chromatin boundaries. Mol Cell Biol *27*, 4058-4069.

Grewal, S.I. (2000). Transcriptional silencing in fission yeast. J Cell Physiol *184*, 311-318.

Grewal, S.I., and Elgin, S.C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Opin Genet Dev *12*, 178-187.

Grewal, S.I., and Klar, A.J. (1996). Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. Cell *86*, 95-101.

Grewal, S.I., and Klar, A.J. (1997). A recombinationally repressed region between mat2 and mat3 loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. Genetics *146*, 1221-1238.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hanada, K., Budzowska, M., Davies, S.L., van Drunen, E., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I.D., and Kanaar, R. (2007). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. Nat Struct Mol Biol *14*, 1096-1104.

Haruta, N., Akamatsu, Y., Tsutsui, Y., Kurokawa, Y., Murayama, Y., Arcangioli, B., and Iwasaki, H. (2008). Fission yeast Swi5 protein, a novel DNA recombination mediator. DNA Repair (Amst) *7*, 1-9.

Hennig, B.P., Bendrin, K., Zhou, Y., and Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. EMBO Rep *13*, 997-1003.

Hickey, A., Esnault, C., Majumdar, A., Chatterjee, A.G., Iben, J.R., McQueen, P.G., Yang, A.X., Mizuguchi, T., Grewal, S.I., and Levin, H.L. (2015). Single-Nucleotide-Specific Targeting of the Tf1 Retrotransposon Promoted by the DNA-Binding Protein Sap1 of Schizosaccharomyces pombe. Genetics *201*, 905-924.

Holmes, A., Roseaulin, L., Schurra, C., Waxin, H., Lambert, S., Zaratiegui, M., Martienssen, R.A., and Arcangioli, B. (2012). Lsd1 and lsd2 control programmed replication fork pauses and imprinting in fission yeast. Cell Rep *2*, 1513-1520.

Holmes, A.M., Kaykov, A., and Arcangioli, B. (2005). Molecular and cellular dissection of mating-type switching steps in Schizosaccharomyces pombe. Mol Cell Biol *25*, 303-311.

Inagawa, T., Yamada-Inagawa, T., Eydmann, T., Mian, I.S., Wang, T.S., and Dalgaard, J.Z. (2009). Schizosaccharomyces pombe Rtf2 mediates site-specific replication termination by inhibiting replication restart. Proc Natl Acad Sci U S A *106*, 7927-7932.

Irelan, J.T., Gutkin, G.I., and Clarke, L. (2001). Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. Genetics *157*, 1191-1203.

Irmisch, A., Ampatzidou, E., Mizuno, K., O'Connell, M.J., and Murray, J.M. (2009). Smc5/6 maintains stalled replication forks in a recombination-competent conformation. EMBO J *28*, 144-155.

Ivanova, A.V., Bonaduce, M.J., Ivanov, S.V., and Klar, A.J. (1998). The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. Nat Genet *19*, 192-195.

Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S.I., and Noma, K. (2010). Centromeric localization of dispersed Pol III genes in fission yeast. Mol Biol Cell *21*, 254-265.

Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev *6*, 1342-1356.

Jackson, V., and Chalkley, R. (1981). A new method for the isolation of replicative chromatin: selective deposition of histone on both new and old DNA. Cell *23*, 121-134.

Jackson, V., and Chalkley, R. (1985). Histone segregation on replicating chromatin. Biochemistry *24*, 6930-6938.

Jacobs, J.Z., Rosado-Lugo, J.D., Cranz-Mileva, S., Ciccaglione, K.M., Tournier, V., and Zaratiegui, M. (2015). Arrested replication forks guide retrotransposon integration. Science *349*, 1549-1553.

Jakociunas, T., Domange Jordo, M., Ait Mebarek, M., Bunner, C.M., Verhein-Hansen, J., Oddershede, L.B., and Thon, G. (2013a). Subnuclear relocalization and silencing of a chromosomal region by an ectopic ribosomal DNA repeat. Proc Natl Acad Sci U S A *110*, E4465-4473.

