



HAL
open science

Implication des factures de remodelage de chromatine de la famille CHD dans les réseaux de régulation transcriptionnelle des cellules souches embryonnaires

Maud de Dieuleveult

► To cite this version:

Maud de Dieuleveult. Implication des factures de remodelage de chromatine de la famille CHD dans les réseaux de régulation transcriptionnelle des cellules souches embryonnaires. Sciences du Vivant [q-bio]. Université Paris Sud - Paris XI, 2010. Français. NNT: . tel-00555030

HAL Id: tel-00555030

<https://theses.hal.science/tel-00555030>

Submitted on 15 Jan 2012

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.

UNIVERSITE DE PARIS-SUD U.F.R. SCIENTIFIQUE D'ORSAY

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS XI ORSAY

École doctorale "Gènes, Génomes, Cellules"

Présentée et soutenue publiquement le 17 septembre 2010

par

Maud de Dieuleveult

Implication des facteurs de remodelage de
chromatine de la famille CHD dans les réseaux de
régulation transcriptionnelle des cellules souches
embryonnaires

Jury

Pr Pierre Capy, Président
Dr Pierre Ferrier, Rapporteur
Dr Christian Muchardt, Rapporteur
Dr Ute Rogner, Examinatrice
Dr Etienne Thevenot, Examineur
Dr Matthieu Gérard, Directeur de thèse

Les remerciements resteront la partie la plus lue et la plus facile à écrire pour moi.

Je tiens en premier lieu à remercier les membres du jury: le Docteur Christian Muchardt d'avoir été mon tuteur de thèse et mon rapporteur, merci de vos conseils avisés, le Docteur Pierre Ferrier d'avoir accepté d'être rapporteur de cette thèse, le Docteur Ute Rogner d'avoir accepté d'examiner cette thèse puis le Professeur Pierre Capy de représenter l'école doctorale GGC et la Faculté Paris XI. Enfin, je remercie le Docteur Etienne Thevenot au CEA/DRT d'avoir accepté d'examiner ma thèse, de sa disponibilité et sa gentillesse lors de mes tentatives de "R". Ce projet n'aurait pas vu le jour sans les financements dont j'ai pu bénéficier: CEA pendant trois ans et FRM pour la quatrième année.

Je tiens aussi à remercier Matthieu Gérard, mon directeur de thèse, de m'avoir fait confiance en 2006 pour initier un projet novateur puis de m'avoir encadrée durant ces quatre années.

Cette thèse n'aurait pas eu cette envergure sans l'aide et le soutien des quatre post-docs qui m'ont guidée. Fayçal, tu m'as tout appris quant au recombineering et à la spermatogenèse. Merci pour nos grandes discussions qui, je l'espère, perdureront. Sébastien, merci de tes conseils et de tes "trucs et astuces" utiles en tout genre. Arnaud, the « very famous one: king of the Bioch », merci de m'avoir enseigné l'art de la biochimie durant cette longue période de mise au point et de m'avoir soutenue quand cela ne marchait pas. Les tricky questions ou "questions à la Arnaud" resteront mémorables ainsi que ta fidèle rapière. Merci de ta disponibilité et de ton écoute. Isabelle, merci de ton aide (manips et manuscrit), de nos discussions et ton soutien pendant la fin de ma thèse et d'avoir toujours goûté avec complaisance et motivation mes inventions culinaires.

Merci Soizik des conseils que tu m'as prodigués lorsque j'ai commencé ma thèse et de ta disponibilité alors que tu rédigeais. Lucie, merci de ta présence, de ta gentillesse, de ton écoute apaisante dans le laboratoire, de nos fous-rires, de nos découvertes musicales au laboratoire, d'avoir supporté mes "post-18h", surtout le vendredi soir, qui finissent dans le bac à vaisselle (sale) et de nos soirées décompression... sans oublier ta voiture et les nombreux trajets Saclay-Versailles!!! En espérant, qu'à mon tour, j'ai pu t'aider au moins à ne pas fuir devant Linux (ça a l'air d'être le cas...) !

Merci à Sylvie de tes bonnes questions, de m'avoir appris toutes les techniques de bases et de me les avoir toujours enseignées avec ta grande patience. Je tenais aussi à te remercier de m'avoir écoutée avec toujours autant d'attention. LN merci de ta bonne humeur et d'avoir été si bon public. Toujours à l'affût d'une bonne blague avec ton complice de toujours, d'une bataille de glace, faisant pendre mes chaussures au plafond ou me poursuivant avec un appareil photo! Grâce à toi le labo a revécu les folles années de "Walnut grove" et pendant encore longtemps nos affiches feront pâlir d'envie les visiteurs du 142. Nombreux sont les sourires qui s'affichent

en regardant naïvement ma photo de "Laura Ingalls" postée par tes soins sur la porte du labo. Merci à toutes deux de votre aide indispensable au laboratoire et à notre projet si ambitieux. Grâce à vous, nous n'avons pas baissé les bras, et les résultats ont été à la hauteur de votre implication.

Merci à tous ceux qui ont participé à cette thèse. Je n'ai été que le fédérateur qui a permis de fusionner idées, savoirs et compétences. Sans vous rien n'aurait pu se faire.

Claire Keime pour son aide pour les analyses RNA-seq,

Arnaud Matel et Jacky Husser du GIPSI au CEA Saclay pour leur aide LINUX et les nombreux debugages en tout genre, je vous promets d'arrêter de faire planter le serveur avec mes analyses,

Jean-Christophe Aude pour le bureau de ministre qu'il m'a prêté pour rédiger au calme mes demandes de quatrième année et mon manuscrit de thèse, ainsi que pour l'aide pour la bio-info,

Arnaud Krebs à l'IGBMC (Strasbourg) pour ma semaine de formation intense aux analyses de ChIP-seq, ainsi que toute l'équipe de Laszlo Tora à l'IGBMC pour son accueil très "schnaps",

Tao Ye pour seqMINER et ses réponses toujours promptes aux mails d'aide en bio-info,

Gaël Yvert pour m'avoir toujours soutenue et pour sa lettre de recommandation très motivée pour l'obtention de la bourse FRM qui a porté ses fruits!,

Patrick Hery, Marine Tavenier, Anne-Sophie Chaplault, Sylvain Thessier qui constituent « THE animal house team », de s'être occupés de mes souris H2AL2 avec soin et surtout de leur présence quotidienne et distrayante lors des coups durs,

Jean-Charles, merci de ton aide lors de mes moments désespérés... "j'ai oublié mes clefs de scooter dans le coffre de mon scooter fermé à clef" ou "je suis en rade devant le CEA sans mon badge", tu es toujours là! Merci pour ton aide spontanée et ta bonne humeur quand la situation semble désespérée...

Robin Piron pour son aide sur LINUX et pour le programme PIRP de comparaison des pics,

Laure Vescovo pour son aide, il y a quelques années, sur les programmes d'alignement de séquence (Mafft) et la phylogénie (MEGA),

Albane pour son aide indispensable pour les figures,

L'équipe du docteur Saadi Khochbin pour la collaboration sur H2AL2,

Le professeur Guiseppe Baldacci de m'avoir permis de faire cette thèse malgré sa désapprobation première, puis son enthousiasme au cours des années suivantes. Merci du

temps que vous m'avez consacré malgré votre emploi du temps chargé et de m'avoir fait confiance pour mener à bien ce projet,

Enfin, merci à mes professeurs de "sciences naturelles" comme il était l'usage de nommer, Madame Bourdillon (Lycée Sainte Marie de Blois), Madame Rougeon et Monsieur Anselme (Lycée Fénelon), Professeur Bardeletti (Université Lyon I) et Messieurs Dayanandan et Hickey (Concordia University) de m'avoir transmis votre passion pour la biologie, le goût du raisonnement scientifique et de votre soutien.

Merci aux personnes du SBIGEM, au 144 comme au 142, de m'avoir accueillie de 2005 à 2010! Merci de votre attention, de votre présence et de vos petits mots d'encouragements glissés au détour d'un couloir...

Aux habitués de la navette de 19h20 avec qui j'ai partagé tous les soirs l'espoir que la RATP n'ait pas impunément décidé de supprimer certains trains, nous permettant ainsi de profiter de la magnifique vue depuis le quai surplombant la N118 (surtout les jours de pluie).

Béa, nous avons bien animé la navette du matin de 7h50 à 8h35. Nos amis du E4 (le chapelier fou, la vieille peau, la flèche, la concierge, la folle, le couscous géant...) vont, je suis sûre, nous (+ nos potins) regretter amèrement...

Mes amis quant à eux sont toujours présents pour me changer les idées après un x^{ième} western blot raté...

Je souhaite ajouter une mention spéciale à mes colocataires successifs, pour leur soutien de tout temps.

Un dernier mot concernant tous ceux qui m'ont entourée lors de ces quatre mémorables années: mes parents, ma grand-mère (et son Larousse de 1928 avec lequel j'ai tenté de rédiger ma thèse), mes quatre frères et sœurs et Venceslas. Un grand merci à vous...

Je tenais à mentionner la phrase qui aura bercé ma thèse durant ces quatre ans dédiée à Arnaud Depaux, Arnaud Martel et Arnaud Krebs,

"Arnaud... tu peux venir s'il te plait? Ca n'a pas fonctionné...je ne comprends pas..."

Merci à tous les trois, car maintenant CA FONCTIONNE!!!

Au Docteur Gabriel Crélin (†), mon arrière grand-père, médecin au Mans

Qui rogat, non errat.

Préambule

Les cellules souches embryonnaires (ES) possèdent d'importantes propriétés d'auto-renouvellement (divisions infinies) et la capacité de se différencier (se transformer) en n'importe quel type cellulaire. Ces propriétés les rendent extrêmement importantes dans le domaine médical, par exemple pour le futur remplacement des cellules et des organes défectueux chez les patients.

D'autre part, ces caractéristiques sont partagées avec certaines cellules cancéreuses qui peuvent, elles aussi, se multiplier à l'infini et, dans certains cas, perdre leur caractère différencié. Ceci suggère qu'il puisse exister un lien mécanistique entre les deux phénotypes.

Ces caractéristiques sont gouvernées par un réseau complexe de protéines régulatrices (Introduction). Parmi ces protéines, les remodeleurs de chromatine sont capables de moduler l'accès à l'information génétique.

La famille des remodeleurs à chromodomaine (Chromodomain Helicase DNA binding protein) constitue une famille de 9 membres (Chd1, Chd2, Chd3, Chd4, Chd5, Chd6, Chd7, Chd8, Chd9) dont certains sont encore peu étudiés. Historiquement, notre laboratoire s'intéresse depuis une dizaine d'année à la protéine Chd4. Parmi les membres Chd, il a déjà été démontré que certaines protéines faisaient partie des réseaux transcriptionnels des cellules ES (Chd4, Chd9), d'autres sont impliquées dans des cancers (Chd5) ou dans des maladies génétiques (Chd7).

Mon projet de thèse a consisté à comprendre la contribution de cette famille de protéines aux réseaux transcriptionnels des cellules ES par une approche comparative (Résultats). Nous tenterons d'expliquer si ces neuf membres ont une action redondante au sein de la cellule ou si le rôle de chaque protéine est spécifique et non compensable (Discussion et Conclusion).

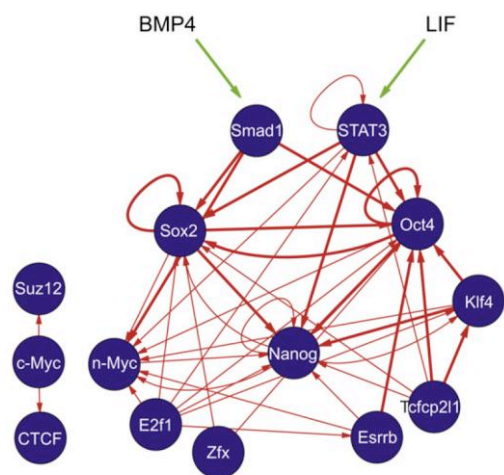


Figure M1: Chen 2008

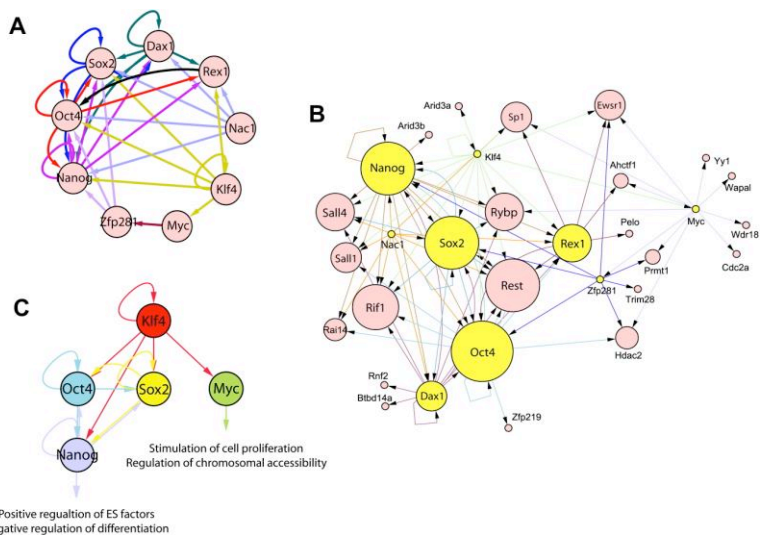


Figure M2: Kim 2008

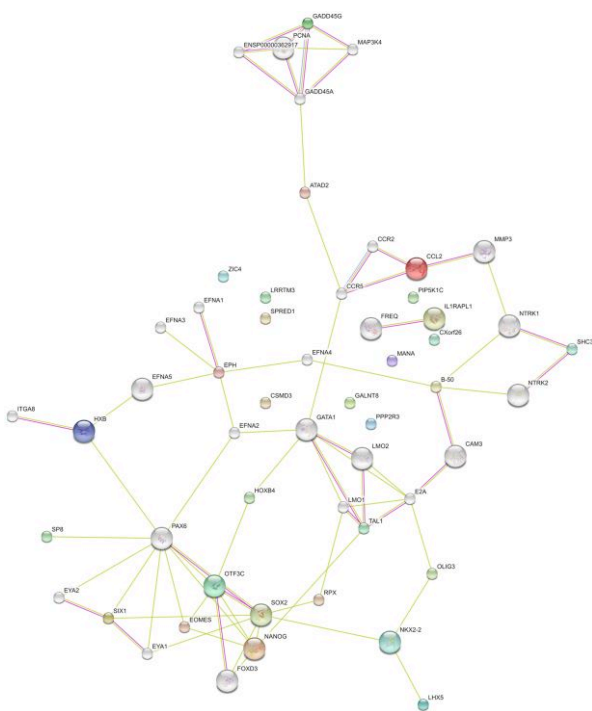


Figure M3: Jung 2010

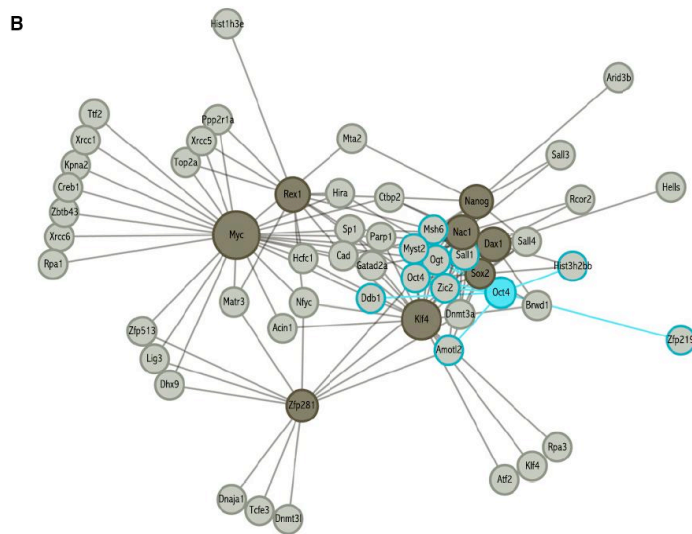


Figure M4: Pardo 2010

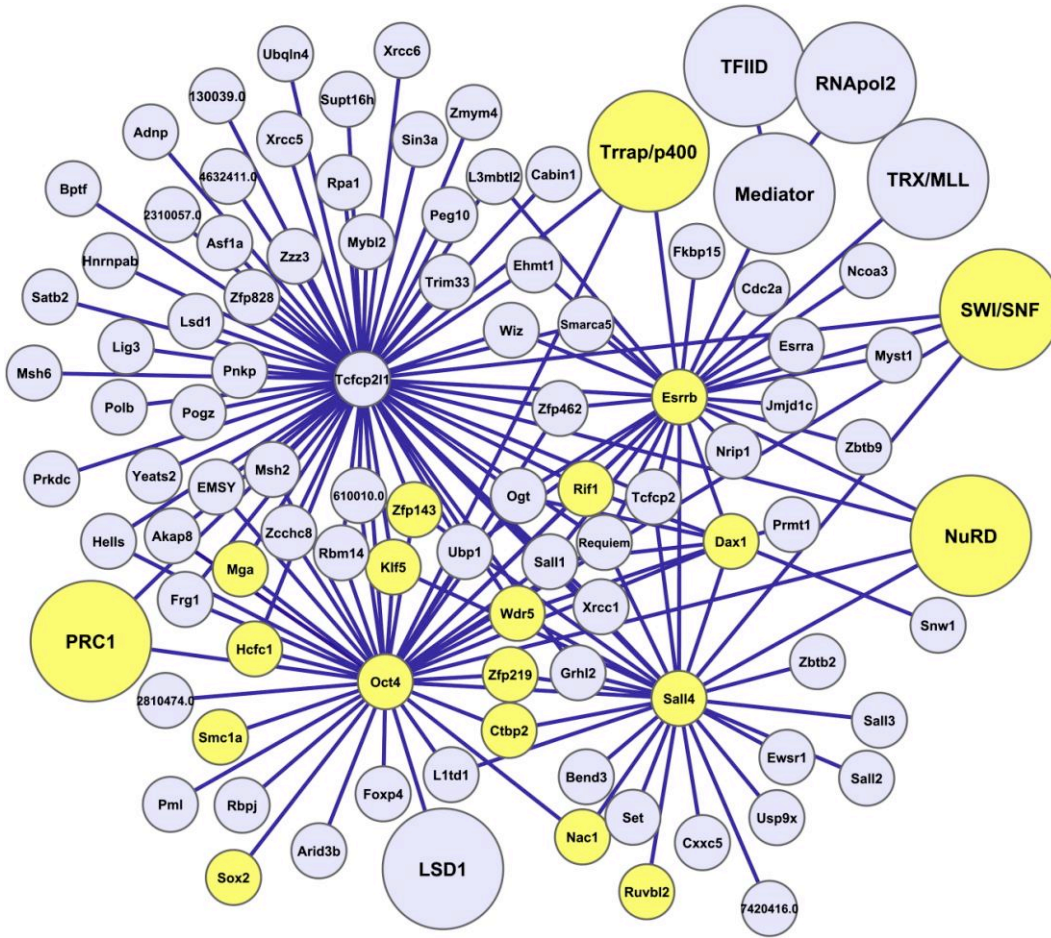


Figure M5: Van den Berg 2010

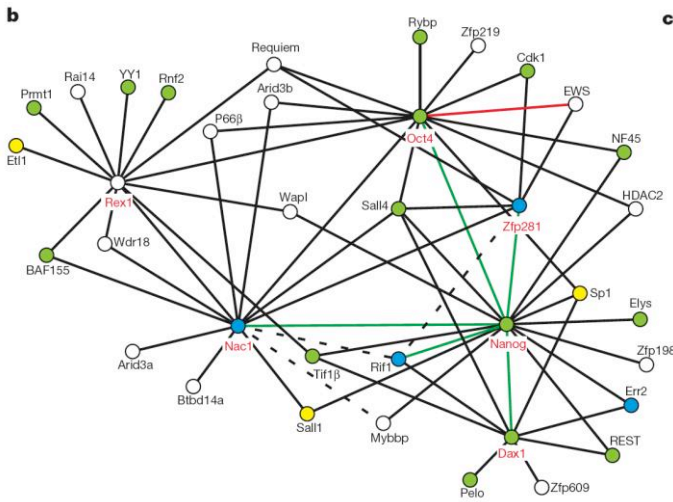


Figure M6: Wang 2006

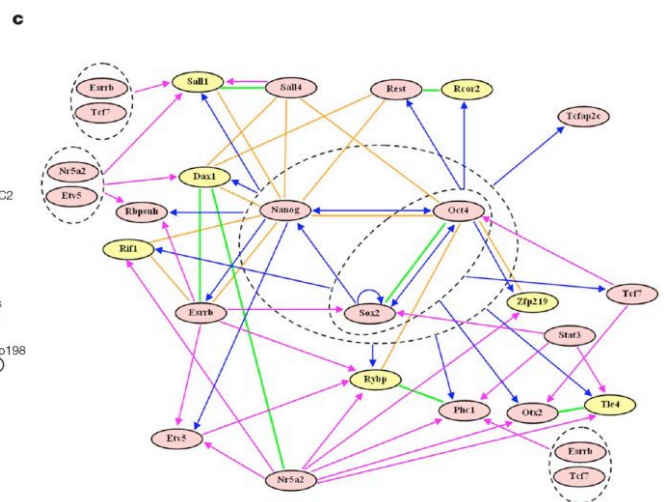


Figure M7: Zhou 2007

Table des Matières

I. Introduction.....	15
Chapitre I. Les caractères des cellules souches embryonnaires	17
A. Origine des cellules ES.....	17
1. Première caractéristique des cellules ES: elles donnent naissance à de multiples lignages	18
2. Deuxième caractéristique des cellules ES: l'auto-renouvellement	20
B. Différenciation des cellules ES	20
C. Divergences entre les cellules ES murines et humaines	21
D. La reprogrammation des cellules ES (iPS).....	21
Chapitre II. Les facteurs clefs responsables du phénotype des cellules ES	25
A. Identification des voies de signalisation	25
1. La voie LIF	25
2. La voie BMP.....	26
3. La voie Wnt.....	27
4. La voie PI3K.....	29
B. Les facteurs importants dans le maintien de la pluripotence des cellules ES	29
1. Oct4	29
2. Sox2.....	32
3. Le complexe Oct4/Sox2	32
4. Nanog	33
5. Le complexe Oct4/Sox2/Nanog.....	34
C. Les facteurs requis pour la pluripotence des cellules ES.....	36
Esrrb	37
Tbx3	37
Dax1	37
Myc:	38

Tcfcp2l1.....	38
Klf4 (EZF) et Klf5.....	39
FoxD3	40
Zfx.....	41
Rif1.....	41
Tcl1	41
Mbd3	41
Ronin.....	41
6. Autres facteurs	42
D. Les facteurs dispensables aux cellules ES mais intervenant dans le maintien de la pluripotence	42
Utf1	42
Eras.....	42
Sall4	42
Usp9x.....	43
Jarid2.	43
E. Identification de nouveaux facteurs protéiques impliqués dans le renouvellement des cellules ES	44
1. Autopsie des facteurs clefs du maintien de la pluripotence	44
2. Un crible ARN interférence pour identifier un nouveau complexe impliqué dans la pluripotence des cellules ES.....	45
3. Un rôle antagoniste pour deux protéines de la même famille	46
F. Identifications de réseaux de régulation en cellules ES	47
1. Un large réseau pour la pluripotence des cellules ES	47
2. Intégration des voies de signalisation LIF et des protéines polycomb aux réseaux de la pluripotence.....	49
Conclusion	51
3. Etude extensive des partenaires d'Oct4	51
4. Des réseaux de régulation à la rescousse de la pluripotence	54

5.	Conclusion	54
G.	Autres mécanismes importants dans la pluripotence des cellules ES.....	55
1.	Rôle des micro ARN dans la pluripotence des cellules ES	55
2.	La méthylation de l'ADN est un lien entre épigénétique et pluripotence....	56
Chapitre III. Pluripotence et épigénétique		59
A.	Présentation des protéines Polycomb et Titorax.....	59
1.	Mécanisme de répression par les protéines Polycomb.....	61
2.	Mécanisme de fonctionnement des protéines Titorax.....	63
B.	Les protéines PcG et TrxG et la pluripotence des cellules ES	63
1.	Les protéines Polycomb dans la pluripotence	63
2.	Différence entre PRC1 et PRC2.....	66
3.	H2AZ.Z et PRC2	66
4.	Le rôle de Jarid2 et du complexe PCR2 dans les cellules ES.....	67
C.	Mécanismes additionnels de répression des gènes du développement ...	68
Chapitre IV. Remodelage de la chromatine et pluripotence.....		73
A.	Les mécanismes du remodelage de la chromatine	76
B.	Les complexes de la super-famille SNF2.....	78
1.	Les protéines de la famille SWI/SNF	79
2.	Les complexes de la famille ISWI (Imitation SWItch).....	81
3.	Les complexes de la famille INO80 (INOsitol 80)	83
4.	Les complexes des familles Rad26 et RAD54	85
5.	La famille Chd	86
II.	Résultats	93
Chapitre 1. Génération de lignées ES étiquetées pour les facteurs d'intérêts		95
A.	Choix d'une étiquette	96
1.	Choix des épitopes.....	96
2.	Choix des étiquettes et mise au point expérimentale.....	96
B.	Adaptation du protocole de « recombineering » pour l'étiquetage des protéines	97

1. Introduction du BAC dans les bactéries permettant la recombinaison homologue 99	
2. Première recombinaison homologue : incorporation des 10kb d'ADN génomique qui permettront de cibler la recombinaison homologue.....99	99
Construction des plasmides de recombinaison homologue pL25399	99
Première recombinaison homologue.....100	100
3. Deuxième recombinaison homologue : incorporation de l'étiquette en fusion avec l'extrémité C terminale de la protéine.....100	100
4. Vérification des clones100	100
Changement d'étiquette101	101
5. Obtention de lignées ES stables étiquetées101	101
Choix de la lignée ES pour l'électroporation.....101	101
Electroporation des plasmides de recombinaison homologues102	102
Chapitre 2. Mise au point d'un protocole performant pour l'immunoprécipitation efficace des protéines étiquetées107	107
A. Test de différents protocoles standard.....107	107
1. Présentation: Les protocoles testés.....107	107
Choix des agents pontants.....110	110
Procédure des tests110	110
Description des différents protocoles111	111
2. Protocole A (extraction des noyaux).....111	111
Principe.....111	111
Bilan.....111	111
3. Protocole B (SDS).....111	111
Principe.....111	111
Bilan.....112	112
4. Protocole C (FA/SDS)112	112
Principe.....112	112
Bilan.....112	112

5.	Protocole D (FA/SDS amélioré).....	113
	Principe.....	113
	Bilan.....	113
	Conclusion du protocole D.....	114
	Améliorations	114
	Insuffisances du protocole D	114
6.	Protocole E (Natif).....	115
	Principe.....	115
	Bilan.....	115
B.	Développement d'un protocole d'immunoprecipitation de chromatine innovant (Protocole F).....	116
1.	Choix du pontage.....	116
2.	Choix des détergents et des sels	117
3.	Choix de la méthode de fractionnement de l'ADN	118
4.	Choix de la méthode d'immunoprécipitation	118
	Le tampon d'immunoprécipitation.....	118
	Les lavages.....	118
	Les éluions.....	119
	La préparation de l'ADN.....	119
C.	Publication	119
	Chapitre 3. Analyse des données CHIP-seq.....	125
A.	Le séquençage	125
B.	Le traitement informatique des données	126
1.	Algorithmes de détection des pics	126
	Quest (Valouev 2008).....	127
	Avantages et inconvénients de Quest	127
	SISSR (Jothi 2008)	127
	Sicer (Zang 2009)	127

Détection des pics avec les différents algorithmes	128
2. Programmes d'analyse utilisés	129
HeatMaps	130
WigMaker3	130
DistributionV5.py	130
Cluster 3.0.....	131
Java Treeview	131
seqMINER	131
GPAT	131
Le programme Pirp	131
C. Etude du complexe NuRD dans les cellules ES	132
D. Etudes des profils de position des protéines Chd	133
1. Présence des Chd aux promoteurs	134
2. Co-localisation avec les marques H3K4me2 et H3K4me3.....	136
3. Autorégulation des Chd.....	137
Chapitre 4. Comparaison des données de liaisons avec des données transcriptomiques afin de proposer des hypothèses fonctionnelles	139
A. Description de la méthode des shRNA	141
B. Construction des vecteurs de shRNA	143
1. Le vecteur pHYPER	143
2. Constructions des shRNA	143
3. Test des shRNA.....	143
4. Etude de la perte de fonction de chaque protéine Chd dans les cellules ES 144	
Etude par shRNA.....	144
RNA-seq comme alternative aux puces	146
Confrontation des données de position et des données de pertes de fonction du complexe NuRD	148
C. Article en préparation	150

Chapitre 5. Projets annexes.....	153
A. Constructions de lignées ES possédant une mutation au niveau du site catalytique pour chaque protéine Chd.....	153
1. Principe de la méthode de construction des lignées possédant la mutation du site catalytique.....	154
Identification du site catalytique	154
Construction des vecteurs de recombinaison homologue	157
Première recombinaison homologue	159
i. Construction du vecteur pL253	159
ii. Recombinaison homologue	159
Deuxième recombinaison homologue.....	159
i. Construction du vecteur pL451AS	159
ii. Le vecteur comportant les bras d'homologie: pL451AS+BH	161
iii. Deuxième recombinaison homologue.....	162
2. Génération de lignées ES mutantes.....	162
Recombinaison dans les cellules ES.....	162
Retrait de la cassette de sélection de résistance à la néomycine.....	165
3. Projets futurs concernant les lignées KR.....	166
B. Etude d'un variant d'histone impliqué dans la spermatogenèse de <i>Mus musculus</i> 167	
1. Etude des variants d'histones chez la souris (travail initié en 2006)	167
2. La famille H2a1/2	168
3. Construction d'une lignée de souris possédant un allèle étiqueté pour le gène H2a2 (NM_026627).....	169
Visualisation de la protéine étiquetée.....	171
Test de fertilité des souris portant l'allèle étiqueté H2AL2-6HFH.....	172
Dans la lignée H2a2 Néo+	172
Dans la lignée H2a2 Δ Néo	172
III. Discussion et conclusion	173
Etude des cibles des protéines chromatiniennes se liant à l'ADN.....	175

Etude des gènes-cibles des protéines chromatiennes.....	176
Etude de la fonctionnalité d'un domaine spécifique.....	177
Un trio infernal	179
Cartographie du/des réseaux.....	179
Lignées de choix.....	180
Conclusion	185
IV. Bibliographie	187

Table des illustrations

Figure 1: Une colonie de cellules souches sur des fibroblastes (cellules nourricières) cultivées en présence de LIF et de sérum de veau fœtal. Grossissement X100	18
Figure 2: Les cellules souches tiré de www.stemcellresearchfoundation.org	19
Figure 3: Reprogrammation des cellules différenciées. (A) provenance originelle des cellules ES (B) reprogrammation de cellules différenciées à l'aide de facteurs clefs essentiels aux cellules ES. Tiré de Kashyap et al (Kashyap et al., 2009)	23
Figure 4: Différentes voies d'action du LIF, tiré de Niwa et al., 2009.	26
Figure 5: KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway de la voie Wnt. La protéine Wnt transmet le signal effecteur via des récepteurs transmembranaires. Puis une cascade d'activation fait intervenir la betacatenine qui migre dans le noyau de la cellule et permet d'activer un certains facteurs de transcription et régulateurs du cycle cellulaire.	28
Figure 6: Modèle d'activation de la voie Wnt présentant l'état inactif de la voie (A) et actif (B) tiré de MacDonald BT et al. 2009. La liaison de la protéine Wnt à ces récepteurs permet le recrutement de l'Axin libérant ainsi la betacatenine qui migre dans le noyau et active des facteurs de transcription.	28
Figure 7: Effet de la dérégulation positive ou négative d'Oct4 dans les cellules ES, tiré de Niwa 2000.	30
Figure 8: Régulation des gènes Tcf1 et Nanog par Oct4 via les histones demethylases Jmjd1a et jmjd2c, tiré de Loh 2007.	31
Figure 9: Prédiction de motifs de liaison à l'ADN de Sox2-Oct4 (a) et Nanog (b) par CHIP-PET tiré de Loh et al., 2006.	33
Figure 10: Boucle de régulation d'Oct4, Nanog et Sox2. Adapté de Marson et al. 2008	35
Figure 11 Réseau de régulation de transcription d'Oct4 et Nanog dans les cellules ES tiré de Loh et al. 2006	36
Figure 12: Réseau de régulation de l'expression du gène Nanog mis en évidence. Tiré de Savarese et al., 2009.....	47
Figure 13: Facteur de co-occupation des différents facteurs impliqués dans la pluripotence, tiré de Kim et al. 2008	49
Figure 14 Représentation de la co-occupation des facteurs importants dans la pluripotence sur le génome murin tiré de Chen et al. 2008.....	51
Figure 15: Différents mécanismes de répression de la transcription par les protéines Polycomb tiré de Sparmann et al. 2006.....	61

Figure 16: Schéma d'un gène bivalent différentiellement exprimé selon certaines voies de différenciation.....	64
Figure 17: Liaisons de SetDB1 aux histones modifiées, tiré de Bilodeau et al. 2009.	70
Figure 18: Représentation schématique des macrodomaines des cinq familles composant la superfamille SNF2.....	74
Figure 19: Conservation du domaine ATPase/helicase au sein de la famille SNF2 autour des positions 40 à 52.....	75
Figure 20: Arbre phylogénétique regroupant l'ensemble des protéines identifiées par BLAST possédant le domaine ATPase sur le génome muri. Méthode utilisée : minimum linkage (Mafft).....	76
Figure 21: Différents modèles de remodelage de l'ADN à l'aide de complexes multi-protéiques pouvant utiliser l'énergie de l'ATP, tiré de Clapier et al. 2009.....	78
Figure 22: Le complexe esBAF dans les cellules ES murines, tiré de de Ho et al, 2010.....	79
Figure 23: Le complexe BAF dans des contextes cellulaires différents, tiré de Ho et al., 2010.	80
Figure 24 Représentation des domaines principaux des 9 membres de la famille Chd	86
Figure 25: Hypothèse de Goldberg et al. proposant Chd1 comme chaperone d'histone pour le variant H3.3 pour les gènes actifs, HIRA comme chaperone d'histone pour H3.3 pour les gènes inactifs, enfin Atrx déposerait H3.3 aux télomères.....	87
Figure 26: Le complexe NuRD, comprenant la sous unité Mi2 β (Chd4) et les autres protéines associées dont le dimère Hdac1 et Hdac2.	89
Figure 27: Expression des Chd dans les différents organes de souris adulte d'après la base de données Unigene. Les valeurs en ordonnées donnent le nombre de transcrits par million. (R.D.: Racine dorsale).....	91
Figure 28: Résumé schématique de la méthode de génération des constructions pour la recombinaison homologue dans les cellules ES	98
Figure 29: Stratégie de génotypage pour le gène <i>Chd8</i> .L'étiquette est indiquée par un bloc rouge	103
Figure 30: Exemple d'obtention de clones recombinants pour la construction Chd8-3FTH. Les clones positifs sont les numéros 3, 6 et 11.....	103
Figure 31: Génotypage des clones électroporés avec les plasmides Cre et PuroR. Les clones positifs possédant l'allèle sauvage et l'allèle étiqueté sans néo sont 2, 3, 8, 10, 13, 14, 16.	104
Figure 32: Expression des protéines Chd étiquetées dans chaque lignée spécifique. L'étiquette utilisée est 3FTH à l'exception de la lignée Chd9 (étiquette 6HFH). Les quantités	

d'extrait cellulaire total déposées sont 40ug pour Chd3, Chd6 et Chd9 et 20ug pour Chd1, Chd2, Chd4 et Chd8. La même membrane de western blot a été révélée avec un anticorps anti-HA (Partie A) et un anticorps anti-FLAG (Partie B).	105
Figure 33: Amélioration de l'efficacité du pontage par d'autres agents que le formaldéhyde	110
Figure 34: Formule chimique des quatre agents pontants utilisés: DTBP, EGS, DSG et formaldéhyde.....	117
Figure 35: Immunoprécipitation de chromatine de la protéine Chd8 étiquetée 3FTH. A: pontage formaldéhyde seul; B: pontage DSG+formaldéhyde. Unb: fraction non liée avec une IP; El: elution; pep: peptide; Tandem: fraction immunoprécipitée par les anticorps anti-HA et anti-FLAG	121
Figure 36: Effet de la digestion MNase sur une région liée par la protéine Chd8. Les lacunes induites par la digestion MNase sont indiquées par des flèches.....	126
Figure 37: Observation des reads séquencés de Chd8 et du contrôle 46C et localisation des régions détectées enrichies par différents programmes. Aucun pic n'a été détecté par Quest, 3 pics larges par SICER (rouge) et 4 pics par SISSR (violet foncé).....	128
Figure 38: (A) Abondance du nombre de pics autour du TSS (en abscisse) détectés par Quest, SISSR et SICER pour la protéine Chd8 sur ses propres gènes cibles. A partir de liste de pic Quest, SISSR et SICER, le centre de chaque pic a été déterminé. Puis à l'aide du logiciel GPAT le TSS le plus proche a été défini. Enfin le nombre de "centre de pic" a été calculé dans un intervalle [-5000; +5000] avec une fenêtre de 100 paires de bases. SICER permet de détecter un plus grand nombre de pics, notamment les pics qui sont plus étalés et moins centrés sur le TSS comme Quest et SISSR. (B) Nombre de gènes communs ou non, identifiés par les algorithmes Quest, SISSR et SICER.....	129
Figure 39: Schématisation de l'algorithme Pirp. A partir de deux listes de pics (Partie A), ce script permet d'identifier si les pics sont recouvrants selon les quatre possibilités (Partie B).....	132
Figure 40: Corrélation des enrichissements aux promoteurs. Chaque point d'une couleur représente un promoteur, il y a au total ~24000 point de chaque couleur. Les valeurs d'enrichissement pour Chd1 sont en abscisse. En ordonnées sont reporté les données pour Chd8 (noir), Chd4 (bleu) et le contrôle la lignée 46C non étiquetée (jaune). En rouge la droite de corrélation $x = y$	134
Figure 41: Abondance de l'ARN polymérase II (Pol2), Chd1 (formaldéhyde), Chd8 (formaldéhyde), Chd4 (formaldéhyde), Chd4 (DSG), Hdac2 (DSG), 46C - lignée contrôle - (formaldéhyde), H3K4me3 et H3K36me3 (Mikkelsen et al., 2007).....	135
Figure 42: Nombres de gènes liés au promoteur par Chd1, Chd4 et Chd8 communs et spécifiques à chacune.....	136

Figure 43: Occupation de Chd1 (A) et Chd8 (B) sur leurs propres cibles. Deux profils d'occupation de Chd1 et Chd8 sont représentés. La position par rapport au centre du pic détecté par SICER est donné en abscisse. En ordonné le nombre de tag séquencé dans une fenêtre de 25 paires de bases.	137
Figure 44: Autorégulation des Chd d'après les données de séquençage obtenues avec la fixation type formaldéhyde. Aucune présence de Chd n'a été détecté sur les gènes Chd7 et Chd8 lors des immunoprécipitations réalisées avec le formaldéhyde.	138
Figure 45: Cartographie du vecteur pHyper	140
Figure 46: Extinction de l'expression d'un gène ciblé par la méthode du shRNA. Le petit ARN en épingle à cheveux est synthétisé par la polymérase III (promoteur H1). L'ARNm est ensuite mûré par la protéine Drosha et exportée vers le cytoplasme grâce à la protéine Exportine 5. L'ARN est ensuite pris en charge par DICER qui clive la boucle reliant la séquence sens et antisens, la dégradation du brin antisens est ensuite prise en charge par le complexe RISC. L'hybridation de l'ARN sur sa cible conduit à la dégradation du grand ARNm vers les P bodies et l'inhibition de la traduction.....	142
Figure 47: Extinction de l'expression de la protéine Chd1 après l'électroporation du vecteur pHyper exprimant le shRNA correspondant. Chd1-3FTH est détecté par western blot et réalisé en triplicat. Le contrôle est le vecteur pHYPER vide électroporé. La partie supérieure du western blot a été révélée par l'anticorps anti-FLAG et la partie inférieure avec l'anti-GAPDH.	144
Figure 48: Description de l'utilisation des shRNA pour déterminer les l'ensemble des gènes dérégulés lors du knock-down de la protéine d'intérêt.....	145
Figure 49: Schéma récapitulatif de la technique de RNA-seq. Les cellules sont cultivées et l'ARN est récupéré, les ARNm sont isolés par complémentarité des chaînes poly A avec des billes oligo dT. Les fragments sont amplifiés et séquencés sur puce Illumina. .	147
Figure 50: Stratégie de construction pour générer une lignée ES possédant une mutation sur son site catalytique: (A) la cassette de selection se trouve en amont de la lysine d'intéret (B) la cassette se trouve en aval de la lysine d'intéret.	155
Figure 51: Introduction de la mutation au niveau du site catalytique. (A) Méthode de génération du bras d'homologie comportant la mutation.....	156
Figure 52: Ensemble des étapes nécessaires à la construction d'une lignée ES possédant une mutation au niveau de la lysine du site catalytique. L'étoile indique le remplacement de la lysine du site catalytique en arginine.	158
Figure 53: Fragment sous-cloné pour la modification du pL451 entre les sites BstB1 et BamH1, contenant le site de recombinaison FRT et les sites de restriction Ascl et Sbf1.	160
Figure 54: Cartographie du plasmide pL451	160

- Figure 55: Génotypage par digestion EcoR1 des clones de la lignée Hdac2-6HFH électroporés avec le plasmide portant la mutation du site catalytique pour le gène Chd4. Les clones ayant correctement recombinés sont les numéros 30, 31, 32, 35 et 36. 162
- Figure 56: Deux possibilités de recombinaison. (1) recombinaison qui n'apporte pas la mutation (2) Recombinaison qui apporte la mutation 163
- Figure 57: Génotypage des cellules ES Hdac2-6HFH électroporés avec la construction Chd4 KR afin de vérifier l'insertion correcte de la construction Chd4KR. A: Description des allèles possibles après électroporation. B: Analyse par Southern blot à l'aide d'une sonde ciblant la cassette de résistance à la néomycine. Le clone 30 montre des tailles aberrantes pour les digestions enzymatiques BglII et BamH1 (identifiées par des étoiles rouges), ce dernier a donc été éliminé..... 164
- Figure 58: Stratégie de génotypage pour distinguer l'allèle Chd4 muté non étiqueté et l'allèle Chd4 muté possédant l'étiquette en C-ter. A: L'allèle qui a incorporé la mutation KR possède également l'étiquette en C-term. B: la recombinaison et l'introduction de la mutation KR a eu lieu sur l'allèle non étiqueté..... 165
- Figure 59: Phylogénie des variants de l'histone H2A. L'histone conventionnelle de référence est histone 1 H2ac NM_178189..... 168
- Figure 60: Phylogénie des différents variant de la famille H2AI1/2. Ce variant est absent chez le tatou (Armadillo), le Galago (bushbaby), le Xenope ou la Chauve-souris (microbat) 169
- Figure 61: Stratégie de construction de la lignée étiquetée pour H2AL2. L'allèle + correspond à l'allèle sauvage. 170
- Figure 62: Expression du variant H2a2 étiqueté dans les spermatides condensées de souris contrôlée sur gel d'acrylamide 18%. Le changement de taille de la protéine correspond à la taille de l'étiquette. 43000 spermatozoïdes ont été déposés sur la piste H2a2 étiquetée, 70000 sur la piste contrôle. La même membrane a été révélée premièrement par un anticorps spécifique (Ref:SAB771; gentiment fourni par l'équipe de Saadi Khochbin à Grenoble) à H2a2 (Partie A) puis l'anticorps anti-HA a été ajouté (Partie B). 171
- Figure 63: Réseau d'occupation des Chd sur leurs promoteurs dans les cellules ES établi à l'aide des données de séquençage Chd1 F, Chd4 F, Chd4 DSG, Chd8 F Chd1 DSG, Chd2 F, Chd2 DSG, Chd3 F, Chd5 F, Chd6 F, Chd6 DSG, Chd8 DSG, Chd9 F, Chd9 DSG. Cette figure est à compléter avec les résultats de séquençage suivants: Chd3 DSG et Chd5 DSG. Les interactions visualisées avec une seule expérience Chip-seq sont indiquées en pointillées et avec deux expériences (F et DSG) en trait plein. 183
- Figure 64: Exemple de résultat de co-occupation des protéines chromatinienne obtenu pour l'analyse de nos propres données et les données publiées grâce au programme Pirp, Cluster et Java Treeview. Ce sont des données préliminaires car les méthodes de

calcul des corrélations et du clustering doivent être améliorées. Cependant, les données les plus robustes sont indiquées en bleu, et un large réseau de protéines partageant les mêmes cibles cellules ES est indiqué en jaune. Cette figure est à compléter avec nos propres données, Chd7, le facteur Esrrb, et des protéines polycomb (Suz12, Ezh2, Ring1B).....184

Index des Tableaux

Tableau 1: Différentes catégories de cellules souches présentant des propriétés différentes	20
Tableau 2: Table non exhaustive des articles majeurs traitant des réseaux importants dans le maintien du phénotype des cellules ES.....	53
Tableau 3: Complexes protéiques Polycomb et Tritorax	60
Tableau 4: Tableau récapitulatif des différents protocoles testés	109

I. Introduction

Chapitre I. Les caractères des cellules souches embryonnaires

Le premier évènement de différenciation dans l'embryon murin est la partition des cellules en deux entités, l'un embryonnaire et l'autre extra embryonnaire (Gardner, 1983). Le composant embryonnaire à l'intérieur de l'œuf est appelé la masse cellulaire interne (Inner cell mass, ICM). C'est la source de tous les tissus du futur embryon, fœtus et au final de l'organisme adulte. Les cellules de l'ICM peuvent être aussi la source des cellules souches embryonnaire ou cellules ES (Brook and Gardner, 1997; Evans and Kaufman, 1981; Martin, 1981). Si les cellules ES dérivent des ICM, ces deux entités ne sont pas complètement identiques. En effet, la propriété des ICM est une durée de vie transitoire alors que les cellules ES peuvent être maintenues en culture de façon quasi illimitée.

A. Origine des cellules ES

L'intérêt de la communauté scientifique pour les cellules ES a commencé par l'identification de cellules cancéreuses, les tératomes. Ces cellules issues de tumeurs cancéreuses (EC) ont attiré l'attention des scientifiques par leurs caractéristiques. Premièrement, elles pouvaient être transplantées d'un individu à un autre. Deuxièmement, elles pouvaient être cultivées *in vitro* et donner naissance à des lignées cellulaires. Enfin, lorsqu'elles étaient introduites dans des blastocytes, elles pouvaient être incorporées dans l'embryon et contribuer aux tissus du fœtus en développement (Brinster, 1974). L'amélioration des conditions de culture *in vitro* a permis à Martin et Evan en 1981 de générer des cellules directement depuis l'embryon sans passer par des tératomes. Possédant les caractéristiques identiques aux EC, ces cellules permettaient la transmission à la lignée germinale, leur statut pluripotent peut être maintenu dans des conditions de culture adéquates et elles peuvent être différenciées en réponse à des stimuli extérieurs (Bradley et al., 1984). Ces cellules sont maintenant connues sous le nom de cellules souches embryonnaires ou cellule ES (Figure 1).

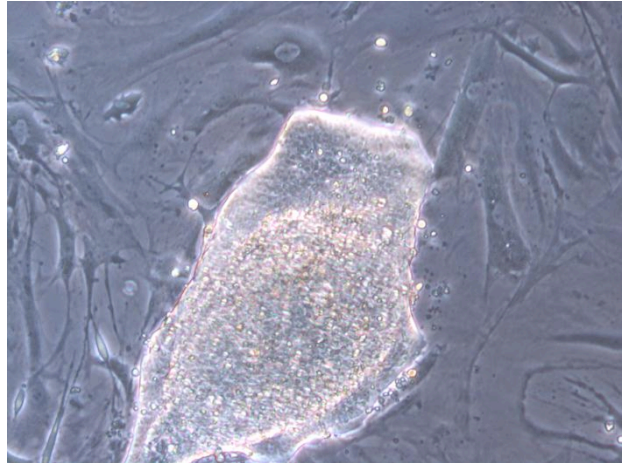


Figure 1: Une colonie de cellules souches sur des fibroblastes (cellules nourricières) cultivées en présence de LIF et de sérum de veau fœtal. Grossissement X100

1. Première caractéristique des cellules ES: elles donnent naissance à de multiples lignages

La valeur des cellules ES tient à leur capacité à contribuer aux trois feuillets embryonnaires: l'endoderme, le mésoderme et l'ectoderme. Ceci peut être démontré in vitro en réponse à des stimuli spécifiques ou bien, de façon plus frappante, lorsque les cellules ES sont introduites dans des blastocystes. Dans ce cas, les nouvelles cellules ES se mélangent aux cellules ES déjà mises en place et colonisent les tissus pour produire un animal chimérique où cellules hôtes et cellules injectées sont intimement associées dans l'organisme (Beddington and Robertson, 1989). Lorsque les cellules ES colonisent la lignée germinale, elles peuvent donc transmettre leurs caractéristiques à leur descendance (Bradley et al., 1984). Cette propriété a été intensément exploitée pour l'analyse de la fonction des gènes in vivo (van der Weyden et al., 2002).

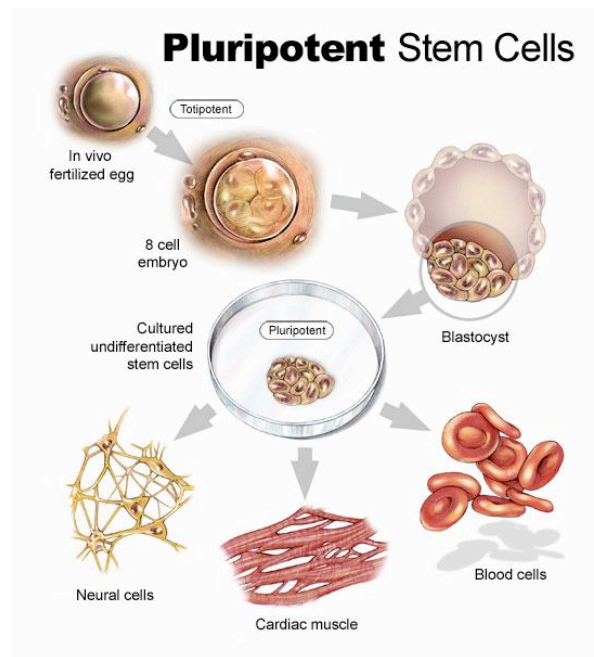


Figure 2: Les cellules souches tiré de www.stemcellresearchfundation.org

La capacité des cellules ES à se différencier pour donner différents tissus embryonnaires est une de leurs caractéristiques et un attribut qu'il est convenu d'appeler "pluripotence" (Figure 2). Les cellules ES ne sont pas totipotentes car naturellement elles ne donnent pas naissance au trophoblaste. Les cellules souches hématopoïétiques ne sont pas pluripotentes car elles ne peuvent donner que certains types cellulaires et sont donc déjà spécifiques d'un certain type de lignage. Afin d'illustrer les différentes catégories de cellules souches dérivant de l'embryon ou de sources diverses chez l'animal post-natal, une classification peut être établie (Tableau 1: Différentes catégories de cellules souches présentant des propriétés différentes).

définitions		
Potentiel	Définition	Exemple
Totipotence	Capacité à donner tous les lignages de l'organisme	Chez les Mammifères seulement le zygote et les cellules issues des premiers clivages sont totipotents
Pluripotence	Capacité à former tous les lignages du corps	Cellules souches embryonnaires
Multipotence	Capacité des cellules souches adultes à donner de multiples types cellulaires d'un même lignage	Cellules hématopoïétiques
Unipotence	Capacité d'une cellule à donner un seul autre type cellulaire	cellule spermatogoniale peut uniquement donner un spermatozoïde
Reprogrammation	Augmentation dans l'échelle de la pluripotence, dédifférenciation	Peut être induit par transfert de noyau, fusion cellulaire ou des manipulations génétiques
Transdifférenciation, plasticité	Notion concernant les cellules souches somatiques qui peuvent élargir leur potentiel pour générer d'autres lignages	C'est un concept controversé chez les Mammifères

Tiré de Jaenisch et Young, 2008

Tableau 1: Différentes catégories de cellules souches présentant des propriétés différentes.

2. Deuxième caractéristique des cellules ES: l'auto-renouvellement

La deuxième caractéristique des cellules ES est la capacité à subir de multiples divisions cellulaires ou divisions infinies sans différenciation et donc de produire des cellules pluripotentes identiques à travers les générations. Cette propriété est appelée auto-renouvellement. Elle est présente grâce à une division cellulaire symétrique. Les cellules ES peuvent s'auto-renouveler continuellement si elles sont dans des conditions de cultures qui préviennent leur différenciation: par exemple, la présence de LIF (Leukemia Inhibition Factor), un facteur de croissance qui est nécessaire pour la prolifération et le maintien des cellules ES à un state indifférencié (Suda, 1987).

B. Différenciation des cellules ES

Bien que les cellules ES soient pluripotentes, elles ne peuvent donner naissance qu'en premier lieu à quelques type cellulaires comme l'endoderme primitif et l'ectoderme primitif car elles possèdent les caractéristiques identiques aux ICM. La différenciation des cellules ES de souris peut être induit par l'expression ectopique de certains facteurs. Par

exemple, l'expression du facteur de transcription Gata6 dans les cellules ES induit leur différenciation en endoderme primitif (Fujikura et al., 2002). De manière similaire, l'expression du facteur de transcription à homéodomaine de type caudal Cdx2 induit une différenciation des cellules ES en cellules du trophoctoderme (Niwa et al., 2005). C'est pourquoi ces deux facteurs, parmi d'autres pouvant induire la différenciation, doivent être fortement réprimés pour maintenir le phénotype indifférencié des cellules ES.

C. Divergences entre les cellules ES murines et humaines

Il a été clairement établi que les facteurs de pluripotence Oct4 et Nanog sont absolument requis pour le maintien des cellules ES chez la souris comme chez l'homme. Cependant, certains mécanismes comme les voies externes LIF, BMP et Wnt ne semblent ne pas être requis chez l'homme (Rao, 2004). Loh et al. ont montré que les cibles des protéines Oct4 et Nanog pouvaient différer de manière importante. En effet, 10% de cibles communes sont retrouvées entre l'homme et la souris. De même Kim et al. ont montré que les données pouvaient largement évoluer selon la méthode utilisée (Chip-on-CHIP, Chip-PET, CHIP-seq) ainsi que la profondeur de séquençage (Euskirchen et al., 2007; Kim et al., 2008a). Ainsi le biais de ces données pourrait expliquer un certain nombre de différences observées.

Il serait dès lors intéressant de déterminer si la différence observée dans les réseaux de régulation murins et humains provient effectivement des différences des deux espèces. En effet d'autres hypothèses peuvent être proposées. Premièrement, la manière d'isoler et de cultiver les cellules peut influencer les lignées ES murines ou humaines. Deuxièmement, il serait intéressant de connaître la variation d'une lignée ES à une autre au sein de la même espèce.

Cependant, les cellules humaines peuvent être reprogrammées avec les mêmes facteurs que les cellules murines (Park et al., 2008; Takahashi et al., 2007; Yu et al., 2007). Cette observation apporte la preuve de réseaux maintenant la pluripotence, communs entre les deux espèces.

D. La reprogrammation des cellules ES (iPS)

Il existe un débat éthique au sein de la communauté scientifique autour de l'utilisation des cellules souches embryonnaires. En effet, ces cellules sont issues d'un embryon et la

question du respect de la vie de cet embryon est alors posée. Cependant les possibilités qu'offrent les cellules ES sont multiples. Les cellules ES ont la capacité unique de se diviser indéfiniment et de pouvoir se différencier en de multiples types cellulaires. Elles apparaissent donc comme très prometteuses comme agents thérapeutiques dans les traitements médicaux du futur.

Plusieurs équipes se sont alors interrogées si la provenance de ces cellules ne pouvait pas être modifiée. En effet, les seules cellules ES humaines autorisées à ce jour sont les cellules de sang de cordon ombilical et qui doivent donc être prélevées à la naissance. Pouvoir obtenir des cellules ES capables de donner le tissu désiré pour un individu malade représente un challenge majeur de la médecine actuelle. Il a été récemment montré qu'il était possible de reprogrammer des cellules différenciées à l'aide des facteurs importants dans la pluripotence des cellules ES (Figure 3). Ainsi, avec l'expression forcée des facteurs Oct4, Sox2, Klf4 et c-Myc avec l'aide de rétrovirus, des fibroblastes peuvent être reprogrammés en cellules souches possédant le caractère pluripotent des cellules ES (Okita et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007). D'après Germain, ces cellules provenant de cellules somatiques sont éthiquement irréprochables (Germain, 2009).

Cette combinaison de facteurs a émergé depuis un crible initial de 24 gènes candidats co-électroporés dans des fibroblastes de souris (Takahashi and Yamanaka, 2006). Parmi eux, Myc peut induire des changements épigénétiques qui sont utiles lors de la reprogrammation (Yamanaka et al., 2008). Il a été reporté que c-Myc induisait une vague d'acétylation (Fernandez et al., 2003) qui pourrait permettre l'accessibilité aux promoteurs des différents complexes transcriptionnels importants dans les lignées ES (Takahashi and Yamanaka, 2006). En effet, via son interaction avec Gcn5, une HAT (Histone Acetyl Transferase), Myc promeut la prolifération cellulaire (Knoepfler et al., 2006). A de nombreuses reprises, l'importance des facteurs Oct4 et Sox2 a été développée dans la littérature et sera plus détaillée au cours du chapitre suivant (Loh et al., 2006; Looijenga et al., 2003; Niwa et al., 2000; Okita et al., 2007; Rodda et al., 2005). Le facteur Klf4 inhibe l'expression de p53 et contribue donc à la survie des cellules ES. Ce facteur contribue aussi à l'expression de Nanog et d'autres facteurs importants dans la pluripotence des cellules ES. Cependant son action peut être antagoniste à c-Myc (Zindy et al., 1998), ainsi la balance de ces facteurs est importante dans l'induction des iPS (induced Pluripotent Stem cells). Il a aussi été montré comme pouvant reprogrammer les cellules de l'épiblaste qui pourtant exprimait déjà les facteurs de la pluripotence: Oct4, Sox2 et Nanog (Guo et al., 2009). De plus des études récentes ont montré l'implication de la protéine Klf4 au sein du réseau de régulation déjà bien connu de Oct4, Sox2 et Nanog (Chen et al., 2008; Kim et al., 2008a; Pardo et al., 2010; van den Berg et al., 2010).

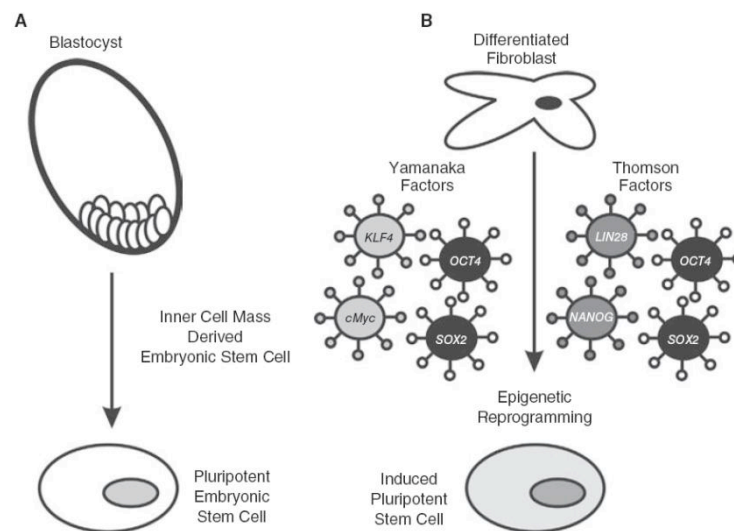


Figure 3: Reprogrammation des cellules différenciées. (A) provenance originelle des cellules ES (B) reprogrammation de cellules différenciées à l'aide de facteurs clefs essentiels aux cellules ES. Tiré de Kashyap et al (Kashyap et al., 2009)

Un an plus tard, une expérience similaire à celle de Takahashi, pionnier pour avoir généré des cellules iPS en 2006, a été réalisée avec des fibroblastes humains (Takahashi et al., 2007). Puis d'autres nombreuses expériences ont permis de générer des iPS à partir de cellules souches neurales humaines à l'aide des facteurs Oct4 et Klf4 (Hester et al., 2009; Kim et al., 2009c; Kim et al., 2008b). Kim et al. ont conclu que la présence d'Oct4 seul était suffisante à induire la pluripotence (Kim et al., 2009a; Kim et al., 2009b). Cependant, le rôle prévalent d'Oct4 dans l'auto-renouvellement des cellules souches reste controversé dans la communauté scientifique. En effet, l'équipe de Rudolf Jaenisch a montré que le facteur Oct4 n'était pas requis pour l'auto-renouvellement des cellules souches somatiques (Lengner et al., 2007) tout comme c-Myc (Wernig et al., 2008). Cependant, le facteur Oct4 est essentiel à la pluripotence. Pour ce faire ils ont construit une lignée de souris où le premier exon du gène Oct4 est entouré de deux sites LoxP. Une hypothèse peut postuler que les facteurs de reprogrammation pourraient être dépendants du type cellulaire différencié de départ. De même, la quantité induite de(s) facteur(s) de transcription important(s) pour la pluripotence joue un rôle dans le déclenchement de la reprogrammation des cellules. Papapetrou et al. ont présenté un système de vecteurs rapporteurs avec des marqueurs fluorescents qui permet de suivre simultanément l'expression de quatre transgènes correspondant à Oct4, Sox2, Klf4 et c-Myc. Ceci, durant l'induction, la maintenance et la différenciation des cellules iPS humaines (Papapetrou et al., 2009). En utilisant ce système, les auteurs ont pu montrer

que l'expression stœchiométrique des quatre transgènes sous le contrôle des promoteurs endogènes est critique pour la reprogrammation efficace des fibroblastes humains.

Conclusion: Caractères des cellules ES

Le caractère pluripotent des cellules ES est sous-tendu par un vaste réseau de régulation qui sera détaillé dans les chapitres suivants. Tout au long de l'étude des réseaux de transcription, la question d'un état préférentiel pour les cellules s'est posée. Le caractère dominant des réseaux transcriptionnels mis en place dans la pluripotence a été démontré par fusion cellulaire. Lorsque des cellules pluripotentes sont fusionnées avec des cellules différenciées les cellules résultantes tendent à devenir pluripotentes (Tada et al., 2001). La capacité du réseau responsable de la pluripotence à se maintenir est supportée par des facteurs de transcriptions mais aussi par une conformation de la chromatine adéquate. Ces deux phénomènes sont donc étroitement liés et s'auto-activent formant une boucle de rétrocontrôle positive (Niwa, 2007a). Cet entretien mutuel stable et efficace est peut être la raison pour laquelle les cellules somatiques reprogrammées ont été obtenues plus facilement qu'estimé (Niwa, 2007a).

Chapitre II. Les facteurs clefs responsables du phénotype des cellules ES

A. Identification des voies de signalisation

1. La voie LIF

Les cellules ES murines ont été isolées pour la première fois depuis un blastocyte de souris (Smith, 2001). Le maintien de l'auto-renouvellement des cellules ES requiert une protéine appelée cytokine leukemia inhibition factor (LIF) (Suda, 1987). La protéine LIF s'attache directement à son récepteur (LIFR) qui contient un long domaine cytoplasmique homologue au récepteur gp130. Le dimère LIF-LIFR recrute ensuite la protéine gp130, pour former un complexe "trimérique" (Zhang et al., 1997). La dimérisation des deux récepteurs entraîne la phosphorylation puis la dimérisation et la translocation du récepteur nucléaire STAT3 (Niwa et al., 1998). La preuve que l'activation de STAT3 est suffisante pour l'auto-renouvellement des cellules ES a été apportée en 1999 (Matsuda et al., 1999). Cependant, les études menées n'excluent pas la possibilité de facteurs synthétisés de façon autocrine qui pourraient contribuer aux capacités d'auto-renouvellement. De plus, il est maintenant connu que la présence additionnelle de sérum contribue à l'auto-renouvellement des cellules ES, montrant que la protéine STAT3 n'est pas suffisante *stricto sensu* à l'auto-renouvellement des cellules ES (Ying et al., 2003a).

Chen et al. ont montré que la protéine STAT3 partage un certain nombre de cibles génomiques avec la protéine Oct4. STAT3 joue donc un rôle clé dans le maintien de la pluripotence dans les cellules ES (page 29). La réduction de l'expression de la protéine Oct4 par ARN interférence a montré que la liaison de STAT3 à ses cibles était affectée, ce qui suggère un rôle d'ancrage sur l'ADN pour la protéine Oct4 (Chen et al., 2008).

Un article paru en 2009 de Niwa explore la voie LIF de façon un peu plus intense et met à jour un réseau de régulation se composant de trois branches effectrices en aval du signal LIF perçu par les récepteurs LIFR et gp130 (Figure 4). En effet, les trois voies effectrices se découpent en une voie Jak avec la protéine STAT3, une voie MAPK et une voie PI3K. L'ensemble de ces voies permet la robustesse de ce système maintenant la pluripotence et l'auto-renouvellement des cellules ES. Ainsi l'absence d'une des voies de

signalisation ne serait pas fatale au phénotype ES, comme par exemple Klf4 (Niwa et al., 2009). Cependant, il est légitime de s'interroger sur les limites de ces compensations mutuelles.

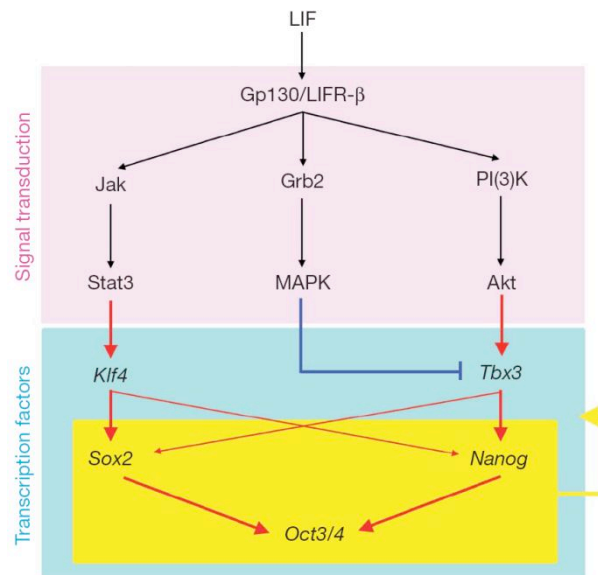


Figure 4: Différentes voies d'action du LIF, tiré de Niwa et al., 2009.

2. La voie BMP

La protéine LIF seule n'est pas capable de maintenir la pluripotence et la présence de sérum de veau fœtal est requise. Les protéines BMPs (Bone Morphogenetic Proteins) apparaissent comme les facteurs clés dérivés du sérum qui agissent de concert avec le LIF pour promouvoir l'auto-renouvellement des cellules et la pluripotence des cellules souches murines (Ying et al., 2003a). Les protéines BMP4, BMP2 ou GDF6 peuvent remplacer le sérum pour la culture des cellules ES et pour la génération *de novo* de lignées (Ying et al., 2003a). Cependant la fonction des protéines BMP/GDF est dépendante de la co-stimulation de la voie LIF. En présence des protéines BMP seules, les cellules s'engagent dans des voies de différenciation non neuronales. En présence de LIF seul, l'auto-renouvellement est assez limité et les cellules ES s'engagent dans la voie de différenciation neuronale. Les protéines BMP suppriment la différenciation neuronale et la combinaison avec le LIF est suffisante pour maintenir l'auto-renouvellement des cellules ES. Cette voie est activée par la fixation de la protéine BMP4 à ses récepteurs, ce qui induit la phosphorylation de Smad1 et active l'expression des gènes de la famille Id (inhibiteurs de la différenciation). Une expérience montre que les cellules ES qui surexpriment les protéines inhibitrices de la

différenciation peuvent maintenir leur auto-renouvellement en absence de BMP4. Il est proposé que l'expression des inhibiteurs de la différenciation ait une contribution critique à la voie de signalisation BMP/Smad (Suzuki et al., 2006). C'est pourquoi, les deux voies de signalisation LIF et BMP ont des rôles-clés dans le maintien de la pluripotence et l'auto-renouvellement des cellules ES (Okita and Yamanaka, 2006).