Jakociunas, T., Holm, L.R., Verhein-Hansen, J., Trusina, A., and Thon, G. (2013b). Two portable recombination enhancers direct donor choice in fission yeast heterochromatin. PLoS Genet *9*, e1003762.

Jia, S., Noma, K., and Grewal, S.I. (2004a). RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science *304*, 1971-1976.

Jia, S., Yamada, T., and Grewal, S.I.S. (2004b). Heterochromatin Regulates Cell Type-Specific Long-Range Chromatin Interactions Essential for Directed Recombination. Cell *119*, 469-480.

Katz, D.J., Edwards, T.M., Reinke, V., and Kelly, W.G. (2009). A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. Cell *137*, 308-320.

Kaykov, A., and Arcangioli, B. (2004). A Programmed Strand-Specific and Modified Nick in S. pombe Constitutes a Novel Type of Chromosomal Imprint. Current Biology *14*, 1924-1928.

Kaykov, A., Holmes, A.M., and Arcangioli, B. (2004). Formation, maintenance and consequences of the imprint at the mating-type locus in fission yeast. EMBO J *23*, 930-938.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. EMBO J *7*, 1537-1547.

Kerenyi, M.A., Shao, Z., Hsu, Y.J., Guo, G., Luc, S., O'Brien, K., Fujiwara, Y., Peng, C., Nguyen, M., and Orkin, S.H. (2013). Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. Elife *2*, e00633.

Kim, H.S., Choi, E.S., Shin, J.A., Jang, Y.K., and Park, S.D. (2004). Regulation of Swi6/HP1-dependent heterochromatin assembly by cooperation of components of the mitogen-activated protein kinase pathway and a histone deacetylase Clr6. J Biol Chem *279*, 42850-42859.

Kim, S.M., and Huberman, J.A. (1998). Multiple orientation-dependent, synergistically interacting, similar domains in the ribosomal DNA replication origin of the fission yeast, Schizosaccharomyces pombe. Mol Cell Biol *18*, 7294-7303.

Klar, A.J. (1987). Differentiated parental DNA strands confer developmental asymmetry on daughter cells in fission yeast. Nature *326*, 466-470.

Klar, A.J. (1990). The developmental fate of fission yeast cells is determined by the pattern of inheritance of parental and grandparental DNA strands. EMBO J *9*, 1407-1415.

Klar, A.J., and Bonaduce, M.J. (1991). swi6, a gene required for mating-type switching, prohibits meiotic recombination in the mat2-mat3 "cold spot" of fission yeast. Genetics *129*, 1033-1042.

Klar, A.J., Bonaduce, M.J., and Cafferkey, R. (1991). The mechanism of fission yeast mating type interconversion: seal/replicate/cleave model of replication across the double-stranded break site at mat1. Genetics *127*, 489-496.

Klar, A.J., and Miglio, L.M. (1986). Initiation of meiotic recombination by double-strand DNA breaks in S. pombe. Cell *46*, 725-731.

Kostriken, R., Strathern, J.N., Klar, A.J., Hicks, J.B., and Heffron, F. (1983). A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. Cell *35*, 167-174.

Krings, G., and Bastia, D. (2005). Sap1p binds to Ter1 at the ribosomal DNA of Schizosaccharomyces pombe and causes polar replication fork arrest. J Biol Chem *280*, 39135-39142.

Krings, G., and Bastia, D. (2006). Molecular architecture of a eukaryotic DNA replication terminus-terminator protein complex. Mol Cell Biol *26*, 8061-8074.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lalonde, B., Arcangioli, B., and Guarente, L. (1986). A single Saccharomyces cerevisiae upstream activation site (UAS1) has two distinct regions essential for its activity. Mol Cell Biol *6*, 4690-4696.

Lan, F., Zaratiegui, M., Villen, J., Vaughn, M.W., Verdel, A., Huarte, M., Shi, Y., Gygi, S.P., Moazed, D., Martienssen, R.A.*, et al.* (2007). S. pombe LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. Mol Cell *26*, 89-101.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P.*, et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813-1831.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lansdorp, P.M. (2007). Immortal strands? Give me a break. Cell *129*, 1244-1247.

Lee, B.S., Grewal, S.I., and Klar, A.J. (2004). Biochemical interactions between proteins and mat1 cis-acting sequences required for imprinting in fission yeast. Mol Cell Biol *24*, 9813-9822.

Lee, J.K., Huberman, J.A., and Hurwitz, J. (1997). Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of Schizosaccharomyces pombe. Proceedings of the National Academy of Sciences of the United States of America *94*, 8427-8432.

Levin, H.L., Weaver, D.C., and Boeke, J.D. (1990). Two related families of retrotransposons from Schizosaccharomyces pombe. Mol Cell Biol *10*, 6791-6798.

Lorentz, A., Heim, L., and Schmidt, H. (1992). The switching gene swi6 affects recombination and gene expression in the mating-type region of Schizosaccharomyces pombe. Mol Gen Genet *233*, 436-442.

Lorentz, A., Ostermann, K., Fleck, O., and Schmidt, H. (1994). Switching gene swi6, involved in repression of silent mating-type loci in fission yeast, encodes a homologue of chromatin-associated proteins from Drosophila and mammals. Gene *143*, 139-143.

Lorenz, D.R., Mikheyeva, I.V., Johansen, P., Meyer, L., Berg, A., Grewal, S.I.S., and Cam, H.P. (2012). CENP-B Cooperates with Set1 in Bidirectional Transcriptional Silencing and Genome Organization of Retrotransposons. Molecular and Cellular Biology *32*, 4215-4225.

Ma, J., Hwang, K.K., Worman, H.J., Courvalin, J.C., and Eissenberg, J.C. (2001). Expression and functional analysis of three isoforms of human heterochromatin-associated protein HP1 in Drosophila. Chromosoma *109*, 536-544.

MacNeill, S.A., and Fantes, P.A. (1997). Genetic and physiological analysis of DNA replication in fission yeast. Methods Enzymol *283*, 440-459.

Magdalou, I., Lopez, B.S., Pasero, P., and Lambert, S.A. (2014). The causes of replication stress and their consequences on genome stability and cell fate. Semin Cell Dev Biol *30*, 154-164.

Mallet, P.L., Larochelle, M., and Bachand, F. (2017). Multiple Transcriptional and Post-transcriptional Pathways Collaborate to Control Sense and Antisense RNAs of Tf2 Retroelements in Fission Yeast. Genetics *205*, 621-632.

Mallette, F.A., Mattiroli, F., Cui, G., Young, L.C., Hendzel, M.J., Mer, G., Sixma, T.K., and Richard, S. (2012). RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. EMBO J *31*, 1865-1878.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bahler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell *151*, 671-683.

Martienssen, R., and Moazed, D. (2015). RNAi and heterochromatin assembly. Cold Spring Harb Perspect Biol *7*, a019323.

Matsuda, E., Sugioka-Sugiyama, R., Mizuguchi, T., Mehta, S., Cui, B., and Grewal, S.I. (2011). A homolog of male sex-determining factor SRY cooperates with a transposon-derived CENP-B protein to control sex-specific directed recombination. Proc Natl Acad Sci U S A *108*, 18754-18759.

Maundrell, K., Hutchison, A., and Shall, S. (1988). Sequence analysis of ARS elements in fission yeast. EMBO J *7*, 2203-2209.

Meister, P., Taddei, A., Ponti, A., Baldacci, G., and Gasser, S.M. (2007). Replication foci dynamics: replication patterns are modulated by S-phase checkpoint kinases in fission yeast. EMBO J *26*, 1315-1326.

Mejia-Ramirez, E., Sanchez-Gorostiaga, A., Krimer, D.B., Schvartzman, J.B., and Hernandez, P. (2005). The mating type switch-activating protein Sap1 Is required for replication fork arrest at the rRNA genes of fission yeast. Mol Cell Biol *25*, 8755-8761.

Meller, V.H., Joshi, S.S., and Deshpande, N. (2015). Modulation of Chromatin by Noncoding RNA. Annu Rev Genet *49*, 673-695.