3. La voie Wnt

La voie Wnt est une autre voie de signalisation cellulaire faisant intervenir des récepteurs extracellulaires et des effecteurs qui peuvent activer un certain nombre de gènes dans le noyau (Figure 5). L'importance de cette voie de signalisation a été soulignée dans de nombreux processus cellulaires. C'est notamment le cas dans les mécanismes de développement (Clevers, 2006; MacDonald et al., 2009), d'induction de la pluripotence ou de cancérisation (Marson et al., 2008a; Reya and Clevers, 2005). Cette voie peut supprimer l'engagement des cellules ES dans la voie de différenciation neuronale (Aubert et al., 2002). Il a été montré que la voie Wnt peut être impliquée dans le maintien à court terme de la pluripotence des cellules ES chez la souris et l'homme (Anton et al., 2007; Sato et al., 2004). En 2004, ces affirmations reposaient principalement sur une drogue inhibitrice de la glycogène synthétase kinase 3 (GSK3) (Chambers and Smith, 2004). Plus récemment, il a été montré que l'activation de la voie de signalisation Wnt augmente l'expression de la protéine STAT3, ce qui suggère un effet synergique avec la voie LIF (Hao et al., 2006). De plus, l'activation de cette voie augmente le niveau d'expression de la protéine c-Myc qui est aussi une cible de STAT3. C'est pourquoi, il est possible que les deux voies de signalisation convergent vers le même gène cible: c-Myc (Cartwright et al., 2005). Il y a plusieurs indices provenant des études sur les cellules souches qui tendent à montrer l'implication de la voie Wnt dans le maintien de la pluripotence des cellules ES. Par exemple, Pereira et al. ont montré que TCF3, un facteur de transcription contrôlé par la voie Wnt, réprime Nanog (Pereira et al., 2006). Un autre exemple a été apporté par l'équipe de Jaenisch: la surexpression d'Oct4 augmente l'activité transcriptionnelle de la β -catenin (Hochedlinger et al., 2005) (Figure 6). Une autre étude met en évidence la régulation par la voie Wnt des facteurs de transcription Oct4, Nanog et Sox2 fortement impliqués dans la pluripotence (Kalmar et al., 2009). Un facteur de la voie Wnt, Wnt3a, peut stimuler la prolifération des cellules ES et affecter la détermination du destin cellulaire (Dravid et al., 2005). Récemment, il a été montré l'importance de cette voie dans la reprogrammation des cellules différenciées (Lluis et al., 2008; Merrill, 2008). Il existe donc un lien évident entre cette voie de signalisation et la pluripotence des cellules souches embryonnaires. Cependant, les données

actuelles ne permettent pas de déterminer l'implication exacte de la voie Wnt dans les réseaux de régulation de la transcription des cellules ES.

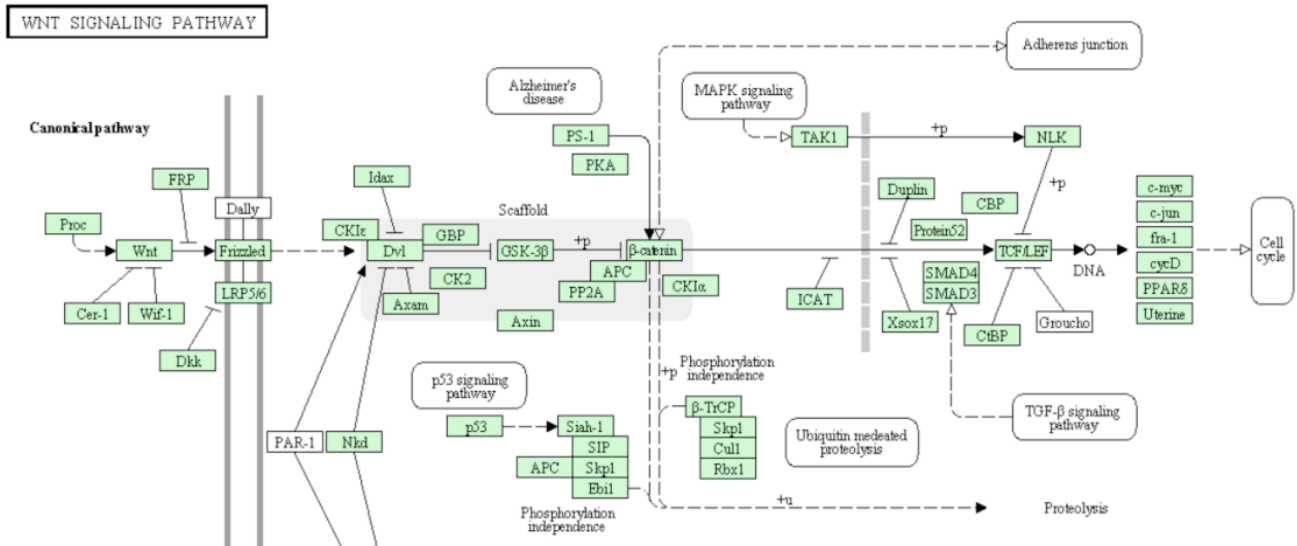


Figure 5: KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway de la voie Wnt. La protéine Wnt transmet le signal effecteur via des récepteurs transmembranaires. Puis une cascade d'activation fait intervenir la betacatenine qui migre dans le noyau de la cellule et permet d'activer un certains facteurs de transcription et régulateurs du cycle cellulaire.

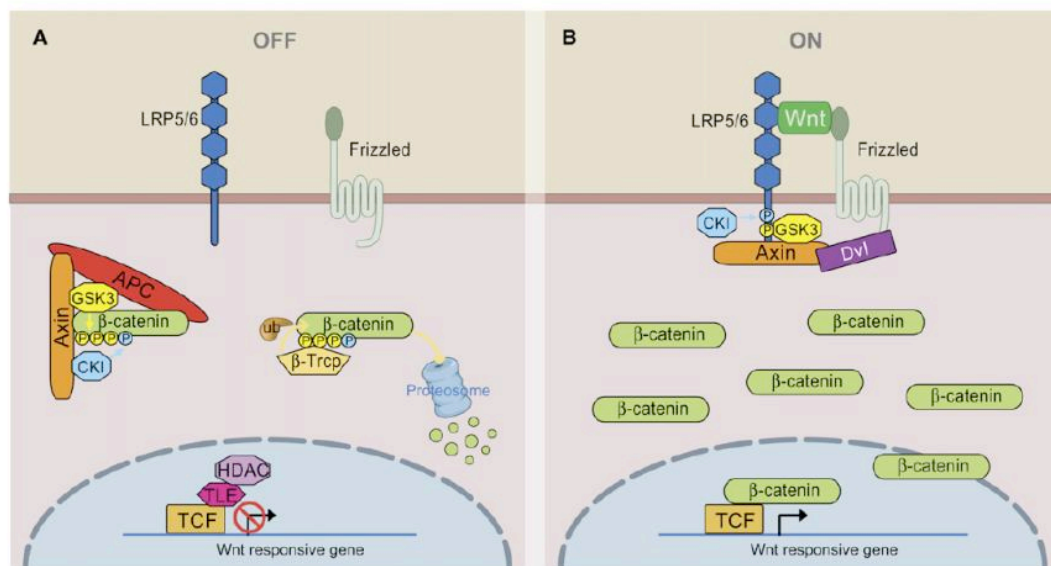


Figure 6: Modèle d'activation de la voie Wnt présentant l'état inactif de la voie (A) et actif (B) tiré de MacDonald BT et al. 2009. La liaison de la protéine Wnt à ces récepteurs permet le recrutement de l'Axin libérant ainsi la betacatenine qui migre dans le noyau et active des facteurs de transcription.

4. La voie PI3K

Des données récentes suggèrent que la voie de signalisation phosphoinositide-3-kinase (PI3K)/Akt joue un rôle-clé pour promouvoir la prolifération, la survie et/ou la différenciation des cellules ES. PI3K augmente la viabilité des cellules ES et la prolifération (Sun et al., 1999) et il a été montré que la surexpression de la protéine Akt est suffisante pour maintenir les cellules ES en absence de LIF (Watanabe et al., 2006). Deux modulateurs de la voie de signalisation PI3K/Akt sont spécifiquement exprimés dans les cellules ES: Eras et Tcl1 (Takahashi et al., 2005). La protéine Eras code pour une forme constitutive de petite GTPase de la famille Ras qui active PI3K pour stimuler la prolifération cellulaire des cellules ES et la tumorigénicité (Takahashi et al., 2003). La protéine Tcl1 augmente l'activation Akt en se couplant à cette dernière, formant un hétérodimère (Teitell, 2005). Le knockdown du gène *Tcl1* dans les cellules ES murines diminue l'auto-renouvellement en induisant une différenciation et/ou une répression de la prolifération (Ivanova et al., 2006; Matoba et al., 2006). Cependant, les mécanismes moléculaires qui contrôlent l'expression des gènes *Eras* et *Tcl1* dans les cellules ES n'ont pas encore été clairement identifiés. La protéine Stk40 a été identifiée lors d'une récente étude comme le lien entre la voie Erk/MAPK impliquée dans la différenciation et plus particulièrement dans ce cas dans la voie endodermique, et Oct4 impliqué dans la pluripotence (Li et al., 2010b). Stk40 est une cible d'Oct4 et est réprimé lorsque Oct4 se lie sur son promoteur (2kb en amont) permettant d'inhiber la voie de différenciation et de maintenir le phénotype ES (Li et al., 2010b).

Alors que ces voies de signalisation ont des senseurs externes qui font intervenir des composants du milieu cellulaire, des facteurs intrinsèques sont aussi requis dans le renouvellement cellulaire. Certains facteurs de transcription (TFs) sont aussi essentiels pour déterminer la capacité des cellules ES à l'auto-renouvellement des cellules ES et à les maintenir indifférenciées.

B. Les facteurs importants dans le maintien de la pluripotence des cellules ES

1. Oct4

La protéine Oct4, encodée par le gène *Pou5f1*, est un facteur de transcription contenant un domaine POU connu pour être indispensable aux cellules ES ainsi qu'aux

premiers stades du développement embryonnaire (Boiani and Scholer, 2005; Nichols et al., 1998; Smith, 2001). Le domaine POU est caractéristique de la famille POU, ainsi nommé pour les premiers gènes qui ont été identifiés PIT1 et Oct1 chez les mammifères, et Unc-86 chez *C. Elegans* (Herr et al., 1988). Les mutations dans ce domaine très conservé conduisent à des maladies ou malformations (Ohta et al., 1992; Pfaffle et al., 1992; Radovick et al., 1992). La protéine Oct4 est très exprimée dans les cellules ES qui se caractérisent par leur phénotype transcriptionnel actif et indifférencié. Mais cette propriété n'est pas partagée par les autres membres de la famille Oct1, Oct2 et Oct6 (Niwa et al., 2000). C'est pourquoi, Oct4 est souvent utilisé comme marqueur des cellules indifférenciées ou tumorales (Looijenga et al., 2003). L'expression du gène Oct4 est très finement contrôlée car une dérégulation positive ou négative peut entraîner la différenciation des cellules (Niwa et al., 2000). En effet, une dérégulation deux fois supérieure à la normale entraîne une différenciation en endoderme et mésoderme (Figure 7). A contrario, une répression de Pou5f1 induit une perte de la pluripotence et une différenciation en trophoctoderme (Niwa et al., 2000).

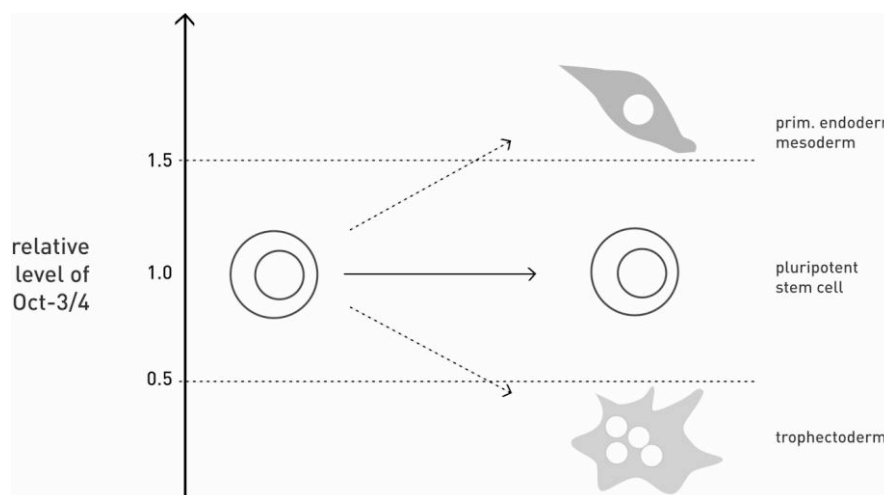


Figure 7: Effet de la dérégulation positive ou négative d'Oct4 dans les cellules ES, tiré de Niwa 2000.

Il a été montré qu'Oct4 s'associe de façon préférentielle avec les séquences comportant le motif ATGCAAAT (Chambers and Smith, 2004). Cependant, Oct4 seul (en l'absence de LIF) ne peut pas maintenir la pluripotence des cellules ES (Chambers and Smith, 2004).

La détermination des gènes liés par Oct4 a été très documentée au cours de ces dernières années (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008a; Loh et al., 2006; Matoba et al., 2006). La comparaison globale des gènes dérégulés lors de la répression d'Oct4 a révélé un certain nombre de gènes dont certains sont des modificateurs

épigénétiques. Loh et al. ont identifié deux histones déméthylases Jmjd2c et Jmjd1 (Figure 8) qui suppriment les marques d'histones H3K9me2 et me3 sur les gènes Nanog et Tcl1 (Loh et al., 2007). L'importance de Nanog dans la pluripotence sera détaillée plus loin (page 33). La protéine Tcl1 est exprimée de façon spécifique dans les cellules ES et régule positivement l'expression de la protéine Oct4 (Matoba et al., 2006) induisant ainsi un rétrocontrôle positif. Cependant, il est probable que l'activation de l'expression d'Oct4 ne se limite pas à cette voie (Niwa, 2007a). D'autres voies de signalisation sont impliquées dans la régulation de la transcription d'Oct4.

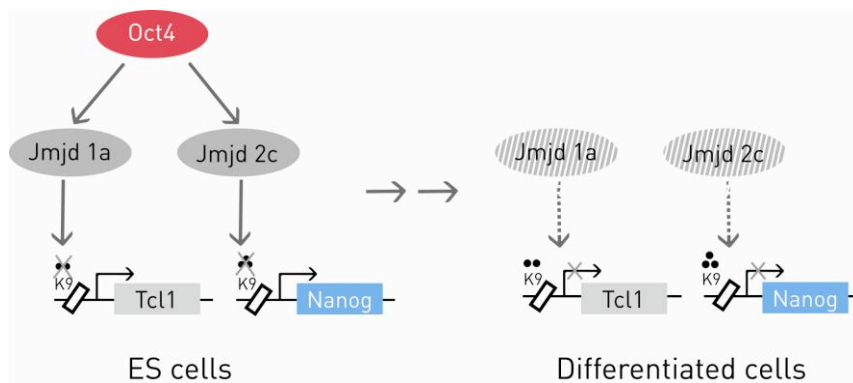


Figure 8: Régulation des gènes Tcl1 et Nanog par Oct4 via les histones demethylases Jmjd1a et jmj2c, tiré de Loh 2007.

Une récente étude a permis de définir l'interactome de la protéine Oct4 (van den Berg et al., 2010). Parmi les protéines interagissant directement avec Oct4 (pas d'interaction via l'ADN) de nouvelles protéines ont été retrouvées comme Chd4, Mbd3 et Rbbp prouvant ainsi qu'Oct4 peut interagir avec le complexe NuRD originalement défini par Zhang en 1999. Cette étude remet donc en question le complexe NODE originellement défini par Liang et al. qui avaient montré que Nanog et Oct4 pouvaient s'associer avec une partie du complexe NuRD (Hdac1/2, Mta1/2, et éventuellement Mbd3 et Rbbp7). Le complexe NODE contenait donc une activité deacetylase dont la suppression entraînait la différenciation des cellules ES et l'augmentation de l'expression des gènes du développement (Liang et al., 2008). Leurs données suggèrent que Nanog et Oct4 s'associent avec un répresseur unique sur leurs gènes cibles afin de contrôler le destin des cellules ES (Liang et al., 2008). L'étude de Poot (van den Berg et al., 2010), basée sur une immunoprécipitation native de la protéine Oct4 étiquetée avec trois épitopes FLAG, permet de confirmer les partenaires décrits précédemment dans la bibliographie (Chen et al., 2008; Kim et al., 2008a). Certains de ces partenaires co-localisent aux promoteurs de gènes importants de la pluripotence comme

Nanog et Oct4 lui-même. De même, Oct4 régule et se lie sur le promoteur de certains gènes (van den Berg et al., 2010). Ce système renforce ainsi l'interdépendance du réseau de transcription sous-tendu dans les cellules ES. Van den Berg et collaborateurs ont décrit une nouvelle carte reportant les interactions de la protéine Oct4. Les interactions avec Sall4, Esrrb, Dax1 et Tcfcp2l1 étant déjà reportées dans la bibliographie, les auteurs ont étendu le réseau de régulation d'Oct4 en cherchant les partenaires de ces protéines (Figure M5).

2. Sox2

Oct4 interagit avec Sox2, un facteur de transcription de la famille SOX. Les 20 gènes de cette famille partagent le motif HMG du gène SRY sur le chromosome Y impliqué dans la détermination du sexe (SOX: Sry-related HMG box). Les embryons Sox2^{-/-} meurent immédiatement après l'implantation (Avilion et al., 2003) et l'inactivation de Sox2 dans les cellules ES murines induit la différenciation en de multiples lignages indiquant son importance dans le maintien de la pluripotence (Ivanova et al., 2006).

Sox2 joue un rôle essentiel dans la transcription de plusieurs gènes cibles d'Oct4 (Niwa, 2001). Les régions liées par Sox2 ont montré qu'elles partageaient un motif dont le consensus est CTTTG(T/A)(T/A) (Chambers and Smith, 2004).

Il a été récemment montré que dans un contexte Sox2^{-/-}, la présence de cette protéine n'est pas indispensable pour l'activation des enhanceurs Sox-Oct. En revanche, Sox2 est nécessaire pour la régulation de multiples facteurs de transcription qui affectent l'expression d'Oct4. De plus, l'expression forcée de Oct4 permet un sauvetage de la pluripotence des cellules ES Sox2^{-/-}. Ces résultats indiquent que la fonction essentielle de Sox2 est de stabiliser les cellules ES dans un état pluripotent en maintenant le niveau requis de l'expression d'Oct4 (Masui et al., 2007). Cependant, il est difficile d'imaginer que le rôle de la protéine Sox2 se limite à maintenir le niveau de transcription de la protéine Oct4 si les nombreuses séquences liées par Sox2 sont prises en compte (Chen et al., 2008).

3. Le complexe Oct4/Sox2

En 2005, Chew et al. ont montré que le complexe Oct4/Sox2 s'autorégule (Chew et al., 2005). Sox2 et Oct4 agissent de façon synergique pour stimuler la transcription (Ambrosetti et al., 2000). La carte génétique des sites de fixation d'Oct4 et Sox2 montrent qu'ils co-régulent un grand nombre de gènes (Boyer et al., 2005). Un exemple est le cas du promoteur du gène Nanog. Les deux sites de fixations d'Oct4 et de Sox2 sont juxtaposés suggérant une étroite collaboration des deux protéines (Chambers and Smith, 2004; Rodda

et al., 2005). Via leurs nombreux gènes cibles, ces facteurs de transcription peuvent exercer un rôle dominant sur le réseau de transcription des cellules souches. Loh et al. ont montré que les deux consensus de séquence ADN déjà connus pour Oct4 et Sox2 étaient juxtaposés (Figure 9). Par ailleurs, la structure 3D du complexe ternaire Oct4-Sox2-ADN a été mise en évidence (Ambrosetti et al., 2000) démontrant l'interaction physique de ces protéines.

Un nouveau facteur de transcription est venu s'ajouter au duo.

4. Nanog

Nanog est un troisième facteur de transcription découvert en 2003 par deux équipes (Chambers et al., 2003; Mitsui et al., 2003). Cette protéine fait partie de la famille Q50 des protéines à homéodomaine. Cette famille a la particularité de posséder un acide glutamique en position 50 à l'intérieur de l'homéodomaine et peut interagir directement avec les nucléotides en amont de la séquence ATTA. Loh et al. ont montré que Nanog pouvait interagir avec la séquence CATT et une séquence consensus reproduite dans la Figure 9 a pu être identifiée (Loh et al., 2006).

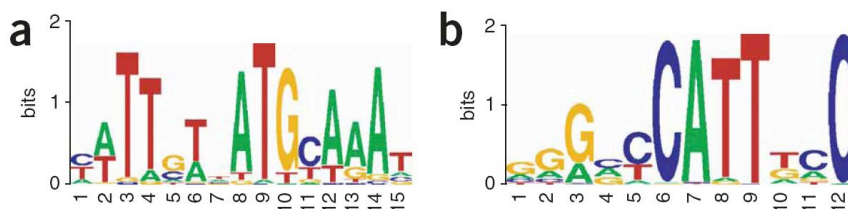


Figure 9: Prédiction de motifs de liaison à l'ADN de Sox2-Oct4 (a) et Nanog (b) par CHIP-PET tiré de Loh et al., 2006.

La protéine Nanog est exprimée dans un nombre restreint de cellules et particulièrement dans les cellules exprimant aussi la protéine Oct4. Nanog n'est pas présente dans l'ovocyte. L'expression est détectée pour la première fois au stade morula à l'intérieur de l'embryon. Le taux maximal de Nanog est atteint dans le blastocyte précoce, mais, de façon intéressante, l'expression décline juste avant l'implantation. Cette signature d'expression dynamique peut indiquer que la réduction de l'expression de Nanog est importante pour éviter une expansion incontrôlée des cellules pluripotentes (Chambers and Smith, 2004). Nanog peut maintenir la pluripotence des cellules ES (Chambers et al., 2003;

Mitsui et al., 2003). En effet, lorsqu'une cassette permettant la surexpression de Nanog est introduite dans une culture de cellules ES, le phénotype ES est conservée en absence de LIF. En retirant la cassette, la dépendance au LIF est restaurée (Chambers et al., 2003). Cependant aucun lien reliant Nanog à STAT3 n'a été démontré (Chambers and Smith, 2004). De même, Ying et al. ont montré que Nanog pouvait bloquer la différenciation neuronale induite par le retrait du LIF et du sérum (voie BMP) dans des cultures des cellules ES. Mitsui et al. ont montré que des cellules ES invalidées pour le gène *Nanog* différenciaient rapidement en cellules proches de l'endoderme primitif alors que la surexpression de Nanog dans les cellules ES humaines induit une différenciation ectodermique (Darr et al., 2006). Finalement Suzuki et al. ont montré que Nanog pouvait inverser la voie de spécification mésodermique. C'est pourquoi Nanog peut bloquer la différenciation à plusieurs stades (endoderme primitif, neural et mésodermique) avec différentes conditions de culture. De même, une surexpression de Nanog peut bloquer les voies de différenciation induite par l'acide rétinoïque (Loh et al., 2006).

La fonction de Nanog requiert la présence d'Oct4. Les cellules ZHBTc4 expriment un transgène codant pour Oct4, qui est réprimé en présence de tétracycline. Lorsque les cellules ont une expression diminuée d'Oct4, elles se différencient en cellules du trophoctoderme. Cet effet peut être compensé par l'électroporation d'ADN complémentaire d'Oct4 (sauvetage) (Niwa et al., 2002) mais pas de Nanog (Chambers et al., 2003). C'est pourquoi la fonction d'Oct4 n'est pas seulement d'activer Nanog, même s'il est établi que la fonctionnalité de Nanog requiert le facteur Oct4. Le maintien de la pluripotence par la surexpression de Nanog suggère que ce dernier est capable de déterminer le niveau d'Oct4. Cependant ce dernier n'est pas indispensable puisque l'expression d'Oct4 est observée en absence de Nanog au niveau des blastocytes (Chambers and Smith, 2004). De même, Oct4 n'est pas nécessaire à l'initiation de l'expression de Nanog. Bien que ces observations soient faites, l'expression de ces deux facteurs ainsi que Sox2 sont intimement corrélées. Cette hypothèse est supportée par l'observation des sites de liaisons d'Oct4 et Sox2 au niveau du promoteur de Nanog (Figure 9).

5. Le complexe Oct4/Sox2/Nanog

Les facteurs de transcription Oct4, Sox2 et Nanog sont capables de s'autoréguler et d'induire leurs expressions respectives. Ce processus est essentiel à l'auto-renouvellement des cellules indifférenciées (Rodda et al., 2005). Il est à présent bien établi que les facteurs Sox2 et Oct4 agissent de concert. Leur site de fixation est donné selon la Figure 9. Le site de fixation de la protéine Nanog identifié par Loh et al. ressemble beaucoup au site commun

d'Oct4 et de Sox2 et il serait intéressant de voir si leurs sites de fixation sont partagés. Des expériences de CHIP-PET ont permis de montrer les gènes cibles communs à Oct4 et Nanog. Cependant, des expériences d'ARN interférence ont montré que les gènes liés n'étaient pas tous dérégulés. Ces données suggèrent reflètent que seulement une catégorie de gènes liés sont régulés par Oct4 et Nanog. Loh et al. proposent l'hypothèse que les gènes non dérégulés peuvent refléter des sites non fonctionnels ou que la redondance des facteurs de transcription ne permet pas d'observer d'effets significatifs lors des tests par ARN interférence. Les données de Loh et al. Boyer et al. ont montré que les facteurs Oct4, Sox2 et Nanog forment une boucle de régulation positive: les trois facteurs de transcription s'autorégulent eux-mêmes (Figure 10). Cette boucle de rétrocontrôle positif permet de maintenir les niveaux d'expression de chaque facteur et donc de maintenir la pluripotence.

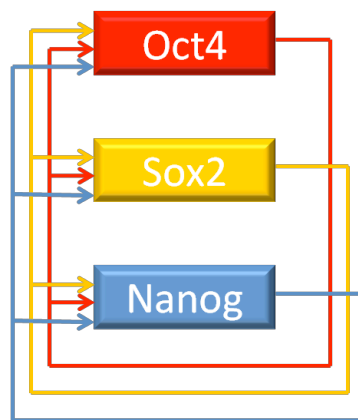


Figure 10: Boucle de régulation d'Oct4, Nanog et Sox2. Adapté de Marson et al. 2008

Cependant, cette boucle doit être finement contrôlée. Niwa propose donc une série de régulateurs positifs et négatifs extérieurs permettant de stabiliser cette boucle de contrôle de maintenance de la pluripotence des cellules ES (Niwa, 2007a).

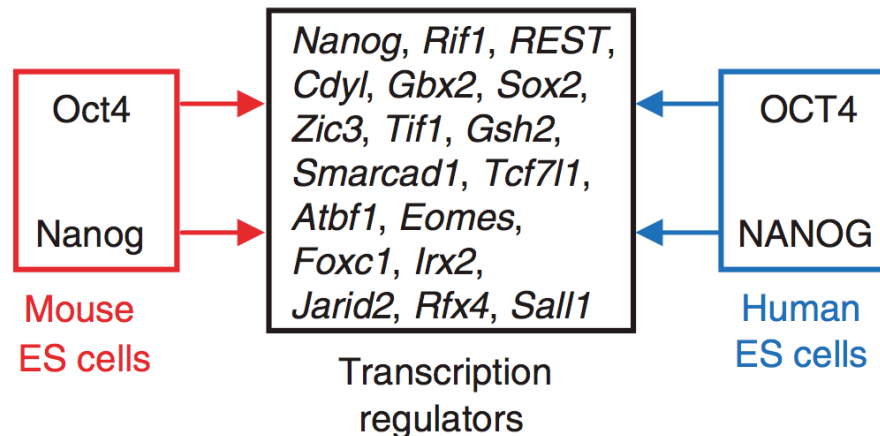


Figure 11 Réseau de régulation de transcription d'Oct4 et Nanog dans les cellules ES tiré de Loh et al. 2006

Il est cependant légitime de s'intéresser à la validité des informations données par ces expériences à grande échelle. Loh et al. ont montré qu'il n'avaient de 9,1% à 13% de recouvrement entre leurs données de CHIP-PET en cellules souches murines pour Oct4 et Nanog respectivement avec les données de Boyer et al. en cellules souches humaines (Boyer et al., 2005; Loh et al., 2006). Loh et al. suggèrent que ces données reflètent les différences des réseaux de régulation de la transcription contrôlés par Oct4 et Nanog entre les différentes espèces (Wei et al., 2005). De plus, les éléments régulateurs ne sont pas toujours placés dans la région proximale du promoteur du gène (Cawley et al., 2004; Wei et al., 2005). C'est pourquoi certains gènes peuvent être omis dans les analyses (Loh et al., 2006). Finalement, la détection des séquences liées est propre à chaque méthode et des variations peuvent apparaître d'une étude à une autre. Malgré cette ambiguïté, 32 gènes ont été trouvés comme cibles d'Oct4 et Nanog chez la souris comme chez l'humain (Loh et al., 2006). Parmi ces 32 gènes, 18 codent pour des facteurs de transcription (Figure 11).

C. Les facteurs requis pour la pluripotence des cellules ES

D'autres facteurs sont requis pour maintenir l'état indifférencié des cellules ES (Niwa, 2007b). Des travaux récents ont permis d'identifier des nouveaux facteurs du réseau de régulation de transcription requis pour maintenir la pluripotence. A travers des cribles génétiques, de nouvelles protéines régulant l'auto-renouvellement des cellules ES ont été

identifiées (Galan-Caridad et al., 2007; Ivanova et al., 2006; Loh et al., 2006). Les principales sont listées par la suite :

Esrrb (Estrogen related receptor, beta ou Nr3b2) est un récepteur nucléaire. Il a été identifié en 2006 comme une protéine essentielle au maintien des cellules ES (Ivanova et al., 2006; Loh et al., 2006). Il régule positivement l'expression du gène Nanog (van den Berg et al., 2008). La répression de Esrrb dans les cellules ES murines induit la différenciation des cellules ES (Niwa, 2007a). La surexpression de cette protéine permet le maintien des cellules ES sans la présence de LIF (Zhang et al., 2008). Esrrb est aussi capable de remplacer la protéine Klf4 lors de la reprogrammation des cellules somatiques (Feng et al., 2009). Le rôle présumé de cette protéine et de la protéine Tbx3 serait de maintenir le niveau de transcription de la protéine Nanog puisque les effets de la répression de ces gènes peuvent être supprimés par une surexpression de Nanog. Plusieurs sites de liaisons d'Oct4 et Nanog ont été retrouvés sur le promoteur de Esrrb (Loh et al., 2006)

Van den Berg et collaborateurs ont montré que ce facteur pouvait se lier aux composants de la machine de transcription comme l'ARN polymérase II ou encore le médiateur (van den Berg et al., 2008). Une hypothèse quant à la fonction d'Esrrb serait un lien entre les facteurs de transcription (Oct4, Nanog, Sox2...) et la machinerie de transcription. Confortant cette hypothèse, Chen et al. ont montré que Esrrb était souvent retrouvée sur les mêmes sites que Nanog et Tcfcp2l1 (Chen et al., 2008).

Tbx3 (T-box transcription factor): l'inactivation du gène déclenche la différenciation en de nombreux lignages issus de l'ectoderme primaire, ce qui montre son importance pour le maintien de la pluripotence des cellules ES (Niwa, 2007a).

Dax1 (nuclear receptor subfamily 0, group B, member 1): Dax1 est un récepteur nucléaire initialement identifié en 1999 comme un facteur anti-testis, c'est-à-dire qu'il joue le rôle d'un antagoniste du gène *Sry* (Goodfellow and Camerino, 1999). Ce récepteur est important pour le renouvellement des cellules ES (Niakan et al., 2006). La protéine Dax1 est exprimée dans les cellules ES et son expression est contrôlée par Oct4 et Stat3 pour maintenir les cellules ES. De plus ces facteurs pourraient se lier et partager les mêmes cibles sur le génome (Sun et al., 2008). Plus récemment, il a été montré que Dax1 interagissait avec Oct4 via le domaine POU (Sun et al., 2009). Par des expériences de GST pull down et de mobilité sur gel, la même équipe a été démontré que Dax1 inhibait l'activité

de liaison à l'ADN d'Oct4. Enfin, par immunoprecipitation de chromatine, ils ont montré que Dax1 empêchait Oct4 de se lier sur ses cibles. De plus, la surexpression de Dax1 provoque la différenciation. Toutes ces données semblent alors indiquer que Dax1 joue le rôle d'un régulateur négatif d'Oct4 dans les cellules ES (Sun et al., 2009).

Myc: Il a été démontré que le facteur de transcription b-Myc est un accélérateur du cycle cellulaire dans les cellules ES murines (Niwa, 2007b). La surexpression d'une forme négative dominante dans ces cellules conduit à l'arrêt en phase G1 (Iwai et al., 2001), indiquant que b-Myc est exprimé en phase G1 et lors de la transition de G1 en phase S par un mécanisme complexe (Joaquin and Watson, 2003). Le facteur Myc, qui comprend un motif hélice-boucle-hélice, est un accélérateur du cycle cellulaire. Il active la transcription de la cycline E pour promouvoir le passage de la phase G1 en phase S (Hooker and Hurlin, 2006). Cartwright et al. ont montré que c-Myc est une cible directe de STAT3 et qu'une forme active dominante de c-Myc (de stabilité accrue) permet de maintenir les cellules ES sans LIF. A contrario, une forme négative dominante de c-Myc empêche l'auto-renouvellement des cellules ES et induit la différenciation. Ces données suggèrent que la transition G1=>S peut contribuer à la pluripotence qui est soutenue par les voies LIF et STAT3 (Burdon et al., 2002). Une récente étude des différents membres de la famille Myc (voir Annexe), montre que les protéines Myc ont des propriétés divergentes mais possèdent la caractéristique commune d'être souvent impliquées dans des cancers, ce qui renforce l'hypothèse de l'implication de cette famille de protéines dans le cycle cellulaire (Gostissa et al., 2009). Enfin, une dernière étude du groupe de Young montre que le facteur c-Myc joue un rôle clef dans la reprise de la transcription après les pauses post-initiation de la Polymérase II pour une large proportion (1/3) des gènes transcrits en ES (Rahl et al., 2010). A la lumière de ces résultats, il est plus aisé de comprendre alors comment c-Myc permet la reprogrammation des fibroblastes en iPS (Takahashi et al., 2007).

Tcfcp2l1 (transcription factor CP2-like 1) colocalise avec Oct4 sur un nombre important de promoteurs et semble être important pour la prolifération optimale des cellules ES. Il est impliqué dans de nombreux processus comme la réplication de l'ADN (interaction avec Polb, Asf1a, Rpa1) ou la réparation de l'ADN (Xrcc1, 5, 6, Msh2...). Il est également impliqué dans la progression du cycle cellulaire, la prolifération cellulaire ou encore le remodelage de la chromatine (interaction avec Hells) (van den Berg et al., 2010). L'inactivation de ce gène affecte la croissance cellulaire mais n'affecte pas l'auto-renouvellement des cellules (Ivanova et al., 2006). Ceci suggère pour Tcfcp2l1 un rôle dans

la progression du cycle cellulaire dans les cellules ES et les processus de réparation de l'ADN (van den Berg et al., 2010). L'interaction de cette protéine avec E2f1 (Ren et al., 2002) un facteur de transcription qui régule le cycle cellulaire sur de nombreux promoteurs (Chen et al., 2008) est consistant avec cette hypothèse.

Klf4 (EZF) et Klf5: En 2005, la revue de Yang présentait deux facteurs Krüppel: Klf4 et Klf5. Ces deux facteurs de transcription appartenant à la même famille Krüppel (Annexe), conservés au cours de l'évolution et très semblables entre eux, contenant un domaine en doigt de zinc (Ghaleb et al., 2005). Toutefois, ces deux facteurs ont des rôles complètement opposés: Klf4 est un inhibiteur de la croissance cellulaire alors que Klf5 stimule la prolifération. Les deux facteurs ont la capacité de lier les mêmes séquences. Un exemple étonnant est le promoteur de Klf4 lui-même qui est activé par Klf4 et inhibé par Klf5 (Dang et al., 2003). Les deux facteurs sont des compétiteurs sur le même régulateur en cis. De plus Klf4 et Klf5 montrent des effets opposés sur la prolifération cellulaire. Dang et al. ont montré que Klf4 est un gène associé à l'arrêt de la croissance cellulaire (Dang et al., 2000). Klf4 maintient l'intégrité du cycle cellulaire et est important pour la réponse aux dommages de l'ADN. A contrario, Klf5 stimule le cycle cellulaire et donc la prolifération cellulaire. L'inactivation du gène *Klf5* est létale pour l'embryon de souris (Shindo et al., 2002) alors que pour les souris *Klf4^{-/-}* la mort intervient peu après la naissance. Cette observation est cohérente avec l'expression de Klf4 dans les cellules qui ont stoppé leur croissance et l'expression de Klf5 dans les cellules qui ont une prolifération intensive. L'implication de ces deux facteurs de transcription a été montrée dans plusieurs cancers sans pour autant que leur rôles puissent être clairement établis. Klf4 et Klf5 font partie de la voie de signalisation LIF et les cellules ES qui sur-expriment ces facteurs montrent une résistance particulière à la différenciation (Bourillot et al., 2009; Li, 2005; Niwa et al., 2009). Il a été montré que Klf5 interagit avec Oct4, Sall4 et Tcf21. Elle se trouve au milieu d'un réseau de régulation et est ainsi qualifiée de "hub" ou foyer par van den Berg et al. (van den Berg et al., 2010). Klf4 possède des propriétés oncogènes (Rowland and Peepers, 2006). Sa surexpression dans les cellules ES de souris empêche la formation de corps embryoïdes (Li, 2005). Klf4 peut se lier au promoteur proximal d'Oct4 et peut activer Oct4 et Sox2 (Nakatake et al., 2006). Une étude récente a comparé trois facteurs Krüppel: Klf2, Klf4 et Klf5 (Hall et al., 2009). Les trois facteurs font partie du réseau d'Oct4 mais ont des propriétés différentes, Klf2 est indépendante de l'activation par la voie LIF, Klf 5 est sensible à l'action du LIF mais indépendamment du récepteur STAT3, alors que Klf4 est sensible au LIF par la voie STAT3. Grâce à cette propriété, la surexpression de Klf2 permet de maintenir le phénotype ES en absence de LIF. La surexpression des facteurs Klf2 et Klf4 a permis de mettre en évidence

le rôle de ces protéines dans l'auto-renouvellement ce qui n'est pas le cas pour Klf5. Il a été montré à de nombreuses reprises que le facteur Klf4 pouvait reprogrammer des cellules différenciées en iPS. Hall et al. ont montré qu'en reprogrammant des cellules préimplantatoires, l'expression induite de Klf2 permet d'obtenir deux fois plus de colonies iPS que Klf4. Cette observation est confirmée par l'expression spécifique aux cellules ES de Klf4, Dppa3, Zfp42 et Dax1 ainsi que la perte d'expression des facteurs Fgf5 et du facteur T (Hall et al., 2009). Cependant, il serait intéressant de poursuivre l'étude des 17 facteurs qui composent la famille Klf afin de mieux comprendre leurs rôles. La caractérisation de leurs partenaires et de leurs cibles seraient des indices-clés afin de comprendre d'une part les mécanismes qui régissent l'auto-renouvellement des cellules ES et la reprogrammation des cellules différenciées.

FoxD3: L'expression du facteur de transcription "forkhead" FoxD3 est détectable dans le blastocyte et plus tard dans l'épiblaste de l'embryon préimplantatoire. Bien qu'une interaction avec Oct4 ait été démontrée, la preuve que FoxD3 soit impliquée dans la pluripotence est resté longtemps indirecte (Guo et al., 2002). Les embryons inactivés pour ce gène survivent jusqu'au stade E6,5 où il apparaît clairement un défaut dans la compartimentation (Hanna et al., 2002). Ce résultat montre que FoxD3 est requis à un stade au-delà de l'établissement et de la maintenance des ICM pluripotentes. Plus récemment, il a été montré que Foxd3 est requis pour le maintien des cellules pluripotentes dans le jeune embryon de souris et lors de la création de lignées ES. Pour comprendre le rôle de Foxd3, Liu et al. ont dérivé une lignée ES à partir de blastocystes qui portaient deux allèles Foxd3 conditionnels (flanqués de séquences LoxP) et un transgène Cre inducible au tamoxifène (Liu and Labosky, 2008). Le traitement par Tamoxifène produit une perte rapide et quasi complète de l'ARNm et de la protéine Foxd3. Les cellules ES Foxd3 déficientes maintiennent un taux de prolifération normale, mais montrent une augmentation de l'apoptose, et des clones dispersés qui ont une capacité affaiblie à se renouveler. Quelles que soient les conditions de culture de ces lignées, les auteurs observent une forte différenciation précoce des cellules ES Foxd3 mutantes en de multiples lignées y compris trophectoderme, endoderme, et mésoderme. Cette modification profonde des comportements biologiques reflète l'altération des facteurs connus pour induire la pluripotence: Oct4, Sox2 et Nanog. C'est pourquoi Liu et al. ainsi que Jung et al. proposent un rôle pour Foxd3 dans la répression de la différenciation, la promotion de l'auto-renouvellement et le maintien de la survie des cellules ES de la souris (Jung et al., 2010; Liu and Labosky, 2008) ([Figure M3](#)).

Zfx: La délétion du gène *Zfx* diminue la croissance cellulaire des cellules ES et augmente le taux d'apoptose (Galan-Caridad et al., 2007). Toutefois, les cellules *Zfx*^{-/-} peuvent donner tous les lignages de cellules. La surexpression de la protéine *Zfx* favorise l'auto-renouvellement des cellules ES. De façon surprenante mais néanmoins intéressante, l'équipe de Reizis a montré que *Tbx3* et *Tcl1*, deux facteurs importants pour la pluripotence des cellules ES, étaient des cibles de *Zfx* (Galan-Caridad et al., 2007).

Rif1 (Rap1 interacting factor 1 homolog): Cette protéine a été identifiée par Loh et al. comme cible commune d'*Oct4* et de *Nanog*. Grâce à une expérience de shRNA, Loh et al. ont pu montrer que l'absence de cette protéine à l'instar d'*Esrrb*, entraînait une différenciation en cellule du trophoctoderme (Loh et al., 2006). *Rif1* est donc requis pour le maintien de la pluripotence des cellules ES et fait partie du réseau de régulation d'*Oct4* et *Nanog*.

Tcl1: Cette protéine a été identifiée par Ivanova et al. (Ivanova et al., 2006) lors d'un crible shRNA comme indispensable au maintien du phénotype ES. *Tcl1* est une GTPase, cofacteur de la protéine *Akt1* (la sérine-thréonine kinase majeure de la voie PI3K), qui renforce son activité kinase et favorise son transport nucléaire (Pekarsky et al., 2000). La surexpression de cette protéine est bien connue dans certains lymphomes (Narducci et al., 1997). Narducci et al. ont montré que *Tcl1* est important pour la fertilité des souris femelles et est requis pour les premiers stades de l'embryogenèse (Narducci et al., 2002). Récemment, Loh et al. ont montré que *Tcl1* était la cible d'une histone déméthylase, *Kdm3a*, qui module l'expression de *Tcl1* en augmentant le taux de méthylation de H3K9me2 (Loh et al., 2007). Un sauvetage du phénotype ES dans un contexte appauvri en *Jmjd1a* a été obtenu grâce à une surexpression de *Tcl1* (Loh et al., 2007).

Mbd3: La protéine *Mbd3* est essentielle pour les cellules ES car les cellules *Mbd3*^{-/-} perdent la capacité à se différencier (Kaji et al., 2006). De plus, son essentialité a été montrée lors d'un crible d'ARN interférence à grande échelle (Fazzio et al., 2008).

Ronin (ou THAP11): Cette protéine a été nommée d'après un samouraï japonais sans maître car elle ne possède *a priori* aucune interaction connue avec le réseau majeur impliqué dans le maintien de la pluripotence. Elle a été identifiée en 2008 par Dejosez et al. par un crible double hybride chez la levure comme cible de la caspase 3 (Dejosez et al., 2008). A l'aide de nombreuses lignées modifiées pour le gène *Ronin*, les auteurs ont pu montrer que cette protéine était essentielle à la génération des ICM, que les cellules sur-

exprimant Ronin pouvait maintenir le phénotype ES en absence de LIF ou d'Oct4. Ceci montre que la voie impliquant Ronin est indépendante (i) de la voie LIF et (ii) du réseau direct d'Oct4.

6. Autres facteurs

D'autres protéines, Nac1, Zfp281 (Wang et al., 2006) [Figure M6](#), Zfp206 (Yu et al., 2009) ou Tcf3 (Pereira et al., 2006) ont été montrées comme interagissant avec Nanog et sont indispensables au maintien de l'auto-renouvellement des cellules ES.

D. Les facteurs dispensables aux cellules ES mais intervenant dans le maintien de la pluripotence

Utf1: Le facteur de transcription de cellules embryonnaires indifférenciées (Undifferentiated transcription factor: Utf1) a été identifié pour la première fois comme un co-facteur spécifique aux cellules ES (Okuda et al., 1998). Les cellules ES murines qui sous-expriment Utf1 ont une prolifération ralentie *in vitro* et une tumorigénicité réduite *in vivo* (Nishimoto et al., 2005). Utf1 est une cible des facteurs Sox2 et Oct4 (Nishimoto et al., 1999) c'est pourquoi ce facteur est considéré comme un lien entre les réseaux de régulation de transcription maintenant la pluripotence et la prolifération des cellules ES.

Eras: La génération de lignées ES invalidées pour le gène *Eras* viables et permettant la transmission à la lignée germinale a montré que cette protéine n'est pas indispensable à la pluripotence des cellules ES. Cependant, la croissance des cellules ES est ralentie par rapport aux cellules ES sauvages et la tumorigénicité des cellules ES est fortement réduite (Takahashi et al., 2003). La surexpression de ce gène dans les cellules ES induit une augmentation de la croissance cellulaire. Eras n'interagit pas directement avec la protéine Raf mais avec un effecteur de Ras, la PI(3)Kinase qui est importante dans la transformation et la propagation des cellules ES. Ces données permettent de conclure que l'oncogène *Eras* est important pour les propriétés de croissance des cellules ES proches de celles des tumeurs.

Sall4 est important pour la prolifération et la stabilité du phénotype des cellules ES (Yang et al., 2010; Yuri et al., 2009; Zhang et al., 2006). Cependant il n'est pas indispensable

au phénotype ES (Yuri et al., 2009). Il fait partie de la famille des protéines *spalt* comprenant Sall1, Sall2, Sall3 et Sall4. Chez *D. Megalogaster*, les gènes *spalt* sont régulés par la voie TGF β (de Celis et al., 1996). Les cellules ES Sall4^{-/-} ont une prolifération ralentie (Sakaki-Yumoto et al., 2006). Une autre étude a montré que Sall4 interagissait avec Nanog pour activer leurs propres expressions (Wu et al., 2006). Cependant, l'expression de Sall4 n'est pas restreinte aux cellules ES et Nanog est toujours exprimée dans les cellules Sall4^{-/-} (Sakaki-Yumoto et al., 2006). Le rôle du complexe comprenant Sall4 et Nanog dans les cellules ES doit être étudié plus en profondeur. Il a été récemment montré que Sall4 est un régulateur important qui gouverne sa propre expression mais aussi l'expression d'Oct4 (Yang et al., 2010). Sall4 et Oct4 ont une action antagoniste qui régule l'expression des gènes de la famille Sall. Ce nouveau mécanisme de régulation d'Oct4 permet d'enrichir les réseaux déjà nombreux qui participent à la pluripotence des cellules ES.

Usp9x est un partenaire de Sall4. C'est un composant essentiel dans la voie de signalisation TGF β : il active Smad4 (Dupont et al., 2009) et il est capable d'interagir avec MLLT4 qui régule la protéine Ras (Radziwill et al., 2003). La mise en évidence de la liaison de Sall4 avec Usp9x montre que l'association des protéines *spalt* avec la voie de signalisation TGF β est aussi conservée chez les mammifères et plus particulièrement dans les cellules souches (van den Berg et al., 2010). D'autre part Usp9x interagit avec la voie Wnt. Premièrement, il interagit avec la protéine Cxx5 qui inhibe la voie Wnt (Andersson et al., 2009) et deuxièmement Usp9x interagit avec la β -caténine (Taya et al., 1999). Il est donc intéressant de noter que les nombreuses voies de signalisation sont intimement mêlées entre elles.

Jarid2. Cette protéine a été nommée à cause du phénotype causé lors de sa délétion par un gene-trap. En effet lors de la délétion de ce gène, la morphologie du sillon neural et des plates neurales a une forme de croix d'où le nom de jumonji qui signifie cruciforme en japonais (Takeuchi et al., 1995). Plusieurs domaines conservés sont présents: jmjN, ARID (un domaine de liaison aux séquences riches en AT) et un domaine jmjC. Le rôle de cette protéine sera développé dans le chapitre suivant.

D'autres protéines semblent être importantes mais non essentielles pour les réseaux de régulation. Par exemple, Ttrap (Herceg et al., 2001; Takahashi and Yamanaka, 2006) est impliqué dans un certain nombre de réseaux de régulation et notamment avec Ep400 et Tip60 (voir page 45).

E. Identification de nouveaux facteurs protéiques impliqués dans le renouvellement des cellules ES

Ce paragraphe est illustré d'exemples de cribles d'identification des protéines essentielles au maintien des caractéristiques des cellules ES

1. Autopsie des facteurs clefs du maintien de la pluripotence

Afin d'identifier d'autres facteurs importants dans la pluripotence des cellules ES, Ivanova et al. ont sélectionné par un crible sur puces d'expression les gènes affectés négativement par la différenciation induite par l'acide rétinoïque. Les auteurs ont confirmé l'importance de ces candidats par la méthode du shRNA (Ivanova et al., 2006). Dix gènes ont été identifiés: Nanog, Oct4, Sox2, Tbx3 (T-box transcription factor protein 3), Esrrb (Estrogen related receptor beta), Tcl1 (T-cell lymphoma breakpoint 1), Dppa4 (Developmental pluripotency associated 4) et Ccnb1ip1 (cyclin B1 interacting protein 1). Par la technique du shRNA, les auteurs ont confirmé la perte de la pluripotence pour la délétion de chaque gène étudié. Puis le retour à la pluripotence par l'expression du gène invalidé a été montré confirmant ainsi le rôle de la protéine dans la pluripotence des cellules ES. De façon surprenante, lors des expériences des shRNA, les voies LIF, BMP et Wnt n'ont pas été affectées alors que la voie ERK1/2 était activée après la délétion de l'ARN messager de Nanog, Sox2, Oct4 et Ccnb1ip1 indiquant une différenciation vers le type throphectoderme (Niwa et al., 2000). Les gènes dérégulés lors des expériences de shRNA se classent dans trois catégories différentes: la première inclut les gènes dérégulés de façon commune à tous les shRNA alors que les deux dernières catégories comprennent d'une part les gènes dérégulés par les shRNA d'Oct4, Sox2 et Nanog et d'autre part Esrrb, Tbx3 et Tcl1. Cette observation pourrait indiquer que certaines voies de régulations sont communes aux facteurs impliqués dans la pluripotence des cellules ES (par exemple les voies les plus en aval) alors que d'autres voies sont plus spécifiques à certains facteurs (les plus en amont).

Ivanova et al. ont posé la question du rôle de Nanog lorsque la différenciation des cellules ES était induite par d'autres facteurs importants pour la pluripotence. Les auteurs ont ainsi pu montrer que la différenciation entraînée par la déplétion des gènes Esrrb, Tbx3, Tcl1 et Dppa4 étaient inhibée par une surexpression de Nanog. Ce n'est pas le cas lors de la différenciation induite par la déplétion d'Oct4 ou Sox2. La surexpression de Nanog opère un sauvetage de l'expression d'Oct4 et Sox2 dans les expériences de shRNA contre Esrrb,

Tbx3 et Tcf1, qui permet de maintenir le phénotype ES. En revanche ceci n'est pas possible dans le cas des shRNA dirigés contre Oct4 et Sox2. Ceci suggère que les protéines contrôlant l'auto-renouvellement sont intégrées dans un réseau transcriptionnel complexe, comprenant plusieurs niveaux de régulation permettant des compensations mutuelles (ex: Nanog et Esrrb). Ainsi, la perte d'un facteur peut dans certains cas être compensée en ajustant le niveau d'expression des autres composants du réseau.

2. Un crible ARN interférence pour identifier un nouveau complexe impliqué dans la pluripotence des cellules ES

La récente découverte d'un certain nombre de facteurs indispensables à la maintenance de la pluripotence des cellules ES a conduit Fazio et al. à tenter d'identifier d'autres facteurs (Fazio et al., 2008). Pour cela, les auteurs ont utilisé un crible d'ARN interférence sur 1008 gènes. Soixante-huit gènes, pour la plupart non connus ou ne possédant pas de rôle connus dans les cellules ES, ont été reportés comme ayant un effet sur la croissance des cellules ES. Trois effets ont été observés: la mort cellulaire, l'altération de la morphologie des cellules ES et l'altération de la morphologie des colonies des cellules ES. Plus spécialement, 7 sous-unités du complexe Tip60-Ep400 ont été retrouvées. Les cellules délétées pour un des facteurs composant ce complexe continuent à exprimer les facteurs de la pluripotence mais perdent certaines caractéristiques propres aux cellules ES: activité phosphatase alcaline faible, perte de la capacité à former des corps embryonnaires et perte de la capacité à former des tératomes. De plus, les cellules invalidées pour les gènes *Tip60* et *Ep400* sont plus allongées et aplaties que les cellules ES normales. Les colonies sont quant à elles plus dispersées et poussent en monocouches. Ces données indiquent donc que le complexe fonctionnel Tip60-Ep400 est nécessaire pour le maintien des caractéristiques des cellules ES.

Une étude transcriptomique sur puce Agilent a permis de montrer que 802 gènes étaient différentiellement exprimés lors de la délétion de Tip60 et Ep400 (intersection des deux jeux de données) dont 128 sous-exprimés (gènes du cycle cellulaire, gènes du métabolisme et requis pour les divisions cellulaires) et 674 surexprimés impliquant un large groupe de gènes exprimés lors de la différenciation. Ces observations allèguent que le complexe Tip60-Ep400 est plutôt impliqué dans la répression des gènes dans les cellules ES. La délétion de Tip60-Ep400 n'affectant pas l'expression des facteurs de transcription majeurs du maintien de la pluripotence (Oct4, Nanog), les auteurs ont postulé que le complexe Tip60-Ep400 intervenait en aval de ces facteurs. En comparant les profils d'expression des ARN lors de l'inactivation des gènes *Tip60* et *Ep400* avec les signatures

déjà connus de Nanog, d'Oct4, Sox2 et Suz12, les auteurs ont pu montrer qu'il existait une corrélation importante entre les profils de Nanog et du complexe Tip60-Ep400. Ces données suggèrent que Nanog et Tip60-Ep400 agissent de concert pour réguler un ensemble de gènes communs.

Fazio et al. ont ensuite identifié les promoteurs cibles par CHIP-CHIP (puce NimbelGen) et ont montré que la protéine p400 se lie à un grand nombre de promoteurs et plus spécifiquement coïncide avec la marque H3K4me3 sur les gènes très exprimés en cellules ES et les promoteurs cibles des complexes Polycomb. Les auteurs ont montré que l'amarrage de la protéine Ep400 à la chromatine requiert la marque d'histone H3K4me3 ainsi que le facteur Nanog, en étudiant l'occupation de p400 sur ses promoteurs cibles dans un contexte pauvre en marque H3K4me3 ou Nanog^{-/-}. Cependant, s'ils partagent de nombreux promoteurs cibles communs, les séquences liées par Tip60-Ep400 et Nanog se chevauchent rarement suggérant une interaction de type indirect. Des données complémentaires montrent que la déposition des histones H2A.Z et H2A.X par le complexe Tip60-Ep400 n'est pas la cause de la perte des caractéristiques ES. *A contrario*, l'acétylation de l'histone H4 sur les promoteurs cibles du complexe est grandement altérée lors de la déplétion de Tip60-Ep400.

Les données fournies par Fazio et al. élargissent la liste de facteurs importants pour la régulation des cellules ES et définissent un rôle important pour le complexe Tip60-Ep400 dans le réseau transcriptionnel de Nanog.

3. Un rôle antagoniste pour deux protéines de la même famille

La protéine Satb1 est une des rares protéines connues pour l'organisation de la structure de la chromatine (Cai et al., 2006). Satb2, une protéine homologue à Satb1 par sa séquence et la conservation des domaines fonctionnels, est impliquée dans le développement de l'embryon (Dobrev et al., 2006). L'équipe de R. Grosschedl, s'est intéressée aux rôles de ces deux protéines dans les cellules souches (Savarese et al., 2009). Les auteurs ont ainsi pu mettre en évidence que l'expression de Satb1 est dominante par rapport à Satb2 dans les cellules ES. L'augmentation de Satb2 est plutôt observée au cours de la différenciation induite par l'acide rétinoïque. Grâce à la génération de lignées ES *Satb1^{-/-}*, il est apparu que cette protéine était nécessaire pour la différenciation des cellules. Une étude plus approfondie a montré que ces cellules ont une capacité fortement diminuée à la différenciation: l'expression des gènes associés à la différenciation (Nestin et Bcl2) n'est pas induite alors qu'au contraire l'expression des gènes importants dans la pluripotence Klf4 et Nanog ainsi que de Satb2, est plus importante que dans les cellules WT. Ces profils de

surexpression sont corrélés. De façon surprenante, ces cellules sont aussi capables d'être cultivées en absence de LIF tout en gardant le phénotype ES. La lignée *Satb2*^{-/-} n'a pas pu être obtenue par les auteurs, ce qui suggère que les cellules ES ne supportent pas l'expression de Satb1 en absence de Satb2. *A contrario*, les cellules ES *Satb1*^{-/-} *Satb2*^{-/-} ont un phénotype similaire aux cellules ES WT. Les auteurs ont ensuite étudié l'induction de l'expression de Satb1 et Satb2 mettant alors en évidence une similarité entre le phénotype *Satb1*^{-/-} et le phénotype correspondant à l'induction forcée de Satb2. Satb2 réprime l'expression de Satb1. En immunoprécipitant Satb1 et Satb2, les auteurs ont montré que ces deux protéines se liaient aux promoteurs de Nanog et Klf4. De plus les marques d'histones présentes à ces loci reflètent parfaitement les défauts d'expression observés sur Nanog et Klf4 au cours de la différenciation. Les auteurs ont donc proposé un schéma de régulation impliquant Satb1 et Satb2 dans le contrôle des gènes importants dans la pluripotence des cellules ES (Figure 12). De manière intéressante, l'absence de Satb1 ou la surexpression de Satb2, facilite la reprogrammation des cellules par fusion cellulaire. Ces données confirment donc bien les rôles opposés de Satb1 et Satb2.

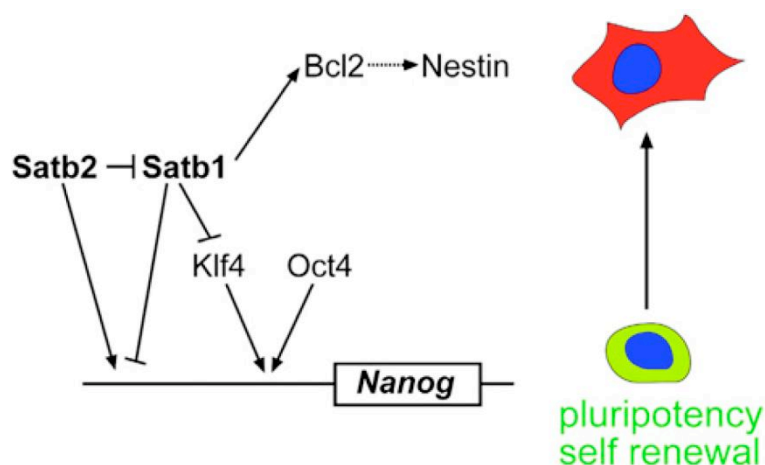


Figure 12: Réseau de régulation de l'expression du gène Nanog mis en évidence. Tiré de Savarese et al., 2009.

F. Identifications de réseaux de régulation en cellules ES

1. Un large réseau pour la pluripotence des cellules ES

Kim et al. se sont intéressés aux facteurs-clefs de la pluripotence formant le noyau des réseaux transcriptionnels ainsi que d'autres facteurs précédemment identifiés comme importants dans la pluripotence des cellules ES: Sall4 (Wu et al., 2006; Zhang et al., 2006),

Dax1 (Niakan et al., 2006) et Rif1 (Loh et al., 2006), trois facteurs qui interagissent avec Oct4 et Nanog: Nac1, Zfp281, Rex1 (Wang et al., 2006) et deux facteurs requis pour la reprogrammation des cellules Myc et Klf4 (Takahashi and Yamanaka, 2006). Afin de déterminer les cibles de ces facteurs, une étiquette biotine a été ajoutée à l'extrémité carboxy-terminale des facteurs de transcriptions d'intérêt. Puis les constructions ont été électroporées stablement dans les cellules ES (Kim et al., 2008a). Les échantillons immunoprécipités grâce aux billes liées à la streptavidine ont été analysés sur puce Affymetrix. Les données des neuf facteurs de transcription montrent que ceux-ci co-localisent sur plus de 14 promoteurs murins et qu'environ 800 promoteurs sont occupés par au moins quatre des neuf facteurs. Cependant, parmi les 6632 promoteurs cibles, 50% d'entre eux sont liés par un seul facteur. En utilisant le logiciel MEME (Bailey et al., 2006), les auteurs ont pu montrer qu'ils retrouvaient dans les promoteurs cibles la séquence consensus déjà identifiée dans la littérature. Puis, en observant le degré de corrélation des cibles entre les facteurs de transcription, les auteurs ont remarqué que les facteurs Nanog, Sox2, Dax1, Nac1 et Oct4 formaient un premier groupe alors que Myc et Rex1 en formaient un second. Cette observation est moins claire pour les facteurs Klf4 et Zfp281. Ces informations ont une pertinence biologique puisque le premier groupe est plus souvent trouvé lié à des gènes impliqués dans le développement alors que le deuxième groupe est retrouvé sur des gènes impliqués dans le métabolisme protéique.

Kim et al. ont cherché à savoir si les marques d'histones H3K4me3 et H3K27me3 pouvaient corrélérer avec les cibles de certains facteurs. Ils ont ainsi montré que Myc, Rex1 et Nac1 corrèlent plutôt avec la marque H3K4me3, Zfp281 corrèle avec la marque H3K27me3 alors que Nanog, Sox2, Dax1, Oct4 et Klf4 corrèlent avec les marques bivalentes. Plus particulièrement, la corrélation remarquable entre le facteur Myc et la marque H3K4me3 est en accord avec l'hypothèse qui postule que l'occupation de c-Myc est associée à de vastes changements dans l'accessibilité de la chromatine. De plus, Kim et al. observent que le marquage H3K4me3 corrèle avec un nombre important de facteurs liés alors que les promoteurs liés par un seul facteur de transcription portera préférentiellement la marque H3K27me3 (Figure 13). Contrairement à leurs attentes, les facteurs de transcriptions se lient à des promoteurs de gènes exprimés ou pas. Les données de Kim et al. ont été confrontées à celles de Wang permettant d'identifier 4 nouveaux foyers (Sall4, Rif1, Rest et Dax1) dans les réseaux de transcription (Figure M2). Ces données révèlent donc que l'ensemble des réseaux de régulation des cellules ES est plus grand et plus étendu par de nombreuses connections insoupçonnées.

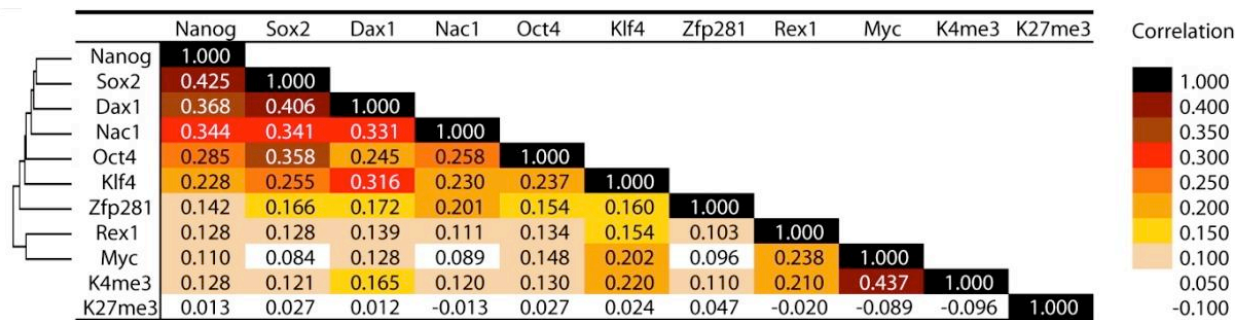


Figure 13: Facteur de co-occupation des différents facteurs impliqués dans la pluripotence, tiré de Kim et al. 2008

2. Intégration des voies de signalisation LIF et des protéines polycomb aux réseaux de la pluripotence

A l'instar de Kim et al., Chen et al. ont étudié 13 facteurs de transcription (FT) dans les cellules ES murines (Chen et al., 2008). Ces facteurs (Nanog, Oct4, Sox2, Klf4, E2f1, Esrrb, CTCF, n-Myc, c-Myc, Smad1, STAT3, Tcfcp2l1, Zfx) ont été choisis grâce à leur importance dans le l'auto-renouvellement des cellules ES. A cette liste ont été ajoutés les co-activateurs p300 et Suz12. Grâce à l'identification des sites de liaison de ces FT sur le génome par la méthode du ChIP-seq, les auteurs identifient la répartition des FTs sur le génome leur permettant de comprendre comment ils sont connectés entre eux (Figure M1). La méthode utilisée a consisté en une immunoprécipitation de chromatine avec les anticorps spécifiques à chaque protéine puis à un séquençage haut débit sur le Genome Analyser de Solexa/Illumina. Les auteurs ont isolé les 500 meilleurs pics et ont identifié les motifs spécifiques de chaque FT. Comme il avait été déjà démontré dans la littérature, ils ont retrouvés les consensus pour Oct2/Sox4 proches de ceux de Nanog et Smad1 qui co-localisent souvent avec Oct4/Sox2 (Loh et al., 2006). A l'instar de Kim et al., les auteurs ont trouvé un grand nombre de loci liés par plusieurs FTs (3583 liés par au moins 4 FTs). Cependant, ces loci liés correspondent pas préférentiellement aux promoteurs qui ne représente que 20% des régions liées par les FTs. En comparant les sites liés par les 13 TFs, Chen et al. ont pu montrer que ceux-ci ségrégent en deux catégories. La première comprend Nanog, Sox2, Oct4, Smad1 et STAT3, confirmant l'implication de la voie LIF avec le cœur des réseaux de régulation. La deuxième catégorie comprend les FTs E2f1, n-Myc, c-Myc et Zfx. Les FTs Tcfcp2l1 et Klf4 partagent des cibles communes avec les deux catégories. Grâce à une expérience de reporteur d'activité luciférase, il a été montré que les régions liées par Nanog/Oct4/Sox2 sont des régions enhanceurs particulièrement efficaces contrairement aux régions liées par Myc peu ou pas efficaces.

Grâce à des expériences d'ARN interférence, Chen et al. ont montré que le recrutement du co-activateur p300 sur la chromatine était dépendant des facteurs Oct4, Sox2 et Nanog mais pas de c-Myc. De plus le motif de liaison fréquemment retrouvé dans les cibles de p300 est très similaire à celui de ces 3 FTs. Une comparaison entre les jeux de gènes liés par les 13 FTs plus Suz12 a montré que 5 classes de gènes pouvaient être différenciées selon les FTs liés. De façon intéressante, il est apparu que Suz12 et E2f1 sont mutuellement exclusifs. En comparant les données d'expression de ces gènes en cellules ES et lors de la différenciation induite par de l'acide rétinoïque, les auteurs ont conclu qu'il existe des combinaisons de profils de liaison de FT spécifiques aux cellules ES, pouvant alors prédire quels gènes seront spécifiquement exprimés en cellules ES. Cependant, il n'existe aucune signature de liaison à l'ADN ("binding patterns") comme exclusivement spécifique aux cellules ES. Afin de proposer un réseau de régulation qui inclue les facteurs de transcription de cette étude, les auteurs ont corrélié les gènes liés par les 13 différents FTs plus le co-activateur Suz12 (nombre de séquences lues) à l'expression du gène dans les cellules ES. D'après les auteurs, les régions génomiques liées par de multiples facteurs sont principalement des régions enhanceurs. L'ensemble des facteurs liés à la même position génomique (locus identique) constituerait un « enhanceosome ». L'existence de cet ensemble protéique avait déjà été suggérée en 2005 (Arnosti and Kulkarni, 2005). L'ensemble Oct4/Sox2/Nanog formerait une partie de cet « enhanceosome ». La différence entre un enhanceur classique et un « enhanceosome » est la densité de FT plus forte sur un « enhanceosome » et les séquences liés qui sont plus courtes par rapport à un enhanceur (Chen et al., 2008). L'importance de la voie LIF et les facteurs BMP (via le serum) avait été précédemment démontrée car les cellules ES peuvent être maintenues indifférenciées en présence de LIF et de sérum. Les travaux de Chen et al. ont mis à jour la co-localisation des facteurs STAT3 (LIF) et Smad1 (BMP) avec les facteurs principaux des réseaux de régulation des cellules ES (Figure 14). Ces observations permettent de comprendre comment les voies LIF et BMP peuvent être liées au maintien de la pluripotence.