Meng, X., and Zhao, X. (2017). Replication fork regression and its regulation. FEMS Yeast Res *17*.

Merrikh, H., Machon, C., Grainger, W.H., Grossman, A.D., and Soultanas, P. (2011). Co-directional replication-transcription conflicts lead to replication restart. Nature *470*, 554-557.

Michael, H., Schmidt, H., Fleck, O., Gutz, H., Liedtke, C., Lorentz, A., and Ostermann, K. (1994). The mating-type region of Schizosaccharomyces pombe contains an essential gene encoding a protein homologous to human modulators of HIV transactivation. Gene *145*, 205-210.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication fork stalling at natural impediments. Microbiol Mol Biol Rev *71*, 13-35.

Mitchison, J.M., and Creanor, J. (1971). Further measurements of DNA synthesis and enzyme potential during cell cycle of fission yeast Schizosaccharomyces pombe. Exp Cell Res *69*, 244-247.

Miyata, H., and Miyata, M. (1981). Mode of Conjugation in Homothallic Cells of Schizosaccharomyces-Pombe. J Gen Appl Microbiol *27*, 365-371.

Mizuguchi, T., Barrowman, J., and Grewal, S.I. (2015). Chromosome domain architecture and dynamic organization of the fission yeast genome. FEBS Lett *589*, 2975-2986.

Mizuguchi, T., Taneja, N., Matsuda, E., Belton, J.M., FitzGerald, P., Dekker, J., and Grewal, S.I.S. (2017). Shelterin components mediate genome reorganization in response to replication stress. Proc Natl Acad Sci U S A *114*, 5479-5484.

Molnar, M., and Kleckner, N. (2008). Examination of interchromosomal interactions in vegetatively growing diploid Schizosaccharomyces pombe cells by Cre/loxP site-specific recombination. Genetics *178*, 99-112.

Mosammaparast, N., Kim, H., Laurent, B., Zhao, Y., Lim, H.J., Majid, M.C., Dango, S., Luo, Y., Hempel, K., Sowa, M.E.*, et al.* (2013). The histone demethylase LSD1/KDM1A promotes the DNA damage response. J Cell Biol *203*, 457-470.

Nakagawa, H. (2002). Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes & Development *16*, 1766-1778.

Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O., and Yanagida, M. (1986). Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J *5*, 1011-1021.

Nakayama, J., Klar, A.J., and Grewal, S.I. (2000). A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. Cell *101*, 307-317.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110-113.

Neelsen, K.J., and Lopes, M. (2015). Replication fork reversal in eukaryotes: from dead end to dynamic response. Nat Rev Mol Cell Biol *16*, 207-220.

Nicolas, E., Lee, M.G., Hakimi, M.A., Cam, H.P., Grewal, S.I., and Shiekhattar, R. (2006). Fission yeast homologs of human histone H3 lysine 4 demethylase regulate a common set of genes with diverse functions. J Biol Chem *281*, 35983-35988.

Nielsen, O., and Egel, R. (1989). Mapping the double-strand breaks at the mating-type locus in fission yeast by genomic sequencing. EMBO J *8*, 269-276.

Nikolov, I., and Taddei, A. (2016). Linking replication stress with heterochromatin formation. Chromosoma *125*, 523-533.

Nimmo, E.R., Cranston, G., and Allshire, R.C. (1994). Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. EMBO J *13*, 3801-3811.

Noguchi, C., and Noguchi, E. (2007). Sap1 Promotes the Association of the Replication Fork Protection Complex With Chromatin and Is Involved in the Replication Checkpoint in Schizosaccharomyces pombe. Genetics *175*, 553-566.

Noguchi, E., Noguchi, C., Du, L.L., and Russell, P. (2003). Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1. Mol Cell Biol *23*, 7861-7874.

Noguchi, E., Noguchi, C., McDonald, W.H., Yates, J.R., 3rd, and Russell, P. (2004a). Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. Mol Cell Biol *24*, 8342-8355.