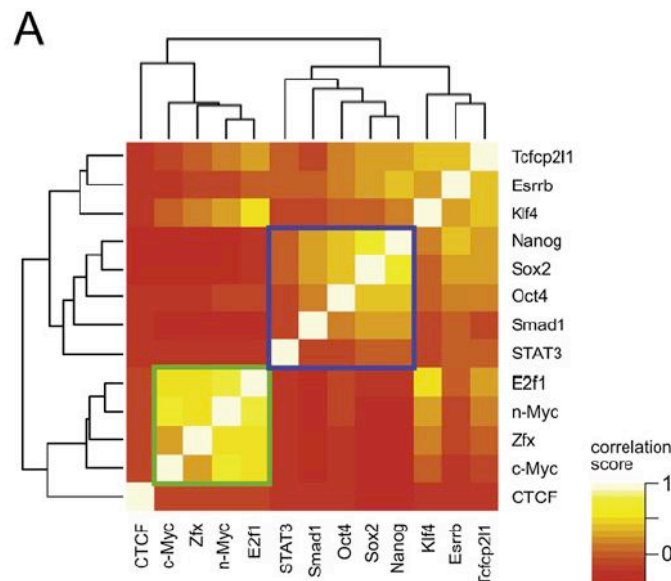


Figure 14 Représentation de la co-occupation des facteurs importants dans la pluripotence sur le génome murin tiré de Chen et al. 2008

Conclusion

Les deux articles de Kim et al. et Chen et al. montrent la puissance d'analyse des études génomiques à grande échelle. De plus le grand nombre de facteurs étudiés permet une vision globale sur l'ensemble des réseaux de régulation présents dans les cellules ES pour maintenir la pluripotence. Les nouvelles technologies de séquençage à haut débit ont permis une avancée sans précédent dans l'analyse des données et a rendu possible de telles études à l'échelle génomique.

3. Etude extensive des partenaires d'Oct4

De nombreuses études ont utilisé les méthodes du ChIP-on-Chip ou ChIP-seq pour identifier les séquences cibles d'Oct4 ou d'autres facteurs importants dans le maintien de la pluripotence des cellules ES. Pour identifier les partenaires d'Oct4, à l'instar de Wang et al. ainsi que de Liang et al., Pardo et al. ont utilisé une purification des complexes protéiques associés à Oct4 suivie d'une identification par spectrométrie de masse (Pardo et al., 2010). Pour ce faire, les auteurs ont utilisé une lignée étiquetée 3*FLAG-Tev-Tev-CBP (Calmodulin Binding Peptide) pour Oct4 sous le contrôle de son propre promoteur ce qui permet de combiner une immunoprécipitation efficace en respectant le taux d'expression naturel d'Oct4. Le vecteur contenant le gène Pou5f1 étiqueté a ensuite été inséré au locus *Hprt* dans les

cellules ES. Trois expériences indépendantes d'immunoprecipitation simple ont permis d'identifier 92 protéines incluant des protéines dont l'association avec Oct4 avait déjà été démontrée directement ou via l'association avec d'autres facteurs. La double purification a permis d'identifier 7 protéines avec lesquelles l'interaction est la plus forte. L'analyse des fonctions associées (www.geneontology.org) aux partenaires d'Oct4 a montré qu'Oct4 est associé avec de nombreux facteurs de transcription. De même, le logiciel PANTHER (www.pantherdb.org) a identifié un certain nombre de protéines impliquées dans la voie Wnt. Parmi les protéines identifiées on y trouve de nombreux domaines dont la fonction est impliquée dans le remodelage de la chromatine, la liaison à l'ADN ou aux nucléosomes arborant des modifications post-traductionnelles. L'expression des protéines liées à Oct4 est assez variée. De façon intéressante, près de 50% des protéines interagissant avec Oct4 sont mieux exprimées en cellules ES que dans des cellules différenciées. En confrontant leur données avec à celles de la littérature, les auteurs ont pu mettre en évidence que les promoteurs des partenaires d'Oct4 sont des cibles préférentielles (51% des gènes) pour les facteurs de transcriptions importants en cellules ES identifiés par Kim et al. Ces observations mettent à jour un réseau d'inter-régulation impliquant d'une part les protéines qui interagissent physiquement avec Oct4 et d'autre part les facteurs de transcription importants dans le maintien de la pluripotence des cellules ES. De façon attendue, la perte de fonction de la plupart de ces protéines (83%) est létale au niveau embryonnaire ou périnatale. De même, l'absence ou l'expression inappropriée des 92 protéines identifiées a été mise en cause dans certaines maladies génétiques et certains cancers. Ces données permettent de souligner l'importance des protéines identifiées dans l'interactome d'Oct4 dans les mécanismes développementaux et l'intégrité des cellules (Figure M4).

D'autres études importantes ont été menées afin de mieux comprendre les réseaux qui maintiennent les caractères des cellules ES et sont regroupées dans le [Tableau 2](#).

Table des articles étudiant les réseaux impliqués dans le maintien de la pluripotence des ES			
Article	Titre	Description	année
Pardo et al.	An expanded Oct4 interaction network: implications for stem cell biology, development and disease	Etiquetage d'Oct4 Purification des 92 protéines associées et identification par spectrométrie de masse Interactome d'Oct4	2010
Junget al.	A data integration approach to mapping Oct4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells	ChIP-on-Chip d'Oct4 dans des cellules embryonnaires humaines de carcinome et incorporations de leurs résultats dans les réseaux déjà connus	2010
Chen et al.	Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.	Etude par ChIP-seq de 13 facteurs de transcription et de deux régulateurs de transcription.	2008
Kim et al.	An extended transcriptional Network for pluripotency of embryonic stem cells	Etude par Chip-on-Chip	2008
Ivanova et al.	Dissecting self-renewal in stem cell with RNA interference	Etude par siRNA de 70 facteurs. Identification de 10 protéines importantes pour les ES	2006
Fazio et al.	An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity	Identification par RNAi de 68 protéines dont l'invalidation affecte le phénotype ES	2008
Boyer et al.	Core transcriptional regulatory circuitry in human embryonic stem cells	ChIP-on-chip d'Oct4, Sox2 et Nanog	2005
Zhou et al.	A gene regulatory network in mouse embryonic stem cell	Identification du réseau Oct4, Sox2 et Nanog et d'autres partenaires par analyse ChIP-chip	2007
Walker et al.	Prediction and Testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment	Identifications des gènes dérégulés lors de la différenciation (-LIF) ou (+RA) et identification de protéines importantes dans le réseau d'Oct4	2007

Tableau 2: Table non exhaustive des articles majeurs traitant des réseaux importants dans le maintien du phénotype des cellules ES.

4. Des réseaux de régulation à la rescousse de la pluripotence

De nombreuses recherches sont actuellement en cours afin de comprendre le rôle de chacun de ces facteurs. Cependant, actuellement, les données de fonction de chacune de ces protéines restent largement inconnues. De plus le manque de recul sur les études actuelles ne permet pas encore d'avoir une vision globale des nombreux réseaux de régulations mis en cause et les liens qui les relient. L'interdépendance des protéines et des réseaux de régulation auxquels elles appartiennent apporte une difficulté supplémentaire pour identifier le rôle de chacune dans le maintien de la pluripotence.

Cependant, quelques données ancrent certains points-clefs. Par exemple van den Berg et al. ont montré qu'Oct4 a un rôle-clef comme recruteur de protéines essentielles pour la pluripotence des cellules ES. Grâce à son motif de liaison de l'ADN, Oct4 joue le rôle d'ancre sur la chromatine et permet le recrutement de facteurs sur des sites particuliers.

Un exemple est donné avec le promoteur du gène Nanog. Plusieurs travaux ont montré l'importance de différents facteurs. Oct4 et Sox2 ont été identifiés comme protéines se liant au promoteur du gène Nanog (Rodda et al., 2005). Le fonctionnement de ce promoteur dépend de l'interaction d'Esrrb et d'Oct4 (van den Berg et al., 2010). Dax1 se lie via Oct4 au promoteur de Nanog (van den Berg et al., 2010). Nac1 se lie aussi au promoteur de Nanog (Kim et al., 2008a). En 2008, les équipes de Ng et Russo ont montré respectivement que Zfp142 et Klf4 régulaient l'activité du gène Nanog par la liaison sur son promoteur. Au final, six protéines sont liées au promoteur du gène Nanog, ce qui reflète une coordination multiple de facteurs de transcriptions qui coopèrent afin de réguler le niveau d'expression adéquat de Nanog. Fait notable, les facteurs de transcription Oct4, Sox2, Dax1, Klf4, Nac1, Esrrb et Nanog semble avoir des motifs consensus des sites de liaisons (déterminés par ChIP) assez semblables, suggérant qu'ils sont souvent présents aux mêmes sites. Ces données permettent de postuler qu'il existe un mécanisme de recrutement dépendant de la séquence ADN et des interactions protéines-protéines, comme observé sur le promoteur du gène Nanog.

5. Conclusion

De nombreux sites ont été trouvés liés par un grand nombre de facteurs (Chen et al., 2008; Kim et al., 2008a). Il est donc légitime se demander si cette séquence est réellement liée par de nombreux facteurs et de façon très spécifique ou alors s'il s'agit de fragments de chromatine ouverte où n'importe quel facteur possédant un domaine d'interaction à l'ADN

puisse venir se fixer. Une hypothèse postule que certaines régions génomiques ouvertes et souvent dépourvues de nucléosomes sont des territoires qui accueillent la plupart des FTs. Ces régions ont été identifiées entre autres par Chen et al. qui les nomment MTL pour Multiple Transcription Factor-Binding Loci (Chen et al., 2008). *A contrario*, certaines régions sont spécifiques de certains facteurs. Actuellement, les mécanismes impliquant ces régions génomiques qui ségrégent dans l'une ou l'autre des catégories ne sont pour l'instant pas élucidés. Le travail de Stamatoyannopoulos et al. sur les sites sensibles à la DNase1 pourrait être utiles afin de mieux comprendre la distribution des FTs sur la chromatine (Hesselberth et al., 2009).

De manière opposée, d'autres régions sont compactées dans l'hétérochromatine et on a particularité de lier peu ou aucun facteur.

G. Autres mécanismes importants dans la pluripotence des cellules ES

1. Rôle des micro ARN dans la pluripotence des cellules ES

Des études publiées en 2003 et 2004, ont montré l'expression spécifique de certains micro ARN (miRNA) dans les cellules souches embryonnaires (Houbaviy et al., 2003). L'inactivation de Dicer chez la souris, une protéine nécessaire à la maturation des ARN, est létale aux stades précoces du développement (neurula) et il n'est pas possible d'obtenir des cellules ES, ce qui suggère un rôle majeur de la machinerie d'ARN interférence pour la survie des cellules ES (Bernstein et al., 2003). Les premiers exemples de miRNA importants dans le maintien de la pluripotence des cellules ES humaines et murins ont été les gènes makorin (antisens de RAF1) et TSIX (antisens de XIST) (Rao, 2004). La dérégulation de ces gènes exprimés aux stades très précoces du développement, altère celui-ci. Afin de mieux comprendre le rôle des miRNA dans les cellules ES, Marson et al. ont proposé d'intégrer certains miRNA importants pour la pluripotence, dans la carte des réseaux de régulations des cellules ES, et ceci à l'échelle du génome entier (Marson et al., 2008b). Cette cartographie des sites de liaisons a été générée par immunoprécipitation de chromatine couplée au séquençage massif (ChIP-seq) de certains facteurs de transcription-clefs tels que Oct4, Sox2, Nanog et Tcf3. Ces données ont montré que les facteurs de transcriptions étudiés occupaient un grand nombre de promoteurs de miRNA. En confrontant ces données avec les données transcriptomiques disponibles sur les cellules ES, les précurseurs neuronaux et les MEFs, Marson et al., ont pu montrer que les miRNA dont le promoteur était lié par Oct4/Sox2/Nanog et Tcf3 étaient 100 fois plus abondants que dans les MEFs et 1000

fois plus abondants que dans les précurseurs neuronaux (Marson et al., 2008b). Cependant, un quart des miRNA liés par les facteurs de transcription sont peu ou pas exprimés en cellules ES. Les promoteurs de ces miRNA sont enrichis en marque d'histone H3K27me3, et la présence de la protéine Suz12 a permis de confirmer la présence des marques répressives. Dans la plupart des cas, ces miRNA sont exprimés spécifiquement dans les MEFs et/ou les précurseurs neuronaux. La présence des protéines polycomb sur les miRNA permettrait une levée rapide de l'inhibition de leur transcription, puis la production rapide de miRNA conduisant à une rapide différenciation grâce à la dégradation des inhibiteurs de différenciation. Les miRNA contribuent ainsi à l'identité des cellules ES par différentes voies:

1) pour ajuster de façon très fine l'expression de certains gènes comme Lefty1 et Lefty2 impliquées dans les réseaux qui contrôlent la pluripotence (Nakatake et al., 2006) ainsi que des *de novo* DNA méthyltransférases (Dnmt3a et Dnmt3b) nécessaires pour la différenciation (Sinkkonen et al., 2008),

2) à maintenir les cellules dans l'état pluripotent qui est un perpétuel équilibre pouvant aisément basculer dans une voie de différenciation.

Cette dernière capacité est particulièrement bien illustrée par l'exemple des cellules où la protéine Dicer est réduite par ARN interférence où les cellules qui sont incapables de se différencier même en conditions adéquates. Sinkkonen et al. ont montré que les défauts de différenciation étaient liés à une méthylation impropre du promoteur d'Oct4, due à l'absence des produits du cluster miR-290 nécessaire à l'expression de *de novo* DNA méthyltransférases. De même, certains miRNA surexprimés au cours de la différenciation induite par l'acide rétinoïque inhibent l'expression de Sox2, Oct4 et Nanog (Tay et al., 2008). Les miRNA sont aussi impliqués dans la régulation du cycle cellulaire et dans la reprogrammation des iPS (Mallanna and Rizzino, 2010).

2. La méthylation de l'ADN est un lien entre épigénétique et pluripotence

De nombreux travaux ont montré que la méthylation de l'ADN dans les cellules de mammifères joue de multiples rôles dans la stabilité du génome, la répression de certains éléments rétroviraux et transposons, l'empreinte génétique, et la régulation de l'expression des gènes au cours du développement. La régulation dynamique de la méthylation de l'ADN se fait grâce aux *de novo* méthyltransférases (Dnmt3a et 3b), la conservation de la méthylation se faisant par l'enzyme Dnmt1. Alors que les cellules ES déficientes pour la méthylation de l'ADN peuvent survivre à un stade indifférencié (Meissner et al., 2005), ce

mécanisme est nécessaire pour la survie des cellules ES au cours de la différenciation. Dans les cellules ES murines, chaque Dnmt joue un rôle spécifique dans l'établissement et/ou la maintenance de la méthylation de l'ADN. Fouse et al. rapportent en 2008 une étude sur la méthylation des promoteurs des gènes dans les cellules ES murines (Fouse et al., 2008). Ils ont isolé les fragments d'ADN méthylés par immunoprecipitation à l'aide d'anticorps dirigés contre les cytosines méthylées ou des colonnes de purification contenant des domaines se liant à l'ADN (issus de la protéine MeCP2) puis identifié les fragments par hybridation sur puce (mDIP-Chip). Dans les cellules ES de souris, les auteurs ont observé que les régions CpG (riches en nucléotides C et G) avaient tendance à ne pas être méthylées, et inversement, que les régions méthylées sont pauvres en CpG. Environ 6000 gènes dont les promoteurs sont méthylés ont été identifiés. Une analyse de la fonction de ces gènes a montré qu'ils sont globalement impliqués dans des mécanismes post-différenciation alors que les 5000 gènes dont le promoteur n'était pas méthylé regroupaient plutôt des gènes essentiels aux cellules ES tels que Oct4, Sox2, Nanog et Stat3.

Les auteurs ont ensuite montré que 87% des promoteurs qui ne possédaient pas les marques épigénétiques H3K4 et H3K27 étaient méthylés et non exprimés en cellules ES. Leurs résultats en ce qui concerne les gènes porteurs des marques H3K4me3 et/ou K27me3 permettent de conclure que la méthylation de l'ADN n'est pas généralement associée à une marque épigénétique spécifique mais consisterait elle-même une marque épigénétique à part entière (plutôt répressive). Venant renforcer cette hypothèse, les auteurs n'ont pas trouvé de corrélation entre la présence des facteurs de transcription Oct4/Nanog ou les protéines PcG et la présence d'ADN méthylé. Dans les cellules ES délétées en Dnmt, les gènes les plus surexprimés sont des gènes spécifiques de la lignée germinale qui sont préférentiellement localisés sur le chromosome X. De plus dans le contexte sauvage, ces promoteurs de gènes ne possèdent souvent pas de marques épigénétiques H3K4me3 et/ou K27me3 ni de site de liaison pour un facteur de transcription important dans la pluripotence. Toutes ces données convergent pour définir la méthylation de l'ADN comme un mécanisme de régulation participant à la pluripotence des cellules ES, indépendant des réseaux déjà connus.

Meissner et al. ont montré quelques mois plus tard que le séquençage de l'ADN méthylé est possible grâce à la méthode de séquençage haut débit de l'ADN génomique traité au bisulfite (Meissner et al., 2008). Si la plupart de leurs résultats confirment les hypothèses de Fouse, leurs données permettent de montrer que la méthylation est plus précisément guidée par les marques épigénétiques que par la séquence d'ADN intrinsèque.

En effet les séquences qui possèdent la marque H3K4me3 et qui ne sont pas enrichies en marque H3K9me2/3 sont de façon très prédictibles des régions non méthylées.

Très récemment, Thomson et al. ont montré que la protéine Cfp1 s'associe aux régions denses en di-nucléotides CpG et non méthylées, dans le cerveau de souris adulte. Cfp1 est capable d'induire la déposition de certaines marques épigénétiques (via vraisemblablement le recrutement des protéines PcG, comme l'avait auparavant suggéré Meissner (Meissner et al., 2008; Thomson et al., 2010).

Enfin, la régulation et la dynamique de la méthylation de l'ADN est non seulement importante pour le phénotype des cellules ES, mais est aussi un mécanisme dérégulés dans de nombreux cancers (Gu et al., 2010).

Chapitre III. Pluripotence et épigénétique

De récentes études ont montré que les cellules ES murines et humaines possédaient de nouvelles caractéristiques épigénétiques. Certaines marques d'histones ont été identifiées uniquement dans les cellules ES. Un exemple particulièrement frappant est la découverte de promoteurs bivalents ou "poised genes" qui possèdent des marques épigénétiques des gènes transcriptionnellement actifs (H3K4me3) et réprimés (H3K27me3). Ces derniers possèdent donc une activité transcriptionnelle très faible ou nulle mais la présence simultanée de ces deux marques est une spécificité de certains gènes des cellules ES. La présence de ces marques est catalysée par deux groupes d'enzymes, les protéines Trithorax et les protéines Polycomb (Tableau 3).

A. Présentation des protéines Polycomb et Trithorax

Les protéines du groupe Polycomb (PcG) et Trithorax (trxG) ont été identifiées initialement chez *Drosophila melanogaster*. Elles inhibent ou activent l'expression des gènes homéotiques (gènes HOX), qui forment une famille de facteurs de transcription spécifiant l'identité cellulaire le long de l'axe antéro-postérieur chez les animaux segmentés (Anderson et al., 1985; Schuettengruber et al., 2007). Chez *D.melanogaster*, le profil d'expression des gènes homéotiques est établi dans les trois premières heures du développement. Il a été montré que les protéines PcG et trxG formaient d'importants complexes multimériques qui ne sont pas requis pour l'initiation de la transcription des gènes HOX, mais pour maintenir leur état de transcription après l'étape initiatrice de transcription. Ce mécanisme permet des états actif et inactif constamment en équilibre dynamique. Un changement PcG/trxG pourrait jouer un rôle important dans les transitions développementales (Buszczak and Spradling, 2006; Maurange et al., 2006; Ringrose, 2006). Chez *D.melanogaster*, des séquences ADN régulatrices capables de recruter les protéines PcG et trxG appelées PREs et TREs respectivement, permettent d'interpréter les marques épigénétiques tout au long du développement de l'animal (Muller and Kassis, 2006; Schwartz and Pirrotta, 2007). La mise en évidence de l'existence des PREs chez les vertébrés a été tardive et a permis de montrer que ces séquences, bien que relativement courtes, sont très conservées (Sing et al., 2009). Cependant cette étude n'exclut pas d'autres moyens de recrutement des PcG. Plus particulièrement, les deux complexes polycomb PRC1 et PRC2 pourraient avoir des moyens de recrutement différents.

Complexes PcG et trxG					
	<i>Drosophila melanogaster</i>	<i>Homo sapiens</i>	<i>Mus musculus</i>	Fonction biologique	
Complexes PcG					
PhoRC	dSfmbt		Mbtd1, L3mbtl2		
	Pho/PhoL	YING-YANG1-2	Ying-Yang1-2	liaison à l'ADN (PREs)	
PRC2	E(z)	EZH1-2	Ezh1-2	Triméthyle H3K27	
	Esc	EED	Eed		
	Su(z)12	SUZ12	Suz12		
	Pcl	PHF1	Phf19, Mtf2		
	N55	RpAp46			
		RpAp48			
PCR1	dRing	RING1A	Ring1a	Ubiquitine H2AK119	
		RING1B	Ring1b		
	Pc	HPC1-3	M33, Mpc2, PC3	Liaisons au lysines méthylées	
	Ph	HPH1-3	Mph1-3		
	Psc	BMI1	Bmi1		
		MEL18	Zfp144, Znf134		
	Scm	SCMH1-2	Scml1-2		
	Pcl	PHF1, MTF2, PHF19	Phf12, Phf19, Mtf2		
		facteurs associés à TBP			
	Complexes trxG				
SWI/SNF	Brm	BRM	Smarca4 Smarca2	liaison aux histones acétylées	
	Osa	BAF250	ARID1B		
	Moirra	BAF170	Smarcc1		
	Snr1	BAF47			
NURF	lswi	SNF2L	Smarca5	liaison aux histones acétylées	
	N38				
	N301	BPTF			
	N55	RpAp46			
RpAp48					
TAC1	Trx		Wbp7		
	dCBP				
	Sbf1				
Ash1	Ash1				
	dCBP				
MLL1-3		MLL1-3			
		WDR5			
		ASH2L	Ash1-2		
		RbBP5			
		CFP1			
largement inspiré de Schuettengruber et al. Cell 2007;128:735-745					

Tableau 3: Complexes protéiques Polycomb et Tritorax

1. Mécanisme de répression par les protéines Polycomb

Les complexes Polycomb (PcG) agissent principalement pour stabiliser une structure de chromatine répressive. Le complexe de répression Polycomb 2 (PRC2) qui inclut les membres Ezh2, Eed, et Suz12 sont des méthyltransférases d'histones spécifiques de la lysine 27 de l'histone H3, une marque épigénétique qui est généralement associée aux gènes transcriptionnellement inactifs (Cao and Zhang, 2004; Schwartz et al., 2006). Bien que la répression transcriptionnelle au cours du développement soit la fonction principale des complexes PcG, peu de choses sont connues sur la façon dont cette répression est réalisée. *In vitro* les complexes PRC1 de la drosophile et de l'homme inhibent le remodelage de la chromatine et un complexe, recomposé *in vitro* avec les protéines PC, PSC, RING et PH, induit la compaction des nucléosomes (King et al., 2002; Levine et al., 2002). Le complexe PRC1 pourrait donc participer à l'activité de répression en formant une structure compacte (Francis et al., 2004). Cependant, les tentatives pour détecter une réduction de l'accessibilité de la chromatine des gènes réprimés par les protéines PcG ont donné des résultats contradictoires (Bender and Fitzgerald, 2002; Fitzgerald and Bender, 2001). Cela indique que, si ce mécanisme participe à la répression des gènes mais n'est pas unique. De plus, l'établissement d'une chromatine figée par les complexes PcG serait difficile à concilier avec le fait que les protéines PC et PH ont une liaison très dynamique avec la chromatine (Ficz et al., 2005).

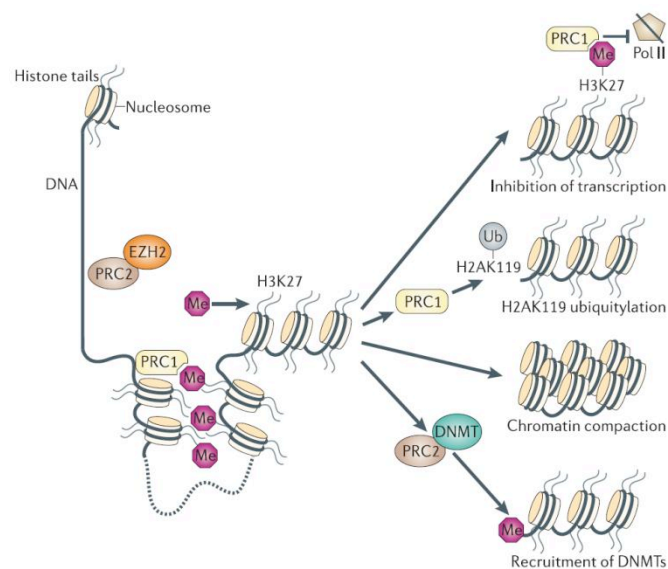


Figure 15: Différents mécanismes de répression de la transcription par les protéines Polycomb tiré de Sparmann et al. 2006

Les complexes PcG pourraient plutôt fonctionner en altérant la liaison de l'ARN polymérase II ou en empêchant l'initiation de la transcription. Dans ce cas, la répression nécessite un rapprochement physique entre les protéines PcG et les promoteurs des gènes cibles. Or les PREs sur lesquels se fixent les protéines PcG peuvent être localisés à une dizaine de kilobases des promoteurs. Différents modèles pourraient expliquer un effet des protéines PcG sur les promoteurs.

Dans un premier modèle, les protéines PcG se positionnent sur la chromatine des PREs jusqu'aux promoteurs. La présence des protéines PcG et les modifications de la chromatine qu'elles induisent inhibent la transcription. Ce modèle peut concorder avec la propagation des protéines PcG sur de longues distances chez les mammifères mais n'est pas conciliable avec la restriction des protéines PcG au niveau des PREs chez la drosophile.

Un second modèle explique le rapprochement des protéines PcG et des promoteurs par la formation de boucles de chromatine. Dans ce modèle, les protéines PcG localisées au niveau des PREs interagissent directement avec les promoteurs. Ce modèle concorde avec l'observation que des cibles différentes des protéines PcG se retrouvent à proximité dans des corps PcG, qui sont des compartiments nucléaires enrichis en protéines PcG. Ainsi, l'association de différentes cibles de protéines PcG au niveau des corps PcG pourrait augmenter leur répression. Le second modèle propose que les protéines PcG recrutent les PREs au niveau des corps PcG. Les promoteurs interagissent avec les protéines PcG et avec d'autres cibles des protéines PcG. L'association entre les cibles des protéines PcG est dépendante des protéines de la machinerie d'ARN interférence qui colocalise avec les protéines PcG au niveau des corps PcG (Bantignies et al., 2003; Grimaud et al., 2006). La machinerie d'ARN interférence pourrait donc être impliquée dans la répression par les protéines PcG. Cela concorde avec le fait que l'homologue humain de la protéine de la machinerie d'ARN interférence AGO1 peut induire la répression transcriptionnelle par de petits ARNs via le recrutement de la protéine PcG EZH2 (Kim et al., 2006).

Des ARNs non codants pourraient également participer au maintien de la répression. En effet, les régions intragéniques de régulation du complexe bithorax sont transcrites de façon dépendante de la position et du stade de développement. Ces ARNs non codants seraient de bons candidats pour participer à la régulation de la transcription. Cependant, pour l'instant, les résultats des différentes études réalisées sur ces ARNs sont contradictoires (Petruk et al., 2006; Sanchez-Elsner et al., 2006). Des études sont donc encore nécessaires pour comprendre les mécanismes permettant la répression transcriptionnelle par les protéines PcG. La [Figure 15](#) regroupe les différents mécanismes possibles pour réprimer les gènes (Sparmann and van Lohuizen, 2006).

2. Mécanisme de fonctionnement des protéines Trithorax

Les protéines Trithorax ont une activité antagoniste de celle des protéines PcG. Elles maintiennent l'état actif des gènes au cours du développement. Une première classe de membres TrxG est composée des protéines à domaine SET comme les protéines TRX et ASH1 de drosophile, ou MLL des vertébrés. Cette classe comprend un ensemble de protéines pouvant modifier les histones, comme MLL1-3 qui sont des Histones acétyltransférases capables d'apporter la modification H3K4me3. Une seconde classe comprend des éléments de complexes de remodelage de la chromatine.

En l'absence de la protéine TRX, les gènes homéotiques sont réprimés par les protéines PcG même dans les cellules où ils ont été activés chez l'embryon. Si la répression par les protéines PcG est altérée et TRX est mutée, l'expression des gènes cibles cellules ES normale. La protéine TRX fonctionne donc comme un anti-répresseur plutôt que comme un activateur. Chez la drosophile, TRX lie les PREs de façon constitutive, que les gènes soient actifs ou non, mais la façon dont elle bloque la répression n'est pas connue (Kahn et al., 2006; Papp and Muller, 2006).

B. Les protéines PcG et TrxG et la pluripotence des cellules ES

1. Les protéines Polycomb dans la pluripotence

En général, la présence de la marque répressive H3K27 et celle de la marque activatrice H3K4me3 sont mutuellement exclusive (Lund and van Lohuizen, 2004; Strahl and Allis, 2000). Cependant Bernstein et al. ont montré que dans les cellules ES murines ces marques colocalisent dans des régions particulières nommés "domaines bivalents" (Bernstein et al., 2006). Les promoteurs bivalents consistent en une large région génomique enrichie en marque H3K27me3 et ponctuellement en marque H3K4me3. Les régions premièrement identifiées par Bernstein et al. en 2006 étaient pour la plupart pauvres en gènes mais contenait des facteurs de transcriptions importants dans le développement. Il a donc été proposé que les deux marques puissent être utilisées pour brouiller la machinerie de transcription et garder les gènes dans un état silencieux ou "latent" en attendant la différenciation (Figure 16).

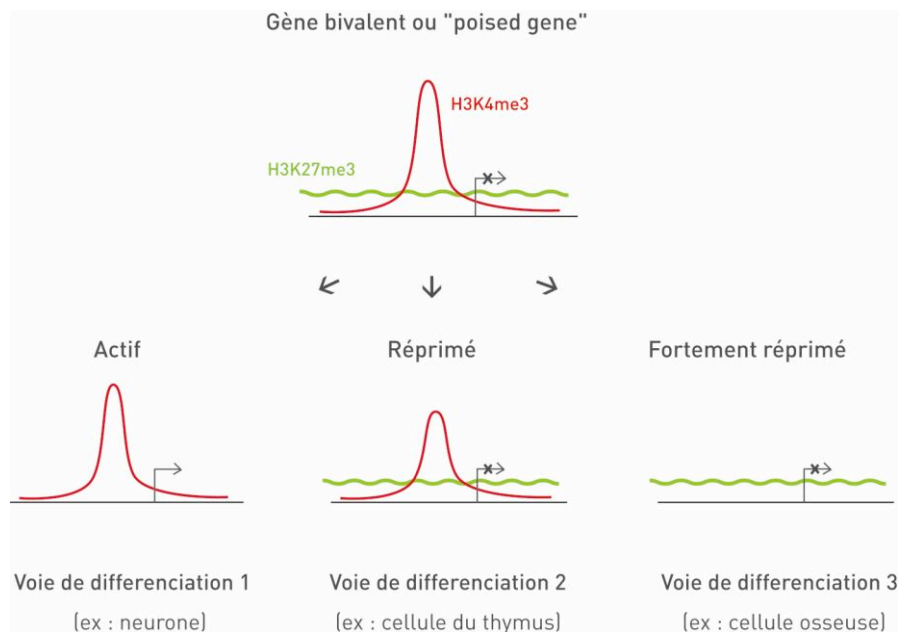


Figure 16: Schéma d'un gène bivalent différentiellement exprimé selon certaines voies de différenciation

En analysant les séquences ADN sous-jacentes, l'équipe de Lander a pu montrer que l'information épigénétique des domaines bivalents était largement supportée par une information génétique codée sur la séquence ADN elle-même (Bernstein et al., 2006). Cette analyse suggère donc que la séquence ADN définit largement le statut épigénétique des cellules souches qui est significativement altéré durant la différenciation, vraisemblablement en réponse aux programmes d'expression des lignages spécifiques des gènes et aux signaux environnementaux. Pionniers sur l'existence des marques bivalentes, ils ont montré que certaines régions génomiques possèdent les marques activatrices et répressives (Bernstein et al., 2006). Ces régions contiennent des facteurs de transcription importants dans le développement (gènes Sox, gènes Pou, gènes Hox, Pax, Wnt8b...etc..). Pour la plupart des gènes se trouvant sur des domaines bivalents, les auteurs ont trouvé que la marque répressive H3K27me3 l'emportait sur la marque activatrice. Ainsi les gènes se trouvant dans les domaines bivalents sont faiblement exprimés. Dans les cellules différenciées, peu de régions bivalentes sont retrouvées, l'une des deux marques (activatrice ou inhibitrice) l'emportant sur l'autre. Les auteurs proposent un modèle où les domaines bivalents sont spécifiques aux cellules souches et disparaissent au cours de la différenciation des cellules ES.

L'équipe de Lander a ensuite montré que les marques d'histones sur l'histone H3K4me3 sont corrélées à ~90% aux promoteurs riches en CpG. En revanche, les régions riches en marques H3K27me3 sont corrélées aux régions peu denses en transposons. Ces

régions excessivement bien conservées au cours de l'évolution, contiennent la vaste majorité des facteurs de transcription tissus-spécifiques et impliqués dans le développement. Les éléments non codants et hautement conservés semblent être des régions régulatrices importantes. Enfin, les données de position des facteurs Oct4 et Nanog sont corrélés avec 50% des domaines bivalents (Bernstein et al., 2006; Boyer et al., 2005). Les gènes qui sont liés par les facteurs de la pluripotence et qui possèdent les marques des promoteurs bivalents sont cependant peu exprimés. Les 50% autres domaines bivalents sont peut-être associés avec d'autres facteurs de transcription qu'il reste à découvrir. Pour cette raison les domaines bivalents sont généralement considérés comme des régions-clés qui maintiennent la pluripotence dans les cellules en réprimant l'expression des gènes de différenciation. Lee et al. ont identifié par ChIP-on-Chip les cibles de Suz12, Eed et H3K27me3. Ils ont montré que régions liées par Suz12 et Eed coïncident entre elles et avec les marques H3K27me3 sur les gènes pas ou peu actifs transcriptionnellement. Le groupe polycomb 2 (PRC2) est donc considéré comme un régulateur-clé car il recrute la marque répressive H3K27 au niveau des promoteurs bivalents (Lee et al., 2006). Boyer et al. ont identifié les cibles communes aux groupes PRC1 et PRC2 par ChIP-on-Chip et ont remarqué la co-localisation avec la marque H3K27me3 (Boyer et al., 2006). De plus, 87% des gènes étaient surexprimés en absence du complexe PRC2 dans un contexte Eed^{-/-}, ce qui renforce l'implication de PRC2 pour la répression génique (Boyer et al., 2006).

C'est pourquoi ces domaines pourraient marquer une structure de chromatine particulière où les gènes seraient prêts à être exprimés après réception d'un signal de différenciation (Niwa, 2007a), comme le propose Szutoritz et Dillon dans leur "Localised Marking Model" (Szutorisz and Dillon, 2005). Ainsi la plupart des gènes posséderaient les marques H3K4me3 mais aussi H3Ac et H4Ac sur leurs promoteurs. Cependant, dans les cellules ES les gènes spécifiques à la différenciation posséderaient en plus la marque H3K27 qui permettrait la répression (Niwa, 2007a). Ces données suggèrent une répression dynamique des différentes voies de développement par des processus épigénétiques requis pour la pluripotence des cellules ES. De façon surprenante lors de la délétion des deux composants Eed et Suz12, la surexpression des gènes de la différenciation est détectée mais n'abolit pas l'auto-renouvellement des cellules ES indiquant que leur fonction n'est pas essentielle dans la pluripotence (Montgomery et al., 2007; Pasini et al., 2007). De même, les cellules ES ne possédant plus la protéine Ring1 (PRC1) sont aussi viables malgré une régression du taux d'H2A ubiquitinyllée (de Napoles et al., 2004). Ces résultats prouvent que les protéines polycomb et les complexes PRC1 et PRC2 ne sont pas requis pour le maintien de la pluripotence.

2. Différence entre PRC1 et PRC2

Afin de mieux comprendre la structure et la fonction des domaines bivalents, l'équipe de Bernstein a identifié par immunoprécipitation suivi d'un séquençage haut débit (ChIP-seq) les sites de liaisons des marques d'histones H3K4me3, H3K27me3 et H3K36me3 et de protéines intervenant dans les complexes PcG PRC1 (Ring1b, puis confirmé par la présence de Bmi1) ainsi que PRC2 (Ezh2 et Suz12) dans les cellules ES murines et humaines (Ku et al., 2008). Les auteurs ont montré que les domaines bivalents ségrégent en deux classes. La première classe comporte les domaines liés par PRC1 et PRC2 alors que la deuxième classe comporte les domaines spécifiquement liés par PRC2 seulement. Il est apparu que les domaines bivalents qui étaient liés par PRC1 conservaient plus efficacement la marque H3K27me3 au cours de la différenciation. Par ailleurs ils semblent également être plus conservés entre les cellules ES espèces et sont associés avec un nombre conséquent de gènes impliqués dans le contrôle du développement. *A contrario*, les domaines bivalents occupés par PRC2 uniquement sont peu stables et correspondent à des assortiments de gènes qui ne sont pas impliqués dans le développement et suggèrent des processus de régulation différents de PRC1. Le complexe PRC1 semble donc avoir une fonction légèrement différente de PRC2. L'identification des sites de liaison spécifiques des complexes PRC2 et PRC1 a permis de mettre en évidence que l'association préférentielle de PRC2 avec les séquences riches en GC. De plus, ces résultats montrent que les domaines liés par PRC2 sont caractérisés par une surreprésentation des motifs répressifs et une forte réduction des séquences associées à l'activation de la transcription. Au contraire les domaines liés par PRC1 n'ont pas de motif particulier si ce n'est qu'ils englobent souvent plusieurs sites liés par PRC2 et sont deux fois plus étendus. Ku et al. suggèrent alors que l'état chromatinien des promoteurs en cellules ES soit par défaut bivalent, ce qui permettrait une rapide activation ou extinction transcriptionnelle du gène (Ku et al., 2008). De nouvelles études doivent être menées afin de comprendre comment la spécificité tissulaire permet de déterminer le recrutement des PcG et trxG sur des promoteurs dont la séquence ADN reste inchangée.

3. H2AZ.Z et PRC2

Creyghton et al. ont montré que le variant d'histone H2A.Z et le complexe PRC2 (Suz12) avaient une occupation commune des promoteurs de gènes bivalents dans les cellules ES (Creyghton et al., 2008). Le rôle suggéré de ce variant, retrouvé fréquemment dans les régions d'hétérochromatine (Dhillon and Kamakaka, 2000; Meneghini et al., 2003) serait d'apporter une souplesse à la chromatine condensée et ainsi permettre une

réactivation rapide. A la lumière du rôle des promoteurs bivalents dans les cellules ES, la présence du variant H2a.Z au promoteur des "poised genes" permettrait de confirmer les hypothèses fonctionnelles proposées. Très récemment, Persson et Ekwall ont proposé un mécanisme de remodelage de la chromatine impliquant H2a.Z chez *Saccaromyces pombe* qui selon eux pourrait aussi être valide dans les cellules ES. D'après les auteurs, la présence de nucléosomes portant le variant H2a.Z maintiendrait un état plutôt réprimé à proximité des promoteurs (avec la présence de marques bivalentes), puis, la protéine Chd1 pourrait remodeler la chromatine en retirant les nucléosomes contenant ce variant permettant la reprise de la transcription par l'ARN polymérase II et un retour à l'état actif. Enfin, l'incorporation du variant H2a.Z à la chromatine par la protéine le complexe Swr1c permettrait le rétablissement de l'état réprimé, créant ainsi un cycle d'activation/répression des gènes "poised" (Persson and Ekwall, 2010). Cette hypothèse est cependant contradictoire avec la détection de dimères H2a.Z/H3.3 puisque H2a.Z serait plutôt associé avec une chromatine réprimée et H3.3 une chromatine active. De plus, il existe un paradoxe entre la détection de ce dimère et le rôle de Chd1 qui déposerait l'histone H3.3 et enlèverait l'histone H2a.Z, les deux protéines n'étant donc jamais en présence l'une de l'autre d'après cette hypothèse. Toujours selon Persson et Ekwall, il s'agirait d'une forme transitoire.

4. Le rôle de Jarid2 et du complexe PRC2 dans les cellules

ES

Dans les cellules ES de souris, le complexe PRC2 n'est pas indispensable pour le maintien de la pluripotence. Cependant, les cellules *Eed*^{-/-} s'engage vers les voies de différenciation plus facilement (Boyer et al., 2006). Les cellules *Suz12*^{-/-} forment des corps embryonnaires désorganisés (Pasini et al., 2007). Malgré de nombreuses études, le moyen par lequel PRC2 atteint ses séquences cibles reste mystérieux. Une unique protéine AEBP2 qui a été co-purifiée avec PRC2 (Cao et al., 2002) possède un domaine en doigt de zinc et pourrait contribuer à la reconnaissance de certaines cibles ADN. De même la sous-unité *Eed* peut reconnaître certaines lysines triméthylées. L'équipe de Reinberg s'est attachée à comprendre comment le complexe PRC2 pouvait être recruté (Li et al., 2010a). Après purification de ce complexe, les auteurs ont identifié deux nouvelles protéines *Jarid2* et *MTF2*. *Jarid2* est une protéine nucléaire essentielle pour l'embryogenèse de la souris. Elle contient un domaine de liaison à l'ADN (ARID), un domaine en doigt de zinc, un domaine jumonji N et un domaine *JmjC*. De façon surprenante, contrairement aux autres protéines possédant un domaine *JmjC*, l'activité déméthylase n'est pas conservée du fait d'une mutation.

Jarid2 interagit préférentiellement avec la sous-unité Ezh2 de PRC2 et stimule son activité méthyltransférase grâce à son domaine en doigt de zinc et JmjC. Les auteurs ont montré que le recrutement de Jarid2 induit le recrutement d'Ezh2 et augmente le niveau de méthylation H3K27me2/3. En utilisant la méthode du ChIP-seq, Li et al. ont montré que Jarid2, MTF2 et Ezh2 possèdent de nombreuses cibles communes dans le génome (Li et al., 2010a). En réduisant le niveau d'expression de Jarid2 par ARN interférence, la présence d'Ezh2 sur ces cibles cellulaires est affectée de façon locale, spécifiquement sur certains gènes. Cependant à l'échelle du génome, le niveau de méthylation global de H3K27me3 reste in affecté indiquant que de nombreuses voies de régulations sont mises en jeu. De manière réversible, l'appariement de Jarid2 sur ses séquences cibles cellulaires ES fortement diminué dans un contexte *Eed*^{-/-}, d'où l'hypothèse proposée par Li et al. de l'importance de l'intégrité du complexe PRC2 pour sa stabilité. L'absence de Jarid2 contrecarre les processus de différenciation des cellules ES. Li et al. concluent en proposant que le complexe PRC2 pourrait être recruté par des PREs putatifs précédemment identifiés par Sing et al. mais aussi par une série de facteurs se liant à l'ADN et jouant le rôle d'ancrage comme Jarid2. Cette interaction a été confirmée par d'autres études (Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). De plus, des ARN non codants pourraient intervenir dans le recrutement de PRC2.

Récemment, Landeira et al. ont montré que dans les cellules *Jarid2*^{-/-}, le recrutement des sous-unités de Ring1B et Mel18 était altéré suggérant un recrutement similaire de PRC1 et PRC2 (Landeira et al., 2010).

C. Mécanismes additionnels de répression des gènes du développement

Bien qu'un grand nombre de facteurs remodelant ou modifiant la chromatine aient été décrits, peu ont été rapportés comme indispensables au contrôle des cellules ES. L'équipe de Young a utilisé un crible shRNA dirigé contre 197 régulateurs de la chromatine pour identifier des régulateurs importants dans le maintien du phénotype ES (Bilodeau et al., 2009). Dans ce crible, la perte des histones méthyltransférases spécifiques de H3K9, ainsi que des composants PcG, ont été montrés comme cause suffisante pour perdre le phénotype ES. La déplétion de la méthyltransférase H3K9 SetDB1 induit un phénotype intéressant. Le shRNA dirigé contre cette protéine induit une perte drastique d'Oct4. Ce résultat est compatible avec le fait que les ICM des blastocytes *SetDB1*^{-/-} sont non viables

(Dodge et al., 2004). Une étude plus poussée par ChIP-seq à l'aide d'un anticorps spécifique a révélé que cette protéine se lie au promoteur de 2232 gènes. La moitié de ces gènes sont actifs (marques d'histones H3K4me2/3, H3K36me3 et ARN polymérase II) alors que l'autre moitié comprend beaucoup des gènes du développement (silencieux en ES). Les gènes dérégulés co-localisent à 38% avec Suz12 (PRC2). Les auteurs ont remarqué que l'essentiel des gènes enrichis en H3K9me3 étaient aussi enrichis en H3K4me3 et pour une moitié en H3K27me3 (Figure 17). Cette remarque est en accord avec l'hypothèse formulée précédemment que la plupart des gènes (actifs ou non) possèdent la marque H3K4me3 au promoteur (Kim 2005, Guenther 2007). SetDB1 ne corrèle qu'avec 22% des sites présentant la marque H3K9me3, ce qui peut s'expliquer par de la redondance avec d'autres méthyltransférases. De façon remarquable, les régions de plus forte densité en marque H3K9me3 corrélaient avec les régions promotrices des gènes liés par SetDB1, ainsi qu'avec les marques H3K4me3 et H3K27me3. Ces régions correspondent à un sous-ensemble de gènes impliqués dans le développement. SetDB1 contribue à la répression d'un sous-groupe de gènes du développement qui avaient été précédemment montrés réprimés par les complexes Polycomb. Ces résultats suggèrent aussi que la perte de SetDB1 entraîne la dérégulation des gènes du développement et peut déstabiliser les cellules ES, expliquant ainsi le phénotype observé lors l'inactivation de SetDB1. Ces résultats montrent qu'un système additionnel aux PcG est impliqué dans la répression des gènes du développement dans les cellules ES et que la marque H3K9me3 n'est pas exclusive de l'hétérochromatine constitutive.

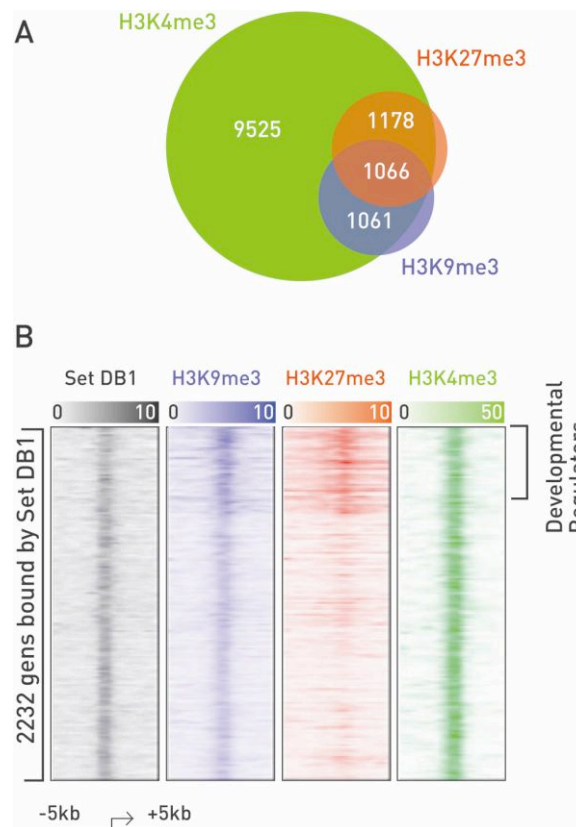


Figure 17: Liaisons de SetDB1 aux histones modifiées, tiré de Bilodeau et al. 2009

Loh et al. ont montré que les protéines antagonistes de SetDB1, Kdm3a et Kdm4c (Histones déméthylases H3K9), étaient aussi indispensables à l'auto-renouvellement des cellules ES (Loh et al., 2007). Kdm3a possède la capacité à déméthyliser H3K9me1 et me2 *in vitro* et est un co-activateur du récepteur androgène (AR) (Yamane et al., 2006). Kdm4c convertit les marques H3K9me3 et K36me3 en me2 (Whetstine et al., 2006). Loh et al. ont montré que ces gènes sont des cibles d'Oct4 dans les cellules ES de souris. De plus, par ARN interférence, les auteurs ont montré la spécificité de chacune de ces enzymes et ont prouvé que le phénotype ES n'était pas maintenu: les cellules électroporées ont des formes de fibroblastes, l'expression des gènes Oct4, Sox2 et Nanog est fortement réduite et l'activité phosphatase alcaline (importante dans les cellules ES) est aussi réduite. Les auteurs ont ensuite montré que Kdm3a régule l'expression de Tcl1 en enlevant les marques répressives au promoteur. Ivanova avait montré que Tcl1 était requis pour la pluripotence des cellules ES (Ivanova et al., 2006). La perte du phénotype ES induit par le RNAi dirigé contre Kdm3a peut être compensé par une surexpression de Tcl1. De manière similaire, les auteurs ont montré que Nanog était une cible de Kdm4c et que cette enzyme est requise pour la déméthylation de son promoteur. La surexpression de Nanog peut restaurer le phénotype ES et compenser la perte de Kdm4c. Les deux mécanismes sont strictement indépendants

et il n'y pas de compensation mutuelle (Figure 8, page 31). Deux enzymes importantes dans la pluripotence des cellules ES ont donc été identifiées, agrandissant la liste des facteurs importants pour la pluripotence. Loh et al. proposent de tester ces nouvelles protéines dans la reprogrammation des iPS (Loh et al., 2007).

Conclusion

Si l'ensemble des informations génétiques est utilisé au moins une fois au cours du développement d'un organisme, tous les gènes doivent être prêts à être exprimés au moment adéquat (Zipori, 2004). Dans les cellules ES pluripotentes, beaucoup de gènes sont faiblement exprimés. Au cours de la différenciation, l'expression de certains gènes se réduit alors que d'autres sont exprimés, déterminant le phénotype final des cellules filles. L'analyse sur puces du profil d'expression de l'ensemble des gènes à l'échelle du génome a montré une grande variété de gènes exprimés à un faible niveau (Carter et al., 2005). Ceci peut être la conséquence de la conformation de la chromatine ouverte au niveau de laquelle ils se trouvent; permettant une expression réduite grâce à des systèmes de répression de la transcription (Roeder, 2005). L'abondance des marques H3K4me3 et des H4Ac dans les cellules ES incline dans de sens (Azuara et al., 2006; Lee et al., 2004). Cette observation dénote une chromatine très dynamique avec de nombreux échanges d'histones (Meshorer and Misteli, 2006). De même, de nombreux sites sont sensibles à la DNaseI ce qui corrèle avec une chromatine de conformation plutôt ouverte (Weintraub and Groudine, 1976) et ceci indépendamment de l'expression des gènes (Meshorer et al., 2006).

Niwa en a conclu que les processus épigénétiques étaient requis pour le bon déroulement de la différenciation des cellules ES mais surement aussi pour la maintenance de la pluripotence (Niwa, 2007b). Une hypothèse pourrait être selon lui, que les processus épigénétiques seraient responsables de l'exécution d'un programme de la maintenance de la pluripotence, qui est lui-même régi par un réseau de facteurs de transcription. La maintenance de la pluripotence des cellules ES serait maintenue par un ensemble de boucle de régulations positives et négatives incluant les modificateurs épigénétiques. Niwa propose même de limiter le rôle des facteurs de transcription à activer des processus d'ouverture ou de fermeture de la chromatine requis pour les changements rapides des statuts transcriptionnels de certains gènes au cours du développement, par exemple pour enlever les marques répressives et les remplacer par les marques de transcription actives (Niwa, 2007a). Les différentes hypothèses créaient donc un dilemme de même nature que l'œuf et la poule: lequel des deux est arrivé en premier? La reprogrammation des iPS à l'aide des

facteurs de transcription permet de montrer que les réseaux de facteurs de transcriptions sont capables de contrôler les réseaux épigénétiques impliqués dans la pluripotence.

Chapitre IV. Remodelage de la chromatine et pluripotence

Introduction

Les réseaux de régulation qui permettent de maintenir la pluripotence des cellules ES sont étroitement liés aux facteurs qui modifient les protéines et aux remodeleurs de chromatine. Par exemple, Oct4 peut interagir avec les remodeleurs de la famille SWI/SNF ainsi qu'avec le complexe NuRD (van den Berg et al., 2010). De plus, les sites de liaison à l'ADN de la famille SWI/SNF (Brg1=Smaca4, Brahma=Smarca2) peuvent être chevauchants avec ceux d'Oct4 (Ho et al., 2009a; Kidder et al., 2009). De même, le complexe PRC1 qui fait partie du groupe polycomb (PcG) couvre plusieurs kilobases autour des promoteurs enrichis en marques d'histones H3K27me3 et H3K4me3 et peut recouvrir certains sites de liaisons d'Oct4 (Boyer et al., 2006; Endoh et al., 2008). Les facteurs de transcription des cellules ES comme Oct4, Nanog, Sox2, Esrrb et Tcfcp2l1 se retrouvent co-localisés (Chen et al., 2008; Kim et al., 2008a). Une hypothèse propose que les facteurs de transcription ne soient pas nécessaires pour diriger les complexes de remodelages ou de modifications de la chromatine vers leurs cibles mais seulement au niveau de l'initiation du remodelage. Cette étape serait ensuite suivie par une modification de la chromatine créant ainsi un marquage et une conformation de la chromatine appropriée à l'ancrage d'autres facteurs. Les complexes de modification de la chromatine contiennent souvent des sous-unités qui reconnaissent des modifications d'histones particulières (Taverna et al., 2007) et qui peuvent donc reconnaître spécifiquement les régions promotrices. La double dépendance des marques d'histones et des facteurs de transcription crée de multiples mécanismes qui permettent d'affiner le recrutement des complexes de modification de la chromatine.

Les complexes de remodelage de la chromatine catalysent un ensemble de changements structuraux qui affectent les nucléosomes et leur distribution sur l'ADN. Certains complexes de remodelage présentent également, *in vitro*, une activité d'assemblage des nucléosomes. Ces complexes sont très bien conservés au cours de l'évolution. On les retrouve chez la levure, la drosophile, l'homme et ils sont généralement très abondants dans les cellules. Ils sont capables de repositionner les nucléosomes pour les espacer régulièrement ou pour rendre l'ADN plus accessible. Ils peuvent également catalyser la dissociation de l'ADN et des histones ou induire le remplacement des histones

par des variants. Ces enzymes sont impliquées dans des processus cellulaires très variés comme la régulation de la transcription, la réplication, la réparation de l'ADN, la cohésion des chromatides-sœurs ou l'intégration des transposons.

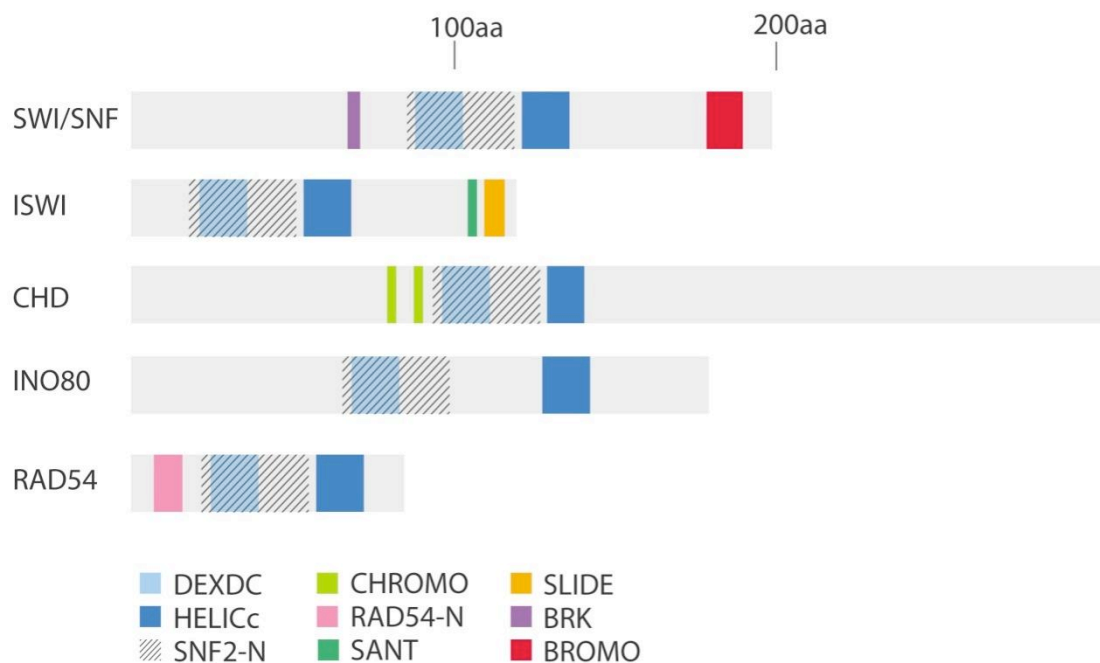


Figure 18: Représentation schématique des macrodomaines des cinq familles composant la superfamille SNF2

Les protéines de la superfamille SNF2 se caractérisent par un domaine de type hélicase et un domaine ATPase typique (Figure 19). Le plus souvent, elles exercent leur activité de remodelage sous forme de gros complexes multi-protéiques (Mohrmann and Verrijzer, 2005) et partagent la capacité à déplacer les nucléosomes sur l'ADN dans des systèmes expérimentaux *in vitro* (Saha et al., 2006). Chez les mammifères, il existe 30 gènes codant les protéines appartenant à la superfamille SNF2, qui peut être subdivisée en cinq familles sur la base de la présence de différents domaines protéiques conservés (Figure 18): SWI/SNF (mating type SWItching/Sucrose Non Fermenting), ISWI (Imitation SWItch), Chd (Chromodomain Helicase DNA-binding), Ino80 et Rad54 (Clapier and Cairns, 2009; Flaus et al., 2006). Les membres présents ont été regroupés dans un arbre phylogénétique (Figure 20). Certaines de ces protéines sont impliquées dans des pathologies humaines telles que des cancers (Denslow and Wade, 2007; Ho and Crabtree, 2010; Reisman et al., 2009).

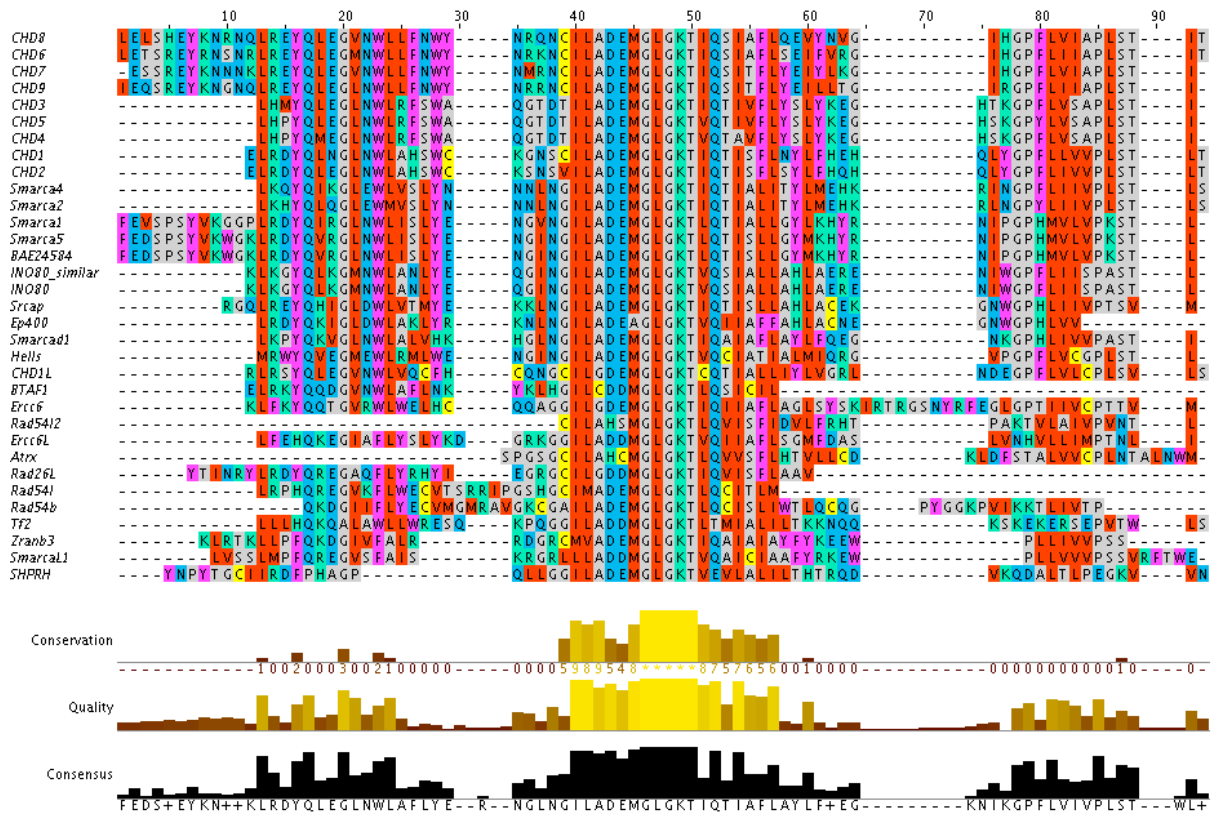


Figure 19: Conservation du domaine ATPase/helicase au sein de la famille SNF2 autour des positions 40 à 52.

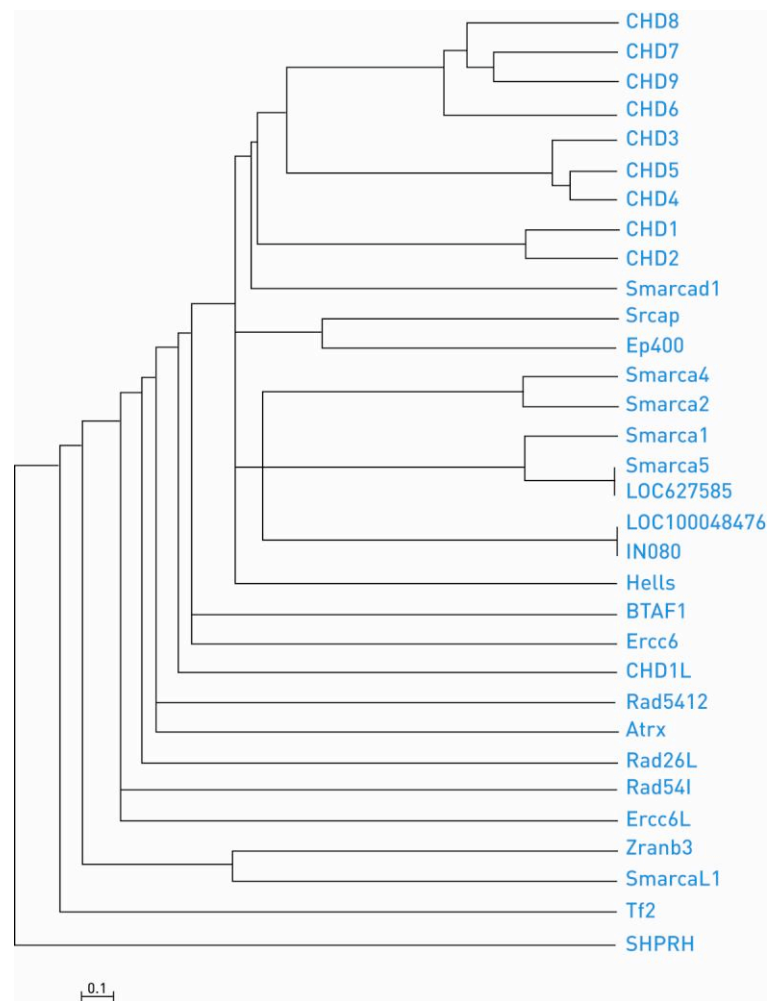
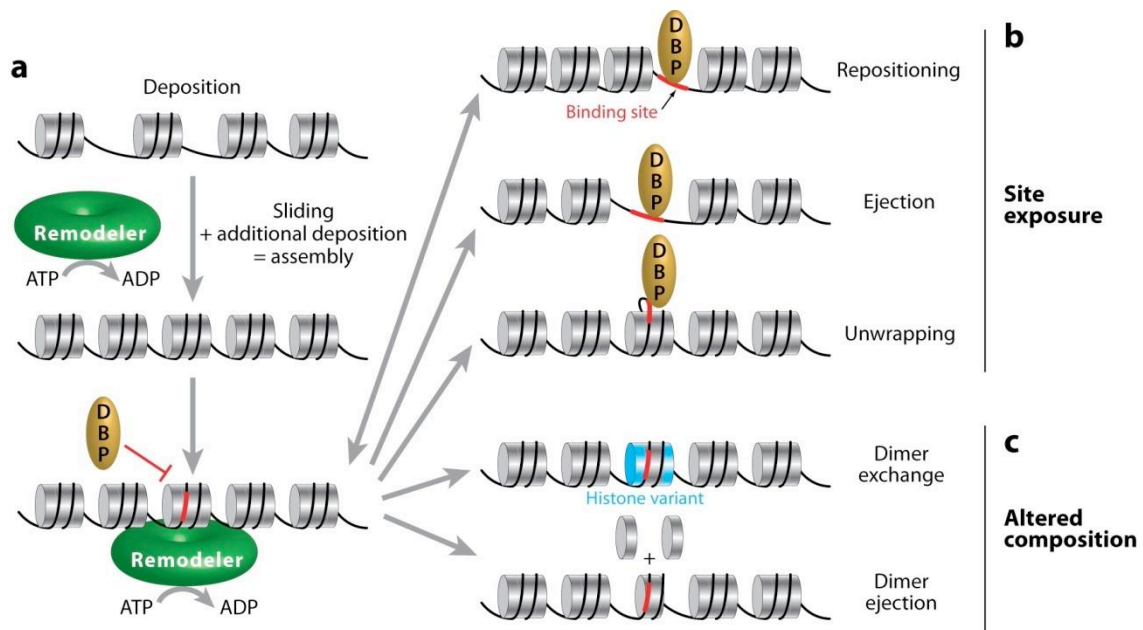


Figure 20: Arbre phylogénétique regroupant l'ensemble des protéines identifiées par BLAST possédant le domaine ATPase sur le génome muri. Méthode utilisée : minimum linkage (Mafft).

A. Les mécanismes du remodelage de la chromatine

Les mécanismes utilisés par les complexes de remodelage de la chromatine pour déplacer les nucléosomes sont encore mal compris. La stabilisation de l'octamère d'histones par ISWI et CHRAC et le fait qu'aucun transfert d'histones ne soit détecté argumente contre un modèle impliquant la dissociation de l'octamère d'histones (Langst et al., 1999; Varga-Weisz et al., 1997). Cette situation pourrait être différente pour le remodelage des nucléosomes par SWI2/SNF2 car ce complexe peut transférer les octamères d'histones d'un fragment d'ADN vers un autre (Lorch et al., 1999; Phelan et al., 2000). Les différents complexes utilisent donc probablement des mécanismes de remodelage différents. Plusieurs scénarios peuvent être envisagés (Figure 21). Les expériences *in vivo* n'ont pas encore pu différencier ces modèles. Dans chacun des cas, l'initiation du remodelage est la même : la machinerie de remodelage entre en contact avec l'ADN et les histones au site d'entrée de

l'ADN dans le nucléosome. Dans le modèle du « spooling », un changement de conformation de la chromatine altère les interactions entre l'ADN et les histones, ce qui libère un fragment d'ADN et laisse libre une partie de la surface du nucléosome (Pazin et al., 1997). Le « morceau » d'octamère d'histones est alors disponible pour capturer un autre fragment d'ADN et le fragment d'ADN libre peut entrer en contact avec un autre octamère d'histones. Il y a alors formation d'une structure ressemblant à un di-nucléosome ou initiation d'un déplacement en trans. D'autres événements peuvent avoir lieu comme le remplacement d'un dimère d'histones H2A-H2B par un dimère H2A.Z-H2B ou l'établissement d'un contact entre le fragment d'ADN libéré et une protéine non-histone. Dans le modèle du « twisting », une rotation de l'ADN à son site d'entrée sur le nucléosome altère les interactions avec les histones. L'histone rentre en contact avec la paire de base voisine et la rotation est propagée à la surface du nucléosome. Une fois que l'hélice de distorsion émerge de l'autre côté du nucléosome, l'octamère s'est déplacé d'une paire de base. Dans le modèle du « bulging », comme dans le modèle du « spooling », il y a dissociation entre l'octamère d'histones et un fragment d'ADN. Le fragment d'ADN est recapturé par l'octamère d'histones mais avec un décalage qui produit une boucle d'ADN à la surface du nucléosome. La boucle peut se propager autour de l'octamère d'histones. Cette propagation induit la relocalisation de l'octamère d'histones sur un fragment d'ADN voisin en cis. Dans le cas du modèle par « twisting », le déplacement du nucléosome est de l'ordre de la paire de base. Dans le cas du modèle par « bulging », le déplacement est de l'ordre d'un tour d'hélice d'ADN autour du nucléosome (10,5bp).