Noguchi, E., Noguchi, C., McDonald, W.H., Yates, J.R., and Russell, P. (2004b). Swi1 and Swi3 Are Components of a Replication Fork Protection Complex in Fission Yeast. Molecular and Cellular Biology *24*, 8342-8355.

Noma, K., Allis, C.D., and Grewal, S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science *293*, 1150-1155.

Noma, K., Cam, H.P., Maraia, R.J., and Grewal, S.I. (2006). A role for TFIIIC transcription factor complex in genome organization. Cell *125*, 859-872.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. Nat Genet *36*, 1174-1180.

Noma, K.i. (2001). Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries. Science *293*, 1150-1155.

Nottke, A.C., Beese-Sims, S.E., Pantalena, L.F., Reinke, V., Shi, Y., and Colaiacovo, M.P. (2011). SPR-5 is a histone H3K4 demethylase with a role in meiotic double-strand break repair. Proc Natl Acad Sci U S A *108*, 12805-12810.

Olsson, T.G., Silverstein, R.A., Ekwall, K., and Sunnerhagen, P. (1999). Transient inhibition of histone deacetylase activity overcomes silencing in the mating-type region in fission yeast. Curr Genet *35*, 82-87.

Opel, M., Lando, D., Bonilla, C., Trewick, S.C., Boukaba, A., Walfridsson, J., Cauwood, J., Werler, P.J., Carr, A.M., Kouzarides, T.*, et al.* (2007). Genome-wide studies of histone demethylation catalysed by the fission yeast homologues of mammalian LSD1. PLoS One *2*, e386.

Ostermann, K., Lorentz, A., and Schmidt, H. (1993). The fission yeast rad22 gene, having a function in mating-type switching and repair of DNA damages, encodes a protein homolog to Rad52 of Saccharomyces cerevisiae. Nucleic Acids Res *21*, 5940-5944.

Persson, J., Steglich, B., Smialowska, A., Boyd, M., Bornholdt, J., Andersson, R., Schurra, C., Arcangioli, B., Sandelin, A., Nielsen, O.*, et al.* (2016). Regulating retrotransposon activity through the use of alternative transcription start sites. EMBO Rep *17*, 753-768.

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Stralfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. EMBO J *31*, 4388-4403.

Reines, D. (1992). Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. J Biol Chem *267*, 3795-3800.

Riley, D., and Weintraub, H. (1979). Conservative segregation of parental histones during replication in the presence of cycloheximide. Proc Natl Acad Sci U S A *76*, 328-332.

Roseaulin, L., Yamada, Y., Tsutsui, Y., Russell, P., Iwasaki, H., and Arcangioli, B. (2008). Mus81 is essential for sister chromatid recombination at broken replication forks. EMBO J *27*, 1378-1387.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T.*, et al.* (2007). Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. Mol Cell *26*, 103-115.

Russev, G., and Hancock, R. (1982). Assembly of new histones into nucleosomes and their distribution in replicating chromatin. Proc Natl Acad Sci U S A *79*, 3143-3147.

Sabouri, N., McDonald, K.R., Webb, C.J., Cristea, I.M., and Zakian, V.A. (2012). DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the S. pombe Pfh1 helicase. Genes Dev *26*, 581-593.

Sanchez, J.A., Kim, S.M., and Huberman, J.A. (1998). Ribosomal DNA replication in the fission yeast, Schizosaccharomyces pombe. Exp Cell Res *238*, 220-230.

Sayrac, S., Vengrova, S., Godfrey, E.L., and Dalgaard, J.Z. (2011). Identification of a Novel Type of Spacer Element Required for Imprinting in Fission Yeast. PLoS Genetics *7*, e1001328-1001318.

Schaak, J., Mao, J., and Soll, D. (1982). The 5.8S RNA gene sequence and the ribosomal repeat of Schizosaccharomyces pombe. Nucleic Acids Res *10*, 2851-2864.

Schmidt, H., Kapitza-Fecke, P., Stephen, E.R., and Gutz, H. (1989). Some of the swi genes of Schizosaccharomyces pombe also have a function in the repair of radiation damage. Curr Genet *16*, 89-94.