AR Clapier CR, Cairns BR. 2009.
Annu. Rev. Biochem. 78:273–304

Figure 21: Différents modèles de remodelage de l'ADN à l'aide de complexes multi-protéiques pouvant utiliser l'énergie de l'ATP, tiré de Clapier et al. 2009

B. Les complexes de la super-famille SNF2

Les ATPases de type SNF2 sont caractérisées par la présence d'un bromodomaine. Les bromodomains sont des modules d'environ 110 acides aminés qui peuvent reconnaître des lysines acétylées sur les queues des histones. Les membres de la famille SWI/SNF contiennent également une multitude de motifs distincts liant l'ADN. Les fonctions de ces motifs ne sont pas connues, mais on peut supposer qu'ils participent au ciblage des complexes de remodelage sur la chromatine. Les membres de la famille SNF2 peuvent être séparés en 2 groupes. Le premier groupe comprend le complexe SWI/SNF de levure et ses homologues (BAP de la drosophile, et BAF des mammifères). Le second groupe comprend le complexe RSC (Remodel the Structure of Chromatin) de levure et ses homologues (PBAP de la drosophile et PBAF des mammifères). Les 2 complexes SWI/SNF et RSC sont très conservés au cours de l'évolution. Cette conservation suggère que leurs différentes fonctions sont également conservées.

1. Les protéines de la famille SWI/SNF

La protéine Brahma (BRM) est la seule ATPase de type SWI2/SNF2 chez la drosophile. Cette protéine est nécessaire à la viabilité (Elfring et al., 1998). BRM appartient à deux complexes distincts : BAP et PBAP. Ces complexes partagent sept sous-unités dont BRM. Ils diffèrent dans la présence de trois sous-unités: BAP contient OSA, tandis que PBAP contient Polybromo et BAP170. La protéine BRM a été identifiée comme un membre du groupe Trithorax, c'est-à-dire qu'elle est nécessaire pour le maintien de l'expression des gènes homéotiques au cours du développement. Cependant BRM pourrait jouer un rôle plus général dans l'activation des gènes car cette protéine colocalise avec l'ARN polymérase II au niveau des chromosomes polytènes chez la drosophile (Armstrong et al., 2002). Les cellules humaines contiennent au moins deux complexes SWI/SNF appelé BAF et PBAF. Les sous-unités ATPases correspondantes, BRG1 (Brahma Related Gene 1) et hBRM, sont très homologues entre elles et avec la protéine Snf2 de levure. Cependant, bien que les protéines BRG1 et hBRM soient identiques à 75%, leurs fonctions sont différentes. En effet, la délétion de BRM chez la souris n'est pas létale tandis que les souris délétées de BRG1 meurent très tôt au cours du développement (Bultman et al., 2000). Ces données sont en accord avec les données du groupe de Crabtree qui indique que BRM n'est pas exprimé en cellules ES et que le complexe esBAF (Figure 22) contient uniquement BRG1 en cellules ES (Ho et al., 2009b).

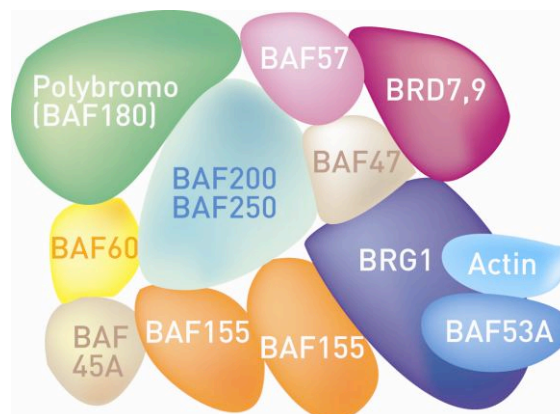


Figure 22: Le complexe esBAF dans les cellules ES murines, tiré de de Ho et al, 2010.

Les effets du remodelage de la chromatine par les complexes SWI/SNF, ont été analysés *in vitro*. Ces effets variés comprennent, un déplacement du nucléosome remodelé d'environ 40bp d'ADN, une augmentation de l'accessibilité de l'ADN aux facteurs de transcription et aux enzymes de restriction, des mouvements de l'octamère d'histones en

trans ou la formation de particules ressemblant à des di-nucléosomes. La diversité des effets de remodelage des nucléosomes par les complexes SWI/SNF *in vitro* suggère qu'*in vivo*, le remodelage pourrait dépendre de la concentration locale des complexes, de la structure de la chromatine à un locus particulier, de la présence de facteurs régulateurs ou encore de la présence d'ADN pouvant recevoir les histones dissociées. De plus la composition du complexe esBAF dépend du contexte cellulaire, ce qui montre une régulation dynamique (Figure 23).

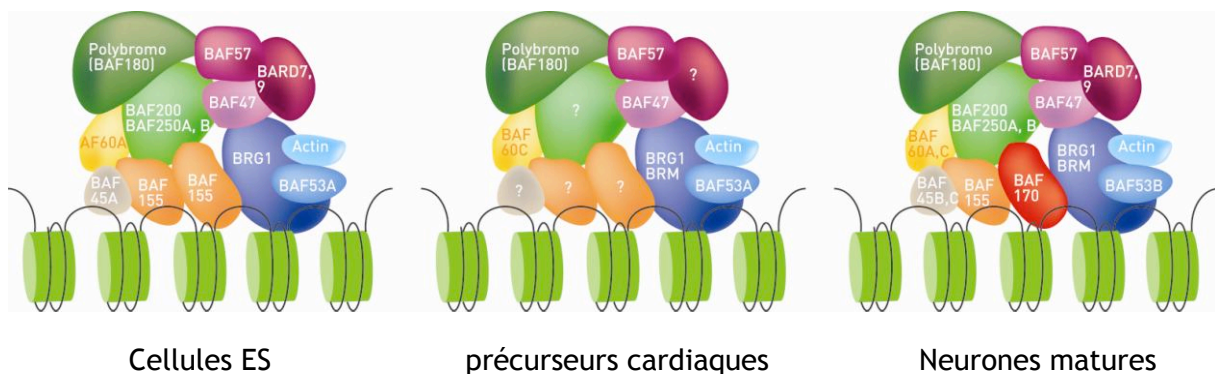


Figure 23: Le complexe BAF dans des contextes cellulaires différents, tiré de Ho et al., 2010.

Le rôle de la protéine BRG1 a été clairement démontré dans des cancers du poumon, de la prostate, du colon et du sein (Medina and Sanchez-Cespedes, 2008). Très récemment, des études à grande échelle ont montré l'implication de la protéine Brg1 dans la maintenance de la pluripotence. Cette implication se fait via le contrôle de gènes importants comme Oct4, Nanog ou encore Sox2 (Kidder et al., 2009). Une autre étude a montré l'importance de la protéine BRG1 dans ce mécanisme (Ho et al., 2009a). Afin de mieux cerner les réseaux régulateurs des cellules ES sous tenus par les complexes SWI/SNF l'équipe de Crabtree a poursuivi son étude par une localisation de la protéine BRG1 à l'échelle du génome grâce à la technologie du ChIP-seq. Le profil de liaison de la protéine Brg1 est alors assimilé à celui du complexe esBAF (Figure 22) auquel elle appartient (Lessard et al., 2007). La protéine Brg1 est retrouvée sur 4% du génome murin et lié à 5630 gènes différents. Ho et al. ont identifié Brg1 et son partenaire BAF155 s'associaient préférentiellement aux promoteurs des gènes en corrélant positivement avec l'expression des gènes liés. Notamment, Brg1 et son partenaire BAF155 se lient sur les promoteurs des protéines dont l'importance a été soulignée dans les chapitres précédents et qui sont les clefs pour le maintien de la pluripotence: Oct4, Sox2, Nanog, Dppa2, Dppa4, Sall4 et Myc. De plus, Brg1 partage un certain nombre de cibles avec ces derniers: 67% des cibles d'Oct4,

74% pour Sox2 et 65% pour Nanog. De plus, Brg1 partage des cibles avec les facteurs Smad1 et Stat3 qui font partie des voies BMP et LIF respectivement, renforçant l'importance de cette protéine au cœur du réseau responsable de la pluripotence. Afin de déterminer plus précisément le rôle de Brg1 dans le réseau transcriptionnel des cellules ES, l'équipe de Crabtree a identifié les gènes dérégulés lors de la perte d'expression de Brg1. Ils ont ainsi pu mettre en évidence que Brg1 avait une activité activatrice pour certains gènes qui ont un rôle dans la pluripotence (Nanog, Oct4, Sox2, Sall4...) et une activité répressive pour les gènes impliqués dans le développement (Ring1, Ezh2, Suz12...) (Ho et al., 2009a). Cette fonction répressive se manifesterait non pas par compétition sur les gènes du développement avec les gènes polycombs mais par répression directe des gènes polycombs eux-mêmes. Cependant ce mécanisme proposé par Ho et al. reste à valider. En effet il subsiste une contradiction entre les données de Fazio et al. qui montre un phénotype précoce (siRNA) et les données de Ho et al. qui observe un phénotype tardif (Fazio et al., 2008), Ref Ho). Bien que Brg1 se lie à des promoteurs-clefs en cellules ES et fasse intimement partie du réseau transcriptionnel maintenant la pluripotence des cellules ES (van den Berg et al., 2010), le rôle de Brg1 doit être défini plus précisément. Le mécanisme qui permettait aux cellules exprimant Cdx2 et s'engageant dans la voie de différenciation trophoblastique, de réprimer l'expression d'Oct4 restait inconnu, jusqu'à ce que Wang et al. montrent que Brg1 est requis pour l'extinction du gène Oct4 (Wang et al., 2010). En effet, la perte d'expression (siRNA) de Brg1 ou de Cdx2 induit une expression d'Oct4 dans le trophoblaste alors que la combinaison des deux donne un phénotype encore plus aggravé. Les auteurs ont ensuite montré que Brg1 et Cdx2 co-localisent et interagissent entre elles et sont capables d'être recrutées sur le promoteur d'Oct4 et d'induire sa répression sans augmenter le taux de méthylation de l'ADN. De futures études pourraient étudier le taux de méthylation et d'acétylation des histones en absence ou présence de Brg1 et/ou Cdx2.

Ce mécanisme est une illustration d'une régulation faite par les enzymes modificatrices de la chromatine. Actuellement, le rôle des remodeleurs de chromatine dans le maintien de l'état pluripotent des cellules ES reste largement méconnu et de futures études sont nécessaires afin d'élucider leur(s) rôle(s).

2. Les complexes de la famille ISWI (Imitation SWItch)

L'ATPase ISWI a été identifiée chez la drosophile à partir de la similarité de son domaine ATPase avec celui de la protéine Brahma. Contrairement à Brahma, ISWI et l'ARN polymérase II ne co-localisent pas au niveau des chromosomes polytènes, ce qui indique que la régulation de la transcription n'est pas la fonction principale d'ISWI (Deuring et al.,

2000). Cette protéine est importante car sa délétion est létale chez la drosophile à un stade larvaire tardif et induit une perturbation de la structure du chromosome X chez les mâles (Deuring et al., 2000).

La protéine ISWI possède une activité de remodelage intrinsèque *in vitro*. Cette activité est amplifiée lorsque cette protéine est présente au sein de complexes. *In vivo*, ISWI est toujours retrouvée au sein de complexes, mais peut être associée à des sous-unités différentes qui régulent son activité. Chez *D.melanogaster*, trois complexes comprenant la protéine ISWI ont été identifiés : NURF, ACF et CHRAC. Le complexe NURF (NUcleosome Remodeling Factor) comprend ISWI, une large sous-unité de 301kDa (NURF301), une protéine pyrophosphatase de 38kDa et l'homologue de la sous-unité p48 de CAF-1. *In vitro*, NURF ne favorise pas l'espacement régulier des nucléosomes mais, au contraire, catalyse la rupture de la régularité (Tsukiyama et al., 1995). Grâce à ce remodelage, NURF facilite l'initiation de la transcription à partir de support chromatinien réprimé *in vitro* en augmentant l'accès des facteurs de la transcription au promoteur (Mizuguchi et al., 2001; Mizuguchi and Wu, 1999).

In vivo, le complexe NURF pourrait jouer un rôle important dans l'expression des gènes régulés par l'ecdysone (une hormone stéroïde) au cours du développement de la drosophile. Chez la drosophile, les récepteurs de l'ecdysone activent les gènes durant la métamorphose de la larve en pupa. Une analyse par puces à ADN des gènes régulés par NURF301 a identifié de nombreux gènes régulés par les récepteurs de l'ecdysone dont l'expression nécessite la présence de NURF301 (Badenhorst et al., 2002). De plus, chez la drosophile, des mutants pour la protéine NURF301 arrêtent leur développement au stade larvaire, ne se métamorphosent pas et finissent par mourir (Badenhorst et al., 2002; Badenhorst et al., 2005). Une mutation de la protéine NURF301, létale au stade embryonnaire chez la drosophile, altère, comme chez les mutants d'ISWI, la structure du chromosome X des mâles. Cela suggère qu'ISWI participe à la structure de ce chromosome au sein du complexe NURF (Badenhorst et al., 2002).

Le second complexe ACF (ATP-utilizing Chromatin assembly and remodeling Factor) contenant ISWI, comprend également, Acf1, une protéine à bromodomaine et à doigt PHD (Plant HomeoDomain). ACF a été identifié comme un complexe requis pour l'assemblage régulier des nucléosomes (Ito et al., 1997). L'interaction entre Acf1 et ISWI module l'activité de ce dernier: la liaison d'Acf1 stimule fortement *in vitro* la mobilisation des nucléosomes par ISWI (Eberharter et al., 2001). Cette interaction peut également induire des changements dans la direction des déplacements des nucléosomes le long d'un fragment d'ADN linéaire (Eberharter et al., 2001). Les modulations de l'activité d'ISWI par Acf1 ne passent pas par l'interaction entre les deux protéines mais par une liaison entre Acf1 et le nucléosome. En

effet, les domaines PHD de Acf1 lient les domaines globulaires des histones et sont importants pour stimuler l'activité d'ISWI, (Eberharter et al., 2004). Il est donc probable que ce soit la liaison des domaines PHD avec les histones qui stimule l'activité d'ISWI.

Le dernier complexe contenant ISWI, CHRAC (CHRomatin Accessibility Complex), est très proche du complexe ACF. En effet, il comprend les sous-unités ISWI, Acf1 et deux autres petites protéines, CHRAC14 et CHRAC16. Le complexe CHRAC a été identifié, chez la drosophile, comme un complexe permettant une augmentation globale, dépendante de l'ATP, de l'accessibilité de la chromatine. Il semblerait que les sous-unités CHRAC14 et CHRAC16 servent à stabiliser les fragments d'ADN nu en les liant au cours du remodelage de la chromatine (Hartlepp et al., 2005). Elles stimulent également l'activité de remodelage du complexe lorsqu'ISWI et Acf1 sont présents en quantité limitante, chez la drosophile et chez l'homme (Hartlepp et al., 2005). Les mutants de la sous-unité Acf1, appartenant aux complexes ACF et CHRAC, présentent trois aspects anormaux chez la drosophile (Fyodorov et al., 2004). Premièrement, la chromatine totale ne présente plus de structure régulière et la distance entre les nucléosomes est globalement diminuée. Deuxièmement, les gènes normalement réprimés par l'hétérochromatine sont réactivés. Troisièmement, le passage par la phase S est plus rapide. Ces résultats suggèrent qu'un des rôles majeurs des complexes contenant Acf1 est d'assembler régulièrement les nucléosomes le long de la chromatine pour contribuer à la répression de la transcription et réguler la réplication.

Des complexes de remodelage contenant des homologues de la protéine ISWI et de composition similaire aux complexes ACF et CHRAC existent chez différentes espèces (levure, xénope, homme). Chez l'homme, on retrouve un complexe hCHRAC et plusieurs complexes ressemblant au complexe ACF de la drosophile.

A l'instar de Smarca4, l'inactivation du gène de Smarca5 (homologue murin d'ISWI) est létale au niveau embryonnaire (Niwa, 2007a). De plus, Pardo et al. ont trouvé que Smarca5 fait partie des remodeleurs qui interagissent avec Oct4 (Pardo et al., 2010). Ces données suggèrent un rôle majeur pour Smarca5 dans les cellules ES.

3. Les complexes de la famille INO80 (INOsitol 80)

La famille des complexes INO80 comprend le complexe INO80 et le complexe SWR1 qui sont caractérisés par des domaines ATPases divisés et par la présence des protéines Rvb1 (RuVB-like helicase 1) et Rvb2. Les sous-unités du complexe INO80 (INOsitol 80) ont été isolées chez *S.cerevisiae* (Shen et al., 2000). Ce complexe comprend quinze sous-unités principales. Le complexe INO80 est très conservé et le complexe purifié chez l'homme

contient les orthologues de Ino80, Rvb1, Rvb2, Arp4, Arp5, Arp8, les2 et les6 ainsi que cinq autres sous-unités uniques. Les complexes Ino80 de la levure et de l'homme présentent une activité de remodelage des nucléosomes dépendante de l'ATP (Jin et al., 2005; Shen et al., 2000). Les protéines Rvb appartiennent à la famille des ATPases « chaperon-like ». Ces protéines sont essentielles pour l'activité de remodelage de la chromatine du complexe INO80 (Jonsson et al., 2004). Il a été montré que Arp4 peut lier les 4 histones *in vitro* et *in vivo* et que Arp8 peut lier les histones H3 et H4 *in vitro* (Shen et al., 2003). Cela suggère que les sous-unités Actin/Arp des complexes de remodelage de la chromatine pourraient avoir des fonctions de chaperons pour l'octamère d'histones et favoriser le réarrangement des contacts histones-histones ou ADN-histones au cours du remodelage de la chromatine.

Il est également impliqué dans les réponses aux dommages à l'ADN. En effet, chez *S.cerevisiae*, la sous-unité Nhp10 permet le recrutement du complexe au niveau des cassures double brin par liaison à l'histone phosphorylée H2A.X (van Attikum and Gasser, 2005). Le groupe de Gasser a montré qu'un mutant de Arp8 présente une diminution de la production d'ADN simple brin (van Attikum and Gasser, 2005). Ce résultat suggère que le complexe INO80 pourrait faciliter la formation d'ADN simple brin qui est une étape importante pour la réparation de l'ADN par recombinaison homologue. Il a été également montré que le complexe INO80 enlève les histones au niveau des cassures d'ADN et qu'il intervient dans la réparation par jonction des extrémités non homologues (van Attikum and Gasser, 2005). L'ensemble de ces résultats indique que le complexe INO80 participe à différentes voies de réparation de l'ADN.

Le second complexe de cette famille, le complexe SWR1 (Swi2/Snf2-Related ATPase1) a été identifié comme permettant l'échange de l'histone H2A dans les nucléosomes pour l'histone H2A.Z (Mizuguchi et al., 2004). En plus de son activité de chaperon de l'histone H2A.Z, le complexe SWR1 montre, *in vitro*, une activité de remodelage des nucléosomes ATP dépendante chez la levure (Mizuguchi et al., 2004). La protéine Swr1 est une ATPase à domaine divisé, essentielle pour l'intégrité et la fonction du complexe SWR1 (Mizuguchi et al., 2004). La fonction majeure du complexe SWR1 est de déposer le variant H2A.Z. H2A.Z et SWR1 sont fonctionnellement liés. En effet, les mutants Swr1 et Htz1 montrent les mêmes phénotypes (Mizuguchi et al., 2004). Cependant, seulement 40% des gènes régulés par l'une de ces protéines le sont également par l'autre, ce qui suggère que le complexe SWR1 possède d'autres fonctions en plus du dépôt de H2A.Z (Mizuguchi et al., 2004). En effet, il semblerait que le complexe SWR1 soit un acteur de la réparation de l'ADN. Il a été montré, *in vitro*, que le complexe SWR1 lie l'histone H2AX et que les complexes contenant la protéine Rvb1 (complexes INO80 ou SWR1) sont recrutés au niveau des cassures double brin *in vivo* (Downs et al., 2004).

Enfin, le complexe humain TRRAP/Tip60, qui comprend des sous-unités homologues à des sous-unités des complexes de levure SWR1 et NuA4 (Nucleosome Acetyltransferase of histone H4), a été impliqué dans l'activation de ATM (une kinase importante dans la voie de signalisation des dommages à l'ADN) et la réparation de l'ADN (Lu et al., 2009; Sun et al., 2005). L'ensemble de ces résultats indique que le complexe SWR1 agit non seulement comme un chaperon de l'histone H2A.Z mais également comme un acteur de la réparation de l'ADN.

Récemment, Ep400, sous-unité remodelatrice du complexe Trrap/Tip60 murin, a été largement identifié comme partenaire de Nanog et donc impliqué dans le maintien de la pluripotence (voir page 45). De plus, la protéine Ino80 a été retrouvée dans deux études d'études des partenaires d'Oct4 dans les cellules ES murines (Pardo et al., 2010; van den Berg et al., 2010). Smarca6 ou Hells est aussi une hélicase appartenant à la famille Ino80 et a été également trouvé comme partenaire d'Oct4 (Pardo et al., 2010; van den Berg et al., 2010).

4. Les complexes des familles Rad26 et RAD54

Plusieurs autres protéines présentent un domaine ATPase de type Swi2/Snf2. Ces protéines possèdent aussi une activité de remodelage de la chromatine. Par exemple, CSB (Cockayne Syndrome protein B ou Rad26), est une protéine remodelant la chromatine qui intervient dans le processus de réparation par excision de nucléotides (Citterio et al., 2000). Shprh (SNF2 histone linker PHD RING helicase) est aussi impliqué dans la réparation (Motegi et al., 2006). La protéine Rad54 est impliquée dans la recombinaison homologue et coopère avec Rad51 dans le remodelage des nucléosomes (Mazin et al., 2010). ARIP4 (Androgen Receptor Interacting Protein 4) est une protéine de la famille RAD54 qui possède une activité de remodelage et qui interagit avec le récepteur des androgènes (Rouleau et al., 2002). Atrx qui fait partie de la famille Rad54, est une protéine impliquée dans la détermination sexuelle et dans une maladie génétique, l' α -talasémie. Elle est impliquée dans de nombreux mécanismes chromatiniens comme la déposition d'H3.3 en coopération avec Daxx (Figure 25). D'autres études ont montré que Atrx co-localise avec H2A et qu'elle est présente sur les régions enrichies en H3K9me2 (Baumann and De La Fuente, 2009; De La Fuente et al., 2004; Wong et al., 2010)

D'autres facteurs possèdent des fonctions encore inconnues comme ZranB3 (Zinc finger Ran-binding domain-containing protein 3).

5. La famille Chd

La famille Chd, caractérisée par la présence d'un double chromodomaine dans la partie N-terminale, est la plus diversifiée : il existe neuf gènes Chd chez l'homme et la souris (Figure 24). Les données publiées montrent que les chromodomaines pourraient servir de module d'interactions avec les histones méthylées (de la Serna et al., 2006) ou avec l'ADN. Des études structurales montrent que les chromodomaines des différentes protéines Chd ont des affinités variables pour les histones méthylées et suggèrent des rôles divergents pour chaque protéine Chd (Flanagan et al., 2007).

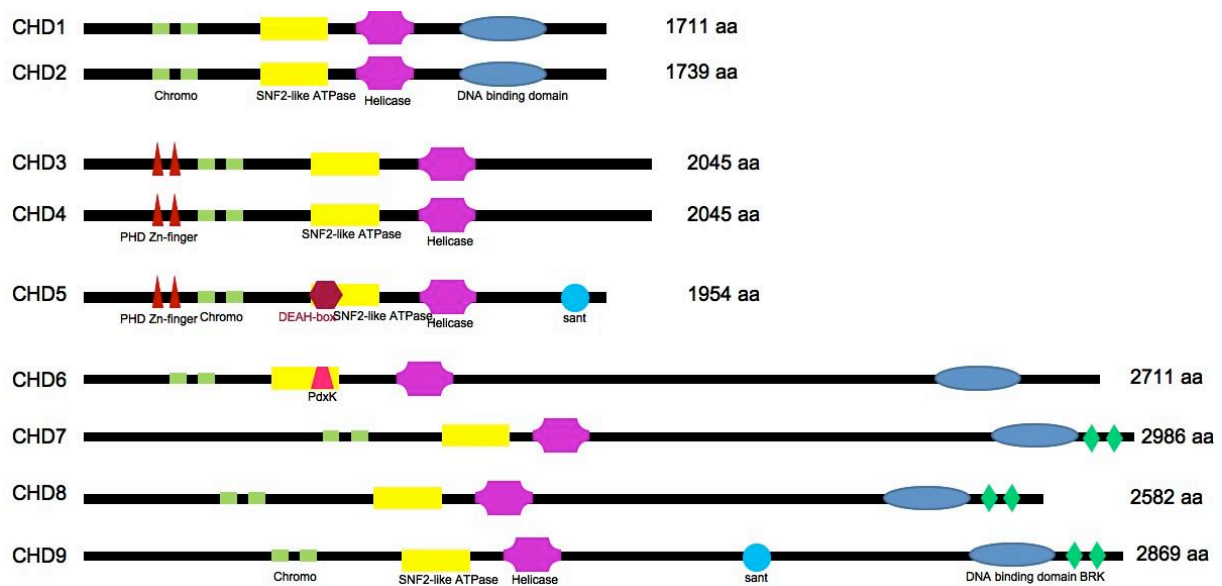


Figure 24 Représentation des domaines principaux des 9 membres de la famille Chd

Les neuf gènes identifiés chez la souris se répartissent en trois sous-familles. Chd1 et Chd2, qui constituent la première sous-famille, possèdent en plus des deux chromodomaines en tandem spécifique de la famille Chd, un domaine de liaison à l'ADN. Chd1 fait partie des complexes de remodelage SAGA et SLIK chez *Saccharomyces Cerevisiae* et peut se lier à l'histone H3 (Pray-Grant et al., 2005; Walfridsson et al., 2007). Les doubles chromodomaines de Chd1 coopèrent pour reconnaître la queue N-terminale de l'histone H3 (Flanagan et al., 2007). Il a été montré en même temps que hChd1 pouvait se lier à la marque H3K4me3 mais cette interaction ne serait pas conservée chez la levure (Sims et al., 2005). Chd1, en association avec la marque H3K4me3, pourrait aussi fonctionner pour maintenir les structures chromatiniennes sur les gènes transcrits (McDaniel et al., 2008; Quan and Hartzog). Chd1 est capable d'assembler les nucléosomes *in vitro* en

l'absence de de l'histone H1 (Lusser et al., 2005). Comme une chromatine dépourvue de l'histone de liaison H1 corrèle avec une chromatine décondensée, il est possible que Chd1 participe à l'assemblage des nucléosomes durant la transcription. Certains auteurs ont montré que Chd1 pourraient aussi jouer un rôle dans la répression de la transcription (Tai et al., 2003) et de la réplication (Biswas et al., 2008). Enfin, Chd1 pourrait incorporer la marque H3K4me3 sur les gènes qui sont transcrits par l'ARN polymérase II lorsqu'elle est affaiblie, par exemple lors d'un traitement au 6-azauracil (Zhang et al., 2005). Elle pourrait ainsi jouer le rôle de chaperon de l'histone H3 (Goldberg et al., 2010; Konev et al., 2007) (Figure 25).

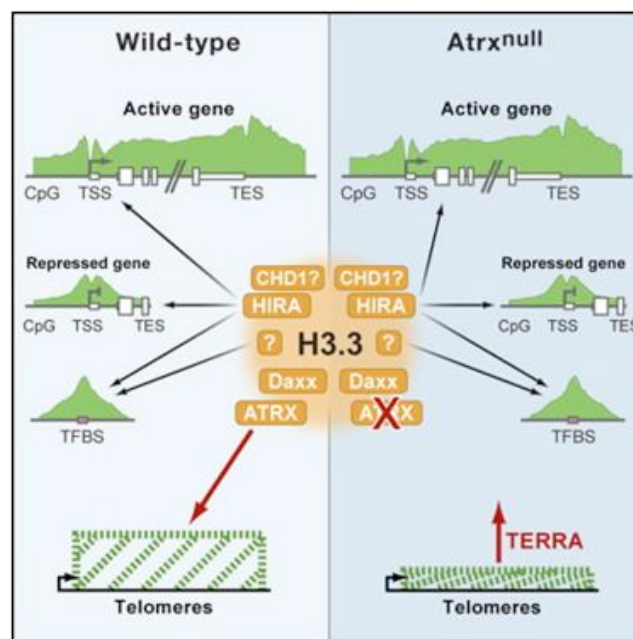


Figure 25: Hypothèse de Goldberg et al. proposant Chd1 comme chaperone d'histone pour le variant H3.3 pour les gènes actifs, HIRA comme chaperone d'histone pour H3.3 pour les gènes inactifs, enfin Atrx déposerait H3.3 aux télomères.

Il a été récemment démontré que Chd1 se lie au promoteur d'Oct4 en maintenant son expression (Gaspar-Maia et al., 2009). La protéine Chd1 est aussi essentielle au maintien de la pluripotence des cellules ES, via un mécanisme favorisant l'ouverture de la chromatine (Gaspar-Maia et al., 2009). Cette hypothèse a été reprise par Persson et al. qui proposent des rôles antagonistes pour Chd1 et H2A.Z qui établissent un équilibre dynamique d'activation et de répression des gènes. D'après les auteurs de cette étude, le rôle de Chd1 serait celui d'une protéine chaperonne d'histones. En effet, Chd1 pourrait contribuer à l'enlèvement de l'histone H2A.Z et à l'incorporation de l'histone H3.3 en étroite collaboration avec l'ARN polymérase II. Cette hypothèse est compatible avec l'observation de

nucléosomes combinant les deux variants H3.3/H2A.Z dans les cellules HeLa, au niveau des promoteurs (Jin et al., 2009). Les marques H3K4me3 et H3K27me3 pourraient fortement contribuer à la formation de ces nucléosomes particuliers (Persson and Ekwall, 2010).

La protéine Chd2 a un rôle direct dans la signalisation des dommages de l'ADN et dans la maintenance de l'intégrité du génome : les souris mutantes pour Chd2 présentent une susceptibilité aux lymphomes (Nagarajan et al., 2009).

Chd3, 4 et 5 qui appartiennent à la deuxième sous-famille, possèdent un double domaine PHD Zn-finger spécifique de leur sous-famille. Chd3 et Chd4 font partie du complexe NuRD (Nucleosome Remodeling and Deacetylation) généralement associé à la répression de la transcription (Denslow and Wade, 2007). Le complexe NuRD ([Figure 26](#)) comprend également les histones déacétylases Hdac1 et Hdac2 et les protéines RbAp46/48. La composition du complexe NuRD peut différer par la nature de la sous-unité Chd (Chd3 et Chd4) et par la présence d'une des trois sous-unités MTA (MeTastasis-Associated protein 1,2,3) dans ce complexe (Bowen et al., 2004; Fujita et al., 2004). Les sous-unités MTA1 et MTA2 sont essentielles pour l'activité déacétylase du complexe NuRD reconstitué *in vitro*. Chd3 et Chd4 ont une activité de remodelage. Elles peuvent induire le glissement des nucléosomes le long de fragments d'ADN (Brehm et al., 2000; Wang and Zhang, 2001). Cette activité intrinsèque est stimulée par les autres sous-unités des complexes NuRD (Wang and Zhang, 2001). La protéine Lsd1 a été immunoprécipitée avec le complexe NuRD et suggère alors une nouvelle composition pour ce complexe (Wang et al., 2009a). Ce dernier complexe pourrait réunir une activité déacétylase et déméthylase. Comme mentionné précédemment, le complexe NuRD a été identifié comme partenaire du facteur de transcription Oct4 (Pardo et al., 2010; van den Berg et al., 2010). Ce complexe pourrait être donc largement impliqué dans les réseaux transcriptionnels qui régulent la balance pluripotence et différenciation.

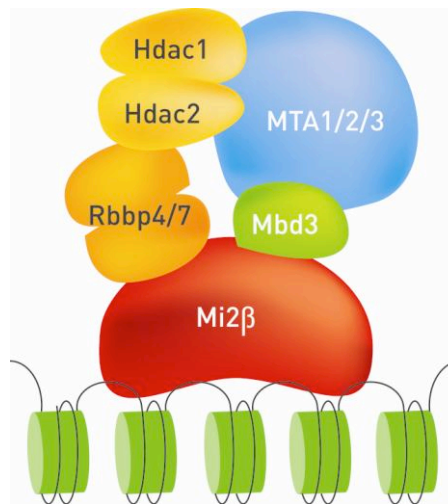


Figure 26: Le complexe NuRD, comprenant la sous unité Mi2 β (Chd4) et les autres protéines associées dont le dimère Hdac1 et Hdac2.

Chd5 est préférentiellement exprimé dans le système nerveux central. Des données récentes montrent que Chd5 est un gène suppresseur de tumeur (Fujita et al., 2008). Ce gène est notamment délété dans les neuroblastomes (Thompson et al., 2003). Il a été récemment trouvé parmi les protéines qui interagissent avec Oct4 dans les cellules ES (Pardo et al., 2010).

La dernière sous-famille, la plus diversifiée et la moins connue, comporte les membres Chd6 à Chd9. Les protéines Chd7, Chd8 et Chd9 possèdent un domaine de liaison à l'ADN ainsi qu'un domaine BRK (Brahma and Kismet) de fonction encore inconnue. Tous les gènes de la troisième sous-famille sont des orthologues de la protéine de drosophile Kismet. Chez la drosophile, le gène kismet a été identifié comme faisant partie du groupe des gènes Trithorax. Ce gène code pour deux protéines : KIS-L et KIS-S. La protéine KIS-L contient un domaine BRK (BReast tumor Kinase) en plus de son domaine ATPase et de ses chromodomaines. La protéine KIS-S ne comprend que le domaine BRK. Chez la drosophile, KIS-L co-localise avec la forme d'initiation de l'ARN polymérase (Srinivasan et al., 2005). Des chromosomes polythènes observés de mutants pour le gène KIS montrent une disparition partielle de l'ARN polymérase II spécifique de l'élongation et du facteur d'élongation SPT6 (Srinivasan et al., 2005). Ces résultats suggèrent que KIS-L favorise la transition de l'initiation vers la phase d'élongation de la transcription.

Chd6 est la protéine la moins caractérisée parmi les Chd. Il a été démontré que Chd6 interagissait avec la sous-unité PA de l'ARN polymérase de la grippe (Huarte et al., 2001), avec certains facteurs de transcription comme p300 et SRC1 (Surapureddi et al., 2002),

avec Nrf2, un autre facteur de transcription (Nioi et al., 2005) ou avec l'ARN polymérase II en élongation (Lutz et al., 2006).

Chd7 est une protéine dont la mutation est responsable de 2/3 des syndromes de CHARGE, acronyme pour Coloboma (yeux), Heart defects (Cœur), Atresia of the choanae (choanale = nez), Retarded growth and development, Genital and/or urinary defects (hypogonadisme), Ear anomalies and deafness (oreilles anormales et surdit ). Ce syndrome a  t  identifi  pour la premi re fois en 1979 par le docteur Hall (Hall, 1979) et r unit pour la premi re des sympt mes qu'on croyait ind pendants. En 1981, une mutation commune   7 patients sur 10 observ s met en cause le g ne Chd7 (Vissers et al., 2004). De multiples mutations au long du g ne Chd7 sont responsables de ces sympt mes (Bosman et al., 2005). Cependant une  tude r cente a montr  que la mutation de Chd7 n' tait retrouv e que dans 67% des cas, ce qui sugg re que des prot ines partenaires de Chd7 pourraient  tre impliqu es dans ce syndrome. Derni rement, il a  t  montr  que Chd7 interagissait avec l'histone H3 sur la lysine 4 mono-, di- et tri-m thyl e (Schnetz et al., 2009). Une autre  tude montre que Chd7 coop re avec le complexe PBAF dans certains tissus pour contr ler la diff renciation des cellules de la cr te neurale (Bajpai et al., 2010). Enfin, Chd8 pourrait interagir avec Chd7 et par l  m me contribuer aux m canismes mol culaires impliqu s dans le syndrome de CHARGE (Batsukh et al., 2010).

Chd8 est aussi connue sous le nom de Duplin pour Duplication inhibitor car il supprime l'effet du g ne siamois chez l'embryon du Xenope par injection d'ARN messenger (Sakamoto et al., 2000). Il a  t  r cemment d montr  que Chd8 interagit avec p53 et module ses propri t s transactivatrices via le recrutement de l'histone H1. Chd8 contrecarre la capacit  de p53   d clencher l'apoptose (Nishiyama et al., 2009). Chd8 contr lerait l'expression de nombreux g nes comme la cyclin E2 et peut interagir avec la polym rase II en cours d' longation (Caldon et al., 2009; Rodriguez-Paredes et al., 2009). De plus, en interagissant directement avec la β -catenin, Chd8 pourrait r guler les g nes cibles de la signalisation Wnt (Kobayashi et al., 2002; Thompson et al., 2008; Yamashina et al., 2006). De m me, Chd8 interagit directement avec WDR5 (Thompson et al., 2008) qui est un composant du complexe MLL m thyltransf rase de l'histone H3 (Song and Kingston, 2008). Plus r cemment, l' quipe de Bochar a montr  que Chd8 r gule l'expression du g ne *HOXA2* (Yates et al., 2010). La m me  quipe a montr  qu'en se liant au r cepteur des androg nes, Chd8 participe   la r gulation de l'expression de leurs g nes-cibles (Menon et al., 2010). De plus, Chd8 s'associe avec la prot ine Stat3 pour activer la transcription de U6, un g ne de l'ARN polym rase III, prouvant ainsi que la polym rase III comme l'ARN polym rase II pouvait faire intervenir des complexes de remodelage afin de faciliter la transcription (Yuan et al., 2007). Enfin, CTCF, un insulateur de chromatine, se lierait   Chd8 pour r guler un

certain nombre de gènes dont le duplex IGF2/H19 soumis à l'empreinte génétique parentale, et participerait à la méthylation CpG de certains locus (Ishihara et al., 2006).

Chd9 (ou CReMM) a été identifiée dans les précurseurs mésenchymateux comme une nouvelle protéine ayant une activité ATPase et liant les motifs riches en A/T (Shur and Benayahu, 2005). Dans les cellules ostéogéniques murines, cette protéine a la particularité de se lier à des promoteurs des gènes spécifiquement exprimés dans le squelette (os ou cartilage) (Shur et al., 2006). Dernièrement, Kim et al. ont identifié le gène Chd9 comme étant lié par des facteurs de transcription majeurs de la pluripotence: Nanog, Dax1, Sox2, Nac1, Oct4 and Klf4 (Kim et al., 2008a; Sharov et al., 2008).

Malgré la grande similarité de séquences entre les protéines Chd qui s'étend de 20% à 60%, les protéines Chd semblent avoir des fonctions indépendantes. En effet, si Chd7 et Chd8 partagent 50% d'homologie de séquence, l'inactivation génique de l'un ou de l'autre est létale embryonnaire (Chd7: (Hurd et al., 2007), Chd8: (Nishiyama et al., 2004)). Ces deux protéines Chd ne semblent donc pas pouvoir se compenser mutuellement. De plus leurs expressions relatives chez la souris adulte (Figure 27) ou dans les cellules ES (voir Annexe) sont différentes, ce qui suggère des rôles particuliers pour chaque protéine Chd.

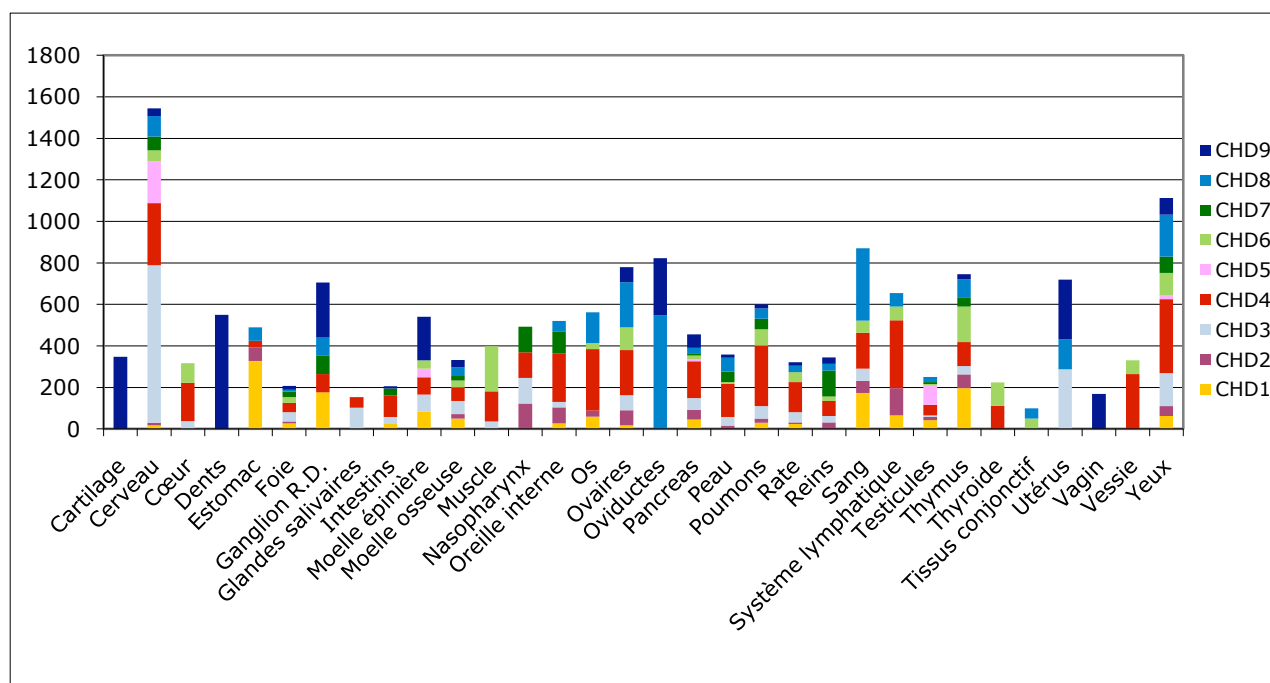


Figure 27: Expression des Chd dans les différents organes de souris adulte d'après la base de données Unigene. Les valeurs en ordonnées donnent le nombre de transcrits par million. (R.D.: Racine dorsale)

L'ensemble des données actuelles connues sur les remodeleurs de chromatine et en particulier sur les Chd laisse entrevoir d'importantes fonctions pour ces différentes protéines dans le contrôle du génome des mammifères. Cependant, leurs rôles respectifs restent encore largement méconnus. L'objectif de ce travail de thèse a donc été d'étudier par une approche génomique globale l'ensemble des protéines Chd afin d'élucider leurs rôles dans le remodelage de la chromatine des cellules ES, d'identifier leurs liens aux réseaux maintenant la pluripotence des cellules ES et finalement de comprendre comment elles se partagent le remodelage de la chromatine. Les évolutions technologiques récentes du séquençage à grande échelle ont rendu possible des approches comparatives pour étudier le rôle des facteurs de transcription et des protéines remodelant la chromatine.

Dans la première partie des résultats, je présenterai la stratégie d'étude retenue pour notre approche expérimentale: étiquetage, recombinaison homologue, établissement des lignées ES (**Chapitre 1**). Puis, je détaillerai les étapes qui nous ont conduits à la mise au point d'un protocole d'immunoprécipitation efficace, nécessaire à la production des échantillons analysés par séquençage haut-débit (ChIP-seq) (**Chapitre 2**). J'analyserai ensuite les données que nous avons obtenues, d'une part pour le complexe NuRD, et d'autre part pour les protéines Chd1, Chd4 et Chd8, de manière comparative (**Chapitre 3 et 4**). Enfin, je conclurai par deux projets supplémentaires que j'ai menés au cours de ma thèse (**Chapitre 5**).

II. Résultats

Chapitre 1. Génération de lignées ES étiquetées pour les facteurs d'intérêts

Objectif du Chapitre

L'étude comparative des gènes cibles des facteurs de remodelage Chd1, Chd2, Chd3, Chd4, Chd5, Chd6, Chd7, Chd8 et Chd9, ainsi que Chaf1a (sous-unité de CAF1) et Hdac2 (deux protéines étudiées en parallèle) nécessite une technique globale et universelle. La génération d'anticorps spécifiques pour chacune de ces protéines aurait entraîné un biais quant à la reproductibilité des expériences. La spécificité différentielle des anticorps n'aurait pas permis de comparer les protéines étudiées directement entre elles. Pour s'affranchir de ces contraintes expérimentales, nous avons choisi d'ajouter une étiquette en fusion avec l'extrémité carboxy-terminale des protéines. Cette étiquette comporte des épitopes contre lesquels des anticorps permettant une immunoprécipitation efficace sont connus. Ils permettront donc d'obtenir des fractions de chromatine enrichies pour la protéine ciblée par chaque lignée de cellule souche. Nous avons choisi d'accroître cette spécificité en combinant deux étiquettes qui permettront successivement deux purifications successives de haute affinité comme cela avait été précédemment montré par l'équipe de Séraphin (Rigaut et al., 1999). Nous avons choisi les épitopes HA et FLAG contre lesquels de bons anticorps immunoprécipitants sont connus.

Pour introduire l'étiquette choisie au locus du gène, nous avons construit des vecteurs de recombinaison homologue. La technique d'obtention de ces plasmides de 10kb par PCR est longue, laborieuse et des mutations sont facilement introduites. C'est pourquoi nous avons choisi une méthode de construction des plasmides basée sur la recombinaison homologue chez *Escherichia coli*, décrite par Liu et al (Liu et al., 2003). Cette méthode a l'avantage d'augmenter la rapidité d'obtention des plasmides tout en limitant les mutations introduites. Les lignées stables de cellules souches murines ont ensuite été obtenues par électroporation des vecteurs et sélection des événements de recombinaison homologue. Le point central de notre stratégie réside en l'étiquetage d'un allèle au locus de chacune des protéines étiquette, ce qui place l'expression de la protéine étiquetée sous le contrôle de son propre promoteur. Cette stratégie permet de respecter le taux de transcription naturel de la protéine.

A. Choix d'une étiquette

1. Choix des épitopes

Notre but est de choisir des épitopes contre lesquels des anticorps immunoprécipitants sont connus et permettent une immunoprecipitation de haute affinité. Nous avons sectionné dans un premier temps 3 épitopes:

L'épitope 6xhistidines (Roche), comporte un enchainement de 6 histidines: HHHHHH

L'épitope FLAG (Sigma), est composé de la séquence aminoacide suivante:
DYKDDDDK

L'épitope HA est composé de la séquence aminoacide suivante: YPYDVPDYA

Le site de clivage de la protéase TEV (Tobacco Etch Virus) est composé de la séquence aminoacide suivante: ENLYFQS. Le site de clivage de cette protéase est généralement donné E-X-X-Y-X-Q-G/S, mais la séquence que nous avons choisie a été montrée comme la plus efficace.

Un plasmide comportant la séquence d'une étiquette possédant les trois épitopes 6histidines-flag-HA (6HFH) nous a été gentiment fourni par le Dr Saadi Khochbin.

2. Choix des étiquettes et mise au point expérimentale

Après de nombreux tests, nous avons constaté que cette étiquette n'était pas optimale puisque les résultats d'immunoprécipitation étaient variables en fonction des protéines. En particulier nous avons pu constater que:

1) **L'épitope 6xHis**. Nous n'avons pas pu utiliser cet épitope pour immunoprécipiter les protéines d'intérêt. Deux approches avaient été envisagées. La première utilisait une colonne de nickel pour retenir les protéines possédant l'épitope. La deuxième approche consistait à utiliser des anticorps dirigés contre cet épitope. Aucune de ces deux approches n'a permis de retenir (même une fraction) les protéines portant l'étiquette.

De plus, plusieurs histidines successives sont susceptibles de perturber la conformation des protéines comme cela pourrait être le cas pour la protéine Hdac2. Par ailleurs cet épitope peut être naturellement présent au sein des protéines murines (domaine en doigt de zinc).

2) **L'épitope FLAG**. Les immunoprécipitations réalisées avec cet épitope ont permis de retenir en partie les protéines qui le possédaient. Cependant la quantité de protéines

immunoprécipitées était insuffisante lorsqu'un seul épitope est présent dans la majorité des cas.

C'est pourquoi, nous avons conçu deux nouvelles étiquettes (Isabelle Hmitou et Matthieu Gérard en collaboration avec Françoise Ochsenbein).

La première étiquette (3FTH) comporte trois épitopes FLAG dont le plus extérieur est séparé des autres épitopes par une hélice alpha. Cette hélice permet d'augmenter l'efficacité d'immunoprécipitation, probablement en éloignant le reste de l'épitope de la protéine. Il devient alors possible de s'affranchir d'une partie de l'encombrement stérique généré par la protéine intrinsèque. Afin de pallier d'éventuels problèmes d'élution à l'aide de peptides compétiteurs, nous avons ajouté un site de clivage par la protéase TEV qui peut remplacer ou compléter l'élution après une IP (voir Annexe pour la séquence peptidique). Un seul épitope HA a été conservé.

La deuxième étiquette comporte un unique épitope FLAG et un épitope HA (FH). Cette étiquette est quasi identique à l'étiquette 6HFH, mais le stretch d'histidine pouvant interagir de façon aspécifique avec la protéine endogène a été enlevé. Cette étiquette a l'avantage d'être moins importante en taille que l'étiquette 3FTH: 3KDa versus 7,6KDa.

En conclusion, nous avons observé expérimentalement que l'immunoprécipitation des protéines était possible avec les différentes étiquettes utilisées. Cependant, nous avons observé que l'étiquette 3FTH possédait la meilleure efficacité d'immunoprécipitation grâce aux triple FLAG et la meilleure élution grâce au site de clivage TEV.

B. Adaptation du protocole de « recombineering » pour l'étiquetage des protéines

Ce protocole adapté du protocole de recombineering de Liu et al. a été mis au point par Fayçal Boussouar dans notre équipe et ensuite amélioré par Sébastien Graziani et moi-même. L'ensemble de ce protocole est disponible dans l'Annexe 1 et est résumé dans la [Figure 28](#).

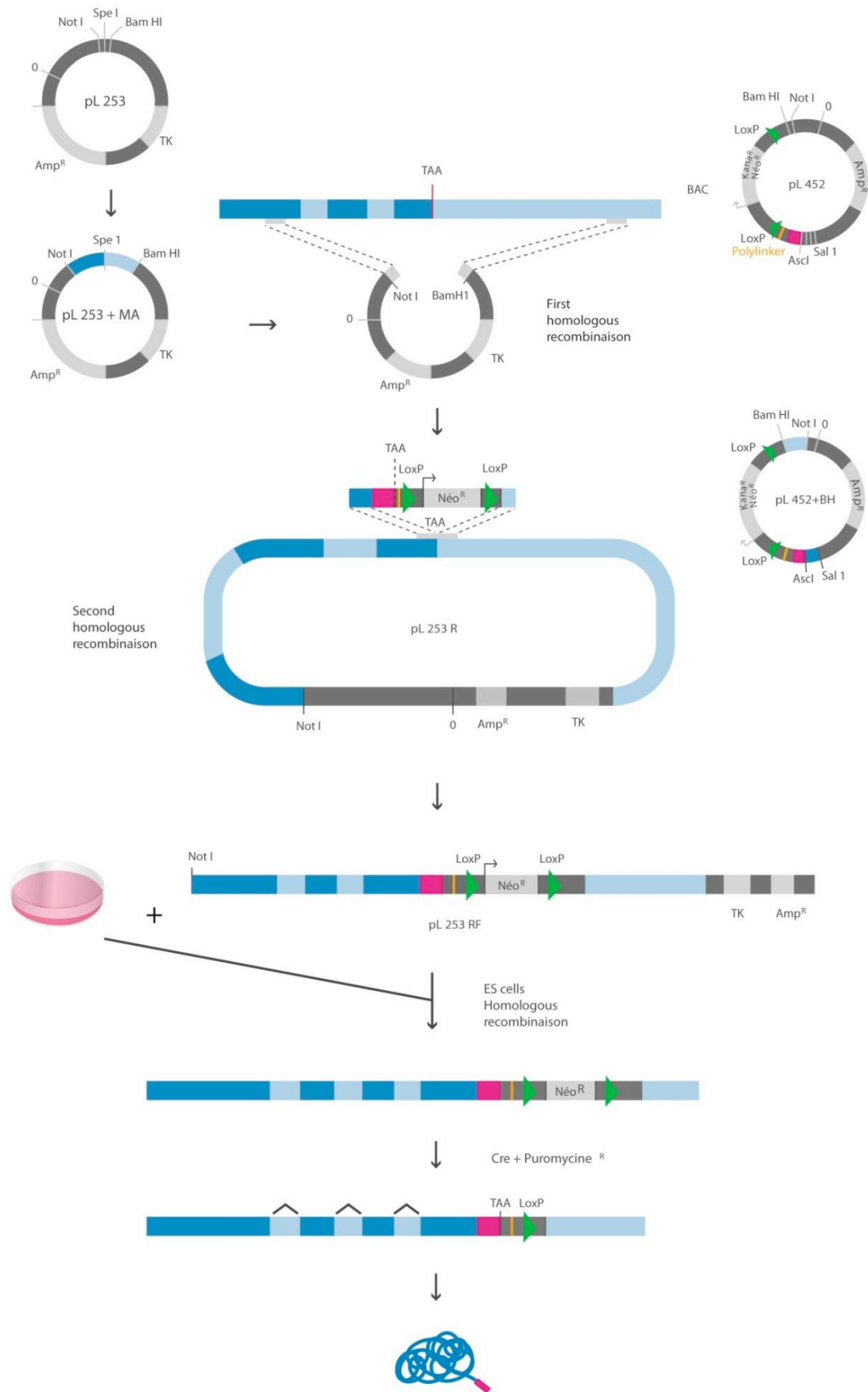


Figure 28: Résumé schématique de la méthode de génération des constructions pour la recombinaison homologue dans les cellules ES

La construction des vecteurs de recombinaison homologue se fait en deux étapes. La première étape comprend l'insertion de 10kb d'ADN génomique de souris correspondant à [-5kb:+5kb] de la région ciblée dans un vecteur plasmidique. Cet ADN est récupéré depuis un clone de BAC. Cette insertion se fait par recombinaison homologue afin de ne pas introduire de mutation. Puis une deuxième étape correspond à l'insertion de l'étiquette et de la cassette de sélection kanamycine/neomycine. Cette cassette à l'avantage de posséder un double promoteur bactérien et murin. En présence de drogue, il est possible de sélectionner les cellules qui possède cette cassette de résistance chez les bactéries et en cellule ES.

1. Introduction du BAC dans les bactéries permettant la recombinaison homologue

La première étape consiste à choisir un clone de BAC qui possède la région [-5kb:+5kb] autour du codon stop de la protéine d'intérêt. Le BAC est purifié et électroporé dans la souche bactérienne d'*Escherichia coli* possédant les protéines de la recombinaison homologue: SW102. Dans cette souche de bactérie, trois gènes Red du bactériophage λ permettent l'insertion de fragments double brin d'ADN. Le gène *exo* code pour une exonucléase qui agit sur l'extrémité 5' du fragment double brin d'ADN pour laisser une extrémité 3' débordante. Le gène *bet* code pour une protéine d'appariement qui se lira à l'extrémité 3' et favorisera son arrimage sur une séquence similaire. Un dernier gène *gam*, code pour une protéine qui inhibe l'activité du complexe RecBCD qui dégrade les ADN double brin linéaires. Les trois gènes *exo*, *bet* et *gam* sont sous le contrôle du promoteur λP_L qui est lui-même sous le contrôle du répresseur $\lambda cI857$ sensible à la température. A 42°C, la répression des gènes est levée et l'expression des protéines de recombinaison homologue est induite.

2. Première recombinaison homologue : incorporation des 10kb d'ADN génomique qui permettront de cibler la recombinaison homologue

Construction des plasmides de recombinaison homologue pL253

Le plasmide pL253 est spécifique à chaque construction. Il comporte deux séquences homologues de 500 pb aux extrémités des 10kb d'intérêt. Ces deux extrémités vont permettre le recombineering (recombinaison homologue). C'est donc ces séquences qui guident l'introduction des 10kb d'intérêt dans le plasmide de recombinaison homologue. Ces séquences de 300 à 500pb sont appelées « mini-arms » ou MA. Des études ont montré que

la recombinaison était possible avec des MA plus petits ~50bp mais nous n'avons pas testé (Liu et al., 2003).

Les procédures du clonage sont détaillées dans Annexe.

Nous avons pu montrer qu'il était possible de regrouper ces deux étapes de clonage en effectuant une double ligation qui ajoute les deux bras d'homologies au vecteur pL253. Cependant l'efficacité de cette ligation est nettement plus faible d'un simple clonage.

Première recombinaison homologue

Cette étape consiste à incorporer par recombinaison homologue dans le plasmide pL253, 10000pb correspondant à [-5kb : +5kb] autour du codon stop de la protéine d'intérêt.

Cette étape est la plus limitative: pour certaines protéines nous n'avons jamais obtenu de clone recombinés. C'est le cas notamment pour Chd7 et Hells (Smarca6). Plusieurs hypothèses peuvent être avancées. La conformation du BAC peut empêcher la recombinaison homologue. Certaines séquences peuvent potentiellement être néfastes.

3. Deuxième recombinaison homologue : incorporation de l'étiquette en fusion avec l'extrémité C terminale de la protéine.

Le but de cette dernière étape est d'introduire dans les 10kb de recombinaison homologue du pL253R, l'étiquette en fusion avec le domaine C terminal de la protéine ciblée.

Le plasmide pL452 comprend l'étiquette qui sera fusionnée à la partie C terminale de la protéine d'intérêt et la cassette de sélection. Ce plasmide a donc été modifié depuis la publication de Liu et al. Chaque plasmide pL452 possède sa propre étiquette c'est pourquoi on distinguera: pL452-6HFH, pL452-FH et pL452-3FTH.

Pour que l'étiquette qui est en fusion avec la C-terminal de la protéine d'intérêt soit exprimée, la séquence finale au niveau du site Asc1 est vérifiée par séquençage.

4. Vérification des clones

Nous avons constaté que la totalité des clones positifs que nous obtenions possédaient en plus de du plasmide final pL253RF, une portion pouvant varier de ½ à ¼ de plasmide non recombiné pL253R. Ceci n'avait pas été décrit dans la publication de Liu et al.

La présence de ce plasmide n'a pas gêné l'électroporation et la recombinaison en cellules ES.

Changement d'étiquette

Deux méthodes sont théoriquement possibles. La première méthode consiste à cloner directement le fragment d'intérêt dans le pL253RF. Premièrement, nous digérons le pL253RF et le plasmide pL452 possédant l'étiquette d'intérêt par Asc1 et Rsr2 (enzymes uniques). La digestion du pL452 est ensuite migrée sur gel et la bande d'intérêt est récupérée. Cette dernière est ensuite clonée dans le pL253RF qui a été préalablement digéré. Cette technique a l'avantage de se faire en une seule étape mais la ligation est difficile. Nous n'avons eu qu'un seul succès pour la protéine Chd8.

La deuxième méthode est plus longue mais possède un taux de réussite vingt fois plus élevé que la première méthode. Cette seconde méthode consiste à modifier le pL452 possédant les deux BH déjà clonés. Il faut ensuite digérer ce plasmide par Asc1 et Rsr2 et remplacer sa séquence par la séquence de l'étiquette désirée (digéré au préalable par Asc1 et Rsr2, voir première méthode). Après avoir obtenu le nouveau pL452, il suffit de procéder à la deuxième étape de recombinaison homologue qui ne pose pas de difficultés particulières.

5. Obtention de lignées ES stables étiquetées

Choix de la lignée ES pour l'électroporation

Nous possédons trois lignées de cellules souches de souris au laboratoire: 46C, AT1 et LTM7.

La lignée 46C établie dans le laboratoire d'Austin Smith (Ying et al., 2003b) est la plus facile à cultiver. Dans cette lignée, le cadre de lecture ouvert (ORF) de Sox1 est remplacé par la séquence codante de la GFP ainsi qu'un gène de résistance à la puromycine sous le contrôle d'une IRES. Le gène Sox1 est impliqué dans le développement du système nerveux central et dont l'expression est nulle en cellules ES (nos propres données). Le marqueur de fluorescence GFP, exprimé au cours de la différenciation neuronale, permet de suivre l'état de différenciation des cellules. La sélection des précurseurs neuronaux se fait soit par FACS soit par sélection puromycine transitoire. Cette lignée est de génotype XY.

La lignée AT1 permet la transmission germinale mais la culture de ces cellules peut se révéler plus fastidieuse. Cette lignée est XY.

La lignée LTM7 est XX et permet donc d'étudier par exemple les conséquences liées à l'inactivation du X.

Afin de pouvoir poursuivre notre étude par de futures investigations quant au devenir de protéines lors de la différenciation, nous avons choisis la lignée cellulaire 46C.

Pour l'étude du variant d'histone H2AL2, l'expression de ce variant étant strictement limité aux derniers stades de la spermatogenèse et donc plus généralement aux testicules, nous avons choisi la lignée cellulaire AT1 qui permet une transmission germinale plus aisément (voir Chapitre 5).

Electroporation des plasmides de recombinaison homologues

Vingt microgrammes μg de vecteur de recombinaison sont linéarisés par Not1, qui possède un site à l'extérieur des 10kb d'intérêt servant à la recombinaison homologue. Le vecteur digéré est ensuite électroporé dans les cellules ES. Les cellules qui ont correctement intégré le plasmide possèdent alors une cassette de résistance à la néomycine. Le lendemain, une sélection positive à la généticine (G418) est appliquée pendant 6 à 7 jours (250mM final). Ceci permet de contre-sélectionner les clones ES qui n'ont pas intégré le plasmide. Pendant les deux premiers jours, une sélection négative au ganciclovir (2 μM final) est ajoutée. L'enzyme thymidine kinase virale présente sur la partie bactérienne du plasmide est capable de phosphoryler certains analogues de la thymidine. Parmi ces analogues, lorsque le ganciclovir est utilisé, il est phosphorylé et devient hautement toxique pour la cellule.

Une semaine plus tard, les clones issus de cellules uniques sont repiqués et cultivés séparément dans des plaques 24 puits avec des cellules nourricières pendant 4 à 5 jours. Une partie est ensuite congelée pendant qu'une autre partie est amplifiée puis lysée. L'ADN de ces clones est ensuite préparé, digéré et analysé par la technique du Southern Blot (Figure 29). Les clones recombinants sont identifiés à l'aide d'une sonde externe à la construction qui a servi à la recombinaison homologue (Figure 30).

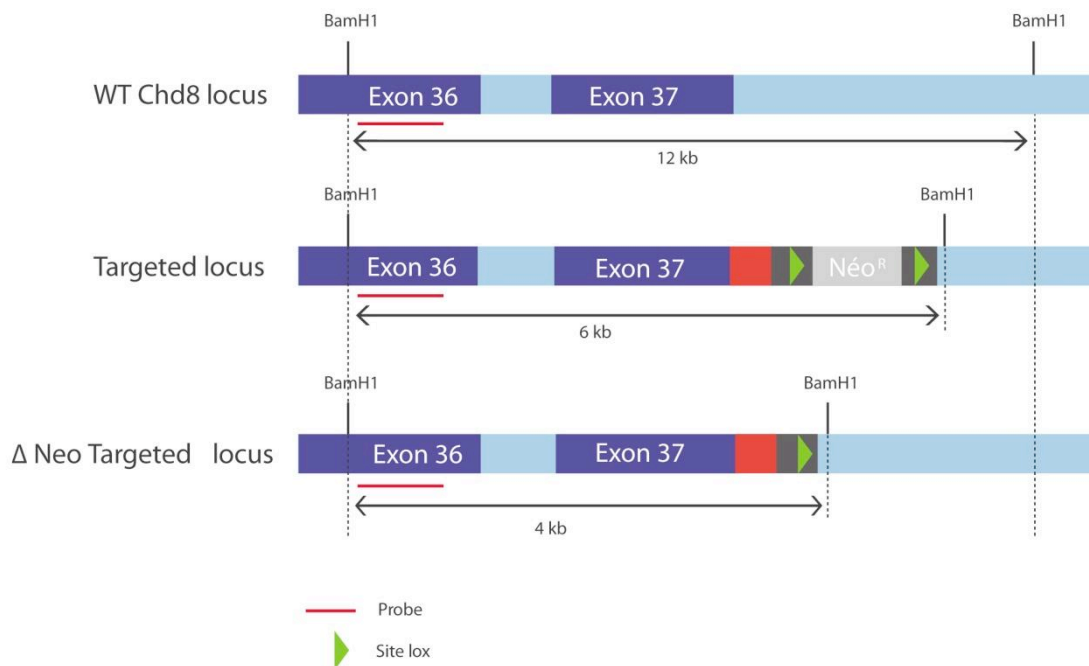


Figure 29: Stratégie de génotypage pour le gène *Chd8*. L'étiquette est indiquée par un bloc rouge

Nous avons obtenu des taux de réussite de recombinaison allant de 33% à 0.5%. Cette disparité est expliquée par le fait que les régions chromosomiques ciblées étaient très différentes et que les séquences ADN environnantes influent sur le taux de recombinaison homologue. Par exemple, une séquence unique sera une cible facilement reconnue alors qu'une région comportant des éléments répétés répartis de façon aléatoires dans le génome sera une cible moins facile car elle aura des cibles homologues multiples.

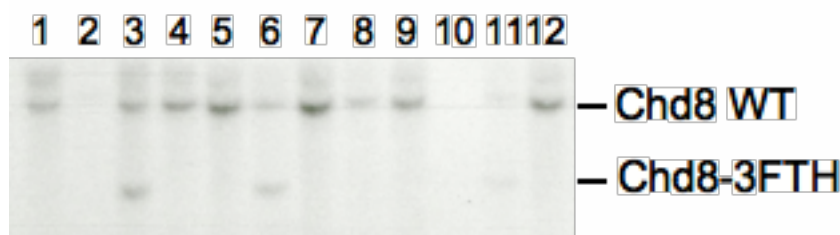


Figure 30: Exemple d'obtention de clones recombinants pour la construction Chd8-3FTH. Les clones positifs sont les numéros 3, 6 et 11.

De façon surprenante, nous avons obtenu un clone qui avait recombiné sur les deux allèles pour le gène *Chd8* avec l'étiquette 6HFH. Compte tenu du fait que le taux de recombinaison était de 33% dans cette expérience, mathématiquement la probabilité

d'obtenir un double recombinant est de 0.10%. Or nous avons contrôlé 192 clones il était donc possible d'obtenir 1 clone ayant effectué une recombinaison sur chaque allèle.

Le caryotype de chaque clone est ensuite contrôlé par comptage de plusieurs figures mitotiques des 40 chromosomes sous microscope.

Pour les lignées Chd8 et Chd4 qui nous ont servis à mettre au point le protocole d'immunoprecipitation de chromatine, nous avons choisi de retirer la cassette de résistance à la néomycine qui réduisait l'expression de 30% à 50%. Cette opération est effectuée par une double électroporation du plasmide exprimant la recombinaise Cre et d'un plasmide de résistance à la puromycine. Un exemple des clones obtenus pour la lignée Chd8 est illustré dans la [Figure 31](#).

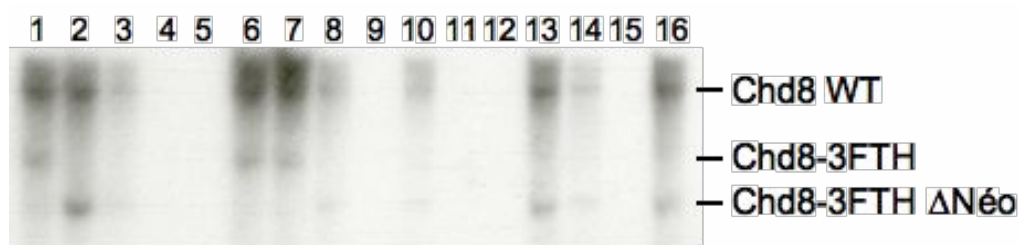


Figure 31: Génotypage des clones électroporés avec les plasmides Cre et PuroR. Les clones positifs possédant l'allèle sauvage et l'allèle étiqueté sans néo sont 2, 3, 8, 10, 13, 14, 16.

Conclusion

Grâce à la technique du recombineering développée par Liu et al et adaptée à l'étiquetage des protéines en carboxyterminal par Fayçal Boussouar au laboratoire, nous avons pu obtenir 18 vecteurs de recombinaison homologue et les lignées ES correspondantes. Cependant, cette technique n'a pas pu fonctionner pour une construction : le gène *Chd7*. L'étape limitative de la technique est la première étape du recombineering pour les raisons développées plus haut. Huit lignées de cellules ES ainsi générées ont été utilisées pour réaliser des expériences d'immunoprécipitation de chromatine afin d'identifier les séquences-cibles de ces protéines. Ces lignées ES possédant un allèle étiqueté pour chaque Chd, ont été regroupées sur la [Figure 32](#).

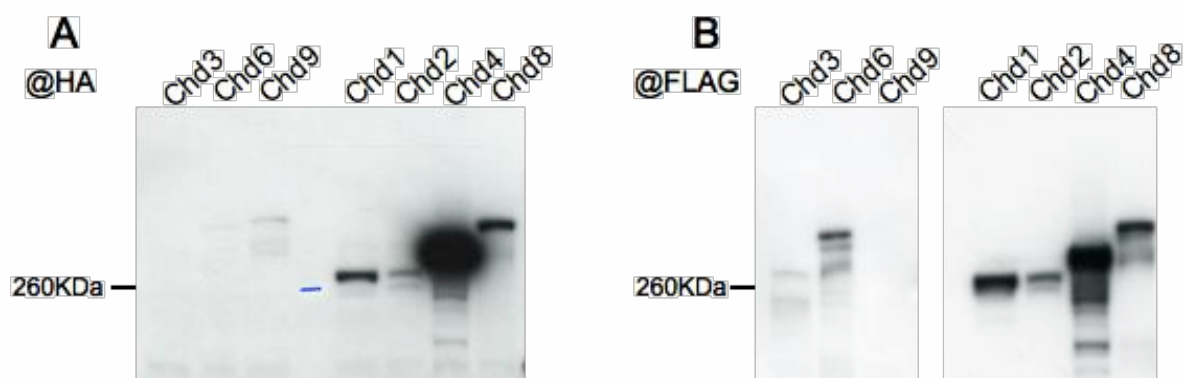


Figure 32: Expression des protéines Chd étiquetées dans chaque lignée spécifique. L'étiquette utilisée est 3FTH à l'exception de la lignée Chd9 (étiquette 6HFH). Les quantités d'extrait cellulaire total déposées sont 40ug pour Chd3, Chd6 et Chd9 et 20ug pour Chd1, Chd2, Chd4 et Chd8. La même membrane de western blot a été révélée avec un anticorps anti-HA (Partie A) et un anticorps anti-FLAG (Partie B).

Chapitre 2. Mise au point d'un protocole performant pour l'immunoprécipitation efficace des protéines étiquetées

Introduction

Afin de mieux comprendre le rôle des protéines Chd dans le remodelage de la chromatine, nous avons purifié et identifié leurs séquences-cibles par immunoprécipitation de chromatine, en utilisant les lignées ES stables obtenues par recombinaison homologue. Chaque lignée ES, étiquetée pour un facteur Chd, est utilisée dans une expérience indépendante. Dans ce but, de nombreux protocoles ont déjà été développés. Les étapes importantes de ces protocoles comprennent une fixation de la protéine à l'ADN, une lyse des cellules puis un fractionnement de l'ADN. Nous avons testé 5 protocoles (nommés de A à E) différents de préparation de chromatine de cellules ES (voir Annexe pour leur descriptif). Ces cinq protocoles étaient très différents quant aux paramètres présentés ci-dessus. Cependant aucun de ces protocoles ne nous a permis d'immunoprécipiter efficacement nos protéines d'intérêt. C'est pourquoi nous avons développé notre propre protocole, en combinant des approches déjà utilisées dans la littérature, ce qui permet d'obtenir une immunoprécipitation de très haute affinité et de fractionner efficacement l'ADN qui sera ensuite séquencé. Ce protocole performant fait l'objet d'une publication.

A. Test de différents protocoles standard.

1. Présentation: Les protocoles testés

Nous avons testé différents protocoles qui sont regroupés ci-après ([Tableau 4](#)). Les étapes majeures y sont données. Les protocoles testés sont regroupés dans leur version intégrale en Annexe 2. Ces protocoles ont été testés sur les 8 protéines Chd qui ont été étiquetées ainsi que la protéine Hdac2, Chaf1a et Atad2 (cette dernière en collaboration avec Fayçal Boussouar à Grenoble).

Résultats

	Protocole A	Protocole B	Protocole C	Protocole D	Protocole E	Protocole final
Quantité de boîtes de 10 cm	4 à 5	4 à 5	4 à 5	12 à 24	12 à 24	24
Crossking des cellules	0,4% F	0,4% F	0,4% F	0,4% F	-	1% F
Recupération des cellules	Grattées	Grattées	Grattées	Grattées	Tripsinées	Grattées
permeabilisation des cellules	Swelling buffer 0,5% NP40 85mM KCl 5mM Pipes	-	-	-	Solution II 0,2% NP40 final	Buffer II 0,3% NP40 final
Lyse des cellules	Tampon de lyse Noyaux 0,1% SDS 1X Triton 150mM NaCl 25mM Tris pH8 3mM EDTA pH8	Lysis buffer 1% SDS 50mM Tris pH8 10mM EDTA pH8	FA/SDS Buffer 0,1% SDS 0,1% Na Deoxycholate 1X Triton 150mM NaCl 50mM Hepes KOH pH7,5 1mM EDTA	FA/SDS Buffer 0,1% SDS 0,1% Na deoxycholate 1X Triton 150mM KCl 50mM Tris pH7,5 1mM EDTA	-	-
Traitement spécifique	-	Dilution de la chromatine dans ChIP dilution buffer 0,01% SDS 167mM NaCl 1,1X Triton 1,2mM EDTA pH8 16,7mM Tris pH8	-	-	Gradient de sucrose pour récupérer les noyaux	Reprise dans le Buffer KN 60mM NaCl 15mM KCl 20mM Tris pH7,5 0,34M Sucrose 1mM CaCl ₂
Fragmentation de l'ADN	Sonication 8 cycles 30"ON 30"OFF	Sonication 8 cycles 30"ON 30"OFF	Sonication 17 cycles 20"ON 40"OFF	Sonication 17 cycles 20"ON 40"OFF	Digestion Mnase dans le tampon MNase 48U/11.10 ⁶ noyaux	Digestion Mnase 80U/10 ⁷ cellules Sonication 4 cycles 20"ON 40"OFF
Tampon d'IP	Tampon de lyse	Lysis buffer	FA/SDS Buffer NaCl	FA/SDS Buffer KCl	Tampon Mnase	Buffer KN

Résultats

	Protocole A	Protocole B	Protocole C	Protocole D	Protocole E	Protocole final
Quantité de billes pour une IP	50ul de slurry billes agarose + 3ug d'anticorps HA ou FLAG	50ul de slurry billes agarose + 3ug d'anticorps HA ou FLAG	50ul de slurry dynabeads + 3ug d'anticorps HA ou FLAG	50ul de slurry dynabeads + 3ug d'anticorps HA ou FLAG		80ul de billes centrifugées pour IP n°1, 40ul de billes pour IP n°2
Préparation des billes	Saturation des billes avec 0,5% BSA 2 heures dans tampon IP S, 1 lavage	Saturation des billes avec 0,5% BSA 2 heures dans tampon IP S, 1 lavage	Saturation des billes avec PBS/BSA 0,1 %	Saturation avec BSA 1% et ARNt 1 heure, lavage avec FA/SDS KCl		2 lavages TEGN
Conditions d'IP	IP overnight 4°C	IP overnight 4°C	IP 2 heures 1300 rpm 21°C	IP 2 heures 1300 rpm 21°C		IP overnight 4°C
Lavages des IP	TSE 25 mM Tris pH7,5 2,5 mM EDTA 0,3 M Sucrose	Low Salt Buffer (150mM NaCl) High Salt Buffer (500mM NaCl) 0,1% SDS 1X Triton	3 lavages FA/SDS NaCl 500mM 1 lavage FA/SDS 150mM NaCl 1 lavage TE	1 lavage Low Salt 1 lavage Low LiCl 2 lavages TE 1 lavage High Salt		10 lavages tampon TEGN 0,01% NP40 10% Glycerol 150mM NaCl 3mM MgCl ₂ 0,1mM EDTA pH8 20mM Tris pH7,5
	Tampon de lavage 10 mM Tris pH8 0,25MLiCl 0,5% NP40 0,5% Na-déoxycholate 1 mM EDTA	Low LiCl Buffer (250mM) High LiCl Buffer (500mM LiCl) 1% NP40 1% Na Deoxycholate	-	OU 2 lavages Low Salt 3 lavages High Salt		
Elution peptide	elution peptide 0,6mg/ml dans tampon IP S	elution peptide 0,6mg/ml dans tampon IP S		Elution peptide 0,6mg/ml dans FA/SDS KCl 4°C overnight ou 15°C 4 heures		Elution peptide 1mg/ml dans tampon TEGN overnight 4°C
Autre type d'élution	Non décrit	Non décrit	Non décrit	(IP FLAG) Elution glycine 10' Glycine 0,1M repeter 1 fois	Non décrit	Elution Tev dans tampon TEGN
Elution SDS	IP elution buffer 1% SDS 0,1M NaHCO ₂	IP elution buffer 1% SDS 0,1M NaHCO ₃	Tampon pronase 0,5% SDS 5mMEDTA 25mM Tris pH7,5	IP elution buffer 1% SDS 0,1M NaHCO ₃		IP elution buffer 1% SDS 0,1M NaHCO ₃

Tableau 4: Récapitulatif des différents protocoles testés

Choix des agents pontants

Le formaldéhyde a été défini comme agent pontant par défaut dans toutes les expériences de préparation de chromatine. Le formaldéhyde permet de fixer les protéines entre elles grâce à un pontage réversible entre les amines primaires des protéines (les lysines et histidines). La capacité du formaldéhyde à ponter les protéines aux acides nucléiques reste encore un sujet controversé dans la communauté scientifique.

Dans un deuxième temps, nous avons choisi de tester le DTBP, dimethyl 3,3'-dithiobispropionimidate, (Fujita and Wade, 2004), le DSG, disuccinimidyl glutarate, et l'EGS, ethylene glycol bis[succinimidylsuccinate], pour améliorer les pontages protéine-protéine (Figure 33 et Figure 34).

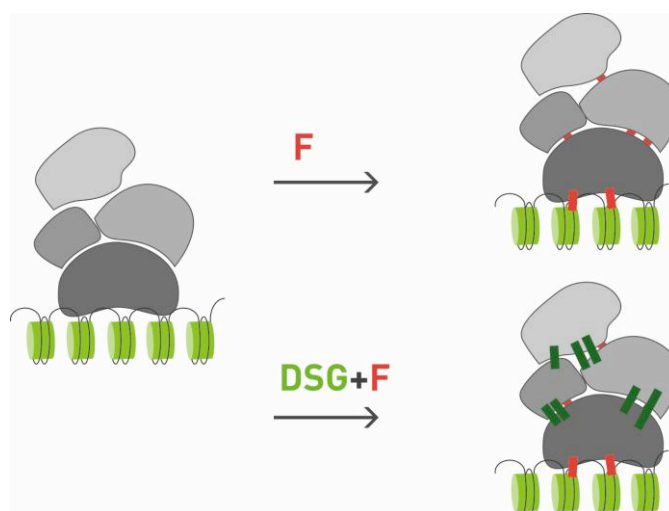


Figure 33: Amélioration de l'efficacité du pontage par d'autres agents que le formaldéhyde

Procédure des tests

Lors de la période de test des protocoles, les procédures suivantes ont été suivies. Lorsque la préparation de chromatine a permis d'avoir des fragments de chromatine majoritairement fractionnés autour de 300 paires de bases, le protocole est retenu. Puis, deux immunoprécipitations à l'aide des anticorps anti-FLAG et anti-HA sont testées en parallèle. Aucun gène cible n'était connu pour les 11 protéines étudiées. C'est pourquoi nous avons choisi de tester la qualité de nos immunoprécipitations dans un premier temps par western blot.

Description des différents protocoles

La première étape commune aux protocoles qui fixent les protéines à l'ADN est une fixation des cellules ES par le formaldéhyde directement dans le milieu D15. La concentration finale de formaldéhyde utilisée est 0,4% final ou 1%.

Durant la description de ces protocoles ces inhibiteurs de protéases ne sont pas mentionnés mais sont toujours ajoutés (Complete, PMSF 1mM final) sauf mention du cas contraire.

2. Protocole A (extraction des noyaux)

Principe

Ce protocole se décompose en quatre étapes. L'étape de pontage qui a été décrite au paragraphe suivant est suivie par une étape de gonflement des cellules dans le Swelling buffer 20' dans la glace. Puis les cellules sont lysées à l'aide d'un tampon contenant des détergents: 0,1%SDS, 1X triton et du sel: 150mM NaCl, 10' dans la glace. La préparation de chromatine se poursuit par une étape de sonication comprenant 8 cycles 30" ON 30" OFF ou 10 cycles 20" ON/ 40" OFF.

Bilan

Les expériences réalisées avec ce protocole ont montré que la fragmentation de la chromatine permettait d'obtenir des fragments de 100-10000 paires de bases, et n'était donc pas optimale pour le séquençage à haut débit pour lequel une taille moyenne de fragmentation de 300 paires de bases est recommandée. De plus les expériences d'immunoprécipitation de chromatine (ChIP) réalisées après la préparation, ont montré l'impossibilité d'immunoprécipiter les protéines étiquetées dans ces conditions.

3. Protocole B (SDS)

Principe

L'originalité de ce protocole repose sur la deuxième étape, c'est-à-dire l'étape de lyse suivant le pontage et la collecte des cellules. La lyse des cellules s'effectue dans un tampon excessivement riche en détergent, contenant 1% SDS. Cette étape rapide de 10 min à

température ambiante, permet aux cellules d'éclater complètement grâce à la dégradation des membranes par le SDS et de libérer la chromatine contenue dans le noyau. L'étape de sonication est faite dans le tampon de lyse cellulaire. La forte concentration en SDS permet d'améliorer la sonication. Seulement huit cycles 20" ON/ 40" OFF sont requis pour une fragmentation de l'ADN nucléaire majoritairement concentrée autour de 300 paires de bases. Les échantillons sont ensuite dilués 10 fois dans un tampon sans SDS pour revenir à une concentration en SDS de 0,1%. De trop fortes concentrations en détergents inhibent en effet les interactions entre anticorps et leur épitopes.

Bilan

Ce protocole est rapide. Il permet une bonne fragmentation de la chromatine (~300pb). Deux lignées de cellules ES contenant les protéines Chd4 et Hdac2 étiquetées ont été utilisées afin de réaliser des expériences d'immunoprécipitation de chromatine à partir de ce protocole. Les résultats ont montrés que l'immunoprécipitation était très faible et moins performante que le protocole C réalisé en parallèle.

4. Protocole C (FA/SDS)

Principe

Ce protocole se déroule en trois étapes. Après la première étape de pontage et de collecte des cellules, la lyse cellulaire se fait directement dans le tampon FA/SDS contenant 0,1% SDS; 0,1% Na Deoxycholate et 1% Triton. Ces trois détergents combinés permettent premièrement la lyse des cellules et deuxièmement une sonication efficace. Ce tampon, à l'instar des autres protocoles, contient 150mM NaCl. L'étape de sonication comprenait 8 cycles de sonication 20" ON/ 40" OFF, mais ce nombre de cycles n'était pas suffisant pour fragmenter la chromatine. Quatre à cinq cycles supplémentaires étaient requis après analyse de l'ADN sur gel.

Bilan

Le protocole C est un bon compromis entre le protocole A et le protocole B: La lyse est faite directement dans 0,1% SDS, 1% Triton ce qui supprime l'étape de lyse 1% SDS du protocole B qui était destructrice pour nos protéines étiquetées. La sonication est plus efficace que dans le protocole A. Les fragments obtenus étaient compris entre 500 et 100 paires de bases ce qui est compatible avec le séquençage à haut débit.

Avec ce protocole de préparation de chromatine, nous avons obtenu les premiers résultats positifs d'immunoprécipitation de nos protéines, notamment avec Hdac2. Cette protéine se trouve parmi les plus abondantes des protéines étiquetées. De plus, sa petite taille de 60KDa comparée aux Chd de 300KDa environ, permet une manipulation plus facile par la technique de western blot.

Cependant, la qualité d'immunoprécipitation étant faible nous avons effectué une série d'améliorations pour aboutir au protocole D.

5. Protocole D (FA/SDS amélioré)

Principe

Ce protocole reprend les mêmes étapes que le protocole C. Cependant nous avons modifié la composition du tampon FA/SDS en remplaçant l'Hepes KOH pH7,5 par du Tris pH7,5 pour homogénéiser les solutions utilisées lors de la préparation de chromatine. En effet, elles sont toutes tamponnées à l'aide de Tris pH7,5. Nous avons aussi remplacé le NaCl par du KCl car nous avons constaté que ce sel permettait une meilleure immunoprécipitation des protéines.

De plus, nous avons mis au point la sonication plus finement. Dix-sept cycles 20"ON/40"OFF étaient requis pour fragmenter la chromatine autour de 300 paires de bases.

Bilan

Ce protocole a été testé avec les protéines Chd1, Chd4, Chd8, Hdac2, Chaf1a et Atad2. Nous avons montré qu'il était possible d'immunoprécipiter Chaf1a et Atad2. Cependant la double immunoprécipitation n'a pas été possible. Dans le cas de la protéine Hdac2 le tandem a été réalisé avec une très bonne efficacité.

Après analyse des fractions immunoprécipitées pour les lignées Chd et Hdac2 sur le Génome Analyser (Illumina), le séquençage s'est révélé négatif. Deux problèmes se sont dressés: 1) pour les protéines Chd que nous n'avons pas réussi à visualiser par Western Blot, nous sommes parvenu à la conclusion que le protocole de la double immunoprécipitation ne fonctionnait pas, 2) Pour la protéine Hdac2, il était clair que le protocole permettait une très bonne immunoprécipitation mais il n'y avait pas d'ADN spécifique associé à la protéine.

Conclusion du protocole D

Améliorations

1. Nous avons montré qu'il était possible d'immunoprécipiter nos protéines avec l'anticorps HA.
2. Ce protocole permettait une bonne immunoprécipitation HA et FLAG de la protéine Hdac2. Nous avons pu mettre en évidence qu'il était possible de faire une double immunoprécipitation avec cette protéine étiquetée.
3. Nous avons aussi pu montrer que le protocole D permettait d'obtenir des fragments d'ADN compatibles avec le Genome Analyser.

Insuffisances du protocole D

Cependant, plusieurs problèmes persistent avec ce protocole.

1. Les protéines de haut poids moléculaire se visualisent difficilement en western blot. Nous avons postulé que la sonication (17 cycles) les dégradait.
2. Malgré une bonne immunoprécipitation en tandem pour la protéine Hdac2, nous n'avons pas obtenu de séquence après le séquençage à haut débit. Nous avons postulé que la protéine ne se liant pas directement à l'ADN, il est nécessaire d'augmenter la fixation en augmentant la concentration en formaldéhyde ou en changeant d'agent pontant. Nous avons essayé d'augmenter la concentration en formaldéhyde à 1% ou 2,5% comme cela avait été précédemment décrit (Thompson et al., 2008) mais la fragmentation de la chromatine s'est révélée laborieuse et la visualisation des protéines par western blot, impossible.
3. Les immunoprécipitations FLAG ont un rendement de 1 à 10% maximum contrairement aux immunoprécipitations HA qui ont un rendement entre 10 à 30% pour la plupart des protéines. Nous avons postulé qu'un seul épitope FLAG n'est pas suffisant pour immunoprécipiter les protéines, particulièrement celles de haut poids moléculaire. C'est pourquoi, nous avons changé les étiquettes des protéines Chd.
4. Les immunoprécipitations FLAG et HA ont un faible rendement qui rend impossible la production d'une double immunoprécipitation pour les protéines de haut poids moléculaire ou faiblement exprimées. Nous avons postulé que la haute concentration en détergents et en sels empêchait la fixation des anticorps sur leurs épitopes. Nous

avons donc opté pour un protocole plus doux qui contienne moins de détergents et de sels.

6. Protocole E (Natif)

Principe

Ce protocole a été précédemment décrit par Umlauf et al. (Umlauf et al., 2004). Il ne comprend pas l'étape de fixation commune à tous les protocoles décrits précédemment. En effet, ce protocole sans agent pontant permet de purifier les noyaux des cellules et de fragmenter l'ADN avec l'aide de la DNase de microcoque. La première étape consiste à collecter les cellules en les trypsinant. Les cellules sont ensuite reprises dans une solution hypotonique qui les fait gonfler. Puis l'ajout de faibles concentrations de NP40 (0,15 à 0,3%) permet la lyse des membranes plasmiques tout en préservant les noyaux. Les noyaux sont ensuite séparés des autres organelles par une purification sur coussin de sucrose. Remis en suspension dans le tampon adéquat, les noyaux sont ensuite digérés par la nucléase de microcoque (MNase). La digestion est stoppée par 1,2mM EDTA final. Le surnageant S1 obtenu à l'issue de cette étape contient les nucléosomes qui ont été relargués très rapidement lors de la digestion. Le culot est resuspendu dans le tampon N-ChIP et les fragments de chromatine libérés (mono et oligonucléosomes) sont collectés dans le surnageant S2.

Bilan

Ce protocole permet de fragmenter l'ADN de façon très fine puisque la majorité des fragments obtenus se situent ~150 paires de base (en fonction de la dose de MNase). Cette taille est idéale avec le séquençage à haut débit. Lors de la préparation de la librairie d'ADN à séquencer, il suffit de découper la bande d'intérêt à 150 paires de bases.

Ce protocole a été testé avec les protéines Chd4 et Chaf1a. Alors que Chd4 était majoritairement présent dans la fraction S1, Chaf1a l'était plus dans la fraction S2. Les résultats ont montrés qu'il était possible d'immunoprécipiter avec une grande efficacité ces deux protéines dans ces conditions plus douces en détergents et en sels.

Les conditions décrites dans ce protocole sont idéales pour immunoprécipiter des protéines seules. Cependant, notre objectif est de séquencer l'ADN associé à nos facteurs. Or, en absence de fixation, les protéines étiquetées peuvent se dissocier de leurs séquences cibles au cours du protocole expérimental. Cette technique d'immunoprécipitation de chromatine fonctionne pour les histones qui interagissent très fortement avec l'ADN.

Nous avons donc élaboré un nouveau protocole qui combine les avantages du protocole E qui comporte des tampons faibles en détergents et en sels tout en permettant une fragmentation de l'ADN compatible avec le séquençage à haut débit, mais nous avons ajouté un pontage fort permettant d'immunoprécipiter l'ADN lié aux protéines d'intérêt.

B. Développement d'un protocole d'immunoprecipitation de chromatine innovant (Protocole F)

La mise au point d'un nouveau protocole d'immunoprécipitation de chromatine s'est révélée indispensable pour palier aux déficiences des protocoles testés.

1. Choix du pontage

Nous avons choisi d'augmenter la concentration en agent pontant en effectuant un pontage avec 1% de formaldéhyde comme cela avait montré dans la littérature pour Chd8 (Ishihara et al., 2006). Certains auteurs ont montré qu'il était possible de réaliser des immunoprécipitations de chromatine avec 2,5% de formaldéhyde (Thompson et al., 2008).

Un certain nombre d'articles ont été récemment publiés décrivant l'efficacité de plusieurs nouveaux agents pontants utilisés dans des expériences d'immunoprécipitation de chromatine pour la détection de gènes cibles de certains facteurs de transcription (Kurdistani et al., 2002; Noma et al., 2006; Nowak et al., 2005; Wang et al., 2009b; Zeng et al., 2006). Ces réactifs font partie de la famille des agents pontants homobifonctionnels possédant une fonction NHS-ester ou imidoester. Ils possèdent des groupes réactifs et réagissent sélectivement avec des amines primaires. Ils sont fréquemment utilisés pour stabiliser les interactions protéine-protéine pour une analyse plus approfondie par immunoprécipitation. Afin de renforcer le pontage des protéines qui ne se fixent pas directement à l'ADN comme Hdac2, nous avons utilisé en plus du formaldéhyde, un agent pontant supplémentaire de cette famille: le DSG (Disuccinimidyl glutarate). Cet agent a déjà décrit dans la littérature et sa caractéristique réside dans sa taille de 7,7Å plus importante que le formaldéhyde qui mesure 2 Å. Cette dernière caractéristique lui permet de lier deux résidus amines plus éloignés que le formaldéhyde. Il a été décrit comme le plus réactif de sa famille.

Nous avons testé de même le DTBP (11,9 Å) avec le protocole D, mais nous n'avons trouvé aucun progrès significatif. C'est pourquoi il n'a pas été retenu dans nos tests suivants.

Un troisième agent pontant, l'EGS (16,1 Å) a été testé avec ce protocole. Sa taille est supérieure et contrairement au DSG le pontage effectué est réversible en utilisant 1M

Hydroxylamine-HCl pH8,5 pendant 3 heures à 37°C. Cependant nous avons constaté que l'efficacité de ce pontage était plus faible que le DSG. Nous avons conservé ce dernier pour nos expériences de ChIP.

Les trois formules chimiques de ces agents sont regroupées dans la [Figure 34](#).

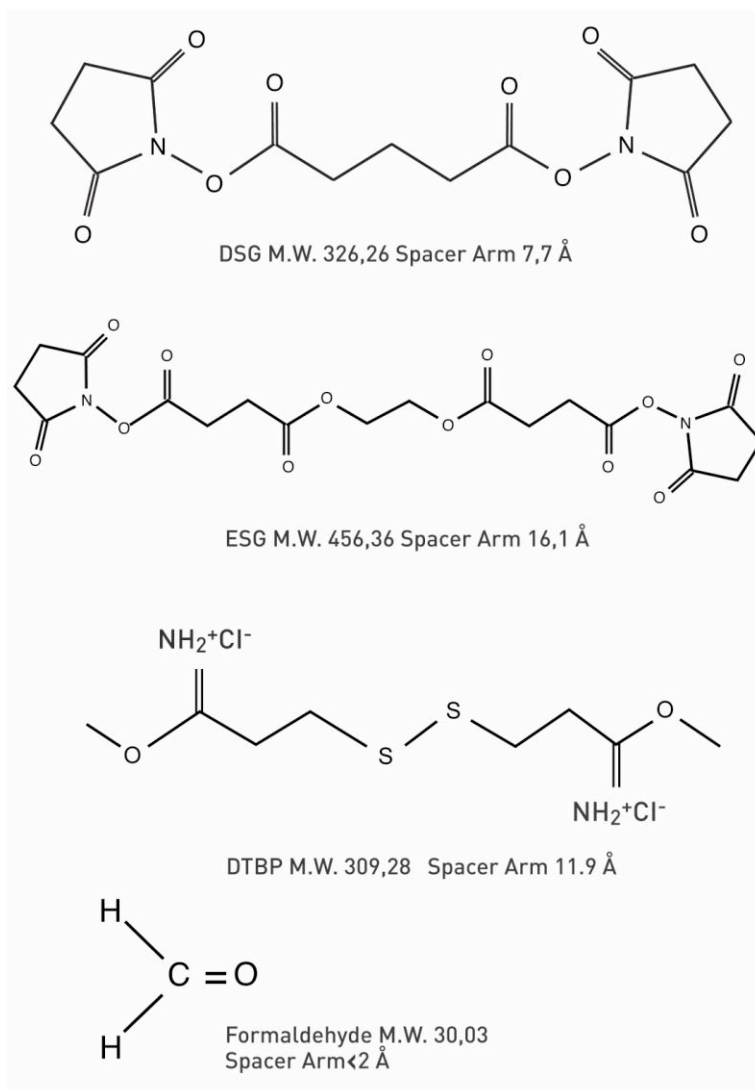


Figure 34: Formule chimique des quatre agents pontants utilisés: DTBP, EGS, DSG et formaldéhyde

2. Choix des détergents et des sels

Les résultats du protocole E montrent une très bonne efficacité d'immunoprécipitation tant avec l'anticorps anti-HA et anti-FLAG. Nous avons donc conservé les conditions de préparation de chromatine et d'immunoprécipitation. Ces conditions de basse concentration en sels et en détergents semblent favoriser l'immunoprécipitation et mieux préserver les

protéines de haut poids moléculaire. D'autres groupes ont amélioré leurs conditions d'immunoprécipitation grâce à la réduction des détergents et des sels est une condition de plus en plus décrite dans la littérature (van den Berg et al., 2010) par exemple.

3. Choix de la méthode de fractionnement de l'ADN

Nos essais montrent que la protéine Chd4 est dégradée par des sonications répétées. Pour préserver les protéines de la sonication, nous avons choisi de fragmenter la chromatine par une digestion par la MNase. Cette enzyme coupe principalement entre les nucléosomes. De plus, l'expérience de notre équipe montre que la MNase est peu sensible aux états de condensation de la chromatine (elle coupe efficacement l'hétérochromatine constitutive). Pour pallier aux potentielles difficultés de sonication à cause de la fixation de la chromatine par les agents pontants, nous avons ajouté 4 cycles de sonication 20"ON/40"OFF.

4. Choix de la méthode d'immunoprécipitation

Les immunoprécipitations sont réalisées sur la nuit afin d'augmenter le temps de contact entre l'anticorps et l'épitope complémentaire.

Le tampon d'immunoprécipitation

Le tampon d'immunoprecipitation est le tampon MNase ou Buffer KN, dans lequel est préparée la chromatine pour la première immunoprécipitation. Ce tampon contient peu de sels et aucun détergent. Le tampon de la deuxième immunoprécipitation est aussi le tampon de lavage qui contient peu de sels et 0,01% NP40 soit moins du dixième de la concentration en détergents du protocole D. Ces deux tampons confèrent donc un environnement peu dénaturant et donc susceptible de dissocier les complexes macromoléculaires. Le manque de spécificité que l'on pourrait supposer est largement compensé par la double purification HA-FLAG de nos protéines.

Les lavages

Les lavages des billes agarose sont faits dans le tampon TEGN qui est un tampon doux en détergent 0,01%NP40, et en sels 150mM NaCl et 3mM MgCl₂ habituellement utilisé pour le lavage des immunoprécipitations natives. De plus il contient 10% de glycérol ce qui préserve les protéines.

Les éluions

Trois types d'éluions sont possibles.

1. Elution peptide: Le peptide correspondant à l'anticorps est incubé à une concentration finale de 1mg/ml avec les billes liées à la chromatine immunoprécipitée dans le TEGN Buffer. Nous avons pu mettre en évidence que l'éluion peptide FLAG est généralement beaucoup plus efficace que l'éluion peptide HA. De plus, l'éluion peptide correspondant à la première immunoprécipitation est toujours plus difficile que la deuxième éluion peptide. En effet, cette dernière a toujours une efficacité proche de 100%.
2. Elution SDS: l'éluion faite à l'aide du tampon d'éluion 1% SDS. Ce tampon permet de décrocher tout ce qui est lié aux billes. Nous utilisons généralement cette éluion après la deuxième éluion peptide pour éluer ce qui reste sur les billes après le tandem.
3. Elution TEV: La mise au point de ce protocole correspond à la création des premières lignées portant l'étiquette 3FTH. La digestion TEV a été testée dans le tampon TEGN et cette éluion suivant l'IP HA s'est révélée être la plus performante. Cependant nous n'avons pas pu éluer la protéine Chd2 après une IP HA. Les limites de cette éluion sont donc des potentielles mutations sur le site de clivage ou bien un repliement de la protéine qui n'est pas favorable à l'accessibilité de la protéase TEV.

Contrairement à ce qui a été montré dans la littérature, nous n'avons pas réussi à éluer les IP FLAG avec 0,1M Glycine-HCl pH 3,5 (Futatsumori-Sugai et al., 2009).

La préparation de l'ADN

La fraction immunoprécipitée avec les anticorps FLAG et HA est séparée en trois: 1/10 pour l'analyse par la technique du western blot, 8/10 pour reverser le pontage et préparer l'ADN pour le séquençage et 1/10 en stock au -80% pour redéposer sur gel d'acrylamide si besoin auquel on ajoute 1/3 de glycérol pour la conservation. L'ADN est ensuite préparé selon les méthodes habituelles.

C. Publication

La technique d'immunoprécipitation, mise au point au laboratoire, fera l'objet d'une publication.

Brièvement, ce protocole consiste en huit étapes.

Les cellules sont fixées à l'aide de formaldéhyde seul ou d'un tandem DSG formaldéhyde. Elles sont grattées et lavées dans le Buffer I (étape 1).

L'étape suivante consiste à les perméabiliser grâce au Buffer II avec une concentration finale de 0.3% NP40 final (étape 2).

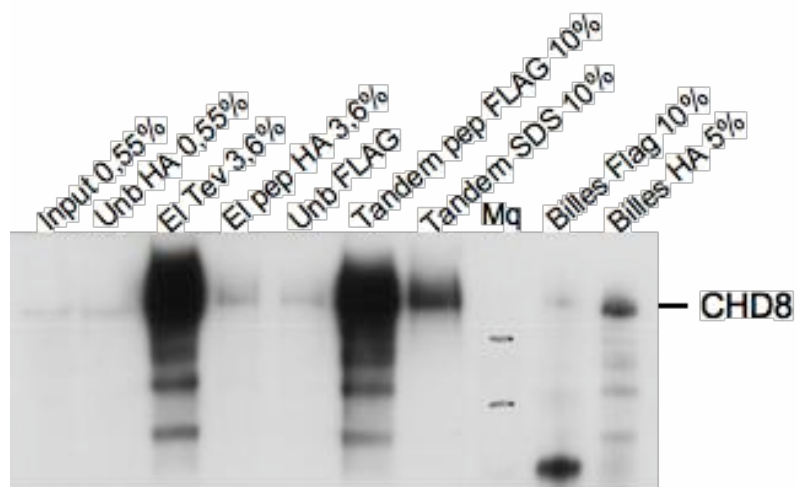
L'ADN est ensuite fragmenté grâce à la DNase de microcoque qui est suivie de 4 cycles de sonication (étape 3).

La chromatine ainsi préparée est utilisée dans des expériences d'immunoprécipitations (étape 4 à 7).

Deux méthodes d'immunoprécipitations sont possibles en fonction de l'étiquette. La première méthode consiste en une IP HA éluée par la Tev puis une deuxième IP FLAG qui est éluée avec le peptide FLAG. La deuxième méthode comprend une IP FLAG qui est éluée par le peptide correspondant et une deuxième IP HA éluée de même. L'éluion peptide FLAG est la plus limitante car 10 à 50% du matériel n'est pas élué par le peptide compétiteur et reste fixé sur les billes FLAG. La première méthode est incontestablement la plus performante mais elle requiert au moins un site HA, un site FLAG et un site de clivage TEV dans l'étiquette. La deuxième méthode, moins efficace, permet d'obtenir du matériel issu de la double immunoprécipitation. Cette méthode est utilisée lorsque les protéines comportent l'étiquette FH ou 6HFH. La dernière étape consiste à préparer l'ADN issu des expériences d'immunoprécipitation (étape 8). Une partie de l'immunoprécipitat est analysé par western blot avec l'ensemble des contrôles prélevés à chaque étape ([Figure 35](#)).

Les analyses par western blot pour les autres protéines Chd sont disponibles en annexe.

A



B

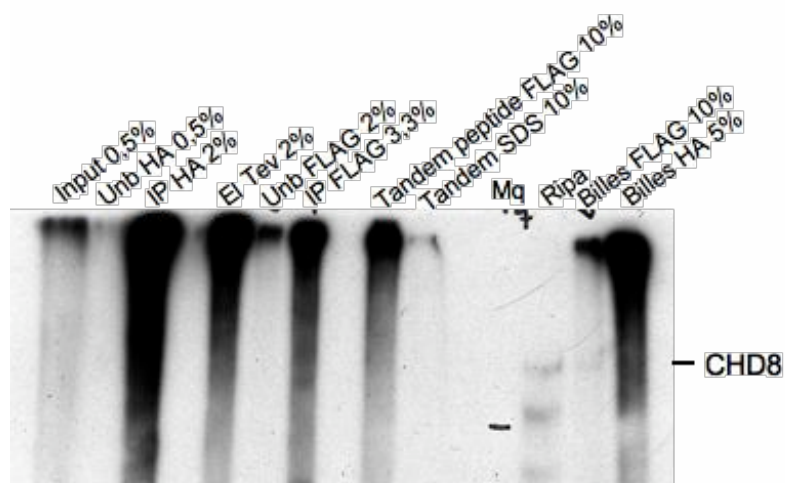


Figure 35: Immunoprécipitation de chromatine de la protéine Chd8 étiquetée 3FTH. A: pontage formaldéhyde seul; B: pontage DSG+formaldéhyde. Unb: fraction non liée avec une IP; El: elution; pep: peptide; Tandem: fraction immunoprécipitée par les anticorps anti-HA et anti-FLAG

Publication technique

Tandem affinity purification of tagged chromatin remodelling factors for ChIPseq applications at single nucleosome resolution

Arnaud Depaux^{1*}, Maud de Dieuleveult^{1*}, Isabelle Hmitou^{1*}, Sylvie Jounier¹, H el ene HumbertClaude¹ et Matthieu G erard¹.

¹ iBiTec-S, epigenetic regulation and cancer team, CEA Saclay, 91191 Gif-sur-Yvette, FRANCE

* Equal contribution

Abstract

Here we describe a chromatin immunoprecipitation protocol that allows the tandem affinity purification of mononucleosomes bound by chromatin remodelling factors, for ChIP-seq applications. This protocol requires the prior establishment of a cell line that expresses a tagged chromatin remodelling factor. We successfully used tandem affinity tags composed of FLAG and HA epitopes.

Proteins are first cross-linked onto DNA with either formaldehyde or a combination of disuccinimidyl glutarate (DSG) and formaldehyde. In this protocol, we describe how to prepare mononucleosomes from fixed cells using low detergent and low salt conditions, to prevent protein denaturation and macromolecular complexes dissociation. Cells are permeabilized by a detergent, and then chromatin is fragmented to mononucleosomes by *in situ* digestion with micrococcal nuclease. Mononucleosomes are then collected by breaking cell and nuclei membranes with a moderate sonication dose. FLAG-affinity purifications are performed using low salt and detergent conditions, including during the washing steps. Elution is performed by either TEV protease epitope cleavage, or by peptide competition. The eluted chromatin is next HA-affinity purified using similar conditions. This tandem affinity purification of chromatin yields highly purified mononucleosomes, from which 150 bp DNA fragments are prepared for high throughput sequencing. This tandem affinity ChIP-seq protocol allows the detection of genome-bound chromatin remodelling factors at a single nucleosome resolution. After tagged cell lines are obtained, the total length of the procedure is 3-4 d.

INTRODUCTION

Development of the protocol

Application

Comparison with other protocols

Experimental design

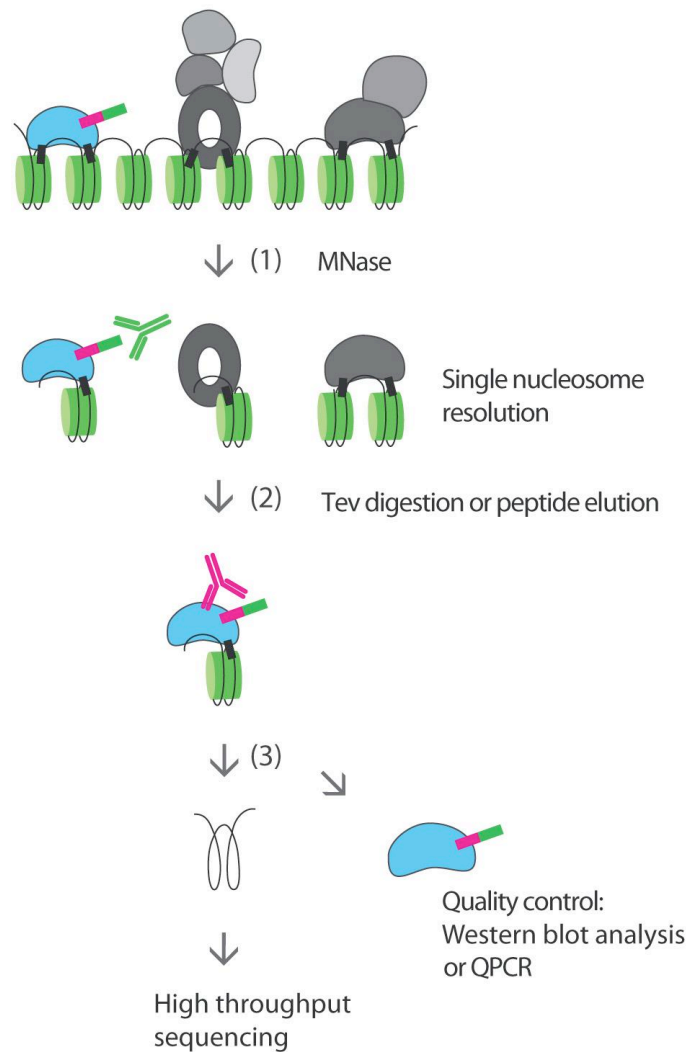


Figure 1: Overview of ChIP experiment for high throughput sequencing. (1) Chromatin preparation and first IP (2) specific elution: Tev digestion or peptide elution (3) Second specific elution followed by quality control (Western blot or QPCR) and high throughput sequencing

MATERIALS

REAGENTS

Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen, cat. no. 41965062).

Fetal bovine serum (Invitrogen, cat. no. 10270-106 500ml).

0.25% trypsin solution (Invitrogen, cat. no. 25200072)

200mM Glutamine (Invitrogen, cat. no. 25030024)

100x non essential amino acids (Invitrogen, cat. no. 11140035).

Leukemia Inhibitor factor (LIF, Euromedex cat. no. L2024-01B).

Penicillin/streptomycine (Invitrogen, cat. no. 15070063).

β -mercaptoethanol (Sigma, cat. no. M6250-250ml) Caution: very hazardous in case of skin contact, ingestion, eye contact or inhalation. Caution should be taken to avoid exposure.

Gelatin from porcine skin, type A (Sigma, cat. no. G1890-100G).

Counting chamber Malassez, Neubauer.

Formaldehyde solution (Sigma, cat. no. F8775)

DSG (Disuccinimidyl glutarate) (Pierce, cat. no. 20593)

DMSO (Sigma, cat. no. D2650)

Micrococcal nuclease 200U/ul, (NEB, cat. no. M0247S)

Complete (Roche, cat. no. 11836145001)

PMSF (Sigma, cat. no. P7626-5G)

M2 anti-FLAG antibody agarose beads (Sigma Aldrich, cat. no. A2220)

Anti-HA agarose beads (Sigma Aldrich, cat. no. A2095)

Pierce centrifuge columns, 0.8ml (Pierce, cat. no. 89868)

Peptide HA 4ug/ul

Peptide FLAG 4ug/ul

Tev protease

RNAse A (Sigma, cat. no. R 4642-50MG)

Phenol Chloroform (Fluka, cat. no. 77617)

Proteinase K (Roche, cat. no. 3115852001)

NuPAGE® LSD (Invitrogen, cat. no.

NuPAGE® sample reducing agent (Invitrogen, cat. no. NP0004

XCell SureLock® Mini-Cell and XCell II™ Blot module Kit CE Mark (Invitrogen, cat. no.)

PVDF membrane (Millipore, cat. no.)

Whatman paper (D Dutcher, cat. no. 036347)

Spectra™ Multicolor Broad range protein ladder (Fermentas cat. no. SM1841)

Acryl/bisacrylamide (Sigma Aldrich, cat. no. A3699)

TEMED (Sigma, cat. no. T9281)

Sodium dodecyl Sulfate (SDS) 10% (,cat. no.

Ammonium persulfate (APS) 10% (Sigma, cat. no.).

NuPAGE® Novex 4-12% Bis-Tris Gel 1.5mm, 10 wells (Invitrogen, cat. no. NP0335BOX)

NuPAGE® MOPS SDS Running Buffer (for bis-Tris gel only) 20X (Invitrogen, cat. no. NP0001)

NuPAGE® Transfert Buffer 20X (Invitrogen, cat. no. NP0006-1)

NuPAGE® Novex 3-8% Tris-acetate Gel 1.5mm, 10 wells (Invitrogen, cat. no. EA0378BOX)

NuPAGE® Tris-acetate SDS Running buffer 20X (Invitrogen, cat. no. LA0041)

HA.11 monoclonal antibody (Covance, cat. no. MMS-101R-1000)

anti-FLAG (Sigma, cat. no.

Secondary antibody: anti-mouse IgG (H+L), HRP conjugate (Promega, cat. no. W4021)

Films (GE healthcare, cat. no. 28906837)

ECL Plus Western Blotting detection Reagents for 3000cm² membrane (Amersham, cat. no. RPN2133)

EQUIPMENT

Dry heater at 100°C

Scapers (D. Dutcher, cat. no. 91822)

Centrifuge, refrigerate and non-refrigerate

Electrophoresis apparatus for agarose gels

Electrophoresis apparatus for NuPAGE® acrylamide gels (Invitrogen, cat. no. EI0002)

Intellimixer (D.Dutcher)

Wheel at RT and 4°C

Incubators for cells lines at 37°C

Chemical hood

Thermomixer (Invitrogen)

Nupage precast gels 1.5mm (Invitrogen, cat. no. NC2015)

Combs 1.5mm, 10 wells or 15wells (Invitrogen, cat. no. and NC3515 respectively)

REAGENT SETUP

Gelatin 1%: 1% gelatin in water, sterile.

PMSF 100X: 100mM PMSF in ethanol. fresh or prepared less than 15 days

Buffer I: 0.3M Sucrose, 60mM KCl, 15mM NaCl, 5mM MgCl₂, 0.1mM EGTA, 15mM Tris-HCl pH 7.5.

Buffer II: 0.3M Sucrose, 60mM KCl, 15mM NaCl, 5mM MgCl₂, 0.1mM EGTA, 15mM Tris-HCl pH 7.5, 0.6% NP40

Buffer KN: 20mM Tris-HCl pH 7.5, 15mM KCl, 60mM NaCl, 0.34M Sucrose, 1mM CaCl₂

Buffer TKNSE: 20mM Tris-HCl pH 7.5, 15mM KCl, 60mM NaCl, 0.34M Sucrose, 4mM EDTA

TEGN: 20mM Tris-HCl pH 7.5, 15mM KCl, 150mM NaCl, 3mM MgCl₂, 0.1mM EDTA, 10% Glycerol, 0.01% NP40

PBS 1X cold

PBS 1X cold with protease inhibitors (Complete, PMSF)

PBS 1X at RT

EDTA 100mM pH8.0

Elution buffer: 1% SDS, 0.1M NaHCO₃

Laemmli buffer (3X)

Tris1M, pH8.8 and Tris1M, pH6.8.

Blocking buffer: PBS 1X, 0.1% Tween, 5% dry skim milk

Wash Buffer: PBS 1X, 0.1% Tween

D10 medium: DMEM with 10% fetal bovine serum.

D15 medium: 15% fetal bovine serum, 50 U/ml of penicillin/streptomycin, 0.1mM non essential amino acids, 0.1mM β-mercaptoethanol, 107 U/ml LIF (Leukemia Inhibitor factor), complete with DMEM.

DSG 250X : 50mg in 306ul of DMSO, final concentration : 163.4ug/ul

PBS 1x : 137mM NaCl, 2.7mM KCl, 4.3mM Na₂HPO₄ , 1.47mM KH₂PO₄, pH 7.4.

FLAG peptide: concentrate at 4ug/ul

HA peptide: concentrate at 4ug/ul

Classical running Buffer: 25mM Tris base, 192mM Glycine, 0.1% SDS.

Classical transfer buffer: 4.8mM Tris base, 38mM Glycine, 0.04%SDS, 20% methanol.

PROCEDURE

ES cell culture **TIMING** 1 week

1| Grow ES cells in D15 in 12 10cm-dishes with feeders cells. The day before the day of chromatin preparation, split ES cells in 25 gelatin coated dishes without feeders.

?TROUBLESHOOTING

Chromatin preparation **TIMING** 1day

2| Change the ES cell medium with fresh D15 medium and incubate during at least 2 hours. During this time, count the cells from one dish. We expect 15 to 25 millions of cells per 10cm-dish.

CRITICAL STEP This step is important for maintaining ES cells undifferentiated.

3| For ES cell crosslink, 2 fixative agents can be used. Formaldehyde crosslink is commonly used and well adapted to DNA binding proteins (most remodeling factors). Crosslink with DSG and formaldehyde is much more efficient for proteins that do not directly bind to DNA.

A. DSG crosslink

- i. Wash ES cells with PBS1X at RT. Prepare DSG 1X (653.6ng/ul) in PBSX at RT. Use 4ml per 10cm dish. Shake gently. Incubate for efficient crosslink during 45min at RT. Shake gently 4-5 times.
- ii. Add formaldehyde to a final concentration of 1%. Incubate 10min at RT. Go to step 4

B. Formaldehyde crosslink

Add formaldehyde to 1% (final concentration) directly in ESC medium. Homogenize by shaking gently. Incubate 10min at RT.

CAUTION: formaldehyde must be handled under a chemical hood

4| Wash twice each dish with 20ml of cold PBS1X. Remove the PBS and add 2ml of PBS1X with protease inhibitors. Scrap the cells and collect them in 6*15ml tube (pool 4 dishes/tube). From here, all steps must be performed on ice.

5| Centrifuge at 1000rpm, 5 min, 4°C. Remove supernatant. Add 1ml of cold PBS1X in each tube and resuspend the pellet with a P1000. Complete up to 10ml with PBS 1X. Repeat this step once. Wash once with 3ml of Buffer 1. Centrifuge at 1000rpm, 5 min, 4°C.

6| Resuspend the cell pellet in 2ml of Buffer 1 with protease inhibitors. Add 2ml of Buffer 2. Mix by inverting. Leave 10min on ice, to allow permeabilization of the ESC membranes. Centrifuge 10min at 2000g, 4°C.

7| DNA is fragmented using MNase digestion and few sonication to obtain fragments ranging between 300-150bp. This fragmentation procedures allows the preservation of protein complexes.

- i. Resuspend the cell pellet in 500ul of TN Buffer. Mix well with 20 up and down with a P1000 pipetman.

CRITICAL STEP Check that no clusters of cells are left in solution as this will lower the quality of MNase digestion. Cells are sticky, beware of lossing material on pipet tips.

- ii. Transfer the content of 15ml tube into a 1.5ml eppendorf tube. 80units of MNase are required to digest the genome of 10 millions ES cells. Dilute the required amount of MNase to 10U/ul final concentration in TN Buffer. The amount of MNase required for digestion of fixed chromatin is higher than for chromatin isolated in native conditions.
- iii. Add the appropriate amount of diluted MNase in each eppendorf tube containing the permeabilized cells. Mix by gentle pipeting.
- iv. Incubate 10min at 37°C in a water bath. Mix by inverting once during the incubation.
- v. Stop MNase digestion by transferring rapidly in ice. Add 20ul of EDTA 100mM pH8, 10ul of complete 50X, 5ul of PMSF 100X.

PAUSE POINT tubes can be left in ice here for 1-2 hours

8| Cool the Diagenode bioruptor with ice. Perform 4 cycles of sonication: 20'' ON /40''OFF. This step will break ESC membrane and release nucleosomes in solution.

PAUSE POINT tubes can be left in ice here for 1-2 hours

CRITICAL STEP We prefer not to freeze chromatin, and use it directly for immunoprecipitation.

CRITICAL STEP Two aliquot of the chromatin should be kept to check DNA concentration, fragments size (20ul) and as input fraction for WB analysis after tandem purification (150ul).

Chromatin immunoprecipitation **TIMING** 4-5 days

Keep an aliquot at each step of the purification. See supplemental information

9| Mix well the beads before pipeting. Pipet the required amount (80 μ l of dry beads (160 μ l of slurry) for the first IP, 40 μ l (80 μ l of slurry) for the second IP) in a 15ml tube. Wash beads twice with 10 ml of TEGN buffer, centrifuge it 1000rpm, 1 min. Perform an additional wash with 10ml of TKNSE buffer for the 1st IP beads. Washed anti-HA and anti-FLAG agarose beads can be kept at 4°C for 1-2 weeks.

10| Incubate the chromatin overnight with 80ul of beads on a rotator at 4°C. We used to split the chromatin and beads mixture in 4*1,5 ml microcentrifuge tubes. Follow the steps indicated in Table 1.

TAG	1 st IP	1 st elution	2 nd IP	2 nd elution
6*His-FLAG-HA	FLAG IP	FLAG peptide elution during day + peptide FLAG elution overnight	HA IP	HA peptide elution
FLAG-HA	FLAG IP	FLAG peptide elution during day + one peptide FLAG elution overnight	HA IP	HA peptide elution
3*FLAG-Tev-HA	HA IP	Tev digestion + <i>facultative step</i> : HA peptide elution	FLAG IP	FLAG peptide elution

Tableau 1: Three different tags described in column 1 have been used for these experiments. 2nd column indicate the antibody used in the first IP. 3rd column show the strategy used for first elution. The fourth column indicates the antibody used for the second IP

11| After overnight incubation, pool the content of the microcentrifuge tubes in a 15ml tube. Wash 10 times with 10ml of TEGN, centrifuge at 1000 rpm, 4°C, for 1 min between each wash.

12| Perform first elution.

A. If the tag contains a Tev protease cleavage site, we perform the first elution with the Tev protease. After the washes, transfer beads into a 1,5 ml tube and add 420 µl of TEGN buffer. Add 30ug of Tev protease and DTT to 0,5 mM final concentration. Allow the Tev digestion to occur for 6 hours at 30°C onto a rotator (850 rpm).

?TROUBLESHOOTING

B. If A is not possible, perform an elution with a competitor peptide

i. Transfer the beads into a 0.5ml microcentrifuge tube. Add competitor peptide to 1 ug/ul (final concentration) and complete with TEGN Buffer up to 400 µl (final volume). Shake onto a rotator/mixer 2h at 4°C followed by 4h at RT.

?TROUBLESHOOTING

ii. Remove supernatant and perform a supplementary peptide elution overnight using the same condition described above.

13| Load the beads and chromatin eluate on a Pierce centrifuge column. Centrifuge 1min at 2000rpm. Discard beads and keep the liquid. Add beads for the second IP to the first eluate and incubate it overnight onto a rotator at 4°C.

CRITICAL STEP Collect an aliquot of the first eluate for quality control (Table S3).

14| Wash beads as described in point 11|. After the final wash, transfer the beads to a 0.5ml microcentrifuge tube. Add competitor peptide to 1ug/ul and complete with TEGN buffer up to 400ul final volume. Shake onto a rotator/mixer 2h at 4°C followed by 4h at RT.

15| Load the beads and eluate on a Pierce centrifuge column. Centrifuge 1min at 2000rpm, at RT. Collect the liquid into a 1,5 ml microcentrifuge tube. Keep the purified fraction on ice (Peptide fraction of tandem purification) and add 400 µl Elution buffer to the beads. Mix well on a rotator, 15min. Reload on the same column. Centrifuge and keep the fraction eluted (SDS fraction of tandem purification).

Steps 16-21 can be carried out in parallel with steps 22-23.

Purification of DNA for high throughput sequencing (input and tandem purified samples)

16| DNA purification is described in step 16 to 19. Fill the sample volume up to 500ul with H₂O. Add 20ul of NaCl 5M to each sample and 15ug of RNase A (only for input sample). Reverse crosslink sample overnight at 65°C.

17| Add 20ul of Tris 1M pH8.0, 10ul of EDTA 0,5 M and 4ul of proteinase K. Incubate 2 hours at 50°C.

18| Add 500ul of Phenol Chloroform, vortex well 30sec.

19| Centrifuge and keep the aqueous phase. Add 50ul of NaAc 3M pH 5.2, 1ml of EtOH 100% and 30ug of glycogen. Leave 1 hour at -20°C for precipitation. Centrifuge 15min at 4°C, 13000rpm. Wash with 70% EtOH. Let the pellet dry 10 min. Resuspend in 20ul of H₂O.

PAUSE POINT DNA Samples can be stored at -20°C.

?TROUBLESHOOTING

20| We quantify DNA with the Quant-iT™ Picogreen® dsDNA Reagent to quantify DNA before sequencing using the recommended protocol given by the supplier. This method can detect dsDNA concentration as low as 10 ng/ml. Examples of that we obtained for our samples are given in table 2. Input samples can be quantified with a Nanodrop.

21| Check input DNA quality by running 1ul and 3ul (out of 20ul) onto a 1.5% agarose gel. Mononucleosomes should migrate at 150. (Figure 2)

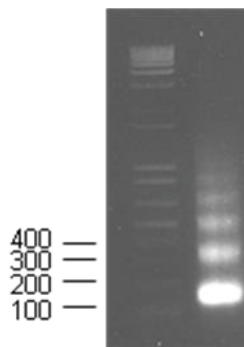


Figure 2 A representative gel for MNase digestion loaded with two different quantities (~600ng). A majority of mononucleosomes (140 bp) is observed.

22| Check the quality of tagged protein immunoprecipitation by Western blot. This control is essential when no DNA sequence target are known for the protein of interest. We used three types of acrylamide gels, depending on the goal of the analysis. The objective of this control is to visualize by Western blot the enrichment of the tagged protein after each immunoprecipitation step.

A. Run acrylamide gel

- i. Perform Western blot analysis with standard procedure described in the literature with the Classical running Buffer.
- ii. For both histone and protein of interest visualization, use NuPAGE® Novex 4-12% Bis-Tris Gel and recommended procedure provided by Invitrogen.

- iii. For DSG conditions use NuPAGE® Novex 3-8% Tris-acetate Gel (high weight complexes).

?TROUBLESHOOTING

B. Transfer acrylamide gels. We use preferentially PVDF membranes.

CRITICAL STEP Transfer one gel per cassette.

- i. For standard acrylamide gels, use Classical transfer Buffer and run the transfer in Invitrogen Xblot cells for 1hour and a half.
- ii. For NuPAGE® Novex gels, use recommended procedures given by Invitrogen.

?TROUBLESHOOTING

23| Reveal western blot. Antibodies concentrations are given in table 3.

	Antibody	dilution
First antibody	Anti-HA Covance	1:1000
	Anti-FLAG	1:2000
	Anti-H3 Millipore	1:8000
Second antibody	Anti-mouse HRP	1:20000
	anti-rabbit HRP	1:20000

Tableau 2 Concentrations of different antibodies used in column 2 are given in column 3. Concentrations are volume to volume.

?TROUBLESHOOTING

Typical result could be seen in Fig 3. Strong signal should be seen in the Tev or peptide elution fractions, tandem peptide eluted fraction and tandem SDS eluted fraction (Figure 3).

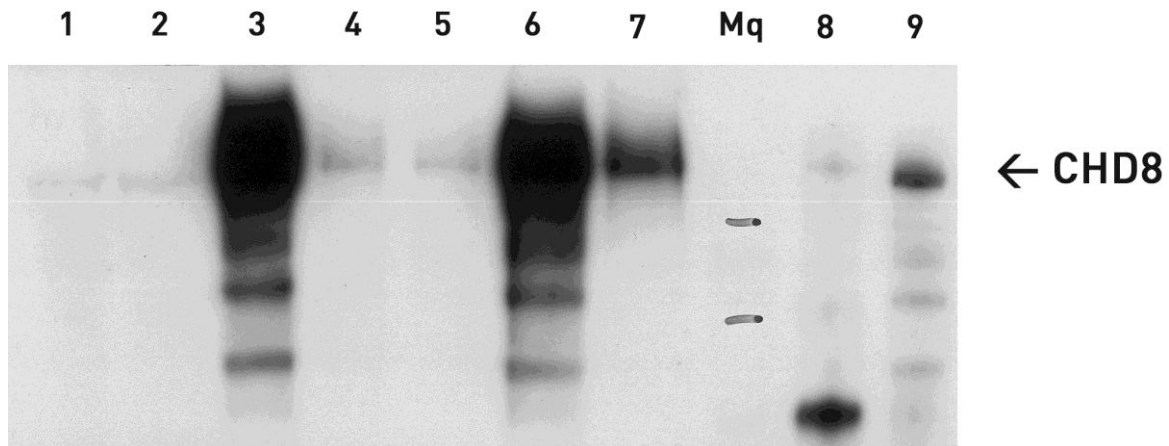


Figure 3: Analysis of purified fraction for Chd8 protein. Lane 1: 0,5% of chromatin input, 2: 0,5% of unbound HA IP, 3: 4% of tev elution, 4: 4% of peptide HA elution, 5: 4% of unbound FLAG IP, 6: 10% of peptide FLAG elution, 7: 10% of SDS elution, Mq: protein ladder, 8: 10% of FLAG beads after peptide elution, 9: 5% of HA beads after peptide elution

TIMING

Step 1, ES cell culture, from beginning of culture up to 24 dishes: ~1 week

Steps 2-9, (day 1) Chromatin preparation: 1 day

Step 10, First chromatin immunoprecipitation: Overnight incubation

Step 11-12, (day2) Preparation of the first elution: 1 hour, incubation time ~6 hours

Step 13, Preparation of the second chromatin immunoprecipitation: 20min, incubation time: overnight

Step 14, (day 3) Preparation of the second elution: 1 hour, incubation time ~6 hours

Step 15, (day 3) SDS elution: 1 hour 30min

Steps 16-21, (day 4) DNA preparation for high throughput sequencing: 4 hours

Steps 22-23, (day 3- day 4) Western blot: 2 days

TROUBLESHOOTING

TROUBLESHOOTING advice can be found in Table 3.

Step	Problem	Possible reason	Solution
1	Insufficient number of cells	Cells are not confluent enough	Report the chromatin preparation to the next day.
11 B	Tev cleavage is not working or inefficient	Mutation, protein folding masking the epitope	Perform elution with a competitor peptide overnight following conditions described step 11 B.
11 B	Peptide elution is not working	Bad tag accessibility, competitor peptide quality	Peptide final concentration is critical In our hands, FLAG peptide is much more efficient (70-80% efficiency) than HA competitor peptide (50% efficiency) on first elution. Check the pH of the peptide solution, peptide quality. Try higher salt condition, higher peptide concentration or add low concentration of SDS (0,1%)
18	No DNA pellet	Insufficient amount of glycogen	Add more glycogen to see it more easily
20 A	Bad run, SDS PAGE	Too much protein concentration in gel lanes	Load less protein onto the gel. Be sure that you use the correct Running buffer Bis-tris gel or tris-acetate gel
20 A	Bad run, SDS PAGE	Hazy bands	Change the running buffer and leammli buffer. For non gradient gels and proteins of high molecular weight, gels can be run at low voltage overnight
20 B	Bad transfer	Transfer is inefficient	Stain you acrylamide gel with brilliant blue to check the level of protein which is the gel. For NuPAGE® Novex 4-12% Bis-Tris Gel , 1h transfert as recommended by Invitrogen is not sufficient in our hands. We have tested successfully 3h at 50V for one gel
21	Bad revelation	Problem with the first antibody	Try a higher concentration of the first antibody. Perform more washes of the membrane. We never had problems with the secondary antibody.

21	H3 signal is unclear	Concentration of H3 antibody is inappropriate	In our hands, the H3 antibody containing solution cannot be re-used like the FLAG and HA antibodies solutions (can be re-use 10 times at least).
21	Signal too weak	Protein of interest is not well expressed	Two factors can be improved: 1) Scale up the protocol by starting from 75 dishes 2) If a selection cassette is present in the cell line, remove it. It can reduce the expression of your protein of interest

Tableau 3 Troubleshooting table

ANTICIPATED RESULTS

Starting from 1-2 mg of chromatin at the beginning of the protocol, DNA concentration should be between 1-2 ng to 30 ng.

SUPPLEMENTARY INFORMATION

	Tag protein sequence
6*His-FLAG-HA	GAP HHHHH GAAGG DYKDDDDK SAAGGYPYDVPDYA
FLAG-HA	GAPGGAAGG DYKDDDDK SAAGGYPYDVPDYA
3*FLAG-Tev-HA	GAPGGAAGG DYKDDDDK GGAADYKDDDDK GGLTPAQLEAAARAG G DYKDDDDK GGENLYFQ SAAGGYPYDVPDYA

Table S1: Amino acid sequence of three tags that we used in our large-scale tagging project. FLAG sequence is indicated in red, HA in green, histidine tag in yellow and Tev cleavage site in blue.

	Sample	Aliquot (ul)	out of (ul)	Proportion
1	input	20	3600	0,55%
2	Unbound first IP	20	3600	0,55%
3	first IP control	10	300	3,30%
4	Tev elution	10	500	2%
5	first peptide elution	10	400	2,50%
6	second IP control	10	300	3,30%
7	Unbound second IP	10	500 800 900	
8	Tandem peptide	40	400	10%
9	Tandem SDS	40	400	10%
10	beads first IP	10	200	5%
11	beads second IP	10	100	10%

Table S3: Amount of each sample that are kept for quality control. Quantities from different fractions (**column “sample”**) are indicated in column **“aliquot”**. The proportion of the fraction/total is indicated in the last column.

Conclusion

Ce protocole de préparation de chromatine et d'immunoprécipitation nous a permis d'immunoprécipiter avec succès l'ensemble des protéines étudiées dans le laboratoire qui inclut en plus des protéines cités 7 autres remodeleurs de chromatine (I. Hmitou) ainsi que par collaborateurs qui étudiaient des protéines possédant une étiquette différente: FLAG et biotin. Ce succès montre que le protocole que nous avons développé est universel pour les étiquettes que nous avons choisies.

Cependant, plusieurs facteurs peuvent influencer l'efficacité, la qualité et la quantité d'immunoprécipitation. Premièrement, les protéines ciblées sont étiquetées au locus, le taux d'expression de chaque protéine étiquetée est donc identique à la moitié du taux d'expression de la protéine endogène. C'est pour cela que nous nous dressons face à deux problèmes pour les protéines les moins abondantes. D'une part elles sont moins concentrées lors de la préparation de chromatine et donc plus difficilement immunoprécipitables. D'autre part, elles sont moins facilement détectables par la technique de western blot.

Deuxièmement, nous avons constaté que l'étiquette choisie influençait l'efficacité d'immunoprécipitation. Ainsi, nous avons montré qu'une digestion par la protéase TEV était plus efficace qu'une élution peptide Flag (sauf dans le cas des protéines CHAF1A et Chd2). Dans l'ensemble, l'étiquette 3FTH est la plus performante.

Les échantillons immunoprécipités à l'aide de notre nouveau protocole ont été purifiés et séquencés au Génoscope (CNG) à Evry par nos collaborateurs, Ivo et Matha Gut. Les analyses de ces fragments d'ADN de 35 paires de bases ont été alignées sur le génome murin de référence (mm9) et nous avons procédé à l'analyse de ces résultats.

Chapitre 3. Analyse des données CHIP-seq

Introduction

L'ADN lié aux facteurs d'intérêt a été séquençé par la technique du séquençage à haut débit développée par Illumina. Le séquençage s'est fait au Centre National de Génotypage à Evry par nos collaborateurs Ivo et Martha Gut. Les fragments de 150 paires de bases isolés de nos expériences de CHIP en tandem ont été séquençés sur un Genome Analyser Illumina. Les séquences lues, appelées "reads", sont ensuite alignées sur le génome de référence de *Mus musculus*. Différents algorithmes sont utilisés pour détecter des zones d'enrichissement en reads. Le terme "pic" est utilisé pour désigner ces zones d'enrichissement. Grâce à une collaboration avec l'équipe de Laszlo Tora à l'IGBMC à Strasbourg, nous avons pu mettre au point une analyse bioinformatique de nos données. Ces analyses nous permettent de localiser ces pics par rapport à des caractéristiques génomiques annotées (gènes, promoteurs). Les outils développés permettent de comparer nos données entre elles, ce qui est pertinent dans le cas d'une comparaison globale de la famille Chd. Cette étude permet de comparer les sites de liaisons des protéines Chd et ainsi de proposer des hypothèses quant à la redondance ou l'exclusivité des Chds. Enfin, les outils permettent de comparer nos données CHIP-seq aux données déjà publiées (RNase Pol2, marques épigénétiques d'histones, facteurs du maintien de la pluripotence). Ces comparaisons nous permettent d'inclure les protéines Chd dans des réseaux de régulations déjà connus.