Schwartz, D.C., and Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell *37*, 67-75.

Seale, R.L. (1976). Studies on the mode of segregation of histone nu bodies during replication in HeLa cells. Cell *9*, 423-429.

Seidman, M.M., Levine, A.J., and Weintraub, H. (1979). The asymmetric segregation of parental nucleosomes during chrosome replication. Cell *18*, 439-449.

Shankaranarayana, G.D., Motamedi, M.R., Moazed, D., and Grewal, S.I. (2003). Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. Curr Biol *13*, 1240-1246.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Singh, J., and Klar, A.J. (1993). DNA polymerase-alpha is essential for mating-type switching in fission yeast. Nature *361*, 271-273.

Styrkarsdottir, U., Egel, R., and Nielsen, O. (1993). The smt-0 mutation which abolishes mating-type switching in fission yeast is a deletion. Curr Genet *23*, 184-186.

Svejstrup, J.Q. (2010). The interface between transcription and mechanisms maintaining genome integrity. Trends Biochem Sci *35*, 333-338.

Szabad, J., Reuter, G., and Schroder, M.B. (1988). The effects of two mutations connected with chromatin functions on female germ-line cells of Drosophila. Mol Gen Genet *211*, 56-62.

Szilard, R.K., Jacques, P.E., Laramee, L., Cheng, B., Galicia, S., Bataille, A.R., Yeung, M., Mendez, M., Bergeron, M., Robert, F.*, et al.* (2010). Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. Nat Struct Mol Biol *17*, 299-305.

Szostak, J.W., and Blackburn, E.H. (1982). Cloning yeast telomeres on linear plasmid vectors. Cell *29*, 245-255.

Tanaka, A., Tanizawa, H., Sriswasdi, S., Iwasaki, O., Chatterjee, A.G., Speicher, D.W., Levin, H.L., Noguchi, E., and Noma, K. (2012). Epigenetic regulation of condensin-mediated genome organization during the cell cycle and upon DNA damage through histone H3 lysine 56 acetylation. Mol Cell *48*, 532-546.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res *38*, 8164-8177.

Thon, G., Bjerling, K.P., and Nielsen, I.S. (1999). Localization and properties of a silencing element near the mat3-M mating-type cassette of Schizosaccharomyces pombe. Genetics *151*, 945-963.

Thon, G., Bjerling, P., Bunner, C.M., and Verhein-Hansen, J. (2002). Expression-state boundaries in the mating-type region of fission yeast. Genetics *161*, 611-622.

Thon, G., Cohen, A., and Klar, A.J. (1994). Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of Schizosaccharomyces pombe. Genetics *138*, 29-38.

Thon, G., and Friis, T. (1997). Epigenetic inheritance of transcriptional silencing and switching competence in fission yeast. Genetics *145*, 685-696.

Thon, G., Hansen, K.R., Altes, S.P., Sidhu, D., Singh, G., Verhein-Hansen, J., Bonaduce, M.J., and Klar, A.J. (2005). The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast Schizosaccharomyces pombe. Genetics *171*, 1583-1595.

Thon, G., and Klar, A.J. (1992). The clr1 locus regulates the expression of the cryptic mating-type loci of fission yeast. Genetics *131*, 287-296.

Thon, G., and Klar, A.J. (1993). Directionality of fission yeast mating-type interconversion is controlled by the location of the donor loci. Genetics *134*, 1045-1054.

Thon, G., and Verhein-Hansen, J. (2000). Four chromo-domain proteins of Schizosaccharomyces pombe differentially repress transcription at various chromosomal locations. Genetics *155*, 551-568.

Toteva, T., Mason, B., Kanoh, Y., Brogger, P., Green, D., Verhein-Hansen, J., Masai, H., and Thon, G. (2017). Establishment of expression-state boundaries by Rif1 and Taz1 in fission yeast. Proc Natl Acad Sci U S A *114*, 1093-1098.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., and Rando, O.J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res *21*, 1851-1862.