A. Le séquençage

La purification de l'ADN immunoprécipité a été décrite au Chapitre 2. L'ADN est ensuite déposé et migré sur gel d'agarose. Une bande se situant à 150pb est découpée et purifiée (article technique). Des adaptateurs sont fixés de part à d'autre des fragments d'ADN purifiés. Ces fragments constituent une banque d'ADN, qui est ensuite amplifiée par une étape de PCR, grâce à des amorces spécifiques des adaptateurs liés. La banque d'ADN amplifiée est ensuite séquençée sur une puce comportant 8 pistes. Chaque piste est

indépendante et permet de séquencer un échantillon. Nos échantillons ont pu bénéficier des améliorations récentes de la technique de séquençage, ainsi 35 paires de bases sont lues (au lieu de 25 pb auparavant) ce qui permet d'augmenter la spécificité des alignements sur le génome. De plus, l'amélioration du séquençage s'est accompagnée d'une augmentation de nombre de séquences lues (appelée «tag») par échantillon environ 17 millions de reads.

A l'issue du séquençage, les reads sont ensuite alignés sur le génome murin de référence: mm9. Lors d'un premier aperçu sur les données, nous avons pu immédiatement observer l'effet de la digestion MNase sur les données de séquençage. En effet, comme nous l'avons vu page 118, l'enzyme MNase digère préférentiellement l'ADN entre les nucléosomes. C'est pourquoi lorsque les nucléosomes sont extrêmement bien positionnés (comme cela a été montré précédemment au niveau des promoteurs transcrits), certaines séquences sont toujours digérées par la MNase et ne sont donc jamais obtenues au séquençage. Cet exemple est illustré sur la [Figure 36](#).

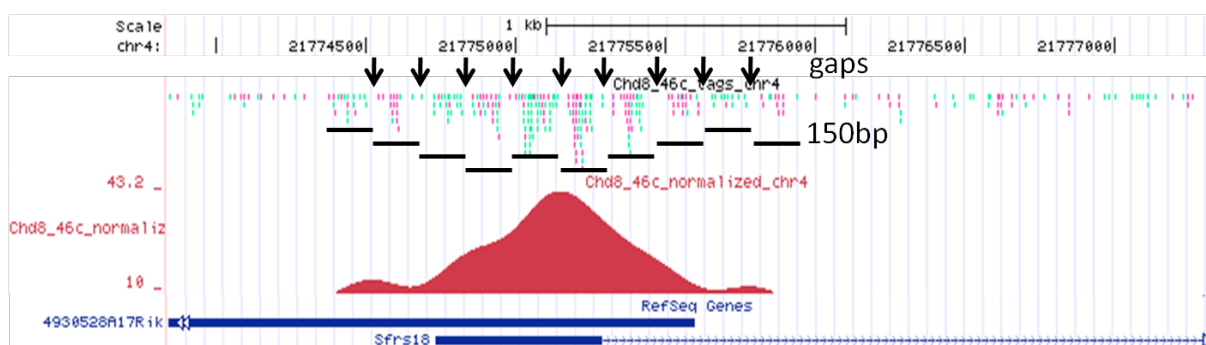


Figure 36: Effet de la digestion MNase sur une région liée par la protéine Chd8. Les lacunes induites par la digestion MNase sont indiquées par des flèches.

B. Le traitement informatique des données

1. Algorithmes de détection des pics

Nous avons sélectionné trois algorithmes différents pour détecter les pics d'enrichissement de nos données: Quest v2.4 (Valouev et al., 2008), SISSR (Jothi et al., 2008) et Sicer (Zang et al., 2009). Les paramètres de détection des pics sont propres à chaque algorithme et sont détaillés ci-dessous.

Il existe d'autres algorithmes disponibles dans la littérature que nous n'avons pas testés: CisGenome, MACS, PeakSeq, etc... (Laajala et al., 2009; Pepke et al., 2009).

Quest (Valouev 2008).

Quest à l'avantage de permettre d'introduire deux paramètres en amont de l'analyse.

Le premier paramètre permet de donner une idée de l'allure du pic détecté. Pour cela, il existe trois modèles sur lequel l'algorithme se base: un pic très étroit (ex: ARN Polymérase III), un pic resserré (ex: ARN Polymérase II, facteur de transcription) ou un pic étalé (ex: modification d'histones).

Le deuxième paramètre donne une indication sur la taille des pics.

Avantages et inconvénients de Quest

L'algorithme Quest est très performant pour détecter des pics concentrés comme c'est le cas pour la Polymérase III. De plus, il tient compte du sens de lecture des reads pour détecter le pic. Cependant, nous avons pu montrer que cet algorithme ne parvient pas à détecter des pics dont l'amplitude est modérée ainsi que les pics comportant des reads sens et anti-sens mélangés. C'est notamment le cas pour la détection du pic de Chd8 sur le promoteur du gène Nanog (Figure 37).

SISSR (Jothi 2008)

Cet algorithme se base sur un scan du génome sur une fenêtre de 20 paires de bases. De plus, cet algorithme prend en compte le sens de lecture sens ou antisens du séquençage et définit le centre du pic lorsque les sens s'inversent.

Sicer (Zang 2009)

Ce dernier algorithme scanne le génome sur des fenêtres dont la taille est définissable par l'utilisateur (par défaut 200 paires de bases). De plus, il est possible d'autoriser un intervalle dont la taille est choisie entre deux fenêtres enrichies. Le dernier paramètre est le FDR (False Discovery Rate ou seuil de détection de faux positifs) plus couramment appelé FDR. Nous avons placé ce seuil à 10^{-9} ce qui permet d'avoir

Cet algorithme compte directement le nombre de reads dans une fenêtre donnée et s'affranchit donc du sens du tag séquencé. Ce qui n'est pas le cas de Quest ou SISSR. Cet

l'algorithme est donc plus efficace pour détecter les pics d'enrichissement des protéines qui ont des pics diffus où les reads sens et antisens se mélangent et lorsque l'enrichissement est peu élevé.

Détection des pics avec les différents algorithmes

Nous avons testé les différents algorithmes sur nos différentes protéines. L'exemple le plus intéressant est illustré dans la [Figure 37](#). Aucun pic n'est détecté avec l'algorithme Quest, 3 pics sont détectés sur le promoteur de Nanog par SISSR et seul SICER parvient à détecter des zones d'enrichissement. C'est aussi le cas pour un grand nombre d'autres gènes ([Figure 38](#)). C'est pourquoi nous avons choisi de conserver l'algorithme SICER pour nos analyses.

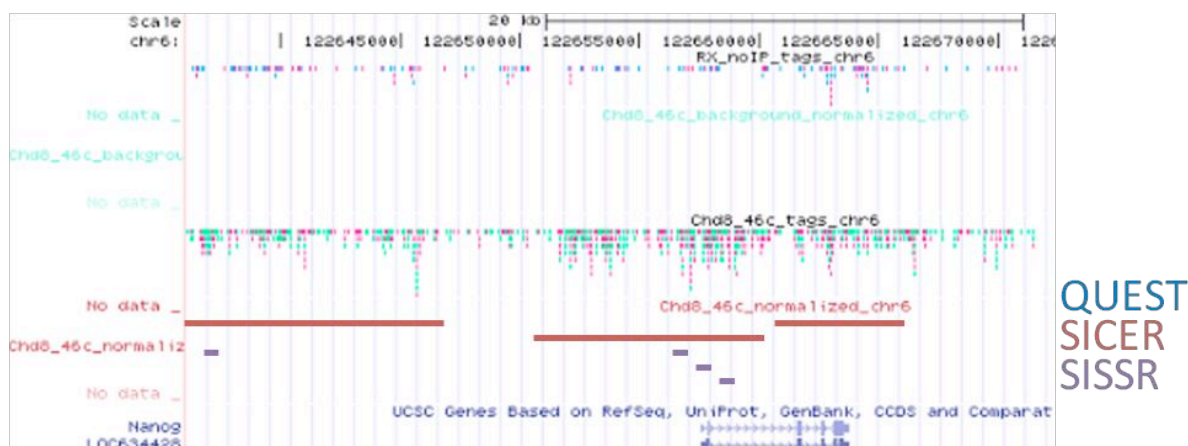


Figure 37: Observation des reads séquencés de Chd8 et du contrôle 46C et localisation des régions détectées enrichies par différents programmes. Aucun pic n'a été détecté par Quest, 3 pics larges par SICER (rouge) et 4 pics par SISSR (violet foncé).

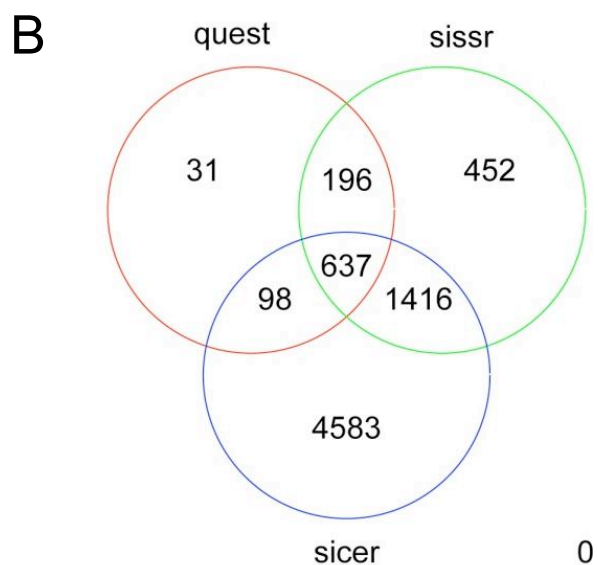
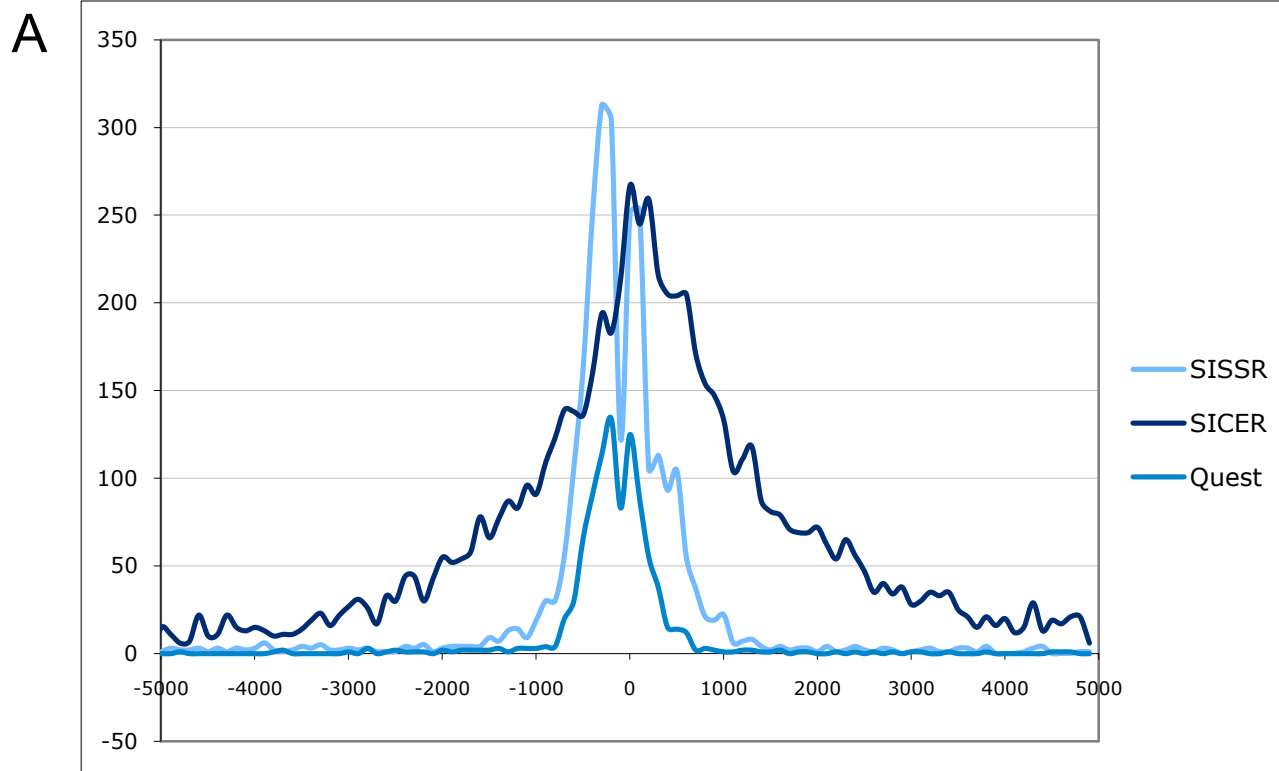


Figure 38: (A) Abondance du nombre de pics autour du TSS (en abscisse) détectés par Quest, SISR et SICER pour la protéine Chd8 sur ses propres gènes cibles. A partir de liste de pic Quest, SISR et SICER, le centre de chaque pic a été déterminé. Puis à l'aide du logiciel GPAT le TSS le plus proche a été défini. Enfin le nombre de "centre de pic" a été calculé dans un intervalle [-5000; +5000] avec une fenêtre de 100 paires de bases. SICER permet de détecter un plus grand nombre de pics, notamment les pics qui sont plus étalés et moins centrés sur le TSS comme Quest et SISR. **(B)** Nombre de gènes communs ou non, identifiés par les algorithmes Quest, SISR et SICER.

2. Programmes d'analyse utilisés

Les 9 protéines Chd ont des homologies de séquences de 25 à 50%. Elles partagent des caractéristiques communes que nous avons vues dans l'introduction. Il est donc

intéressant de comparer leurs séquences de liaison afin de comprendre leurs interactions et leurs rôles respectifs.

HeatMaps

Nos collaborateurs ont développé un algorithme qui permet, à partir d'une liste de positions donnée, de connaître l'enrichissement d'autres facteurs choisis (qui peuvent provenir de nos propres données ou des données déjà publiées).

Les programmes sont utilisés dans l'ordre suivant:

- a. Construction des profils de densité avec l'algorithme WigMaker3.jar (développé par Arnaud Krebs et Tao Ye, IGBMC)
- b. Construction de la liste contenant les profils de densités sur les listes de position avec l'algorithme distributionV5.py (développé par Arnaud Krebs et Tao Ye, IGBMC)
- c. Organisation et regroupement des profils de pics similaires (cluster) par le logiciel Cluster 3.0 (de Hoon et al., 2004)
- d. Visualisation de la heat map par le logiciel Java Tree View (Saldanha, 2004)

Détails des différents programmes utilisés.

WigMaker3

Ce programme découpe le génome en fenêtres de 25 paires de bases. A partir des fichiers de séquençage, il compte le nombre de tag dans chaque fenêtre. Il crée ensuite un dossier de sortie avec un fichier par chromosome, comportant le découpage du chromosome en fenêtre de 25 paires de bases. On appelle ces fichiers des fichiers de densité. Ce dossier de fichier est donc spécifique à chaque protéine et à chaque séquençage.

DistributionV5.py

Ce programme se base sur la liste de position en entrée. Après avoir trouvé le centre du pic par simple soustraction, il étend la sélection à -5000 et +5000 paires de bases et récolte les densités d'après le dossier de fichiers de densité fourni par l'utilisateur. Il crée ensuite un tableau où les données collectées sont ajoutées les unes en dessous des autres. Il y a autant de lignes que de positions entrées. Il y a 400 colonnes correspondant aux 10000

paires de bases divisées par le pas de 25 paires de bases. A cela s'ajoute une colonne d'index.

Cluster 3.0

A l'origine, ce programme a été créé pour le CHIP-on-CHIP où les sondes se trouvaient en ordonnée et les différentes expériences de CHIP en abscisse. Il organise les lignes d'un fichier en fonction des paramètres donnés par l'utilisateur afin de regrouper celles qui ont des profils communs.

Java Treeview

Ce programme a été conçu pour prendre naturellement la suite du programme Cluster 3.0. Il permet de visualiser les fichiers de sortie de Cluster 3.0 en attribuant à chaque valeur correspondant au nombre de tag lus dans la fenêtre de 25 pb, une couleur selon un dégradé défini par l'utilisateur. L'image obtenue est appelée HeatMap.

seqMINER

L'ensemble de ces programmes est maintenant rassemblé dans une interface graphique (Ye et al., 2010).

GPAT

Le programme GPAT (Genomic Position Annotation Tool) a été élaboré par nos collaborateurs à Strasbourg (Krebs et al., 2008). Ce logiciel permet à partir d'une liste de pics d'identifier le gène le plus proche selon des critères définis par l'utilisateur.

Le programme Pirp

Le programme Pirp (Programme d'identification de recouvrement de pics) a été élaboré avec l'aide de R. Piron.

On considère les ensembles des pics détectés dans chaque expérience. On néglige l'allure de ces pics et on n'identifie chacun d'entre eux que par une "position" de début et une "position" de fin (on réduit le pic à ses bornes). Le programme permet de confronter deux

ensembles A et B et d'identifier les pics qui se recouvrent (Figure 39), que l'on considère comme communs à A et B. En sortie, le programme donne trois ensembles de pics : l'ensemble $A \cap B$ des pics communs à A et B, l'ensemble des pics uniquement présents dans A, et l'ensemble des pics uniquement présents dans B.

En utilisant ce programme dans un script, on peut alors construire successivement à partir de N ensembles $\{A_1, A_2, \dots, A_N\}$ correspondant à des expériences les $C(2, N)$ ensembles $A_i \cap A_j$ (intersections de 2 ensembles, i allant de 1 à n et j de 1 à i), les $C(3, N)$ ensembles $A_i \cap A_j \cap A_k$ (intersections de 3 ensembles) ... jusqu'à l'ensemble $A_1 \cap A_2 \cap \dots \cap A_N$.

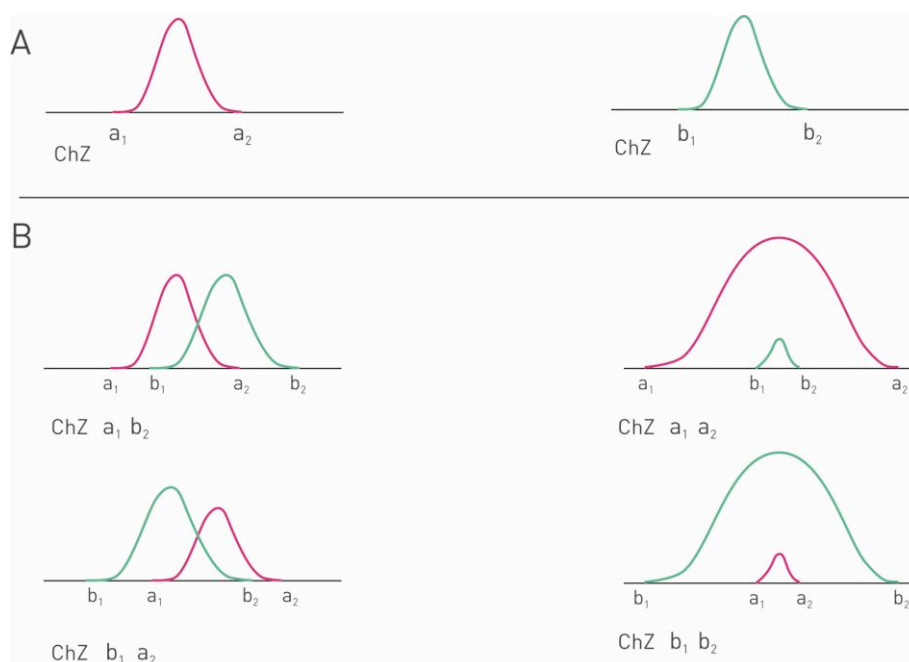


Figure 39: Schématisation de l'algorithme Pirp. A partir de deux listes de pics (Partie A), ce script permet d'identifier si les pics sont recouvrants selon les quatre possibilités (Partie B).

C. Etude du complexe NuRD dans les cellules ES

Le complexe NuRD détaillé dans la Figure 26 (page 89), comprend plusieurs protéines dont Chd4 et Hdac2 que nous avons étiquetées et obtenu les séquences cibles par ChIP-seq. Afin de déterminer les séquences cibles du complexe, les pics communs à Chd4 et Hdac2 sont déterminés.

Une première analyse a consisté à identifier les pics communs aux expériences d'immunoprecipitation de la protéine Chd4 avec le pontage formaldéhyde (Chd4 F), de Chd4 avec le pontage DSG+formaldéhyde (Chd4 DSG), d'Hdac2 avec le pontage formaldéhyde (Hdac2 F) et d'Hdac2 avec le pontage DSG+formaldéhyde (Hdac2 DSG). Cette opération correspond à l'étude de l'ensemble:

Chd4 F \cap Chd4 DSG \cap Hdac2 F \cap Hdac2 DSG

Cette identification se fait grâce à l'algorithme Pirp décrit ci-dessus. 222 pics sont trouvés en commun. Une analyse GPAT permet d'identifier à partir de la liste des 222 pics, 89 gènes. Parmi ces gènes, il est intéressant de noter la présence des gènes Myc et Dppa2, tous deux des gènes impliqués dans la pluripotence des cellules ES. La présence des protéines Chd4 et Hdac2 identifiées par le ChIP-seq en condition DSG et formaldéhyde seul permet d'affirmer que le complexe NuRD est présent sur les séquences régulatrices composées des promoteurs et enhanceurs en 5' et 3' du TSS, de ces gènes aux positions identifiées.

Une deuxième analyse consiste à regrouper les séquences liées pour les expériences de ChIP-seq pontées par la formaldéhyde seule ou le DSG+formaldéhyde. Ce regroupement est effectué pour la protéine Chd4 d'un coté et la protéine Hdac2 de l'autre. Cette analyse correspond à l'étude de l'ensemble:

(Chd4 F U Chd4 DSG) \cap (Hdac2 F U Hdac2 DSG)

L'intersection des deux listes est ensuite déterminée. La liste Chd4 contient 22535 pics, correspondant à 7035 gènes. La liste Hdac2 contient 13739 pics correspondant à 6114 gènes. 3823 gènes (correspondants à 6030 pics) sont retrouvés en commun entre Chd4 et Hdac2. La liste de ces gènes sera donnée en table supplémentaire avec l'article. Parmi ces gènes ont trouve les gènes responsable du maintien de la pluripotence décrits dans la littérature ([Tableau 2](#)). Le complexe NuRD se lie aux gènes de la pluripotence. Le rôle du complexe NuRD est ensuite détaillé dans le Chapitre 4.

D. Etudes des profils de position des protéines Chd

En juillet 2010, l'analyse se limite aux protéines Chd1, Chd4 et Chd8 pour lesquelles nous avons obtenu les résultats du séquençage Illumina.

1. Présence des Chd aux promoteurs

La première étape de l'analyse a consisté à connaître l'enrichissement des Chd aux promoteurs. Pour cela, le logiciel promoteur profile a été utilisé (Figure 40). Ces données permettent de montrer que Chd1 et Chd8 sont globalement assez présents aux promoteurs et leur présence est en faveur d'une corrélation. Chd4 est quant à lui moins présent aux promoteurs et lorsqu'il s'y trouve, il est beaucoup moins abondant.

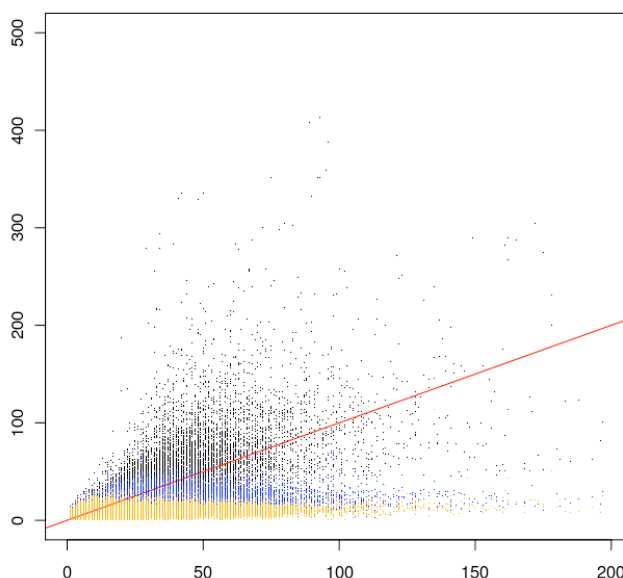


Figure 40: Corrélation des enrichissements aux promoteurs. Chaque point d'une couleur représente un promoteur, il y a au total ~24000 point de chaque couleur. Les valeurs d'enrichissement pour Chd1 sont en abscisse. En ordonnées sont reporté les données pour Chd8 (noir), Chd4 (bleu) et le contrôle la lignée 46C non étiquetée (jaune). En rouge la droite de corrélation $x = y$.

Cette observation est confirmée par la Figure 41, qui présente les données de séquençage de l'ARN polymérase II (Goldberg et al., 2010), de Chd1 F, Chd8 F, Chd4 F, Chd4 DSG, Hdac2 DSG, 46C (lignée contrôle non étiquetée), H3K4me3 et H3K36 (Mikkelsen et al., 2007) sur les promoteurs des 2000 gènes les mieux exprimés dans les cellules ES (nos propres données). Chd1 et Chd8 sont très présentes au niveau des promoteurs alors que Chd4 est globalement absent. Malgré la co-localisation précédemment observée, ces données permettent de montrer que Chd1 et Chd8 n'ont pas le même type de liaison sur les promoteurs des cellules ES. En effet si Chd1 semble excessivement bien corrélé avec la polymérase II et avec les gènes bien exprimés, Chd8 ne semble pas être corrélé avec la polymérase II. En effet, Il est moins présent que Chd1 sur les gènes très exprimés et est plus présent que Chd1 sur les gènes moins exprimés.

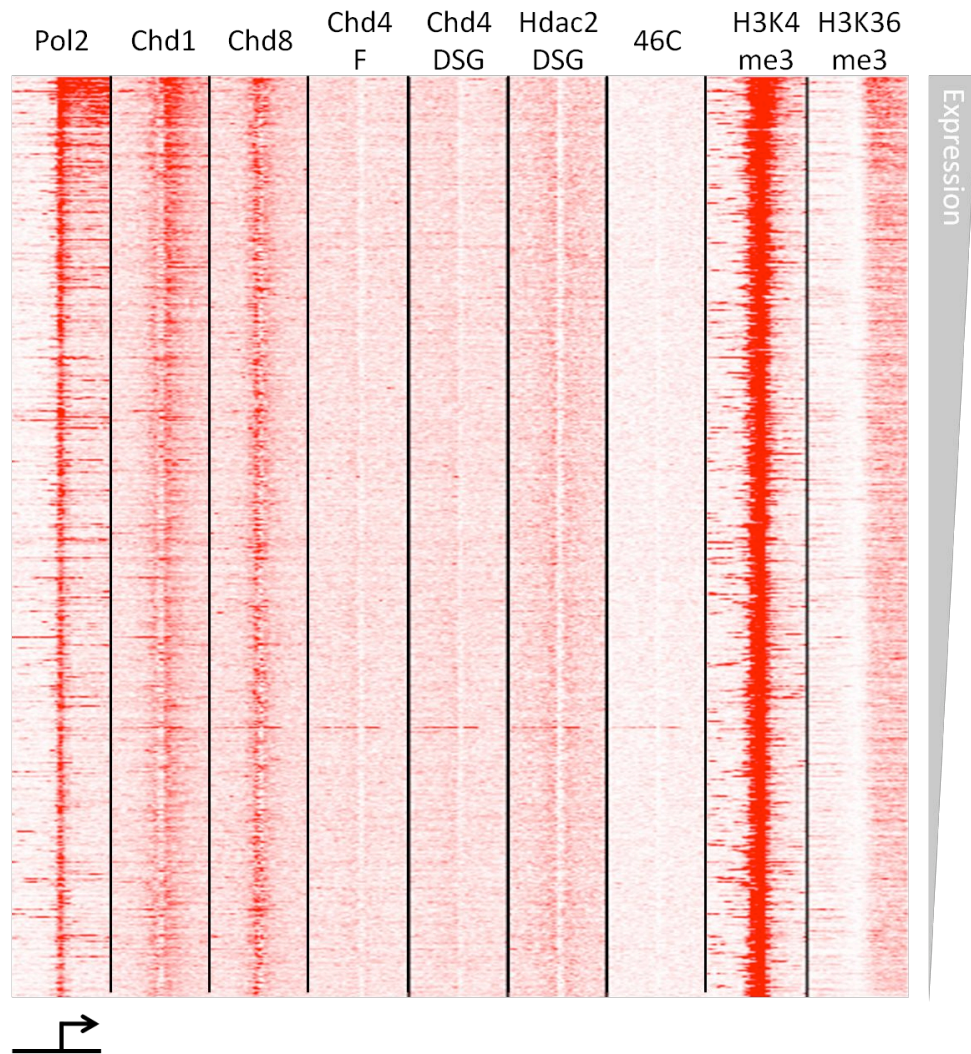


Figure 41: Abondance de l'ARN polymérase II (Pol2), Chd1 (formaldéhyde), Chd8 (formaldéhyde), Chd4 (formaldéhyde), Chd4 (DSG), Hdac2 (DSG), 46C - lignée contrôle - (formaldéhyde), H3K4me3 et H3K36me3 (Mikkelsen et al., 2007)

L'analyse par le logiciel GPAT permet de mettre en évidence un certains nombres de gènes qui sont liés par les Chd sur l'intervalle [-5000; +5000] des TSS de gènes murins. Ces listes de gènes liés ont été comparées entre elles (Figure 42) et il a été mis en évidence que 1364 gènes étaient communs aux 3 protéines Chd. Chd4 partage peu de cibles avec Chd1 et Chd8 qui partagent la plupart de leurs cibles (4489).

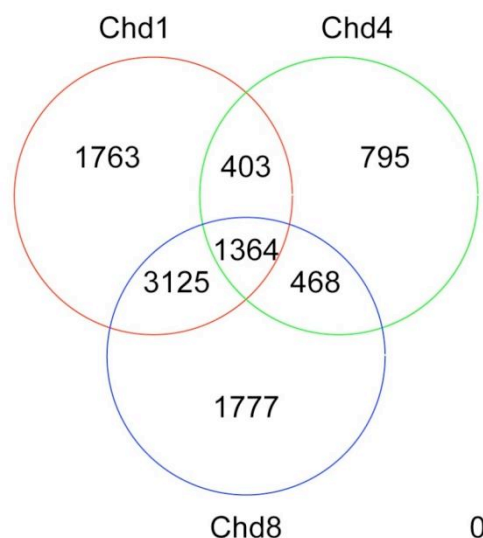


Figure 42: Nombres de gènes liés au promoteur par Chd1, Chd4 et Chd8 communs et spécifiques à chacune.

Enfin, une observation au cas par cas sur les promoteurs montre que lorsque Chd1 et Chd8 se lient au promoteur d'un gène, le site de fixation peut être légèrement décalé. Ceci suggère que leur mode de recrutement est différent.

2. Co-localisation avec les marques H3K4me2 et H3K4me3

L'analyse de l'abondance de Chd1 et Chd8 sur leurs propres séquences cibles a été visualisée grâce à une analyse Java Tree View. Deux profils de liaisons ont été identifiés. Le premier montre une co-localisation de Chd8 avec la marque H3K4me3, H3K4me2 ainsi que l'ARN polymérase II et une anti corrélation avec la marque H3K4me1 (Figure 43). De façon surprenante, il n'a été trouvé aucune corrélation avec CTCF ce qui avait été pourtant décrit par Ishihara et al. (Ishihara et al., 2006). Ce même profil a été identifié avec Chd1. Le deuxième profil correspond aussi à une corrélation avec la marque H3K4me3 mais moins abondante.

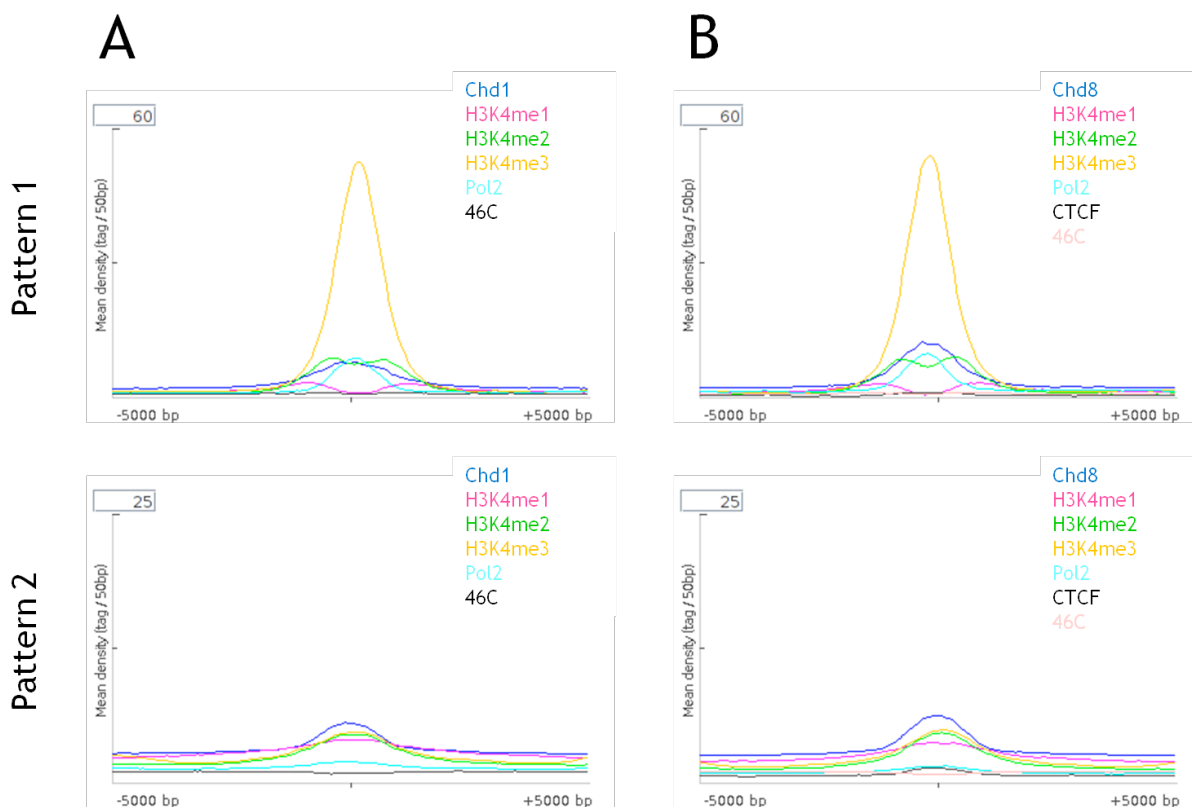


Figure 43: Occupation de Chd1 (A) et Chd8 (B) sur leurs propres cibles. Deux profils d'occupation de Chd1 et Chd8 sont représentés. La position par rapport au centre du pic détecté par SICER est donné en abscisse. En ordonné le nombre de tag séquencé dans une fenêtre de 25 paires de bases.

3. Autorégulation des Chd

Après détection des pics d'enrichissement par SICER, le logiciel GPAT est utilisé pour identifier les gènes liés. Parmi ces gènes, certains Chd sont retrouvés. Les protéines Chd forment donc elles-mêmes un réseau de régulation (Figure 44).

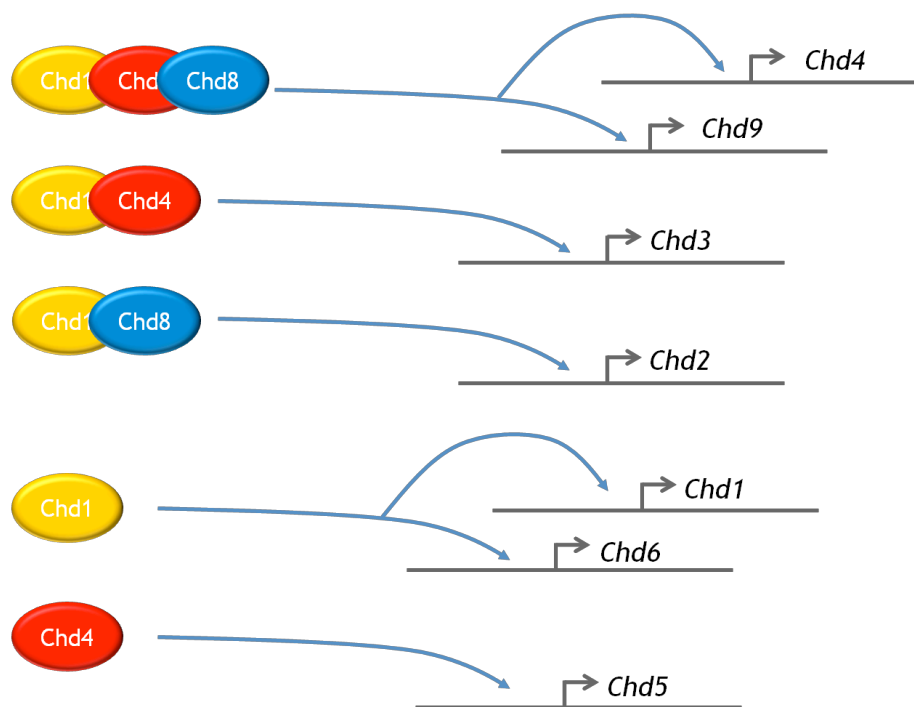


Figure 44: Autorégulation des Chd d'après les données de séquençage obtenues avec la fixation type formaldéhyde. Aucune présence de Chd n'a été détecté sur les gènes Chd7 et Chd8 lors des immunoprécipitations réalisées avec le formaldéhyde.

Conclusion

Ces analyses bioinformatiques nous ont permis de mettre en évidence des rôles inédits pour les protéines Chd. L'analyse préliminaire présentée ici, est un échantillonnage de chaque sous-famille Chd (sous-famille 1: Chd1, sous-famille 2: Chd4, sous-famille 3: Chd8). Des similarités mais aussi des rôles divergents apparaissent. La spécificité de liaison de Chd4 est étonnamment singulière et l'analyse des données de Chd3 et Chd5 permettra de savoir si ce profil est spécifique à Chd4 ou englobe la sous-famille 2 des Chd.

Cette étude se poursuit actuellement (septembre 2010) avec l'analyse de toutes les protéines Chd. Plus spécifiquement nous nous intéressons à leur profil de liaison, aux similarités et différences que l'on peut retrouver entre ces protéines.

Enfin, nous tentons de savoir si les protéines Chd sont connectées aux réseaux de transcription régulant la pluripotence des cellules ES (voir étude sur le complexe NuRD détaillé dans la publication numéro 2).

Chapitre 4. Comparaison des données de liaisons avec des données transcriptomiques afin de proposer des hypothèses fonctionnelles

Introduction

Les données issues du ChIP-seq nous ont permis de déterminer des zones d'enrichissement des facteurs Chd qui correspondent aux régions liées de manière directes (pontage entre la protéine d'intérêt et l'ADN) ou indirectes (pontage de la protéine d'intérêt à l'ADN via une ou plusieurs autres protéines). Certaines de ces régions liées correspondent à des régions enhanceurs, promotrices ou transcrites de gènes (Chapitre 3). Afin de comprendre le rôle des protéines Chd sur les gènes que nous avons détectés par la méthode du ChIP-seq, il serait intéressant d'observer l'effet de l'absence des protéines d'intérêt sur la transcription gènes liés. Nous postulons que si la protéine liée a un rôle sur le contrôle transcriptionnel des gènes, son absence entrainera une dérégulation de ces derniers.

Afin d'inhiber l'expression de la protéine d'intérêt, nous avons choisi la méthode de l'ARN interférence (RNAi). Cette méthode basée sur l'interférence d'un petit ARN avec l'ARNm de la protéine d'intérêt permet d'obtenir une diminution de 50 à 95% d'expression de la protéine. La perte d'expression est induite par l'électroporation d'un plasmide codant pour ce petit ANR et possédant une cassette de sélection puromycine. Elle est transitoire, et après 48 heures de sélection, les ARN sont collectés et analysés sur puce Illumina par un collaborateur du CNG, Robert Olazo (CNG, Evry).

Trois effets sont observés. Premièrement, l'absence de la protéine d'intérêt augmente l'expression d'un gène. Deuxièmement, l'absence de la protéine diminue l'expression d'un gène. Dans ces deux cas, l'absence de la protéine d'intérêt a un impact sur l'expression d'un ensemble de gènes. L'absence de protéine n'altère pas l'expression du gène dans une troisième hypothèse.

L'ensemble des gènes dont l'expression est altérée selon un seuil fixé, constitue les gènes dérégulés. L'ensemble des gènes dérégulés est recoupé avec les gènes liés. Les gènes liés et dérégulés sont appelés gènes-cibles. Ces gènes sont à la fois liés par la protéine d'intérêt et leur expression est dérégulée lors de l'absence de la protéine. La

protéine a un rôle dans la régulation de la transcription de ces gènes. Un certain nombre de gènes sont dérégulés mais pas liés par la protéine. Ces gènes sont donc des gènes-cibles indirects. Il est aussi possible que l'absence de la protéine d'intérêt, n'ait pas d'effet sur l'expression du gène lié. Dans ce cas, le rôle de la protéine est indirect et d'autres expériences devront être menées pour déterminer ce rôle.

Grâce au recoupement des données de positions et des données d'expression, nous pourrons alors définir les rôles des protéines Chd dans les réseaux de régulations de la transcription.

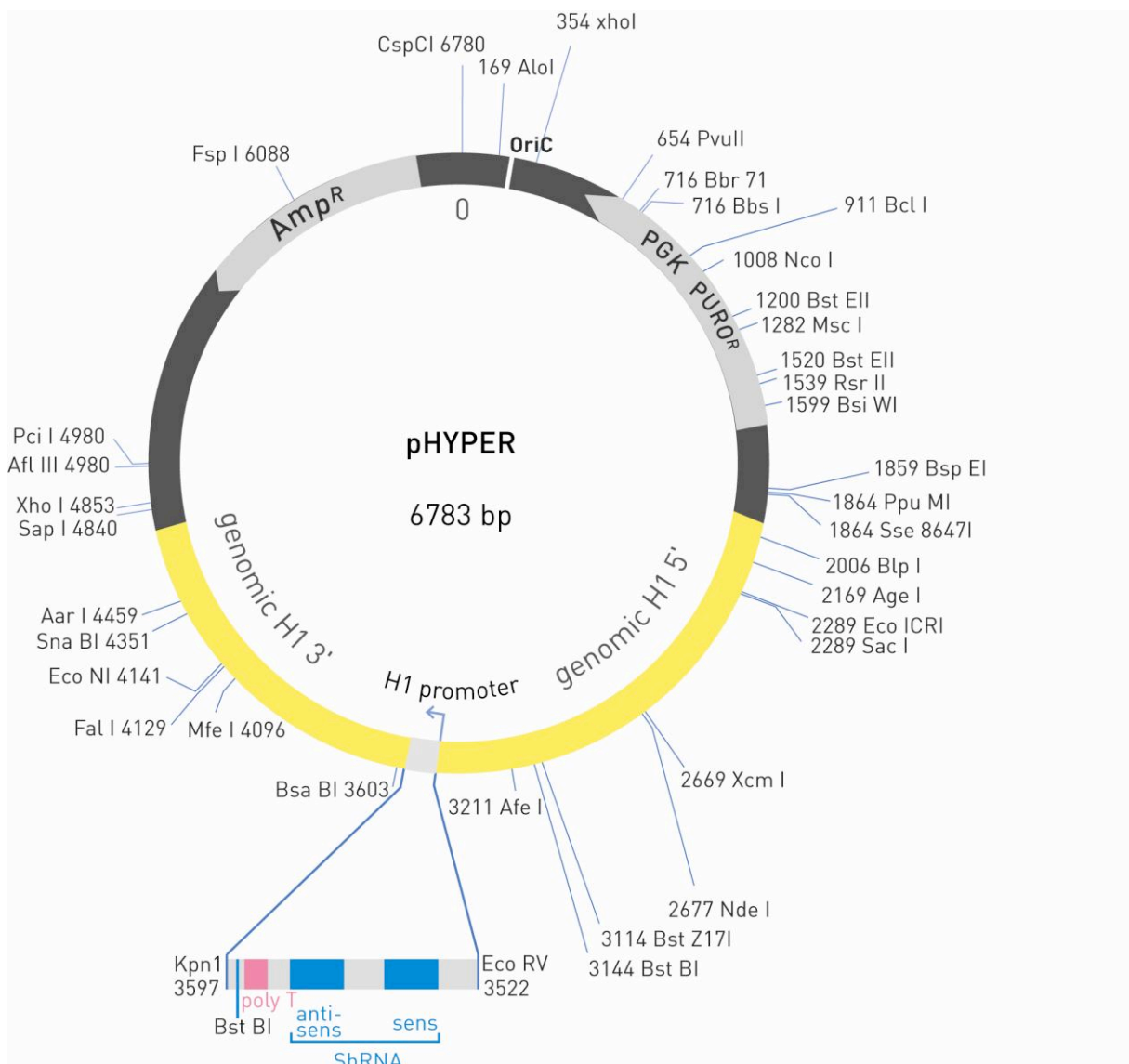


Figure 45: Cartographie du vecteur pHyper

A. Description de la méthode des shRNA

Cette méthode consiste à introduire dans la cellule ES un plasmide ([Figure 45](#)) qui exprime une séquence d'environ 70 paires de bases dont 21 sont complémentaires à une partie de l'ARN messager de la protéine ciblée. La séquence clonée est déterminée à partir de l'ADN complémentaire de la protéine d'intérêt. La séquence complémentaire est ajoutée après un linker. L'ensemble de cette séquence est ensuite clonée dans un vecteur pHYPER décrit par Berlivet et al (Berlivet et al., 2007). Après vérification du vecteur par séquençage, le vecteur pHYPER recombinant est électroporé dans les cellules ES. Une boucle en épingle à cheveux appelée shRNA est transcrite dans le noyau et migre dans le cytoplasme. Le linker est ensuite clivé par la protéine DICER. La séquence de RNAi est libérée dans le cytoplasme et complémente l'ARN messager de la protéine d'intérêt. L'appariement de la séquence interférence à l'ARNm est reconnu comme une séquence ARN double brin par des enzymes du complexe RISC, qui induit la dégradation de l'ARN messager cible. L'absence d'ARNm codant pour la protéine ne permet plus de néo-synthèse. Le pool de protéines cibles cellules est diminué alors rapidement. En fonction de la demi vie de la protéine observée et l'efficacité du shRNA, une réduction de 50 à 95% est observée. L'ensemble de ces étapes est regroupé dans la [Figure 46](#).

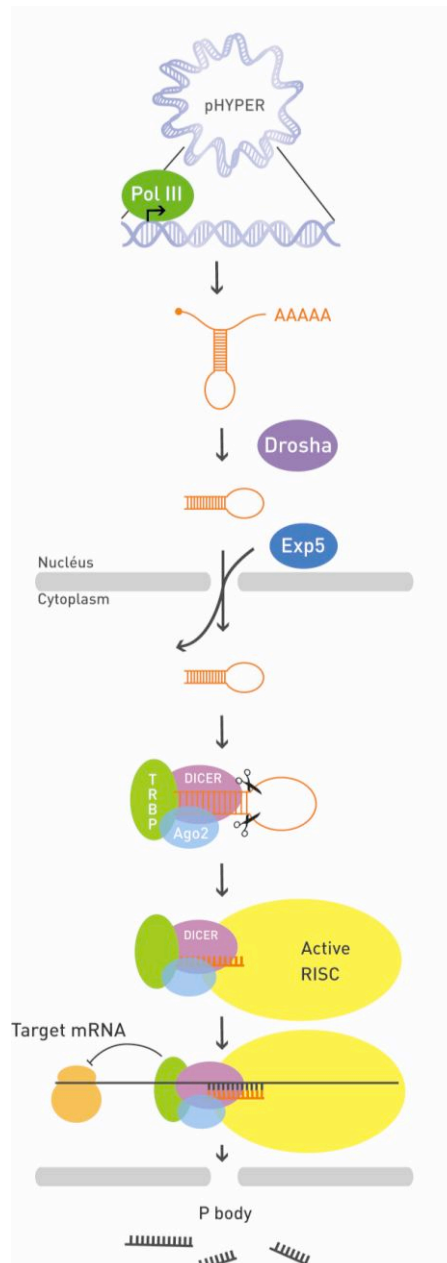


Figure 46: Extinction de l'expression d'un gène ciblé par la méthode du shRNA. Le petit ARN en épingle à cheveux est synthétisé par la polymérase III (promoteur H1). L'ARNm est ensuite mûré par la protéine Drosha et exportée vers le cytoplasme grâce à la protéine Exportine 5. L'ARN est ensuite pris en charge par DICER qui clive la boucle reliant la séquence sens et antisens, la dégradation du brin antisens est ensuite prise en charge par le complexe RISC. L'hybridation de l'ARN sur sa cible conduit à la dégradation du grand ARNm vers les P bodies et l'inhibition de la traduction.

B. Construction des vecteurs de shRNA

1. Le vecteur *pHYPER*

Le vecteur utilisé est le vecteur pHYPER mis au point au laboratoire et décrit précédemment dans la littérature (Berlivet et al., 2007). Ce vecteur possède 2500 paires de bases du promoteur du gène H1. Ce dernier est reconnu par l'ARN polymérase III des cellules ES et permet donc une transcription efficace. Ce vecteur possède un gène β lactamase sous le contrôle d'un promoteur bactérien qui permet une résistance à l'ampicilline chez E.Coli. Ce vecteur possède une cassette PGK-puro, exprimant un gène de résistance à la puromycine (puromycine transacetylase) sous le contrôle du promoteur murin du gène de la phospho glycérate kinase, qui permet aux cellules ES électroporées d'acquérir la résistance à la puromycine.

2. Constructions des shRNA

Les séquences d'ARN interférence (RNAi) ont été choisies avec le logiciel DSIR (Vert et al., 2006), <http://biodev.extra.cea.fr/DSIR/DSIR.html>.

Les constructions de shRNA ont été élaborées pour chacun des huit protéines Chd1, Chd2, Chd3, Chd4, Chd5, Chd6, Chd8, Chd9 auxquelles nous avons ajouté les protéines Chaf1a et Hdac2. Les amorces d'environ 70 paires de bases contenant l'ARN sens et anti-sens, sont clonés dans le vecteur pHYPER et vérifiées par séquençage en collaboration avec Sophie Chantalat au CNG (Evry).

3. Test des shRNA

Les cellules utilisées pour l'électroporation sont les cellules qui ont été étiquetées. Ces lignées ont l'avantage de nous permettre de suivre la présence de la protéine ciblée par western blot. Ainsi, nous pouvons suivre précisément l'extinction de la protéine correspondant au shRNA électroporé.

Quinze microgrammes de plasmide pHyper ont été électroporés dans 5 millions de cellules ES. Les cellules sont sélectionnées 48 heures ou 72 heures avec une sélection de 2 μ g/ml puromycine. La résistance est transitoire car le plasmide circulaire n'est pas intégré au génome murin et est progressivement perdu par les cellules au-delà de 72 heures.

4. Etude de la perte de fonction de chaque protéine Chd dans les cellules ES

Etude par shRNA

Le test de quatre pour chaque protéine étudiée nous a permis de déterminer le shRNA le plus performant. Nous avons donc alors transfecté le vecteur codant ce shRNA dans la lignée ES étiquetée correspondante ainsi que le pHYPER seul (contrôle négatif), en triplicat. Une partie des cellules est conservée pour préparer les ARN (RNeasy Lipid tissue Kit, Qiagen) et l'autre partie est lysée afin de contrôler l'efficacité de l'inactivation par western blot (Figure 47).

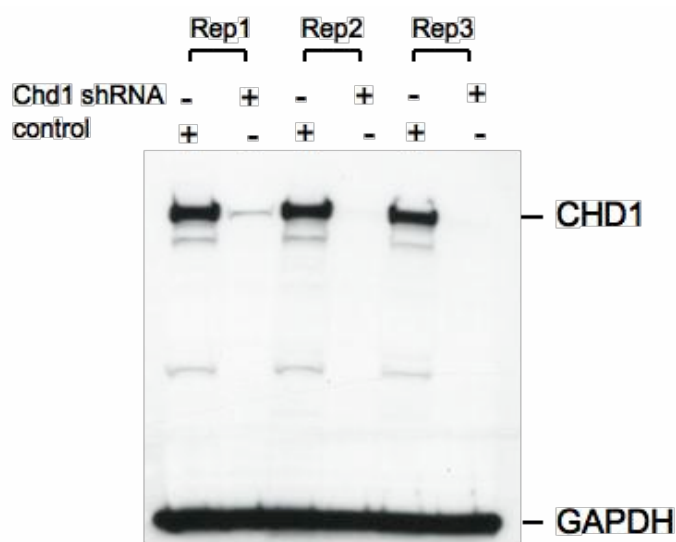


Figure 47: Extinction de l'expression de la protéine Chd1 après l'électroporation du vecteur pHyper exprimant le shRNA correspondant. Chd1-3FTH est détecté par western blot et réalisé en triplicat. Le contrôle est le vecteur pHYPER vide électroporé. La partie supérieure du western blot a été révélée par l'anticorps anti-FLAG et la partie inférieure avec l'anti-GAPDH.

Les échantillons d'ARN sont ensuite envoyés au CNG où leur qualité est contrôlée. Ils sont ensuite hybridés sur puce d'expression illumina (Figure 48). Ces puces comportent 48000 sondes contenant 24000 transcrits Refseq et 24000 EST Riken.

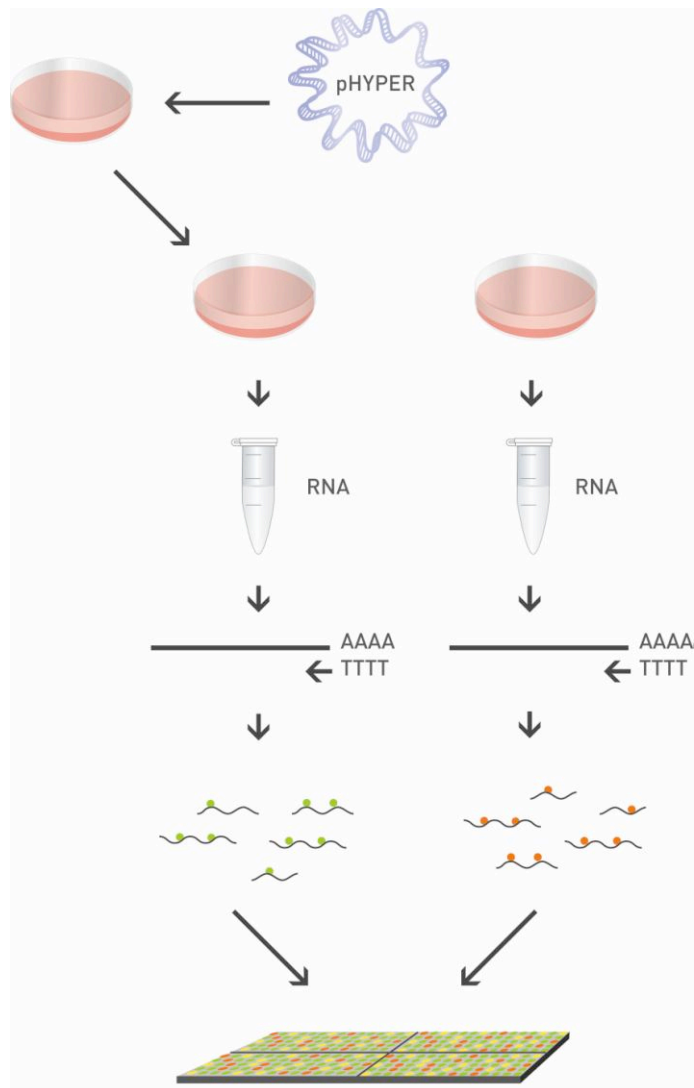


Figure 48: Description de l'utilisation des shRNA pour déterminer les l'ensemble des gènes dérégulés lors du knock-down de la protéine d'intérêt

Les résultats sont ensuite triés selon:

1. Le coefficient de dérégulation ou "fold change" (FC) donné par:

$$\text{valeur}_{\text{shRNA protéine cible}} / \text{valeur}_{\text{shRNA contrôle}}$$

2. La p-value associée au FC qui représente le degré de confiance dans ce coefficient.

Les gènes retenus sont ceux qui présente une p-value d'au moins 0,05% et dont le FC dépasse le seuil de 1,2.

RNA-seq comme alternative aux puces

Le complexe NuRD, décrit dans l'introduction, se compose d'un ensemble de protéines comprenant les protéines Chd4 et Hdac2. Nous avons dérivé au laboratoire des lignées de souris mutantes pour Hdac2 (allèle genet rap, équivalent KO) (*Hdac2*^{-/-}) et mutante pour Chd4. Cette lignée porte un allèle affaibli ou hypomorphe, pour Chd4 (*Chd4*^{hypo/hypo}) résultant de deux mutations entre les deux chromodomains (voir les informations supplémentaires de la publication numéro 2). Une technicienne du laboratoire (S. Jounier) a dérivé des lignées ES à partir de ces souris et nous avons obtenu une lignée *Hdac2*^{-/-} et plusieurs de génotype *Chd4*^{hypo/KO} *Hdac2*. Cet allèle permet de travailler avec un allèle affaibli de *Chd4* alors que l'homozygote invalidé est létal. Les cellules ES *Chd4*^{-/-} meurent ou entrent en différenciation très rapidement.

Ces deux lignées ont été utilisées dans des expériences de RNA-seq (Figure 49). Brièvement, les ARN sont préparés (RNeasy Lipid tissue Kit, Qiagen). Une étape de reverse transcription permet d'obtenir les ADN complémentaires. Les ARN sont séquencés sur le génome analyser, Illumina par M. Gut (CNG, Evry).

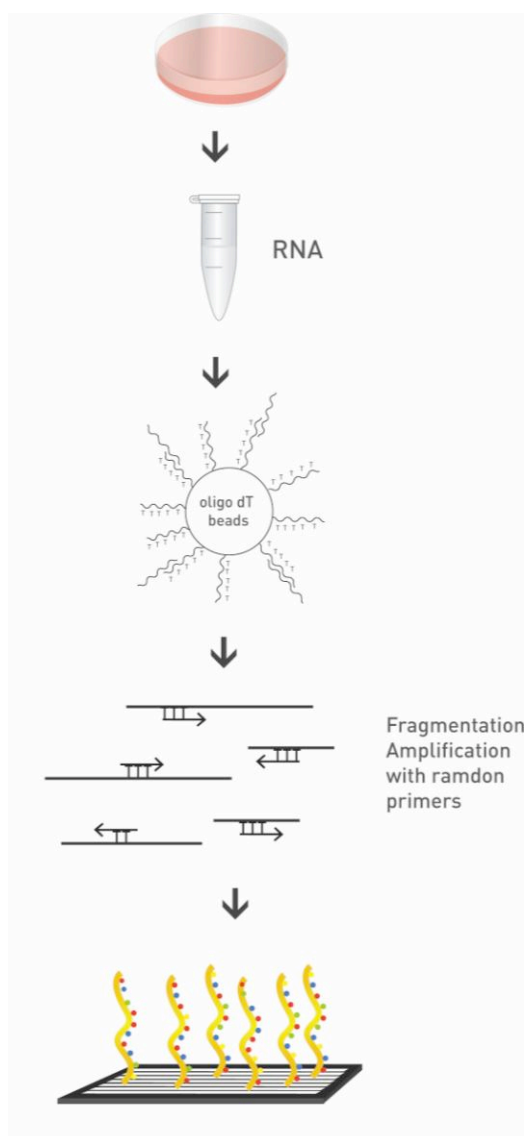


Figure 49: Schéma récapitulatif de la technique de RNA-seq. Les cellules sont cultivées et l'ARN est récupéré, les ARNm sont isolés par complémentarité des chaînes poly A avec des billes oligo dT. Les fragments sont amplifiés et séquencés sur puce Illumina.

Les données de séquençage ont été analysées avec le logiciel ERANGE. Les analyses bioinformatiques ont été réalisées par Céline Keime (IGBMC, Strasbourg).

Les analyses ont permis de montrer que 224 gènes sont dérégulés (avec un seuil FC de 1,7) dans la lignée *Chd4*^{hypo/hypo}. Alors que dans la lignée *Hdac2*^{-/-} 890 gènes sont dérégulés. Enfin, 106 gènes sont dérégulés dans les lignées *Chd4*^{hypo/hypo} et *Hdac2*^{-/-}.

Confrontation des données de position et des données de pertes de fonction du complexe NuRD

Ces analyses sont décrites dans la publication en préparation numéro 2.

Les données ChIP-seq ont été analysées par le programme SICER qui nous permettait de détecter le plus grand nombre de régions (Chapitre 3). Quatre listes sont étudiées pour le facteur Chd4 et Hdac2 avec les pontages réalisés avec le formaldéhyde seul et l'ensemble DSG+formaldéhyde. Les données de positions enrichies sont cartographiées sur le génome murin afin d'identifier les gènes proches de ces pics. Pour chaque liste, l'attribution d'un gène à un pic se fait grâce au logiciel GPAT (Krebs et al., 2008).

Parallèlement, les analyses RNAseq ont consisté à identifier les gènes dérégulés selon deux valeurs: le "fold change" (FC) et la p-value associée. Le seuil choisi est 1,7 correspondant à la dérégulation de Chd4 dans le contexte *Chd4^{hypo/hypo}*.

Il est donc très intéressant de rapprocher les données RNAseq qui correspondent aux données de transcriptions et les données de positions. Les gènes qui seront liés par le complexe NuRD et dérégulés dans les lignées *Chd4^{hypo/hypo}* et *Hdac2^{-/-}* seront des gènes cibles du complexe NuRD.

Trois approches ont été choisies:

La première approche a consisté à identifier, dans la liste des gènes dérégulés pour Chd4, ceux qui étaient identifiés par le ChIP-seq Chd4 formaldéhyde, puis ceux qui étaient retrouvés dans le ChIP-seq Chd4 DSG. Puis de même avec la protéine Hdac2 avec le pontage formaldéhyde+DSG. Les données ont été ensuite recoupées entre elles. Les 26 gènes en communs entre Chd4 (F+DSG) et Hdac2 (F+DSG) ont été identifiés et les FC ont été comparés. Le petit nombre de gènes en commun (26) suggère que les protéines Chd4 et Hdac2 ont des rôles individuels indépendamment du complexe NuRD. Parmi ces 26 gènes, 23 sont dérégulés dans le même sens. C'est-à-dire que lorsque un gène est up régulé pour *Chd4^{hypo/hypo}*, il l'est aussi pour *Hdac2^{-/-}* (14 gènes). Le même effet est observé pour les gènes down-régulés (9 gènes). Ces gènes sont co-régulés par Hdac2 et Chd4. De plus, ces gènes sont liés par les protéines, ils sont donc des gènes cibles du complexe NuRD. Cependant sur trois gènes on observe des corrélations opposées. Pour les gènes *Lefty2* et *Grhl2*, la mutation de Chd4 entraîne une dérégulation positive et l'absence de Hdac2 entraîne une dérégulation négative. Le dernier gène *Ntn1* est down-régulé dans la lignée *Chd4^{hypo/hypo}* et up-régulé dans la lignée *Hdac2^{-/-}*.

Nous observons que les protéines Chd4 et Hdac2, ne se lient pas sur les mêmes sites liés aux gènes *Ntn1*, *Lefty2* et *Grhl2*. Les deux protéines jouent chacune un rôle dans la transcription de ces gènes mais pas de la même façon. Nous pouvons donc postuler que plusieurs complexes de régulation de la transcription interviennent. Ces complexes ont des rôles opposés.

Une deuxième analyse peut être conduite en partant des pics communs par le script Pirp (Chapitre 3, page 131). Une analyse GPAT permet d'identifier 89 gènes qui sont liés par Chd4 F et DSG ainsi que Hdac2 F et DSG. Ces gènes sont des liés par le complexe NuRD.

En comparant cette liste avec la première, on ne retrouve qu'un seul gène en commun. C'est-à-dire qu'il n'y a qu'un seul gène (*Tns3*) qui possède à la même position un pic pour les quatre conditions ChIP-seq et qui soit dérégulés (négativement) dans le KO Hdac2 et le mutant hypomorphe de Chd4. Ces conditions d'analyses sont donc très restrictives.

Une dernière analyse consiste à rassembler les listes formaldéhyde et DSG pour chaque protéine (Chapitre 3). En effet les fixations sont complémentaires et certaines régions peuvent être renforcées dans une condition ou une autre. C'est particulièrement le cas pour Hdac2 où peu de gènes liés sont trouvés avec le pontage formaldéhyde alors que la condition de fixation DSG permet d'obtenir 20 fois plus de gènes liés. Cette observation est consistante avec le fait que la protéine Hdac2 ne se fixe pas directement à l'ADN. Les pontages protéines-protéines ainsi renforcés par le DSG, l'ADN reste fixé à la protéine Hdac2 lors de l'expérience de CHIP. Cette analyse a permis de détecter 3823 gènes liés par Chd4 et Hdac2. Parmi ces gènes, la plupart des gènes essentiels aux cellules ES sont retrouvés. Cependant, la dérégulation de ces 3823 gènes n'est pas homogène. Elle peut être faible ou inexistante dans les lignées *Hdac2^{-/-}* ou *Chd4^{hypo/hypo}* ou au contraire être significative. Ceci est peut être dû aux nombreux protéines régulatrices qui sont recruté aux promoteurs de certains gènes et l'absence d'un seul complexe régulateur n'est pas suffisant pour induire une dérégulation visible.

C. Article en préparation

Nucleosome Remodelling and Deacetylation Complex (NuRD) is extensively involved in the transcription regulatory networks that control ES cell pluripotency

S Jounier*, M de Dieuleveult*, C Keime*, M Gut*, F Boussouar*, A Depaux*, H Humbertclaude, R Olaso, S Chantalat, A Krebs, T Ye, R Guérois, S Legras, L Carrière, A Martel, Y Duffourd, P Héry, V Guiraud, I Davidson, M Werner, L Tora, M Lathrop, I Gut, M Gerard

* equal contribution

Results

Efficient tagging of *Chd4* and *Hdac2* genes by recombineering and homologous recombination in mouse embryonic stem cells.

We used an epitope-tagging tandem affinity strategy to identify the target genomic sequences of NuRD chromatin remodelling factor. We chose to tag the Chd4 (Mi-2b) and Hdac2 subunits, which carry nucleosome remodelling and histone deacetylase activities, respectively. A sequence encoding a 3x FLAG epitope followed by a TEV cleavage site and a single HA epitope was introduced at the 3' extremity of Chd4 coding sequence, in frame with the C-terminal end of the protein, using the recombineering method (Liu et al. 2003). This fusion was then integrated into the *Chd4* locus of ESCs by homologous recombination (Figure S1). An important implication of this strategy is that the resulting Chd4 fusion protein (named Chd4-3FTH) is expressed at normal level from one of the two endogenous *Chd4* alleles, and properly regulated during the cell cycle. A similar strategy was used for Hdac2, except that the tag consists of a single FLAG followed by a single HA epitope, resulting in the *Hdac2-HF* allele. We obtained several independent ES cell lines carrying the *Chd4-3FTH* and *Hdac2-HF* alleles. We confirmed by Western blot the expression of the corresponding fusion proteins (data not shown). Chd4-3FTH and Hdac2-HF ESC lines grew normally in culture, were morphologically indistinguishable from the 46C ESC line (Ying and Smith 2003) from which they are derived, and expressed normal levels of mRNA encoding ESC markers *Pou5f1*, *Nanog*, *Sox2*, *Sall4*, *Klf4*, *Zfp42*.

Mapping the genome-wide distribution of Chd4 and Hdac2 by tandem-affinity chromatin immunoprecipitation

We prepared chromatin from *Chd4-3FTH*, *Hdac2-HF*, and 46C cells, which do not express any tagged protein and serve as a control. We selected two different strategies to preserve Chd4 and Hdac2 proteins association with chromatin: i) conventional fixation with formaldehyde and ii) treatment with

the bifunctional crosslinking agent disuccinimidyl glutarate (DSG), followed by formaldehyde fixation (Nowak et al. 2005). After fixation, cells were collected and permeabilized with a detergent in order to allow the penetration of micrococcal nuclease in ESC nuclei in the subsequent step. We set up a protocol based on micrococcal nuclease (MNase) fragmentation of genomic DNA, rather than sonication, because the size of the enzymatically fragmented DNA (around 150bp for mononucleosomes) gives a better resolution in ChIPseq experiments than sonication (around 300 bp). Previous studies showed that MNase –based preparation of chromatin for ChIP applications is an interesting alternative to sonication (Schmid et al. 2004). MNase digestion does not preferentially degrade open chromatin regions such as promoter and active genes (Schones et al. 2008). Furthermore, compact chromatin regions such as pericentromeric constitutive heterochromatin are digested by MNase with the same efficiency as euchromatin regions (SC and MG, unpublished data). Chd4-3FTH bound chromatin was purified in the first step by HA-affinity purification, which was performed from $5 \cdot 10^8$ cells, using low detergent and low salt conditions, in order to preserve protein complexes from dissociation (Drane et al.). After elution by TEV protease, the chromatin was purified onto FLAG-affinity beads, and subsequently eluted by FLAG peptide competition. Hdac2-HF bound chromatin was first purified onto FLAG-affinity beads, eluted by FLAG peptide competition, and then purified onto HA-affinity beads. Elution was then performed by HA peptide competition. Aliquots of the eluted fractions were analyzed onto an SDS polyacrylamide gel, and the presence of tagged proteins revealed by Western blot. This analysis revealed high levels of enrichment of the tagged proteins in purified chromatin fractions. DNA was prepared from tandem-affinity purified chromatin fractions and processed for high throughput sequencing onto an Illumina genome analyzer.