Tuzon, C.T., Borgstrom, B., Weilguny, D., Egel, R., Cooper, J.P., and Nielsen, O. (2004). The fission yeast heterochromatin protein Rik1 is required for telomere clustering during meiosis. J Cell Biol *165*, 759-765.

Umesono, K., Hiraoka, Y., Toda, T., and Yanagida, M. (1983). Visualization of chromosomes in mitotically arrested cells of the fission yeast Schizosaccharomyces pombe. Curr Genet *7*, 123-128.

Vagin, D.A., Khasanov, F.K., and Bashkirov, V.I. (2006). [The role of recombinational repair proteins in mating type switching in fission yeast cells]. Genetika *42*, 487-493.

Vengrova, S., and Dalgaard, J.Z. (2004). RNase-sensitive DNA modification(s) initiates S. pombe mating-type switching. Genes Dev *18*, 794-804.

Vengrova, S., and Dalgaard, J.Z. (2006). The wild-type Schizosaccharomyces pombe mat1 imprint consists of two ribonucleotides. EMBO reports *7*, 59-65.

Wang, D., and Hawley, D.K. (1993). Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. Proc Natl Acad Sci U S A *90*, 843-847.

Wang, G., Ma, A., Chow, C.M., Horsley, D., Brown, N.R., Cowell, I.G., and Singh, P.B. (2000). Conservation of heterochromatin protein 1 function. Mol Cell Biol *20*, 6970-6983.

Wang, J., Tadeo, X., Hou, H., Tu, P.G., Thompson, J., Yates, J.R., 3rd, and Jia, S. (2013). Epe1 recruits BET family bromodomain protein Bdf2 to establish heterochromatin boundaries. Genes Dev *27*, 1886-1902.

Weintraub, H. (1976). Cooperative alignment of nu bodies during chromosome replication in the presence of cycloheximide. Cell *9*, 419-422.

Wilkinson, C.R., Bartlett, R., Nurse, P., and Bird, A.P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res *23*, 203-210.

Willer, M., Hoffmann, L., Styrkarsdottir, U., Egel, R., Davey, J., and Nielsen, O. (1995). Two-step activation of meiosis by the mat1 locus in Schizosaccharomyces pombe. Mol Cell Biol *15*, 4964-4970.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S.*, et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Xie, J., Wooten, M., Tran, V., Chen, B.C., Pozmanter, C., Simbolon, C., Betzig, E., and Chen, X. (2015). Histone H3 Threonine Phosphorylation Regulates Asymmetric Histone Inheritance in the Drosophila Male Germline. Cell *163*, 920-933.

Xie, J., Wooten, M., Tran, V., and Chen, X. (2017). Breaking Symmetry - Asymmetric Histone Inheritance in Stem Cells. Trends Cell Biol *27*, 527-540.

Yennek, S., and Tajbakhsh, S. (2013). DNA asymmetry and cell fate regulation in stem cells. Semin Cell Dev Biol *24*, 627-642.

Yu, C., Bonaduce, M.J., and Klar, A.J. (2012). Going in the right direction: mating-type switching of Schizosaccharomyces pombe is controlled by judicious expression of two different swi2 transcripts. Genetics *190*, 977-987.

Zaratiegui, M., Castel, S.E., Irvine, D.V., Kloc, A., Ren, J., Li, F., de Castro, E., Marín, L., Chang, A.-Y., Goto, D.*, et al.* (2011a). RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. Nature *479*, 135-138.

Zaratiegui, M., Vaughn, M.W., Irvine, D.V., Goto, D., Watt, S., Bahler, J., Arcangioli, B., and Martienssen, R.A. (2011b). CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. Nature *469*, 112-115.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. (2008a). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat Struct Mol Biol *15*, 381-388.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008b). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Livres :
Alfa, C. (1993). Experiments with fission yeast: A laboratory course manual. Cold Spring Harbor Laboratory Press

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning a laboratory manual. Cold Spring Harbor Laboratory Press Vol. 2 1989

Waddington, C. H. (1957). The Strategy of the Genes. Geo Allen & Unwin, London.