We used SICER software (Zang et al. 2009) to identify genomic regions significantly enriched for Chd4 and Hdac2. We identified 14030 Chd4-bound regions from ChIPseq experiments performed with chromatin fixed by formaldehyde alone, and 15183 Chd4-bound regions with chromatin fixed by the combination of DSG and formaldehyde. 6678 regions (about 30%) were identified in both experimental conditions. In total, we obtained 22535 independent genomic regions bound by Chd4. We can thus estimate that Chd4 is bound to 1-2 % of the mouse ESC genome. A global comparison of Chd4-bound regions using Seqminer software (Ye et al., Tora's laboratory, manuscript in preparation) revealed a high level of overlap between DNA regions bound by Chd4 in either fixation conditions (Figure 1A). Examination of sequencing reads repartition onto individual genes reveals that Chd4 binding to DNA is highly similar following either fixation method (Figure S2). However, peak amplitude is sensitive to the fixation method, which causes many regions to be differentially detected by SICER at high stringency (FDR 10^{-9}). Therefore, we concluded that the fixation protocol does not strongly influence the identification of Chd4 target genomic sequences. Using the GPAT software (Krebs et al. 2008), we could ascribe the 22535 Chd4 regions of enrichment (or peaks) to 7035 genes.

We detected 13739 regions bound by Hdac2. In contrast to Chd4, the nature of the fixation agents (formaldehyde alone or combination of DSG and formaldehyde) considerably affect Hdac2 peak detection onto the mouse ESC genome (Figure 1B). 12038 peaks were detected using the DSG-formaldehyde condition, and 1375 with formaldehyde alone. Only 326 peaks were shared in ChIPseq data obtained from these two experimental conditions. This divergence is likely due to the fact that formaldehyde mainly induces protein-DNA crosslinks, whereas DSG causes protein-protein crosslink (Nowak et al. 2005). In addition, DSG can crosslink 2 proteins located at a larger distance from each other than formaldehyde. Hdac2, which covalently modify histone lysine residues, is not expected to be in contact with DNA. Therefore, we expect Hdac2 binding onto chromatin to be well stabilized by DSG, but poorly by formaldehyde alone. In contrast, Chd4, which is the nucleosome remodelling subunit of NuRD, is expected to be in contact with DNA during the remodelling process, when histone-DNA interactions are altered. This property would explain why formaldehyde alone is a good fixative agent for Chd4, but not Hdac2.

Chd4 has both Hdac2-associated and Hdac2-independent binding sites onto the mouse ES cell genome

A global comparison of Chd4 and Hdac2 binding onto the ES cell genome is shown in Figure 1. The majority of Chd4-bound genomic regions are also enriched for Hdac2 (Figure 1A). The width and pattern of Chd4 and Hdac2 binding is variable, showing on one side discrete binding regions of 500-1000 bp and, on the other side, wide regions of enrichment spanning up to 10 kb. In two-third of Chd4-bound regions, Hdac2 binding closely mirrors Chd4 fixation. However, the intensity of Chd4 and Hdac2 binding onto the genome is variable. A first category of regions shows strong enrichment for both Chd4 and Hdac2. A second, larger category displays a weaker enrichment for Hdac2, compared to Chd4, and a third category shows no significant enrichment for Hdac2 (Figure 1A). One third of Chd4 peaks were found in this last category. Hence, our data show that Chd4 is often present onto the genome independently of Hdac2.

Analysis of Hdac2 peaks detected by SICER shows that most genomic regions bound by Hdac2 are also bound by Chd4 (Figure 1B). However, only one fourth of the Hdac2-bound regions show a comparable binding intensity for both Hdac2 and Chd4. Indeed, Chd4 binding intensity is considerably weaker than the intensity observed for Hdac2, as shown by the heat map (Figure 1B), suggesting that Hdac2 and Chd4 are not present at stoichiometric proportions at most loci. In addition, we observed that 22% of Hdac2-bound regions are not associated with Chd4 (or at very low levels). This is expected from the knowledge that Hdac2 belongs to protein complexes different from NuRD, such as the Sin3 containing complexes (Laherty et al. 1997).

Among the 22535 Chd4 and 13739 Hdac2-bound regions, 6030 were found to be bound by the two factors, using a high stringency for peak detection (SICER, FDR 10^{-9}). We refer to these regions as

loci bound by NuRD. These NuRD-bound regions are in average equally enriched for both Chd4 and Hdac2 (Figure 2). The NuRD binding pattern is variable: a minority of regions display sharp peaks of enrichment (1 kb or less), and a majority of loci show broader regions of enrichment, sometimes larger than 10 kb.

We further analyzed Chd4 and Hdac2 bound genomic regions in comparison with histone methylation marks and RNA polymerase II ChIPseq data, to integrate the binding of these enzymes with transcription genomic landmarks. 26% of Chd4 and 38% of Hdac2 binding regions were found at less than 5 kb from the transcription start site (TSS) of genes, as revealed by the presence of RNA polymerase II and histone H3 trimethylated at lysine 4 (H3K4me3) (Figure S3). We also compared the binding of Chd4 and Hdac2 with the binding of Ezh2 and Suz12, which belong to the Polycomb Group Complex 2 (PRC2) that catalyze the repressive H3K27me3 histone modification marks. This analysis revealed a low proportion of common target sequences between NuRD subunits and PRC2 (Figure S4).

Recent data showed that many developmental genes that remain silent in ES cell are marked by the association of H3K27me3 and H3K4me3 modifications at their promoter (Bernstein et al. 2006). We analyzed the proportion of genomic loci bound by Chd4 and Hdac2 that carry this bivalent marking at their promoter: we found that 4% of Chd4 and 3% of Hdac2 binding loci correspond to promoters carrying bivalent H3K27me3 and H3K4me3 modifications (data not shown). Hence, we concluded that there is no particular relationship between NuRD binding and the presence of this bivalent histone mark at promoters.

We also analyzed Chd4 and Hdac2 binding in relationship to gene expression levels. We compared Chd4 and Hdac2 binding to the pattern of RNA polymerase II, H3K4me3 and H3K36me3, at the level of the 2000 genes that are the most expressed in ESC, which were classified according to mRNA expression level (Figure S5). While the enrichment in RNA polymerase II and H3K36me3 are proportional to gene transcription levels, no specific change in the binding pattern of Hdac2 and Chd4 could be detected across the 2000 tested genes. A similar analysis performed with the 10 000 most expressed genes showed a similar absence of correlation between Chd4 and Hdac2 binding and gene expression levels (data not shown). In conclusion, our data show that Chd4 is often, but not always, bound in association with Hdac2, and that Chd4/Hdac2 binding to the genome do not correlate with gene silencing.

Chd4 and Hdac2 are extensively associated with pluripotency transcription factors onto the ESC genome

A series of independent studies have reported the association, in soluble protein complexes, of NuRD with Oct4, Nanog as well as several other transcription factors involved in the maintenance of ESC regulatory networks such as Esrrb, Sall4 and Tcfcp2l1 (van den Berg et al.). These studies suggested that NuRD might be recruited onto chromatin by these TFs, and play a role in the regulation of

pluripotency transcription networks. To test this hypothesis we analyzed the genomic target regions of Chd4 and Hdac2 in comparison with the binding sites previously identified for Oct4, Nanog, Sox2, Esrrb, Tcfcp2l1, and other factors involved in the maintenance of ESC pluripotency and proliferation. Strikingly, our data comparison revealed that 75% of Oct4 genomic binding loci are also bound by Chd4 (Figure 3A).

Large scale analysis of pluripotency factor binding onto the 6030 NuRD-bound regions revealed a very strong association with Nanog: more than 90% of NuRD-bound regions are also bound by Nanog (Figure 4A). Oct4, Sox2 and Klf4 were also often present onto NuRD binding loci. A lower proportion of NuRD enriched regions were also bound by N-Myc and C-Myc (Figure 4A). Finally, NuRD binding regions were also frequently bound by Esrrb, Tcfcp2l1, and E2f1 (Figure 4B).

We analyzed in detail the binding of tagged Chd4 and Hdac2 onto the Nanog locus, whose transcriptional regulation was studied in a previous study (Levasseur et al. 2008). DNaseI hypersensitive sites (HS), which are strong indicators of cis-acting DNA regulatory elements, were mapped within 160 kb around the Nanog locus. We compared the position of HSs with the binding of Chd4 and Hdac2, as well as several transcription factors that control ESC transcriptional networks (Figure 5). Chd4 and Hdac2 bind several regions upstream and downstream of Nanog. Strikingly, Chd4 and Hdac2 bind all previously identified HSs loci. At the Nanog gene, Chd4 and Hdac2 were strongly enriched at the levels of the -5 kb (relative to TSS), proximal promoter, and +9kb regions, which all correspond to previously identified HSs (Levasseur et al. 2008) (Figure 5). Furthermore, Nanog -5 kb and promoter proximal regions bind, in addition to NuRD, combinations of Oct4, Nanog, Esrrb and Tcfcp2l1 TFs (Figure 5). A major Chd4 and Hdac2 binding region is located at the level of several HSs located in the -48 to -37 kb upstream region of Nanog. These regions also contains strong binding sites for Esrrb and Tcfcp2l1 TFs (Figure 11B).

In conclusion, Chd4 and Hdac2 have several strong binding regions within the Nanog locus, that coincide with binding sites for Oct4, Nanog, Esrrb and Tcfcp2l1. These data strongly suggest that these four TFs, which were shown in other studies to interact in solution with NuRD, are responsible for the recruitment of NuRD onto chromatin. We noted that the regions of enrichment for Chd4 and Hdac2 are always wider than the regions bound by the TFs (Figures 5, 6 and S9-S13), which suggests that once recruited by a DNA-bound TF, NuRD can move onto adjacent nucleosomes.

A Chd4 hypomorphic mutation impairs the maintenance of ESC self renewal and pluripotency

We generated mouse lines carrying loss of function alleles for Chd4 and Hdac2. Hdac2-deficient mice were obtained by injecting a gene-trap embryonic stem cell (ES) line into blastocysts. Southern blot and 5'-rapid amplification of cDNA ends (RACE) revealed that the pPT1- β geo gene trap vector, which contains a splice acceptor site, the LacZ coding sequence, and the neomycin resistance gene, is inserted downstream of exon 10 (Figure S6). The resulting Hdac2-lacZ gene is predicted to encode a

fusion protein between HDAC2 amino acids 1 to 244 and beta-galactosidase. This fusion protein is missing the C terminal half of HDAC2 (total length of 488 amino acids for the wild-type protein). The HDAC2-betaGal fusion protein is expected to be inactive for deacetylase activity, because the deleted C-terminal domain includes a conserved tyrosine residue essential for deacetylase activity (Finnin et al. 1999). Heterozygous mice carrying the *Hdac2* mutant allele were obtained at the expected frequency, and were indistinguishable from their wild-type littermates. We derived *Hdac2*^{-/-} ES cell line from blastocysts, obtained from *Hdac2*^{+/-} intercrosses. We obtained a single *Hdac2*^{-/-} ESC line out of 25 derived from blastocysts, which is well below the expected 25% frequency. Once established, however, this *Hdac2* deficient ESC line behaved normally and was morphologically identical to the control 46C ES cell line. This suggests that loss of *Hdac2* function is detrimental to the derivation of ESC lines, but not to the maintenance of pluripotency and proliferation genetic programs.

We studied the function of *Chd4* by generating two different loss of function alleles: a null *Chd4* allele (*Chd4*^{KO}), which is deleted of exons 8 and 9, and an hypomorphic allele (*Chd4*^{Hyp}), that carry two amino-acid substitutions (I505V and R614C) in the first chromodomain of *Chd4* (Figure S6). Embryos that are homozygous for the *Chd4* null allele die early in development, between embryonic day 4 and 7, showing that *Chd4* has essential functions early in development (data not shown). In contrast, embryos homozygous for the hypomorphic allele survive until embryonic day 13. At this stage, homozygous *Chd4*^{Hyp/Hyp} embryos display a major failure in neural tube closure (Figure S7). Hence, the *Chd4*^{Hyp} allele still carries part of the *Chd4* wild-type function. We collected blastocysts from intercrosses between *Chd4*^{KO/+} animals and between *Chd4*^{KO/+} and *Chd4*^{Hyp/+} mice. We successfully derived three independent ES cell lines carrying one null and one hypomorphic *Chd4* allele (*Chd4*^{KO/Hyp} genotype), but none homozygous for the null allele. We further tried to derived ES cells carrying two null *Chd4* alleles by transfecting *Chd4*^{KO/Hyp} cells with the Cre recombinase (the hypomorphic allele has loxP sites inserted in intron 7 and 9). We could never obtain *Chd4*^{KO/KO} ES cells, despite repeated attempts. We concluded that *Chd4* function is essential for ESC survival, and that a single copy of the hypomorphic *Chd4* allele is sufficient for ESC survival and proliferation. Western blot analysis revealed that the level of *Chd4* protein was similar in *Chd4*^{KO/Hyp} and wild-type ES cells (data not shown). However, all three *Chd4*^{KO/Hyp} ESC lines behaved differently than wild type in culture. While *Chd4*^{KO/Hyp} ESC have a normal morphology in culture when passaged every day, they show signs of differentiation two days after the last passage (Figure S7). All three independently derived *Chd4*^{KO/Hyp} ESCs are prone to spontaneous differentiation despite the presence of feeder cells and leukemia inhibitory factor (LIF) in culture medium. In contrast, *Hdac2*^{-/-} ESCs show both robust growth and ES phenotype, similarly to our reference 46C ES cell line (Figure S7). Transcriptomic analysis of spontaneously differentiated *Chd4*^{KO/Hyp} ESC showed that, 96 hrs following initiation of differentiation, these cells express 10-fold and 20-fold less *Nanog* and *Oct4* mRNA, respectively, than undifferentiated cells. In addition, *Chd4*^{KO/Hyp} spontaneously differentiating cells express numerous

genetic programs normally silent in ES cells and observed in post-gastrulating embryos. For instance Homeobox genes of the *HoxA* and *HoxD* clusters, which are normally expressed in limb buds and in posterior regions of the developing mouse embryo during late gastrulation and organogenesis, are activated in *Chd4*^{KO/Hyp} differentiating ESCs. We concluded that ESC carrying a single *Chd4* hypomorphic allele are less robust in the maintenance of their pluripotency program than wild-type ESCs.

To analyze the potential impact of I505V and R614C mutations, a structural model of CHD4 tandem chromo domains was built by homology with human and *S. cerevisiae* CHD1 structures (Flanagan et al. 2005; Flanagan et al. 2007) (Figure 13A and B). This study (see supplementary results) revealed that the mutations are unlikely deleterious to the structural integrity of the chromodomain, but since they lay in the most conserved face of mouse *Chd4*, an important function of the tandem chromo domains may be impaired. Interestingly, the detailed analysis of Mi-2 chromo domains in *Drosophila melanogaster* (Bouazoune et al. 2002) showed that the domain strongly interacts with DNA but not with methylated histone motifs (CHDM_DROME in Figure 13A, 40 % identical to CHD4_MOUSE). The second chromo domain was found a major contributor to the DNA binding property, fully consistent with both conservation and electrostatic analyses (Figure S8). The conserved surface exposing both positively charged and aromatic residues suggests that this property is shared among all members of the CHD4 functional group shown in Figure S8. Accordingly, a reasonable hypothesis is that I505V/R614C double mutation encoded by the *Chd4* hypomorphic allele impairs the DNA binding properties of the tandem chromo domain in mouse *Chd4*.

Chd4 and Hdac2 are involved in both transcriptional activation and repression

Using the RNA-Seq methodology (Mortazavi et al. 2008), we analyzed the transcriptomes of two reference wild-type ESC lines, as well as *Hdac2* and *Chd4*-deficient ES cells. We used for this purpose RNA prepared from *Chd4*^{Hyp/KO} and *Hdac2*^{-/-} cells. We first compared the transcriptome of two wild-type reference ESC lines against the transcriptome of ESC carrying a single hypomorphic *Chd4* allele. We identified 224 differentially expressed genes (Table 2). Of these, 156 (70%) were overexpressed, and 68 (30%) repressed in *Chd4* hypomorphic ESCs. This relatively modest change in transcription probably reflects the fact that *Chd4* loss of function is only partial in these cells. *Hdac2* loss of function resulted in the deregulation of 890 genes, 561 (63%) and 329 (37%) of which were overexpressed and repressed in homozygous mutant ESC, respectively. For both *Chd4* and *Hdac2* deficient cells, the number of overexpressed genes exceeds the number of repressed genes, suggesting a more frequent role of these proteins in gene repression. However, the number of genes that are repressed following loss of *Chd4* or *Hdac2* function is surprisingly high, considering the common view that NuRD is a protein complex mainly involved in gene repression. Among the 890 genes deregulated in *Hdac2* KO cells, 106 were also deregulated in *Chd4* hypomorphic ESCs. 60 genes were

overexpressed in *Chd4* hypomorphic and *Hdac2* deficient ESC, whereas 37 genes were repressed in both cell lines. 9 genes had opposite transcriptional changes in response to Chd4 and Hdac2 loss of function. Most of the genes deregulated upon Chd4 and Hdac2 loss of function were bound by Chd4 and Hdac2, suggesting a direct role of these proteins in transcription regulation (Data not shown).

Figures S9-S13 show examples of genes that are deregulated upon loss of Chd4 and/or Hdac2 functions. In these figures, we analyzed the genomic binding profiles of Chd4, Hdac2, and several pluripotency-related TFs (data from Chen et al. 2008) in comparison with RNA-Seq data obtained from control, Chd4 hypomorphic and Hdac2 KO ESCs.

The *S100a6*, *Trf* and *Gjb5* genes are induced in *Chd4* and *Hdac2* mutant ESCs (Figure S9). In contrast, *Zfp42* (*Rex1*), *Trh* and *Fxyd6* are downregulated in both mutant ESCs (Figure S10). Many genes bound by both Chd4 and Hdac2 were however affected by the loss of only one of these two proteins. For instance, *Dnmt3L*, *Epas1* and *Chd4* genes were all overexpressed in *Chd4* hypomorphic ESCs, but not in *Hdac2* KO ESCs (Figure S11). Similarly, *Trim26* and *Pde9a* were up-regulated in *Hdac2* KO ESCs, but unaffected in *Chd4* mutant cells (Figure S12). Many genes, such as *Otx2*, *Mycn* (that encodes N-Myc) and *Fbxo15* (previously shown to be regulated by Nanog), were found to be severely down-regulated by *Hdac2* loss of function (Figure S13). Finally, a few genes were deregulated in the opposite direction in our mutant ESCs. For instance, *Lefty2* is markedly overexpressed in Chd4 hypomorphic ESCs, and strongly repressed in *Hdac2* deficient cells (Figure S12). These data suggest that NuRD role(s) in chromatin remodelling can affect transcription positively or negatively, depending on the genomic context. Furthermore, our data show that the two NuRD subunits, Hdac2 and Chd4, do not have same importance at every target gene. A careful analysis of Chd4 and Hdac2 profile onto the target genes described above also revealed partially, but not completely overlapping profiles, suggesting that Hdac2 and Chd4 might dissociate from each other during chromatin remodelling and/or transcription.

Chd4 and Hdac2 are enriched onto actively transcribed Oct4 target genes

Our data show that Oct4 is often bound to Chd4 and Hdac2 regulated genes. We next analyzed at a larger scale the repartition of Hdac2 and Chd4 enzymes onto Oct4 genomic binding regions. Using Seqminer software, we identified four and six patterns for Chd4 and Hdac2 binding onto Oct4-bound loci, respectively (Figure 3B and 3C). For Chd4, category 1 displays a high level of Chd4 enrichment and a wide repartition around Oct4 DNA binding site, whereas group 2 shows high levels of Chd4 associated with a sharper region of enrichment (Figure 3B). Finally, group 3 shows little or no Chd4 binding, and group IV a moderate enrichment for Chd4. We used the GPAT software to identify the genes associated to these Oct4 and Chd4-bound regions. We next compared the average rate of expression of genes from each category. This analysis revealed that genes bound by Oct4 and Chd4

tend to be more expressed than genes bound by Oct4 alone (Figure 3D). The difference between the four categories disappears in differentiating ESCs (Figure 3F).

Analysis of Hdac2 binding onto Oct4-bound loci revealed a strong correlation between the level of Hdac2 binding and the rate of expression of the corresponding genes (Figure 3E). Oct4-bound regions from category 5 display a high level of Hdac2 enrichment, associated with a wide repartition of the enzyme around Oct4 DNA binding sites. The genes that belong to this category 5 have the highest level of expression of all categories (Figure 3E). In contrast, genes from category 4, which show little or no Hdac2 binding around Oct4, have the lowest average level of expression of all categories. The difference in expression rate between the six categories is reduced in differentiating ESCs.

In conclusion, our data show that Chd4 and Hdac2 binding onto Oct4-bound regions correlates positively with active transcription in ESCs.

Chd4 and Hdac2 are required for the proper transcriptional regulation of key pluripotency-associated transcription factors

We described earlier how Chd4 and Hdac2 bind to several regulatory elements located within the *Nanog* locus, suggesting that these enzyme might play an important role in transcriptional regulation of this locus (Figure 5). Our RNA-Seq and microarray hybridization data both suggest that *Nanog* expression is reduced in Chd4 hypomorphic mutant ESCs (data not shown). We thus performed quantitative PCR (qPCR) experiment to compare *Nanog* mRNA levels in two reference ES cell lines (46C and AT-1), and in the three independently derived *Chd4*^{Hyp/KO} ESC lines. We carefully selected *Chd4*^{Hyp/KO} ESC cultures of normal growth and appearance, and excluded cultures displaying overt signs of differentiation. *Nanog* mRNA levels in *Chd4*^{Hyp/KO} ESC were comprised between 22% and 54% of *Nanog* level in 46C ESCs, showing that *Nanog* was severely downregulated in Chd4 hypomorphic mutant ESCs (Figure 6C).

We next analyzed in detail the binding of NuRD to the *Pou5f1* locus, which encodes Oct4. Both Chd4 and Hdac2 occupies a 20 kb genomic region located upstream of *Pou5f1* and extending until the TSS (Figure 6A). This genomic region also includes numerous binding sites for Oct4, *Nanog*, *Esrrb* and *Tcfcp2l1*. All these TFs interacts in solution with NuRD (van den Berg et al.), and are therefore very likely responsible for the massive recruitment of NuRD to the the *Pou5f1* regulatory regions. QPCR analyses of mRNA prepared from *Chd4*^{Hyp/KO} ESC revealed reproducible lower levels (between 60% and 80%) of Oct4 mRNA when compared to wild-type reference ESC lines (Figure 6B). In contrast to *Nanog* and *Pou5f1*, *Lefty2* was strongly overexpressed in all three *Chd4* hypomorphic ESC lines (Figure 6D). Note that *Lefty2* is overexpressed in *Chd4*^{Hyp/KO} ESC, but severely down-regulated once spontaneous differentiation has been initiated in *Chd4*^{Hyp/KO} cells (Table S1). This result further supports our visual observations that the three *Chd4*^{Hyp/KO} ESC lines used in these qPCR experiments

were not differentiating ESC, and that the lower amounts of *Nanog* and *Oct4* transcripts are not a mere consequence of a contamination of our ESC cultures with differentiating cells. We concluded from these experiments that *Chd4* has a critical role in the control of *Nanog* and *Oct4* mRNA levels in ESCs.

Other pluripotency-related genes were similarly deregulated in NuRD deficient cells. *FoxD3*, which is involved in the regulation of ESC self-renewal and pluripotency (Liu and Labosky 2008), was markedly downregulated in both *Chd4* and *Hdac2* mutant ESCs (Figure 7). In addition, our RNA-Seq and microarray data revealed that *Chd4* hypomorphic cells had lower levels of *Esrrb* and *Tcfcp2l1* mRNA than control cell lines. Both genes are heavily bound by *Chd4* and *Hdac2*, as revealed by our ChIPseq data, suggesting a direct role for these enzymes in transcriptional regulation (Figure 7). In contrast to *Chd4* hypomorphic ESCs, *Pou5f1* mRNA level was unaffected in *Hdac2*-deficient ESCs, whereas *Esrrb* was overexpressed in these cells (Figure 7). Hence, as previously shown for other genes, *Chd4* and *Hdac2* can have opposite transcriptional effects on pluripotency genes. Another example of such differential regulation by *Chd4* and *Hdac2* occurs at the *Mycn* locus, which encodes N-Myc, a TF involved in the control of ESC proliferation; *Mycn* was severely down-regulated upon *Hdac2* depletion, but only subtly affected in *Chd4* hypomorphic ESCs (Figure S13). Not every pluripotency-related genes bound by *Chd4* and *Hdac2* were affected by *Chd4* and *Hdac2* loss of function. For instance, the expression of *Sox2*, which is also bound by *Chd4* and *Hdac2*, was unaffected in both mutant cell lines (data not shown).

In conclusion, transcriptional alterations of pluripotency-related genes occurs widely in *Chd4* and *Hdac2* deficient cells.

- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., and Lander, E.S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): 315-326.
- Bouazoune, K., Mitterweger, A., Langst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. 2002. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *Embo J* **21**(10): 2430-2440.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., Loh, Y.H., Yeo, H.C., Yeo, Z.X., Narang, V., Govindarajan, K.R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.K., Clarke, N.D., Wei, C.L., and Ng, H.H. 2008. Integration

- of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**(6): 1106-1117.
- Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* **24**(12): 1253-1265.
- Finnin, M.S., Donigian, J.R., Cohen, A., Richon, V.M., Rifkind, R.A., Marks, P.A., Breslow, R., and Pavletich, N.P. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401**(6749): 188-193.
- Flanagan, J.F., Blus, B.J., Kim, D., Clines, K.L., Rastinejad, F., and Khorasanizadeh, S. 2007. Molecular implications of evolutionary differences in CHD double chromodomains. *J Mol Biol* **369**(2): 334-342.
- Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**(7071): 1181-1185.
- Krebs, A., Frontini, M., and Tora, L. 2008. GPAT: retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* **9**: 533.
- Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. 1997. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89**(3): 349-356.
- Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. 2008. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev* **22**(5): 575-580.
- Liu, P., Jenkins, N.A., and Copeland, N.G. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**(3): 476-484.
- Liu, Y. and Labosky, P.A. 2008. Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. *Stem Cells* **26**(10): 2475-2484.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**(7): 621-628.
- Nowak, D.E., Tian, B., and Brasier, A.R. 2005. Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* **39**(5): 715-725.
- Schmid, M., Durussel, T., and Laemmli, U.K. 2004. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**(1): 147-157.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.
- van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**(4): 369-381.
- Ying, Q.L. and Smith, A.G. 2003. Defined conditions for neural commitment and differentiation. *Methods Enzymol* **365**: 327-341.
- Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**(15): 1952-1958.

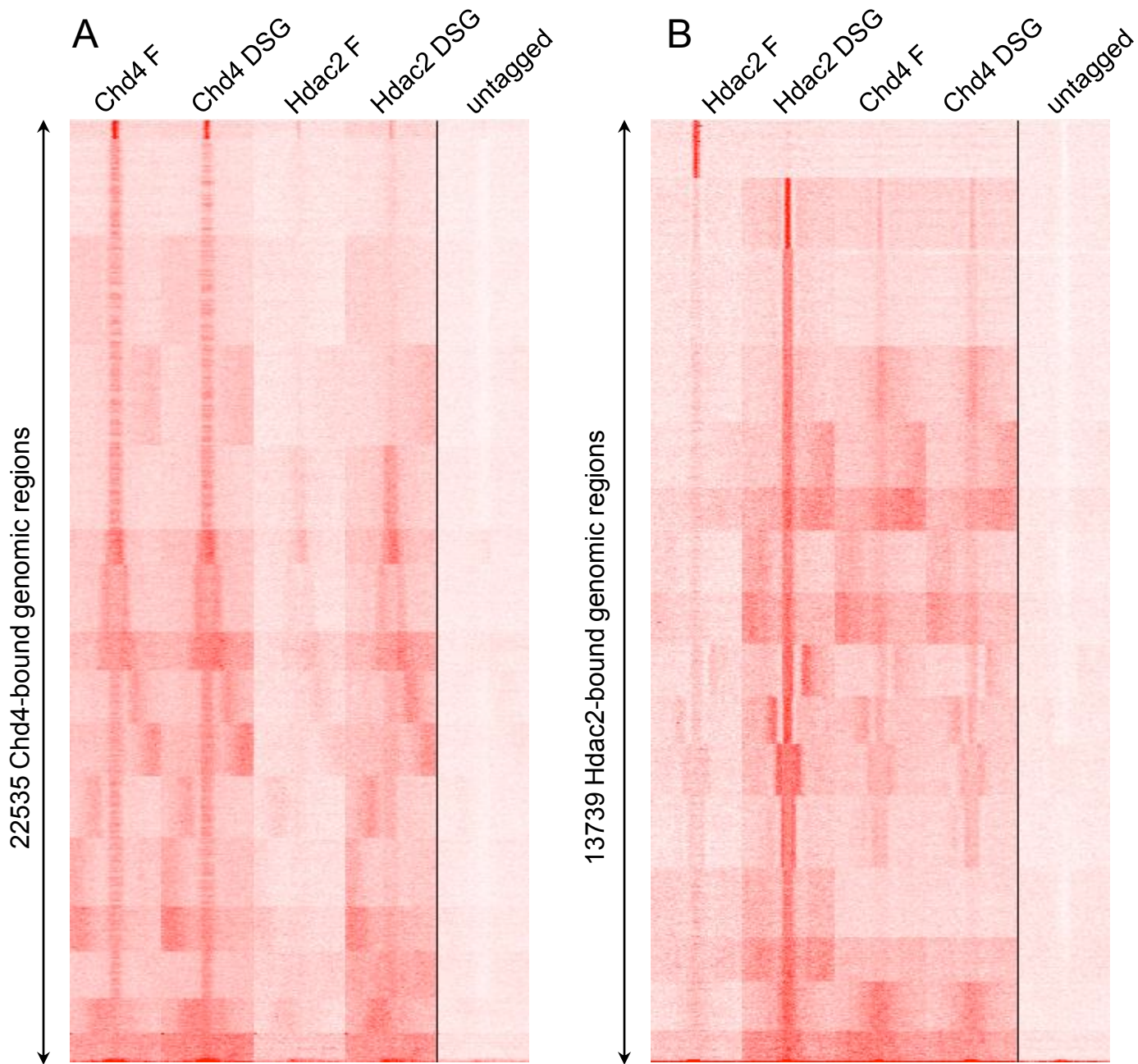


Figure 1

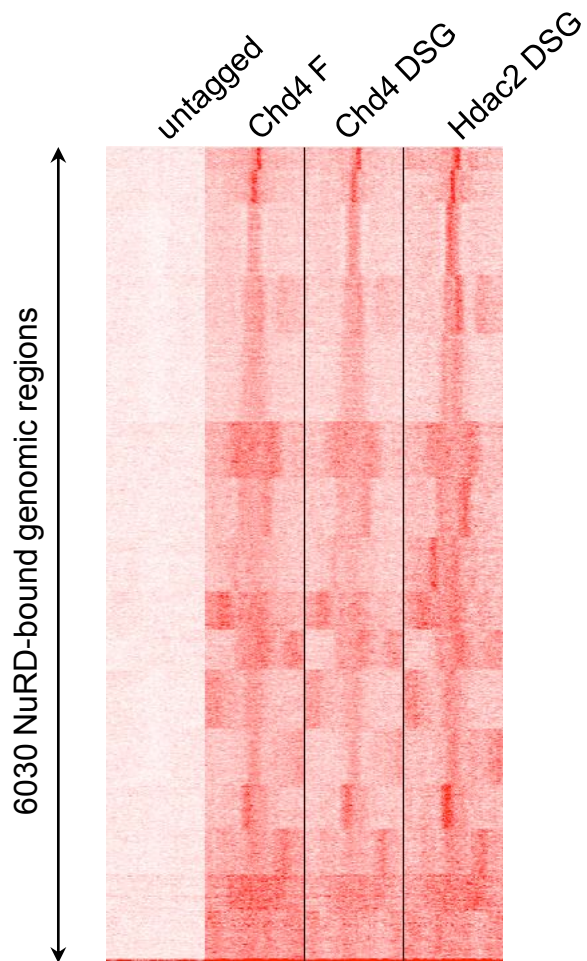


Figure 2

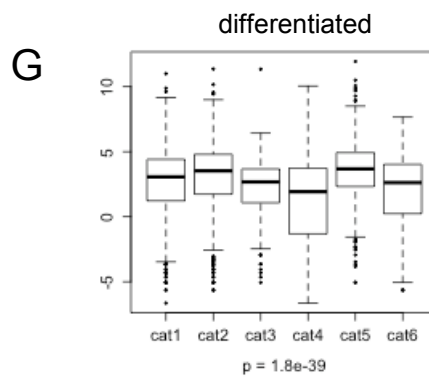
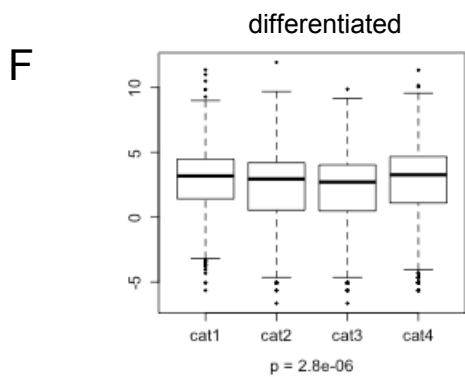
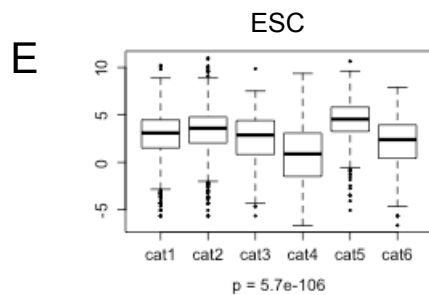
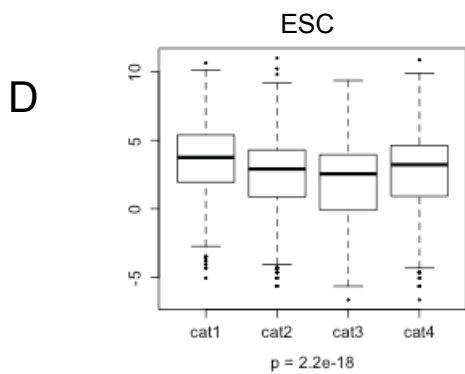
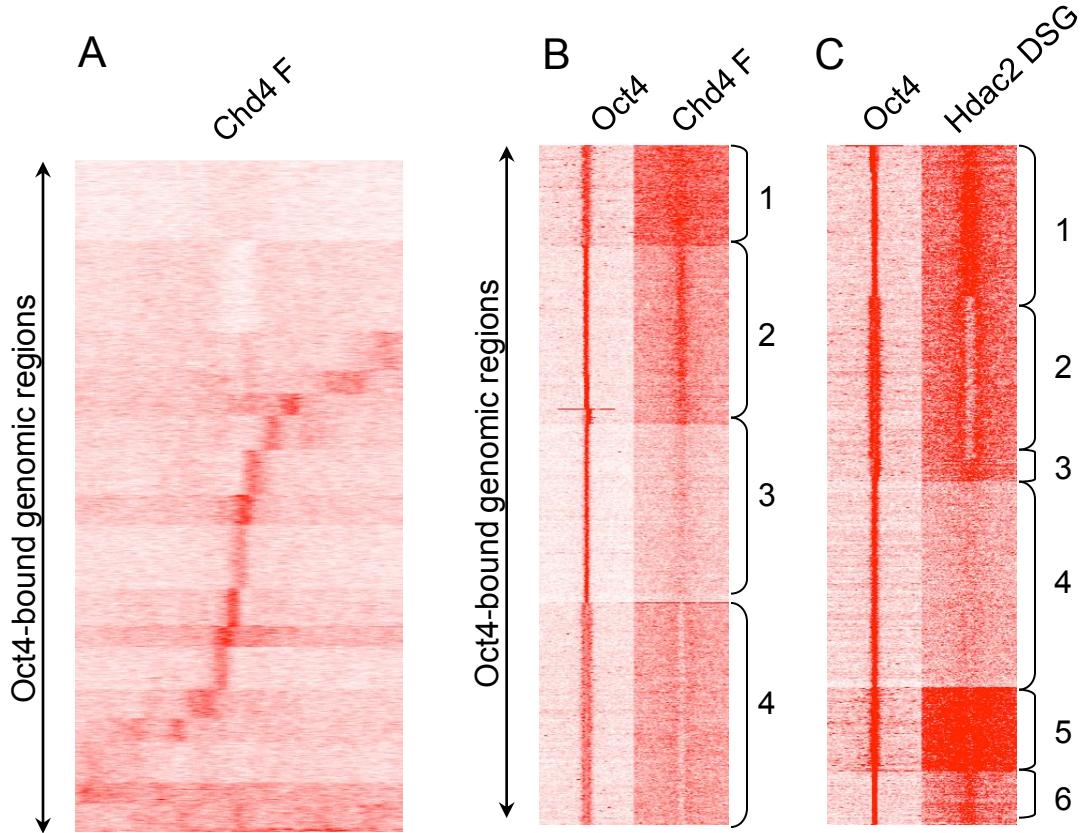


Figure 3

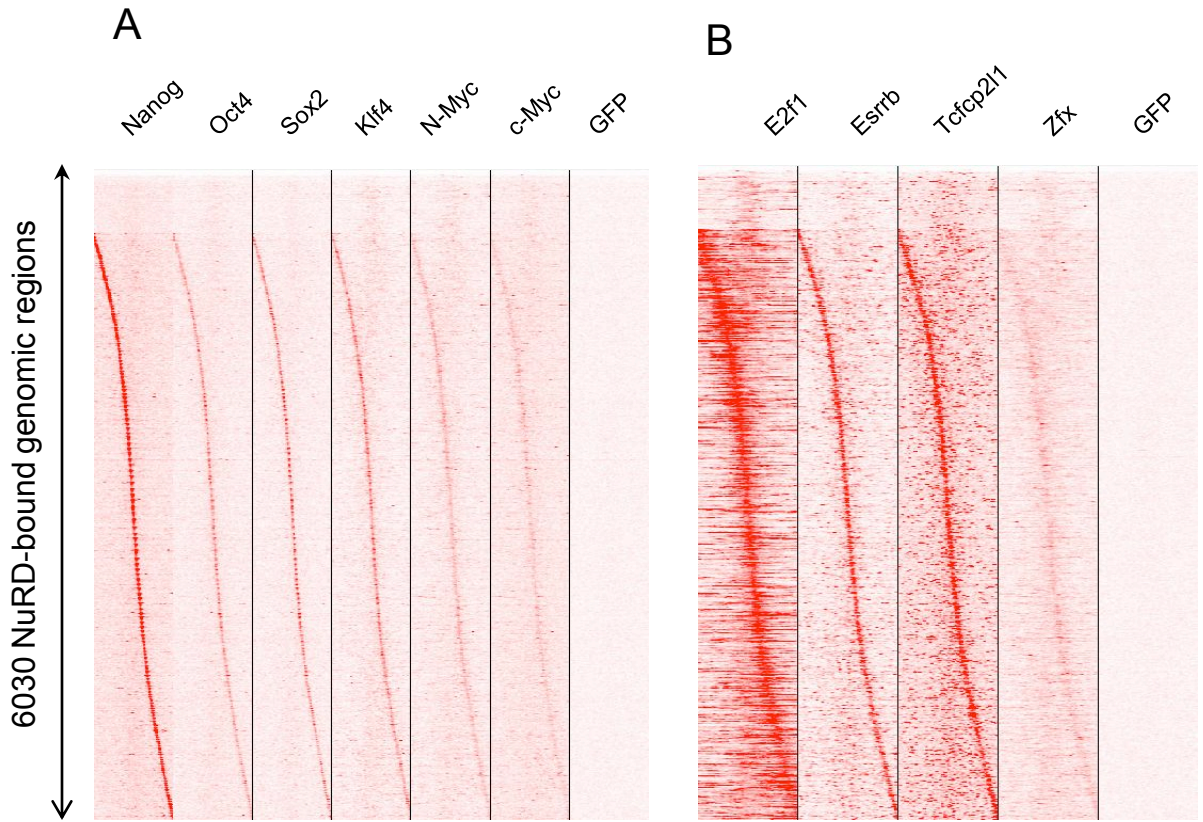


Figure 4

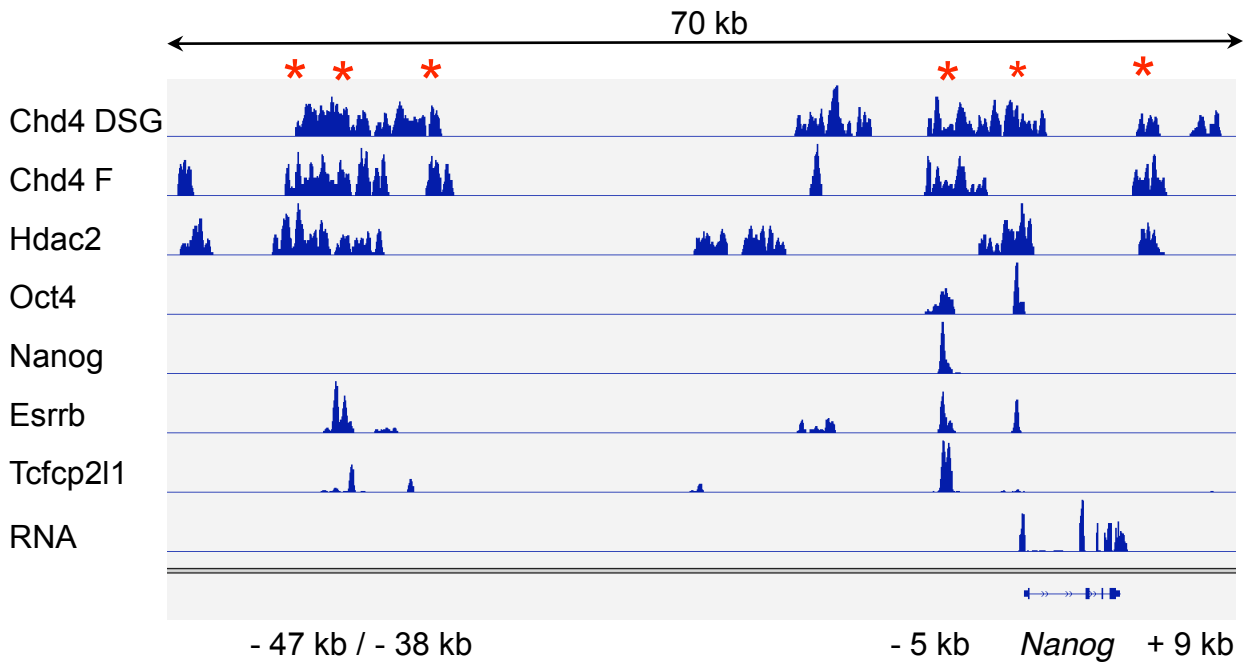


Figure 5

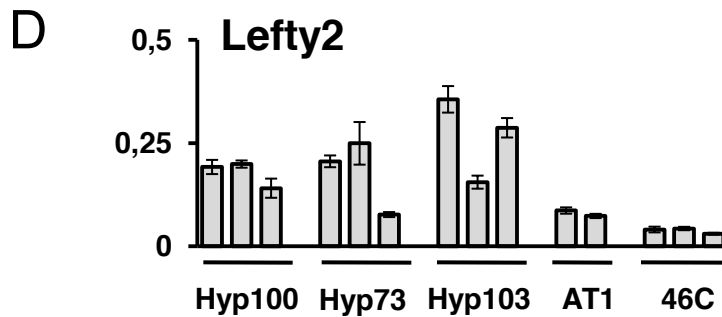
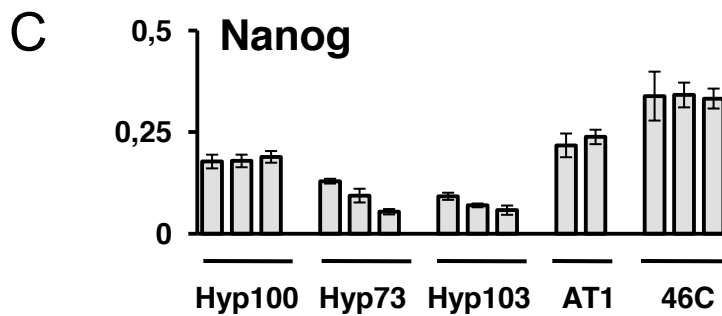
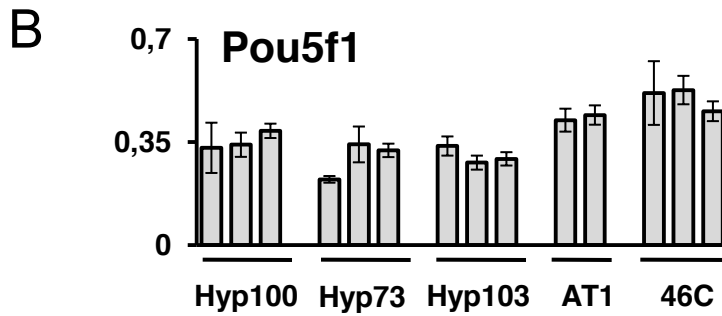
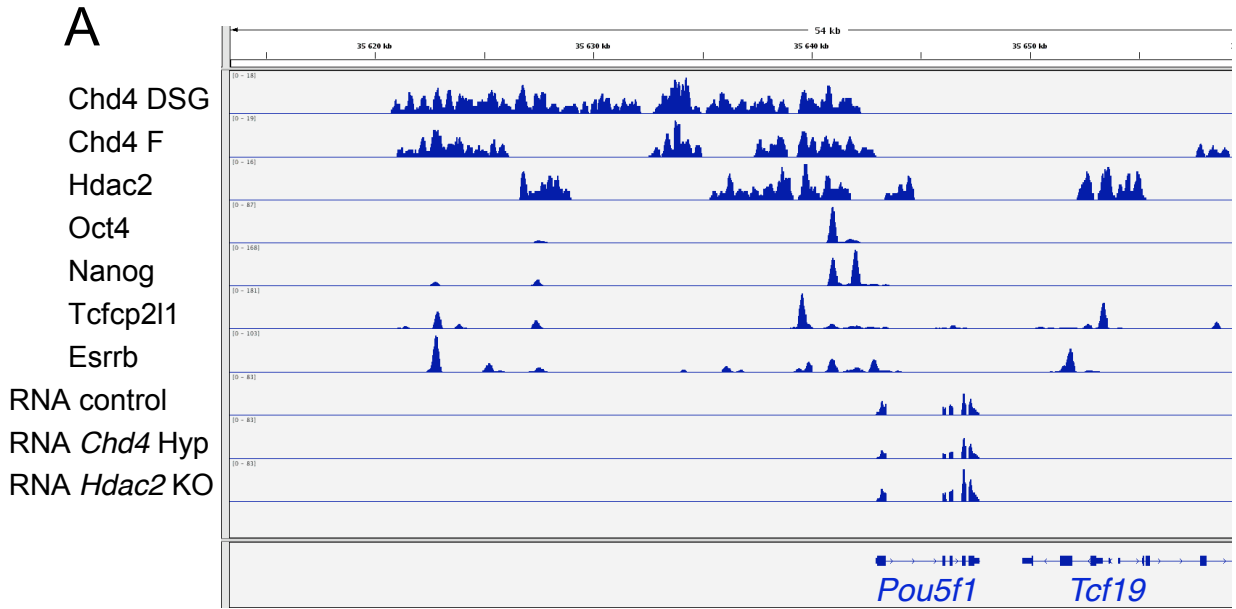
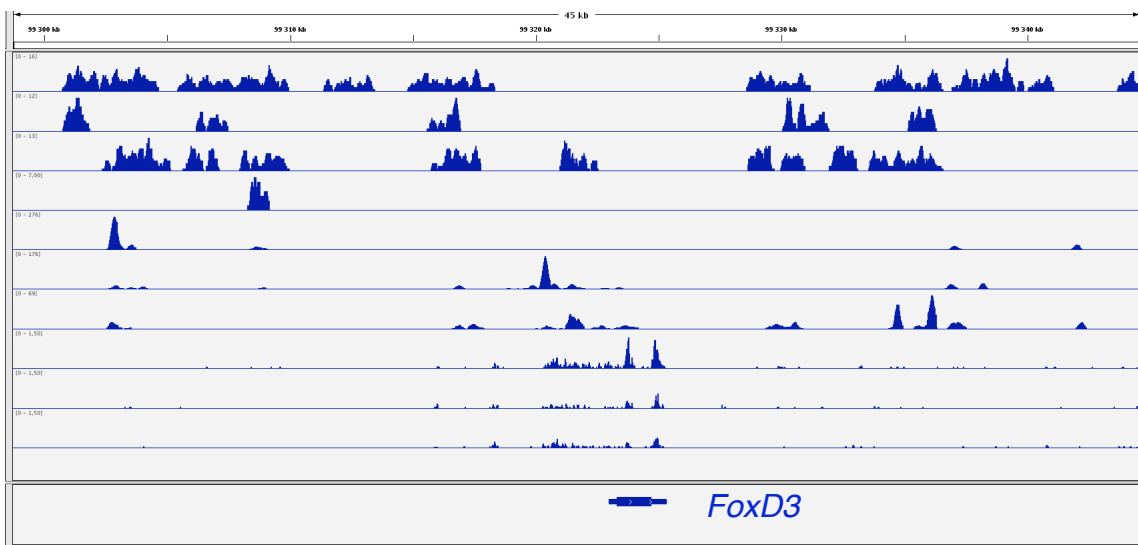
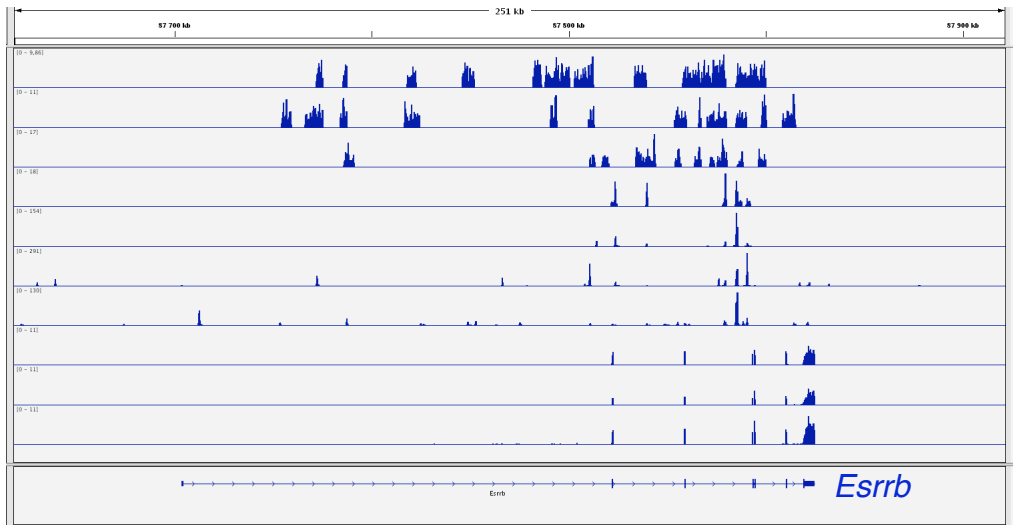


Figure 6

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

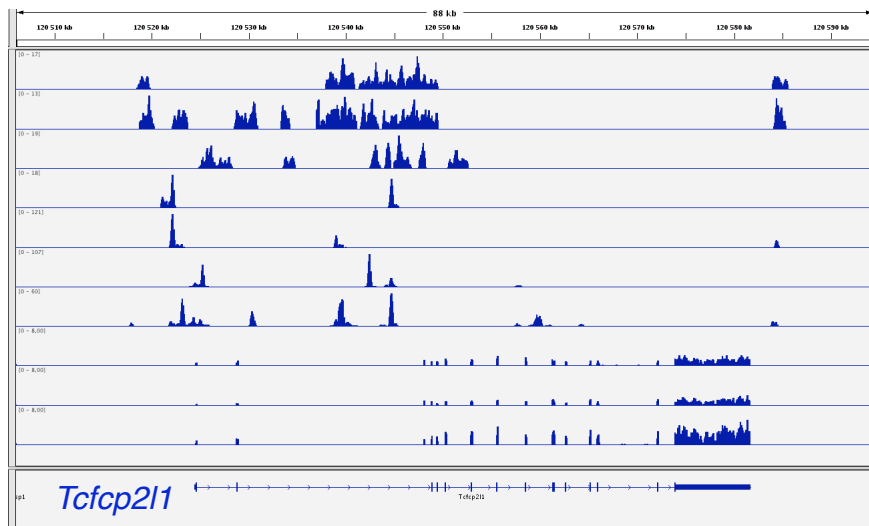


Figure 7

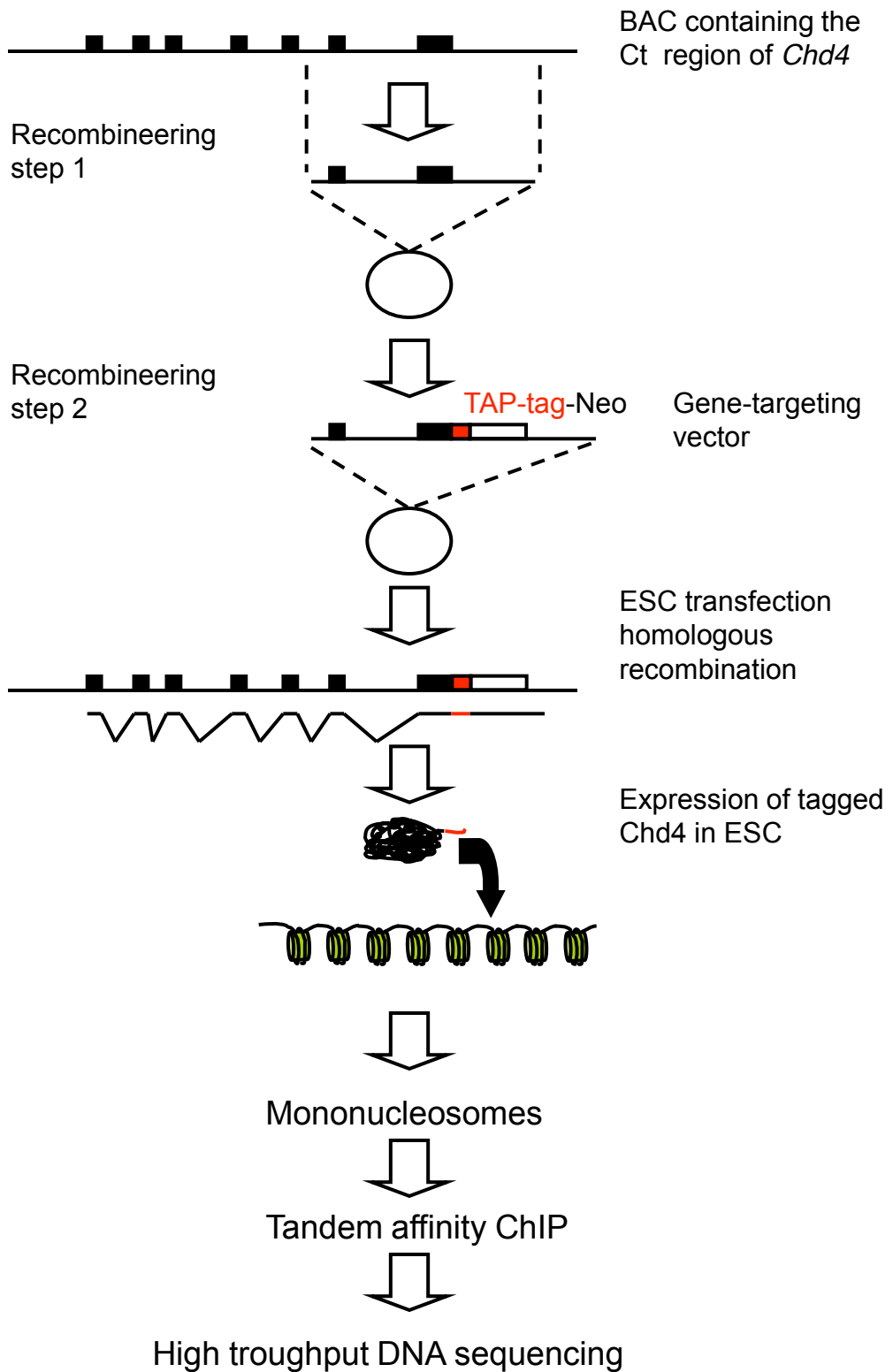


Figure S1

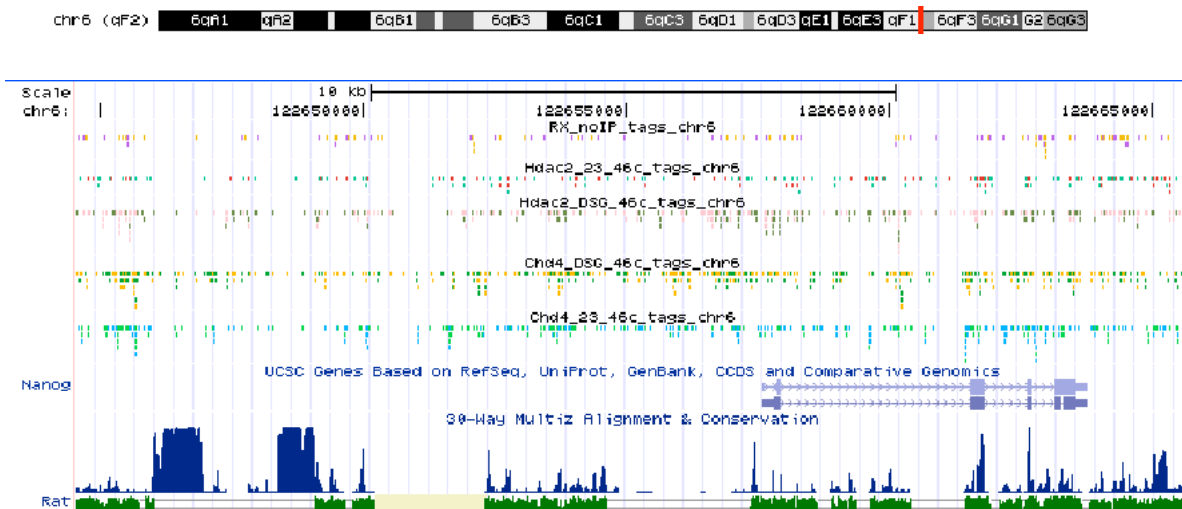


Figure S2

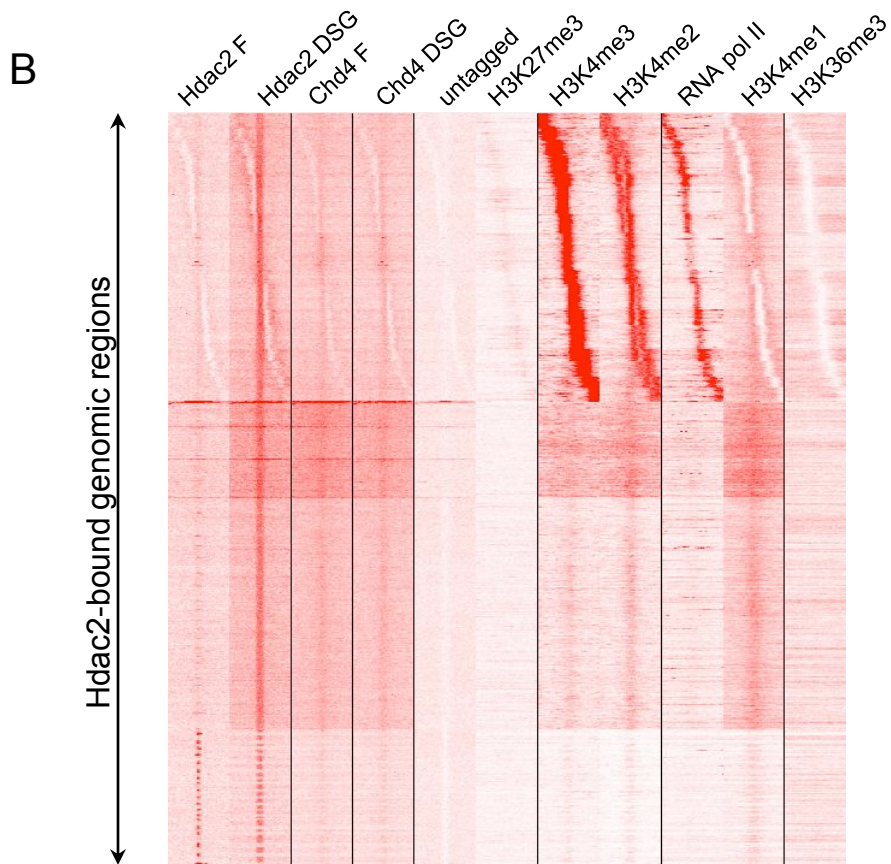
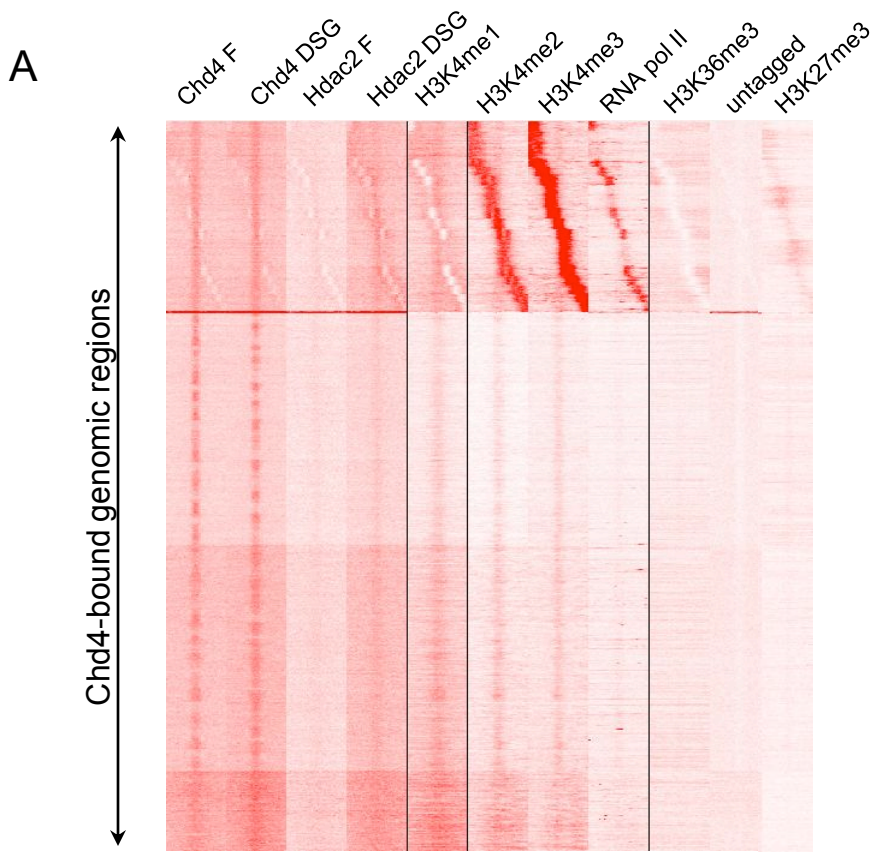


Figure S3

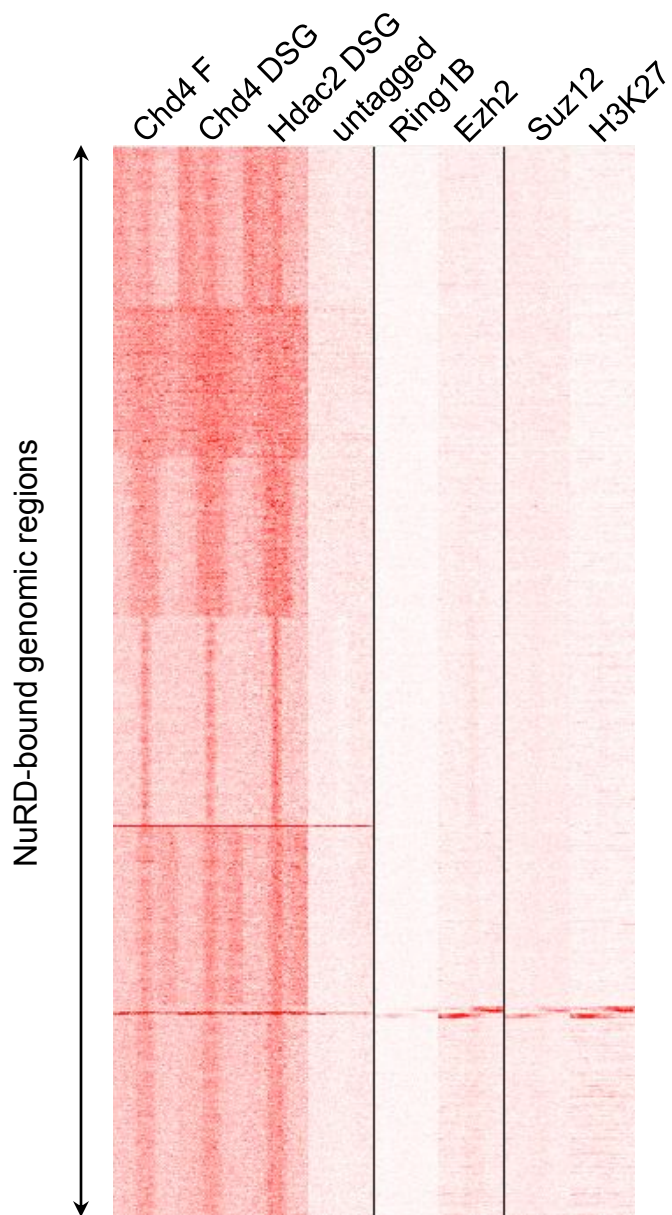


Figure S4

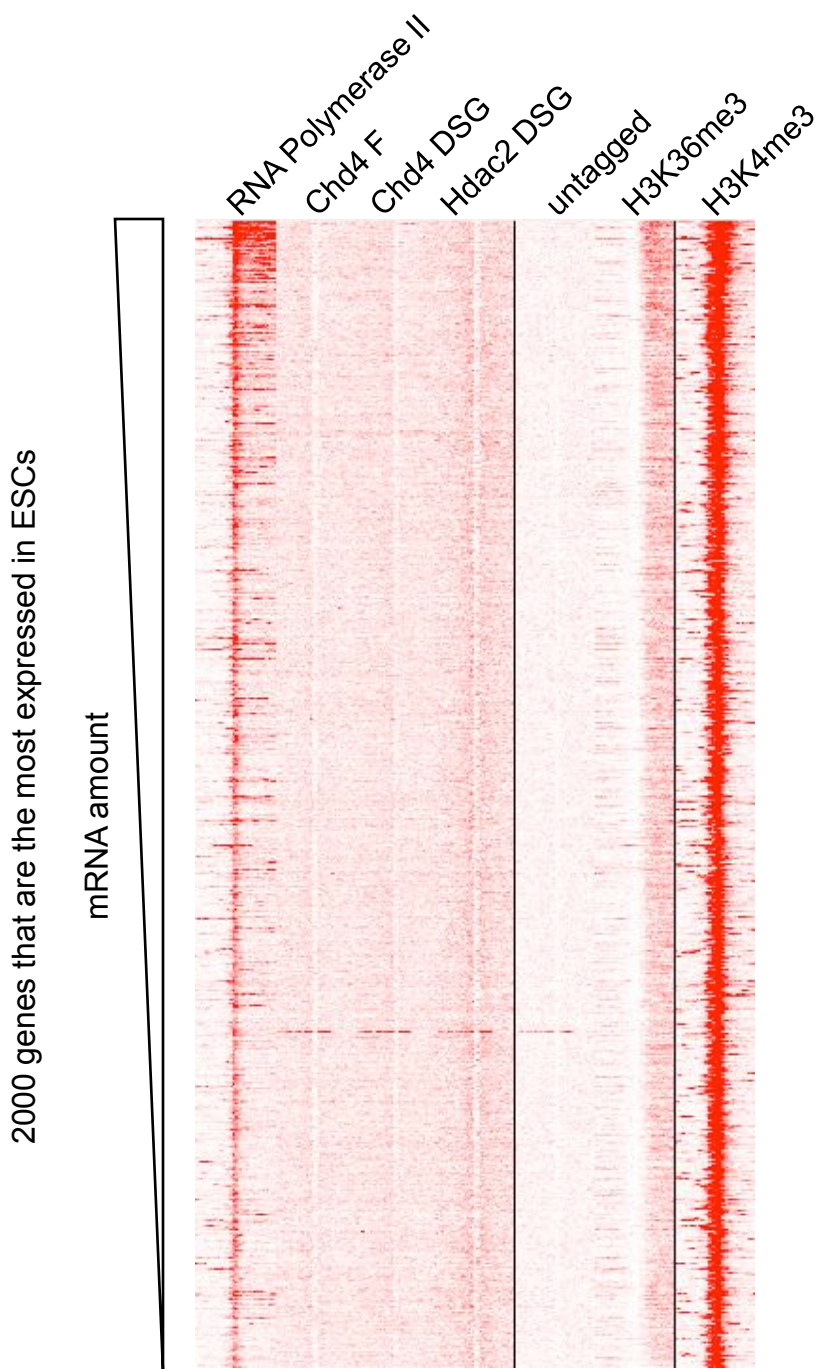
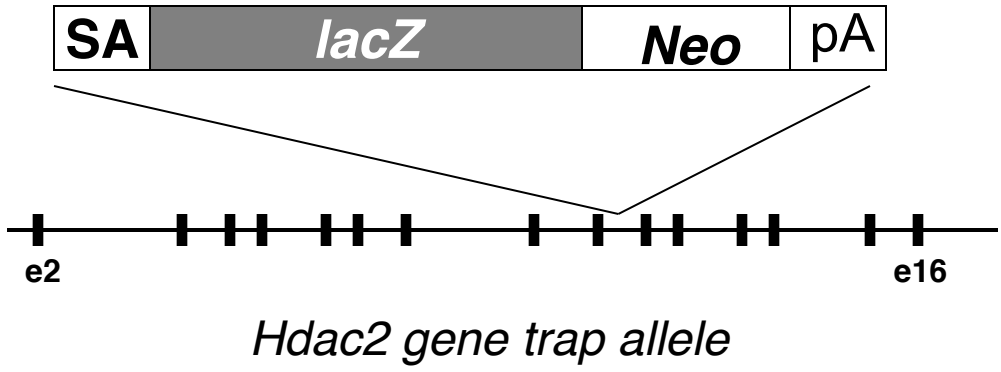


Figure S5

A



B

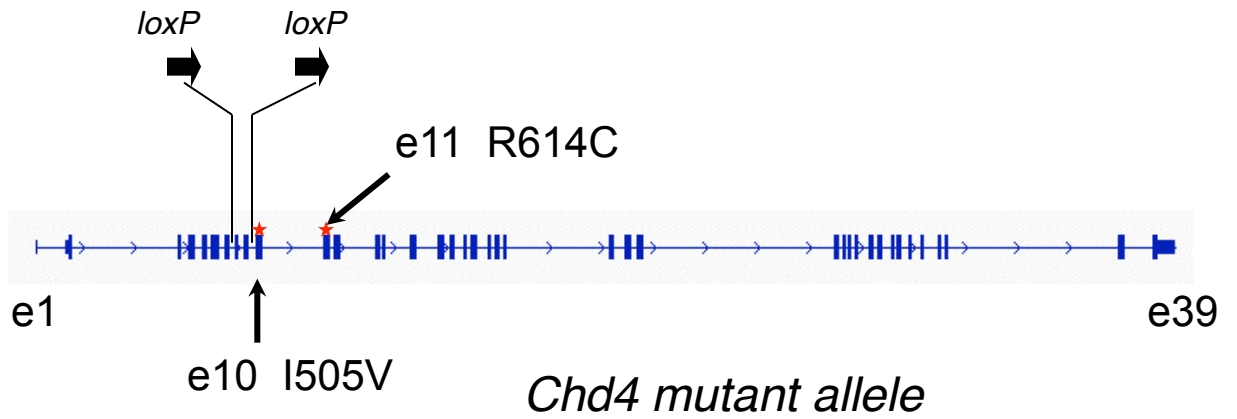


Figure S6

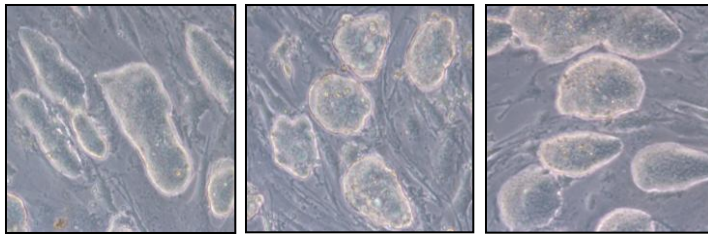
A



Chd4^{+/+}

Chd4^{Hyp/Hyp}

B



control

Chd4^{Hyp/KO}

Hdac2^{KO/KO}

Figure S7

A

g|51701319|sp|Q6PQ02|CHD4_MOUSE
 g|292622418|Danio_reio(fish)
 g|47274371|Tetraodon_nigrovirdis(fish)
 g|889112735|sp|C12873|CHD3_HUMAN
 g|51701343|sp|Q8TD10|CHD5_HUMAN
 g|165993|sp|Q22516|CHD3_CAEL
 g|115939069|Stromylococytus_purpuratus(sea_urchin)
 g|221120608|Hydra_magnapapillata
 g|156379809|Nematostella_vectensis(sea_anemone)
 g|260812970|Branchiostoma_floridae(amphioxus)
 g|198432553|Ciona_intestinalis(sea_squirt)
 g|292606963|Schmidtea_mediterranea(planarian)
 g|242000110|xodex_scapularis(tick)
 g|170013510|Tribolium_castaneum(red_flour_beetle)
 g|125089877|sp|Q16102|CHD3_DROME
 g|13124018|sp|Q97159|CHD4_DROME
 g|242023690|Pedicularis_humana(ouse)
 g|156545320|Nasonia_vitripennis(wasp)
 g|158300661|Anopheles_gambiae(mosquito)

```

KGGVORILLIWRWGPPSHTFVFRPPDADNTRSPKPLEGRERQFFVWQGMSWHSWSELEQLLHCQVFRNORNDMDEPPSGDFGGDEEK
KGGVORVLLWRWGEAPPMPVFRPADLPAEADPPPMIGRAEERFFVWGNKSWHSWSELEQLLNCQVFRNORNDMEPPNLEMGAEQDE
KGGVORILLWRWGEPPDRIFVFRPADLPAEADPPPMIGRAEERFFVWGNKSWHSWSELEQLLNCQVFRNORNDMEPPSLDYGSGGDD
KGGVORILLWRWGEPPRYAVFRPADLPAEADPPPMIGRAEERFFVWGNKSWHSWSELEQLLNCQVFRNORNDMEPPPLDYGSGGDD
KGGVORILLWRWTEPPAPFMVGLRGGDVEERSPKPLEGRERQFFVWQWAGLSWHSWSELEQLLNCQVFRNORNDMEPPRFDYSGGDED
AORIEKILLWRWKEISYPERLECKEAEASKPPPKRMEPRERFFVWQWYLAWQGEWLESLMDVDFALVLRMWRVSENPPIFEEST...
KNRPEKCLSWRTELVYPPMTEEELRELEKPKELANRERELFKWYMSWHSWSELEQLLNCQVFRNORNDMEPPPEVDDGSGQEDL
AGRVORILLWRWNTPEMPPEEVEEVEEAEKLRQKPKRERFFCFVGMSEFWHSWSELEQMFVHSLHRNFRKNDMEPPFDDEEKDTE
PGKVERILLWRWFLTHPTKEEII...AAAGKDEQYKVFVYANMSWNGHWIDELQLEKHSILRFFCRKNDMEPPMEEGTVEEG
RGRVERILLWRWVFTDPAQDITLADLNMTCVILFVQVPRTRQFLVWGRMSWHDGWTLEQLEIHPMFRFAFRKNDMEPPPDDDDDDSD
KGGCORILLAWWRDAPYEVADNRGEGHSG...KKLWGYKQREFLIKHAFS WYDGEVISELOMEIYHPAMFRVTRRNDMEPEPTQLEDDDEW
KAKVHKILLWRWMEPPKGDDELHDTKSTEVGGHGRQRAPMREFFVFRDLSWNGEWLTELQLDVHHPILLRNYFKKNDMEPPAPEEGSTY
KGGVORILLWRWAEVKGDSGSDVYKKEEGKSTSKKPKRERFFVWHDMSWHSWSELEQLLNCQVFRNORNDMEPPPLDYGSGNSG
KGGIARILLWRWVVDONNGNEADKQDQNGK...KRRKRERFFVWHDMSWHSWSELEQLLNCQVFRNORNDMEPPPLEESDDEMD
PGKAEILLWRWALDRSWEI...RTSKRERFFVWHDMSWHSWSELEQLLNCQVFRNORNDMEPPLEEDDQD
TQKAEILLWRWALDRSWEI...RTSKRERFFVWHDMSWHSWSELEQLLNCQVFRNORNDMEPPLEEDDQD
GKGVSKILLWRWADLPSAEKPEEESTTT...SKRAQPRRERFFVWAEKSWHSWSELEQLLNCQVFRNORNDMEPPKLEAIDDS
KGVKILLWRWKEVMDPSEESTSKAS...SSGKPRERFFVWDRS WYDGEVISELOMEIYHPAMFRVTRRNDMEPPKLEAIDDS
-KGVKILLWRWTKPILNPDEBSTSKGAA...AAGSTRERFFVWHEKSWHSWSELEQLLNCQVFRNORNDMEPPKLEAIDDS

```

g|51701319|sp|Q6PQ02|CHD4_MOUSE
 g|292622418|Danio_reio(fish)
 g|47274371|Tetraodon_nigrovirdis(fish)
 g|889112735|sp|Q12873|CHD3_HUMAN
 g|51701343|sp|Q8TD10|CHD5_HUMAN
 g|165993|sp|Q22516|CHD3_CAEL
 g|115939069|Stromylococytus_purpuratus(sea_urchin)
 g|221120608|Hydra_magnapapillata
 g|156379809|Nematostella_vectensis(sea_anemone)
 g|260812970|Branchiostoma_floridae(amphioxus)
 g|198432553|Ciona_intestinalis(sea_squirt)
 g|292606963|Schmidtea_mediterranea(planarian)
 g|242000110|xodex_scapularis(tick)
 g|170013510|Tribolium_castaneum(red_flour_beetle)
 g|125089877|sp|Q16102|CHD3_DROME
 g|13124018|sp|Q97159|CHD4_DROME
 g|242023690|Pedicularis_humana(ouse)
 g|156545320|Nasonia_vitripennis(wasp)
 g|158300661|Anopheles_gambiae(mosquito)

```

SRKRKNDK...KFAEMEERFYRIGKPEWMMIHRILLNHS-VDKKGVHYLLIWRDLPYDGAWESEEDV...EIQDYLDFKISLW
D...KSKCKRRNKDP...FYARMEDEKGRFGVLEWLFIRILLNHS-VDKKNNVHYLLIWRDLPYDGAWESEEDM...DVPDFETKQHW
ENALLKSEKRRKADP...QYAVLEDKYRYGKPEWMMIHRILLNHS-VDKKIGCHVLYKWKDLTYDGAWESEEDM...DVPDFETKQHW
G...KSDKRRKVPD...HYAEMEERFYRIGKPEWMTVHRILLNHS-VDKKGNVHYLLIWRDLPYDGAWESEEDM...NIPVEEEKHSW
G...KSEKRRNKDP...LYAKMEERFYRIGKPEWMMIHRILLNHS-FDKKGVHYLLIWRDLPYDGAWESEEDM...DIPYDNLKQAW
...LSRHSDHDP...YKLRERFYRIGKPEWMMIHRILLNHS-SYAKSQDILYKWKELSYEHWSEEDM...DIANWEDAIIKW
QSGK...IEKKKEKND...HNLERFYRIGKPEWMMIHRILLNHS-QYGVTFDLYKWKELSYEHWSEEDM...DIANWEDAIIKW
EEEEDEVD...EILALIEKGFVGVPEWMMIHRILLNHS-TDYGNIIEFFIKWILP...DHSWEEEF...VRKIEGKNOIYFW
EDVNDVE...HKNRLDGRFFQYKPEWLQVHRVILNHR-EVKKIQLYKWKELSYEHWSEEDM...DIDEDKIEI
IEDGCHSPSTAKSKAKEQ-DENNLAEARFYGNPEWLQVHRVILNHR-VLDGSEEFYKWKVDPVPSSTWSEED...PIL...NHQIS
...NDSTQDND...DLEKRFYRIGKPEWLTIRILLNHR-FGHTHTDII...HTHTEE...IPDFQOHIKSW
...TDKQLEIKYRYGKPEWLAIRILLNHRKVPSSQLQVLYKWKHLANDEKWEPEK...DIPGMKDEIKKK
-KGRAREKVA...DPHNLEERFYKWSRPEWMMIHRILLNHR-THRNSIWLILKWRDLADECWTEQLD
DKSNREKHSK...NYQNLIEKRYRIGKPEWLQVHRVILNHR-TLKGSTLYKWKRELSYDGAWESEED...PILQKATDQW
QRSKRLLSLGGDY...KDAELIEKRYRIGKPEWLQVHRVILNHR-QMDGRTLVLYKWKRELSYDGAWESEED...SIVGLKQAIYWM
...GNLHERFYRIGKPEWLQVHRVILNHS-EENGGTMYLYKWKRELSYDGAWESEED...SIPGLNQAIALKQ
TRYKRIQRHKDKYGMKANDDAEVLEERFYKNGVYKPEWLQVHRVILNHR-TANDGSTMYLYKWKRELSYDGAWESEED...DIDGLRQADYKQ
YRMKKIEFBA...NDNLEEKYRYRIGKPEWLVHRVILNHR-TMDGRRTLVLKWKRELSYDGAWESEED...DIPGLKQAIYV
GRVKRLRDVYDKNLALR--DEYNLEERFYRIGKPEWLQVHRVILNHR-LQDRGRLVLYKWKRELSYDGAWESEED...DIPGLKQAIYV
NRYKRIQRHRETNCQL--NETLEEKYRYRIGKPEWLQVHRVILNHR-TMDGRRTLVLKWKRELSYDGAWESEED...DIPGLKQAIYV

```

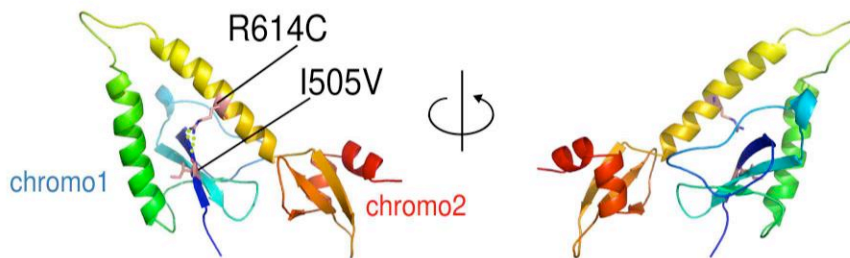
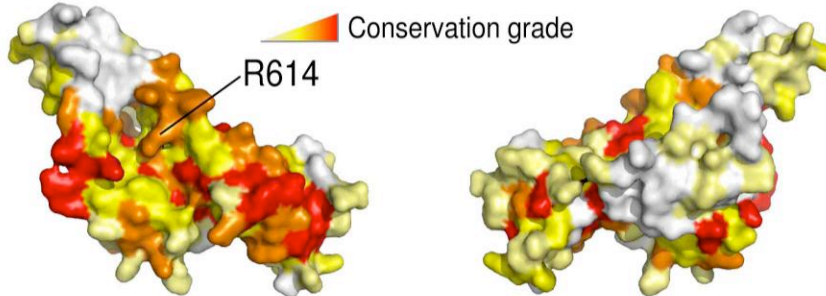
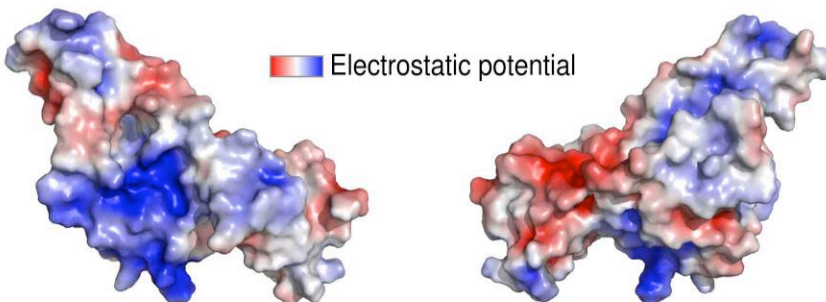
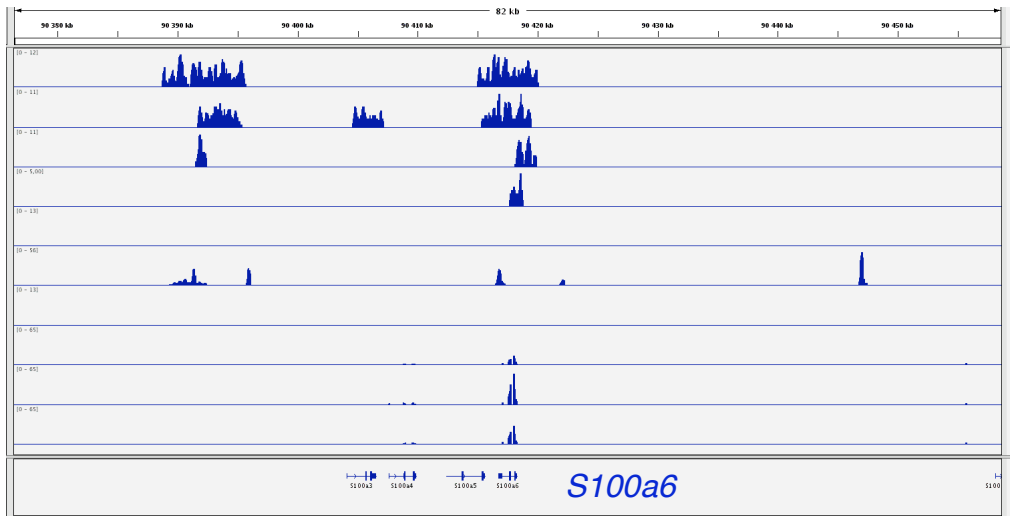
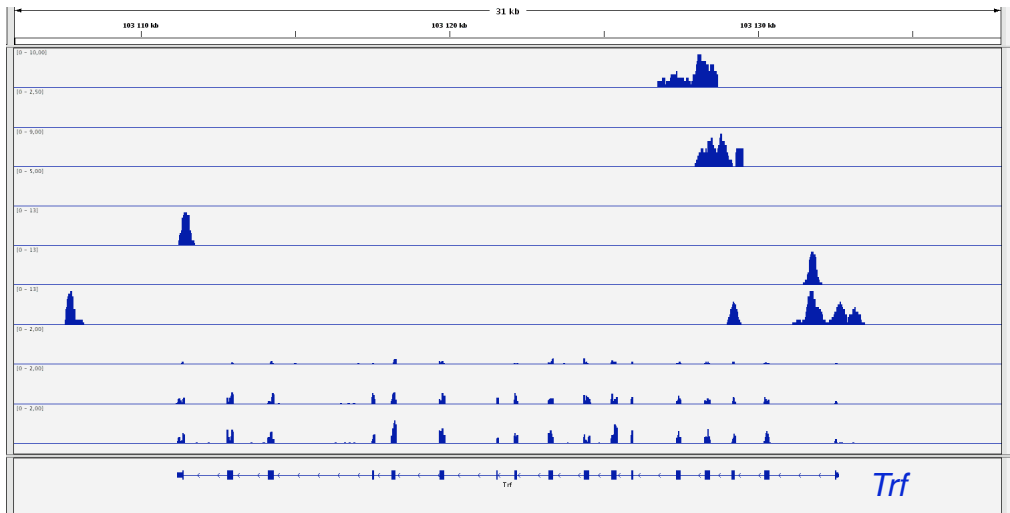
B**C****D**

Figure S8

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

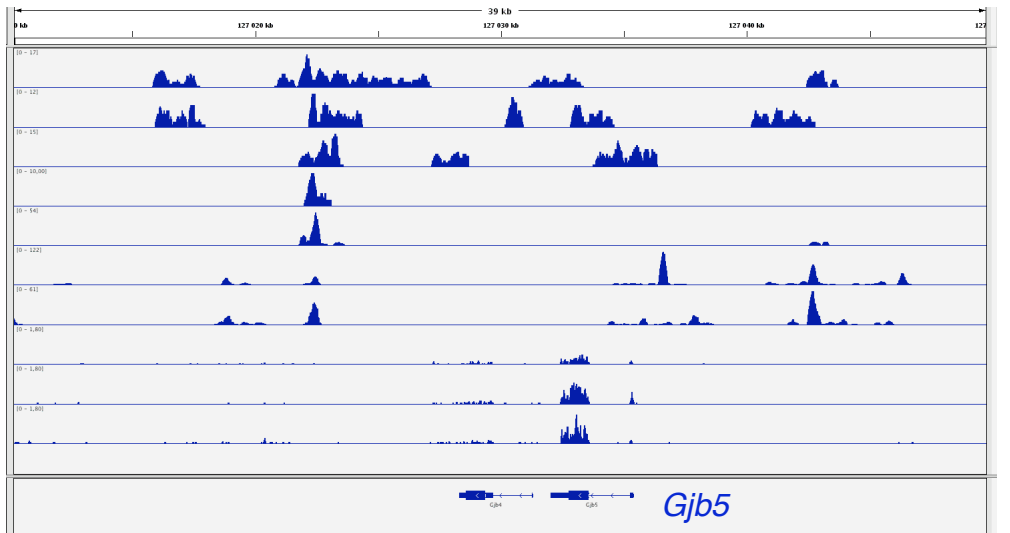
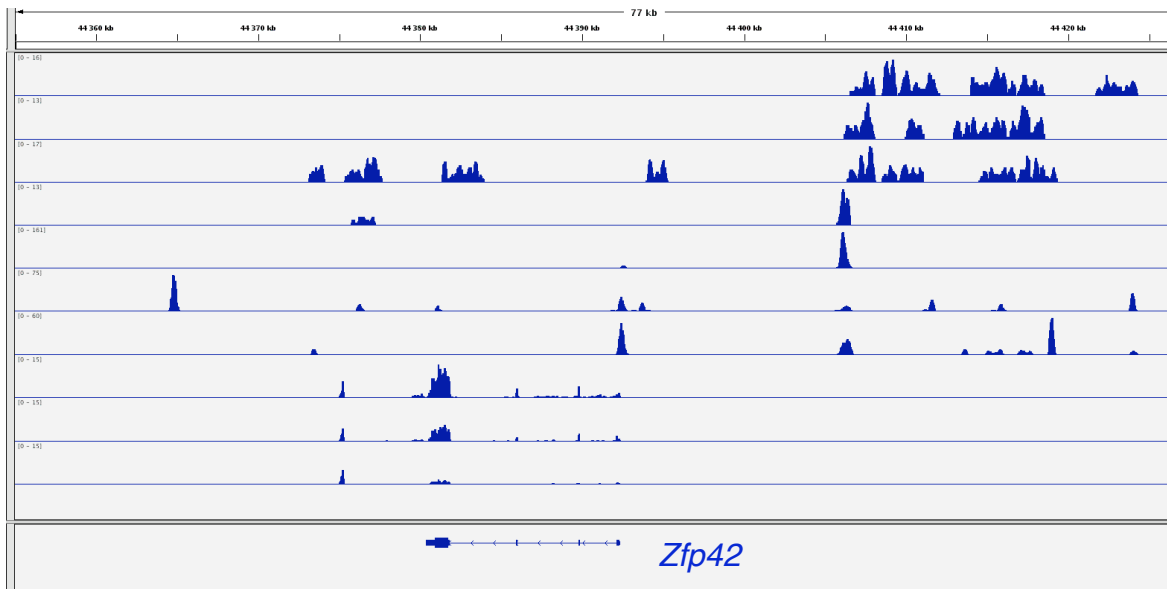
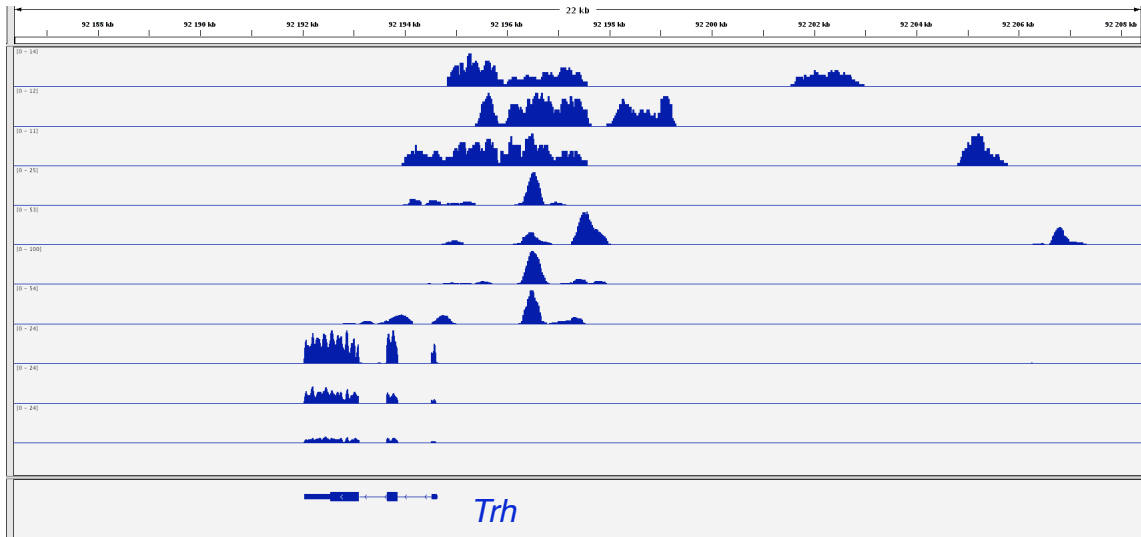


Figure S9

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

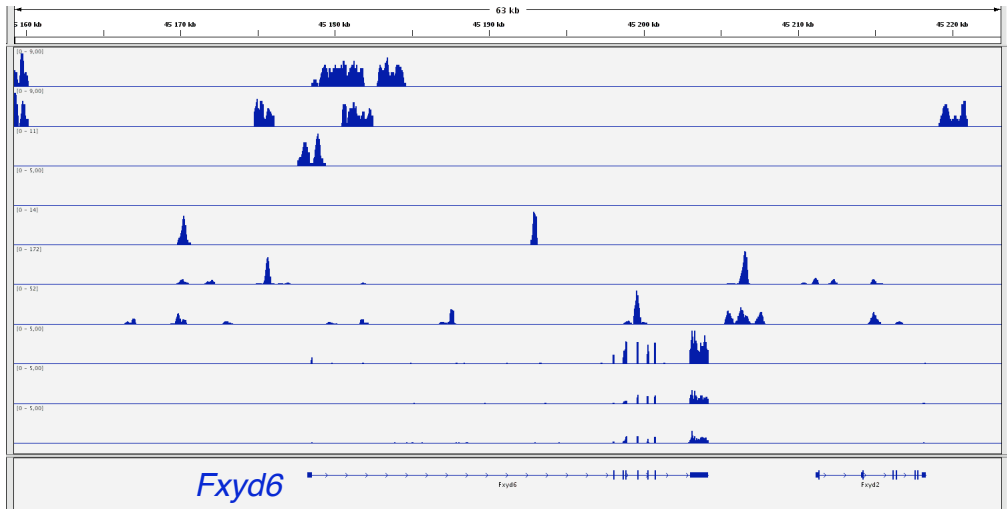
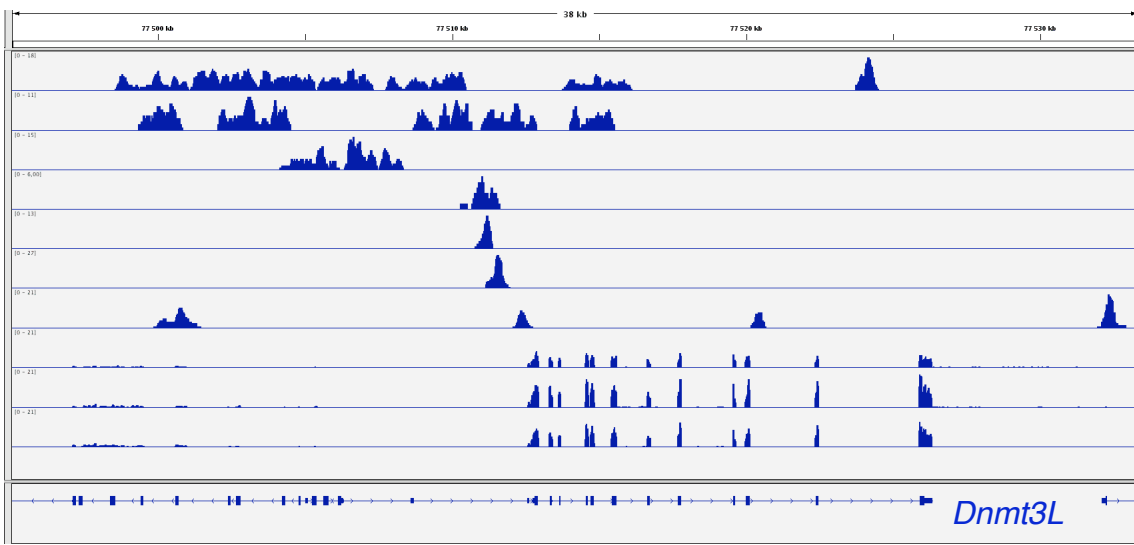
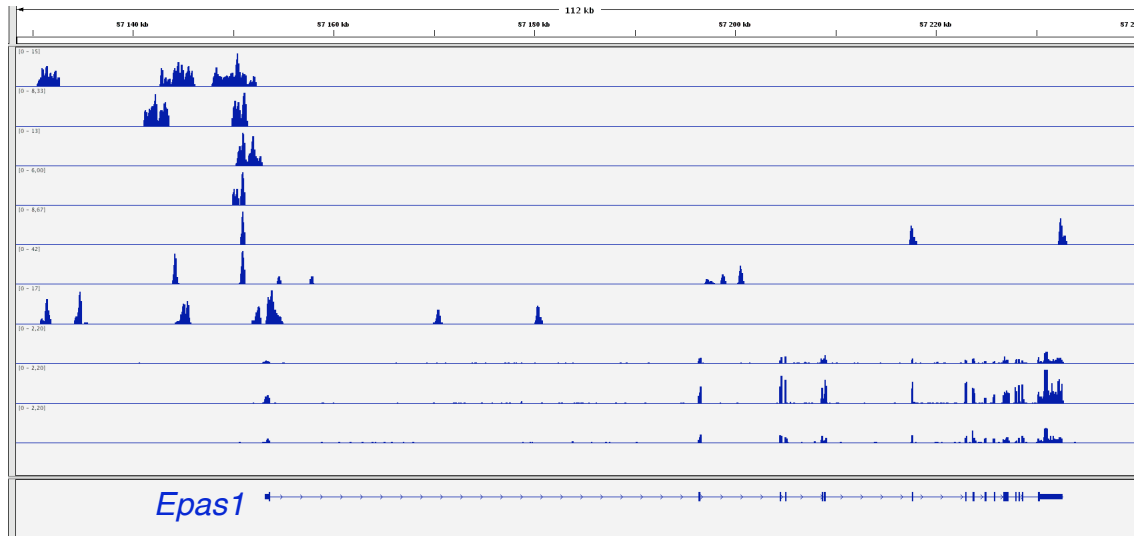


Figure S10

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

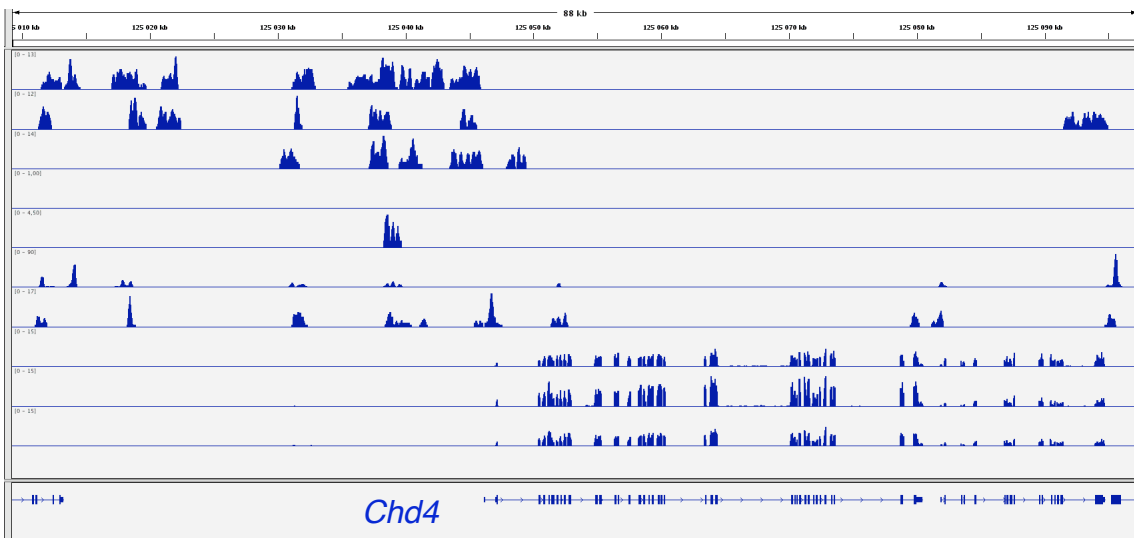
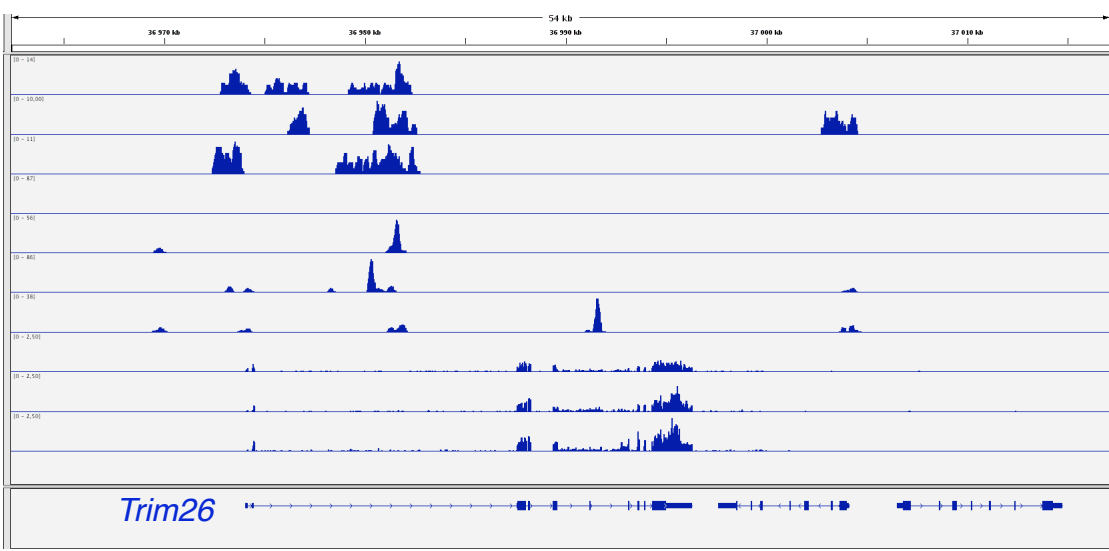
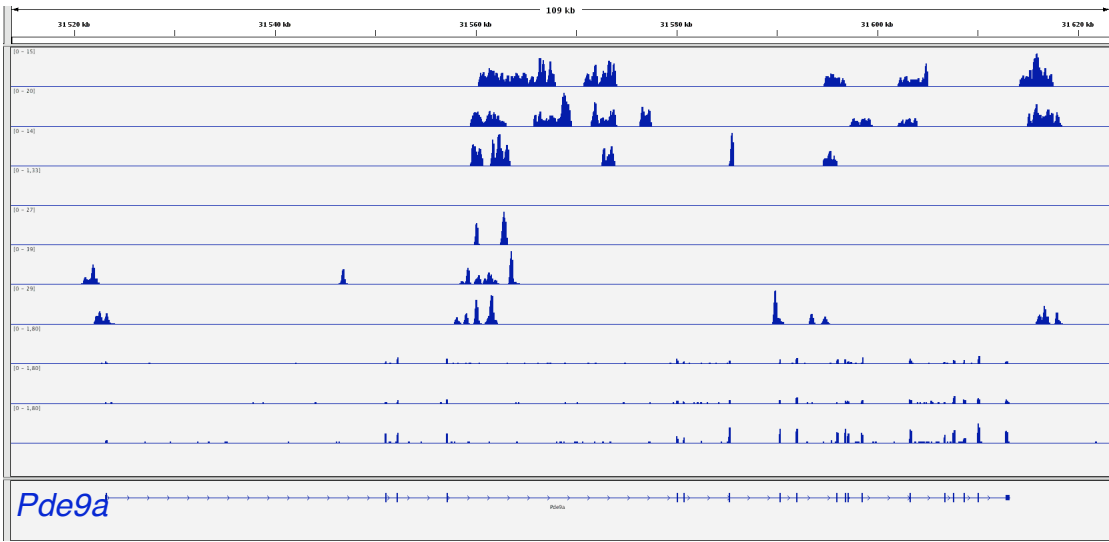


Figure S11

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp211
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
Chd4 F
Hdac2
Oct4
Nanog
Tcfcp211
Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

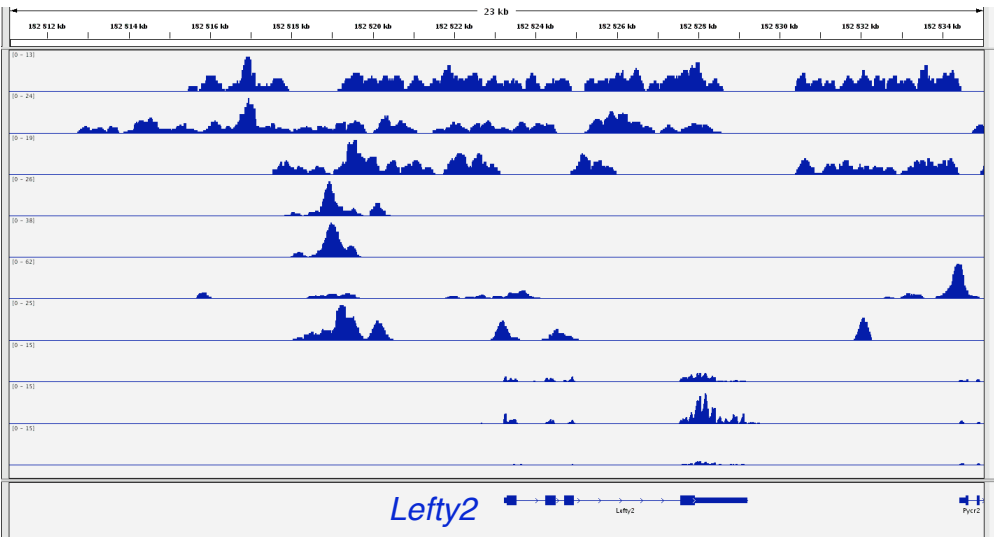
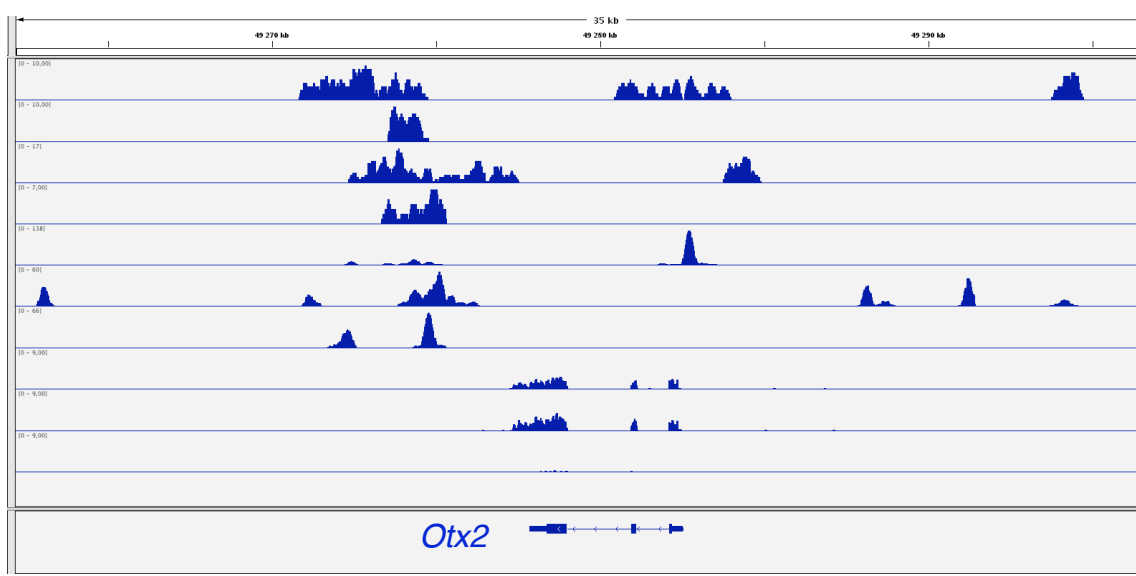
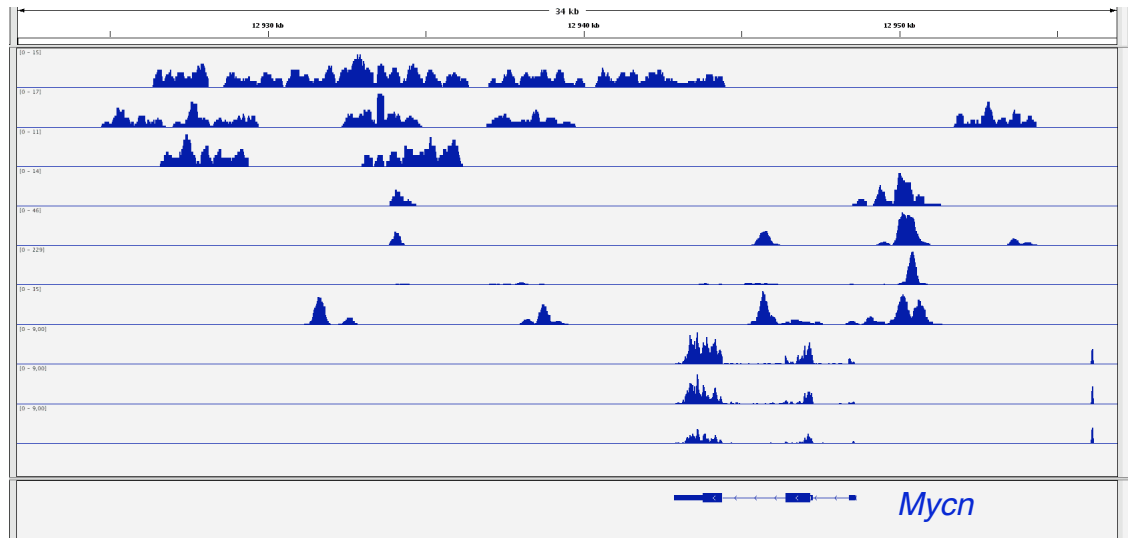


Figure S12

Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO



Chd4 DSG
 Chd4 F
 Hdac2
 Oct4
 Nanog
 Tcfcp2l1
 Esrrb
 RNA control
 RNA *Chd4* Hyp
 RNA *Hdac2* KO

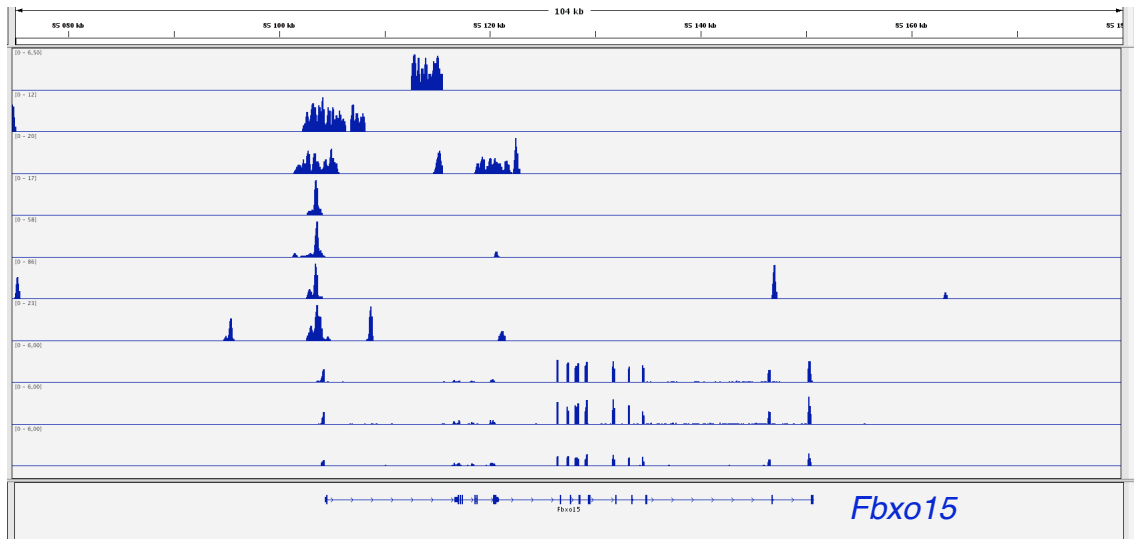
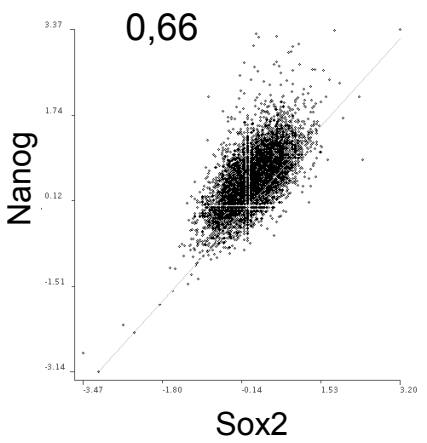
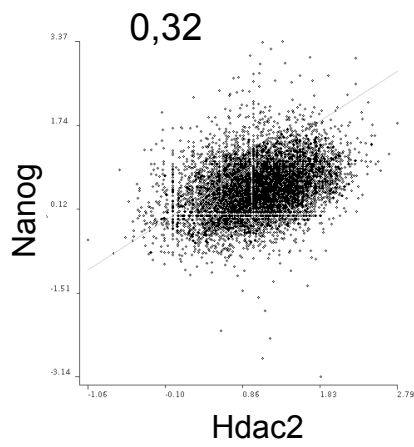
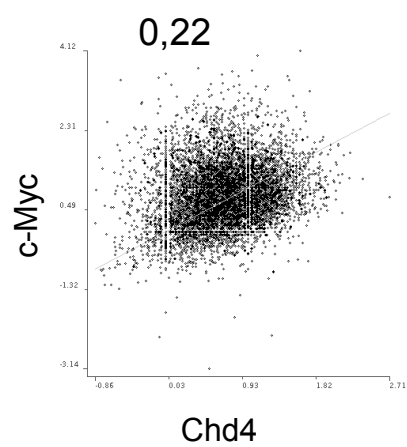
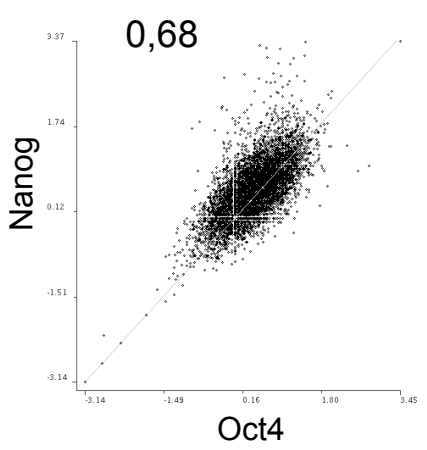
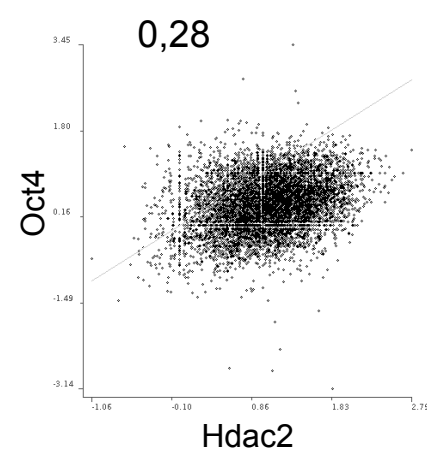
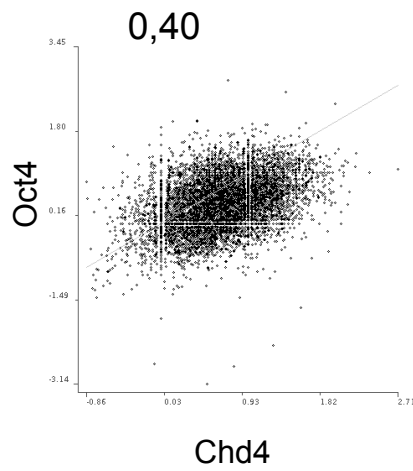
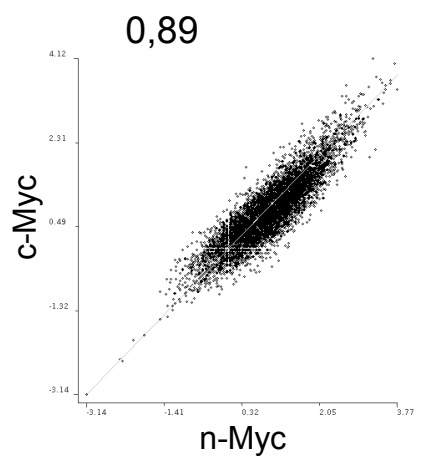
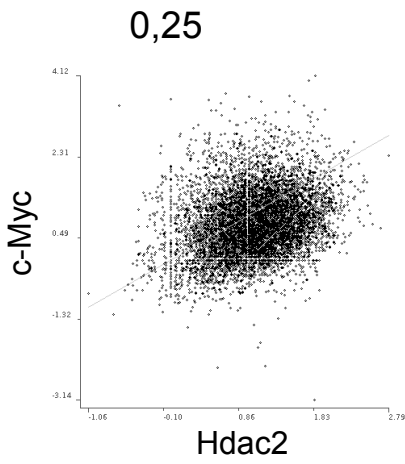
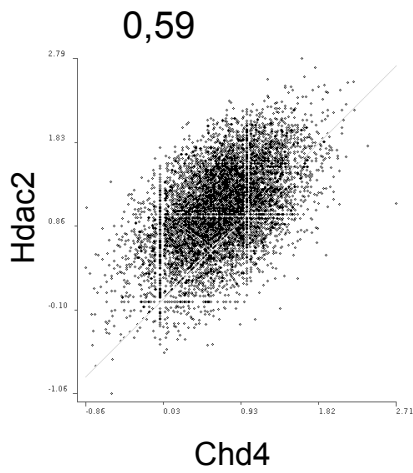


Figure S13



Coefficients de corrélation (test de pearson) sur les 10000 promoteurs des gènes les mieux exprimés en ES

Conclusion

La méthode d'ARN interférence nous permet d'identifier les gènes qui sont dérégulés lors de l'absence de la protéine ciblée. Cette dérégulation peut se faire de manière directe ou indirecte. Les cibles directes sont liées à la fois par la protéine et dérégulés. Les cibles indirectes sont dérégulés lors de l'absence de la protéine cible mais la protéine ne se lie pas directement à ces gènes. Pour différencier ces deux aspects, nous avons confronté les données de pertes de fonction (shRNA) aux données de position (ChIP-seq). Cette confrontation nous permet de distinguer les cibles directes de la protéine d'intérêt aux cibles indirectes. Nos résultats ont montré que le complexe NuRD possède de nombreuses cibles et que toutes ne sont pas dérégulées lors de la perte de fonction. La production des échantillons correspondants à l'ensemble des Chd a été réalisée en juin et juillet 2010 et les résultats sont actuellement en cours d'analyse (septembre 2010).

Chapitre 5. Projets annexes

A. Constructions de lignées ES possédant une mutation au niveau du site catalytique pour chaque protéine Chd

Introduction

Les résultats obtenus par la méthode du ChIP-seq montrent que les protéines Chd se lient à de nombreuses régions d'ADN. Cependant les régions liées par les protéines Chd ont des tailles très variables: 150pb à quelques milliers de paires de bases. C'est notamment le cas pour les protéines Chd4 et Hdac2 par exemple qui sont souvent enrichies sur des larges régions de plus de 1000 paires de bases. En 2005, au cours d'une étude sur ISW2p, une enzyme de remodelage de la chromatine à domaine ATPase chez *Saccharomyces cerevisiae*, Gelbart et al ont montré que les données de position obtenues par ChIP-on-ChIP, ne corrélaient pas avec les données obtenues par perte d'expression (Gelbart et al., 2005). Ils ont ainsi mis en évidence la mobilité de la protéine ISW2p qui se fixait de façon labile sur ces cibles. La méthode de détection des séquences cibles mise au point ne permettait pas de rendre compte des données de position de la protéine ISW2p. Afin de la stabiliser sur ses sites de liaison à l'ADN, les auteurs ont inactivé le site catalytique de la protéine. Cette mutation permettant un recrutement normal d'ISW2p sur ses gènes cibles. Cependant, en l'absence d'activité enzymatique, il n'y a pas de remodelage possible, ce qui a pour résultat la stabilisation de la protéine sur ses cibles. Gelbart et al. ont ainsi montré que pour ISW2p mutant, les données de positions de la protéine ISW2p corrélaient avec les gènes dérégulés.

Nous avons, en 2008, anticipé qu'une telle situation pourrait également s'appliquer aux protéines Chd. Dans le but de stabiliser les protéines Chd qui possèdent un domaine ATPase très semblable à la protéine ISW2p, nous avons choisi d'inactiver leur site catalytique. Pour cela nous avons construit des lignées ES stables qui possèdent un allèle étiqueté et dont le site catalytique a été inactivé. Nous avons initialement prévu de tester cette combinaison pour chaque protéine Chd. La méthode employée pour la construction de ces lignées a été basée sur la recombinaison homologe développée par Liu et al et décrite

au Chapitre 1. Cependant, j'ai adapté cette méthode afin de pouvoir introduire une mutation ponctuelle au sein du gène ciblé.

1. Principe de la méthode de construction des lignées possédant la mutation du site catalytique

Identification du site catalytique

Gelbart et al ont montré en 2005 que l'inactivation du site catalytique d'ISW2p permettait de stabiliser la protéine sur ses cibles. L'inactivation est réalisée à l'aide d'une simple mutation qui remplace une lysine en une arginine. Cette lysine, localisée en position 215 dans la protéine ISW2p, est conservée chez tous les membres de la famille SNF2 (Figure 19). Nous avons donc pu par homologie de séquence déterminer la lysine faisant partie du site catalytique des protéines Chd. Comme décrit précédemment pour l'introduction de l'étiquette, nous avons identifié une région de 10 kb [-5000: +5000] pour cibler la mutation par recombinaison homologue chez *E. Coli*.

La cassette de résistance à la néomycine est introduite dans un intron. La mutation doit donc être introduite sur l'un des bras d'homologie. Pour cela, il faut vérifier si la lysine se trouve plus proche de l'extrémité 5' ou 3' de l'exon dans lequel elle s'inscrit. Si la lysine se trouve plus proche de l'extrémité 5' de l'exon, la cassette de résistance sera placée dans l'intron situé en amont. Si la lysine se trouve plus proche de l'extrémité 3', la cassette de résistance sera placée dans l'intron aval (Figure 50).

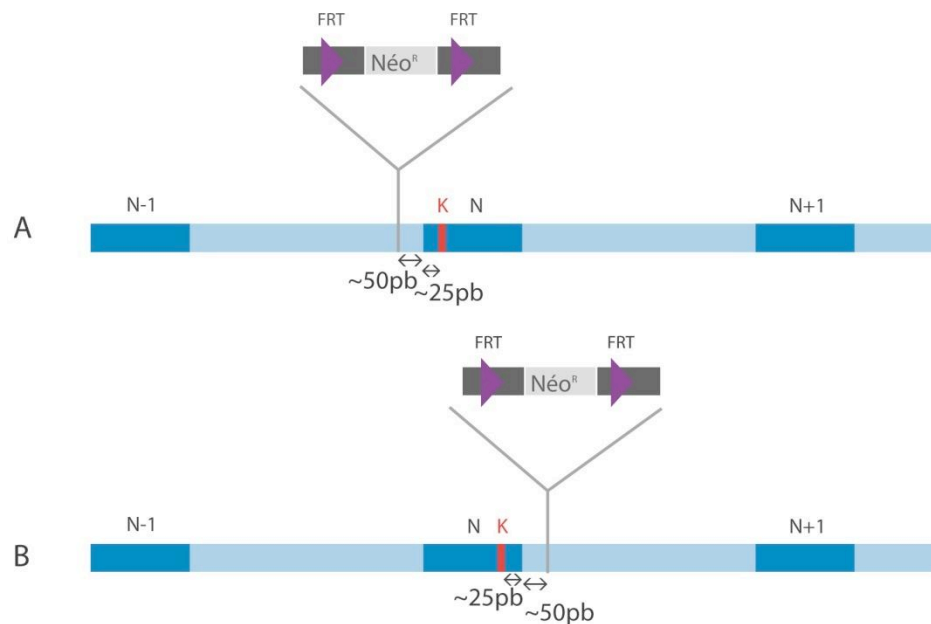


Figure 50: Stratégie de construction pour générer une lignée ES possédant une mutation sur son site catalytique: (A) la cassette de sélection se trouve en amont de la lysine d'interet (B) la cassette se trouve en aval de la lysine d'interet.

Dans les paragraphes suivants, nous allons suivre l'exemple du gène *Chd8*. Dans cet exemple, la lysine du site catalytique se trouvait en 3' de l'exon. La cassette de résistance à la néomycine a donc été placée dans l'intron suivant.

Le consensus des séquences d'épissage CAG|GTRAG au début de l'intron et CAG|G en fin d'intron a été donné par Stephens and Schneider (Stephens and Schneider, 1992). Cette séquence est nécessaire à l'épissage des introns. Afin de ne pas perturber l'excision de l'intron dans lequel se trouve la cassette Néo^R, la séquence de la cassette ne doit pas interférer avec les séquences d'épissage. Cependant la distance entre la mutation du site catalytique doit être assez faible pour minimiser le risque de recombinaison homologue entre la cassette et la mutation. Tenant compte de ces deux paramètres, nous avons décidé de placer la cassette de résistance à la néomycine à 50 paires de bases de la frontière intron-exon. Cette distance permet une proximité entre la cassette, nécessaire pour la sélection des cellules ES, tout en permettant un épissage correct de l'intron dans lequel se trouve la cassette Néo^R.

Afin de pouvoir génotyper facilement les constructions qui ont incorporé la mutation, j'ai ajouté une astuce dans la construction.

Afin d'introduire la mutation, plusieurs changements d'acides nucléiques sont nécessaires. Ci-dessous sont indiqués les codons suivants possibles pour le motif GKT → GRT:

Glycine : GGN

Arginine: CGN, AGR

Thréonine: ACN

Une méthode permettant de contrôler la mutation des vecteurs de recombinaison homologue est le séquençage. Cependant cette méthode est moins rapide que la plupart des autres méthodes de génotypage. C'est pourquoi j'ai choisi d'introduire, en plus de la mutation de la lysine en arginine, un site de restriction BsiWI qui permettra un génotypage simple et rapide. Ce site est créé en remplaçant le codon qui code pour la thréonine (ACT) par un codon synonyme c'est-à-dire différent mais qui code pour le même acide aminé (ACG) (Figure 51).

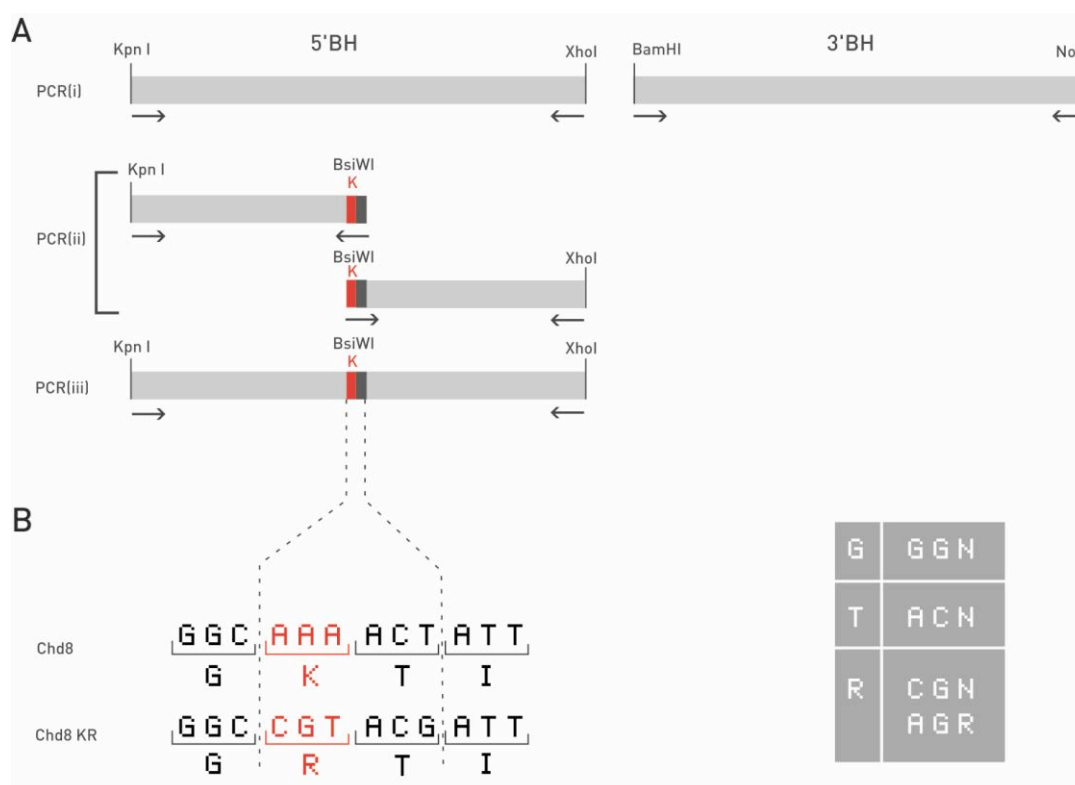


Figure 51: Introduction de la mutation au niveau du site catalytique. (A) Méthode de génération du bras d'homologie comportant la mutation

Le site ainsi créé combine d'une part la mutation de la lysine en arginine et permet d'autre part un géotypage rapide.

Construction des vecteurs de recombinaison homologue

L'ensemble des étapes de construction est regroupé dans la [Figure 52](#).

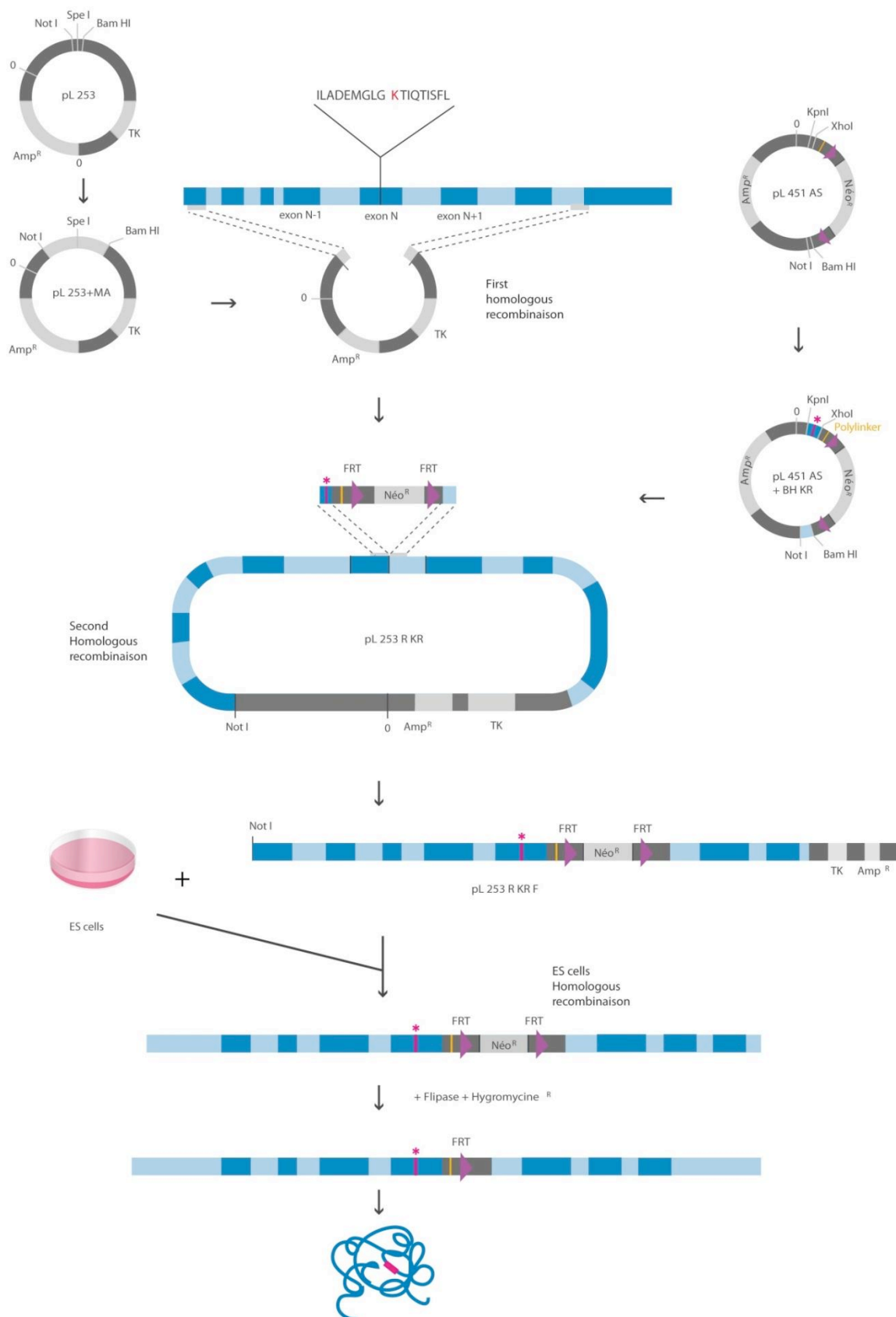


Figure 52: Ensemble des étapes nécessaires à la construction d'une lignée ES possédant une mutation au niveau de la lysine du site catalytique. L'étoile indique le remplacement de la lysine du site catalytique en arginine.

Première recombinaison homologue

i. Construction du vecteur pL253

La première recombinaison homologue est basée sur le même principe que celui décrit au chapitre 1. Dix mille paires de bases d'homologie sont identifiées de part et d'autre de la mutation que l'on veut substituer. Ce fragment est extrait d'un BAC et cloné dans le plasmide pL253 par recombinaison homologue.

ii. Recombinaison homologue

La première recombinaison homologue permet d'incorporer dans le pL253R, les 10000 paires de bases du BAC qui serviront à la recombinaison homologue en cellule souche. L'induction des protéines de la recombinaison homologue se fait pendant 15min à 42°C. Les bactéries contenant le BAC d'intérêt sont électroporées avec le pL253+MA. Les bactéries sont ensuite sélectionnées sur Ampicilline pendant 24 heures. La recombinaison du BAC est vérifiée par digestion enzymatique du plasmide.

Deuxième recombinaison homologue

i. Construction du vecteur pL451AS

Le plasmide pL451 est similaire au plasmide pL452 décrit au chapitre 1. La différence est la présence de sites FRT au lieu des sites LoxP de part et d'autre de la cassette de résistance à la néomycine. Cependant il existe un site LoxP juste après le dernier site FRT.. Dans notre cas, ce site n'est pas requis et la présence d'un site LoxP pourrait induire des recombinaisons aspécifiques. En effet les cellules ES que nous allons utiliser sont déjà étiquetées et ont été électroporées avec la recombinase Cre afin de supprimer la cassette de résistance à la néomycine (voir [Figure 29](#), [Figure 31](#), page 104). Ces cellules possèdent donc un site LoxP en C-terminal du gène étiqueté. J'ai donc modifié le plasmide pour retirer le site LoxP et j'ai ajouté des sites de restriction Asc1 et Sbf1.



Figure 53: Fragment sous-cloné pour la modification du pL451 entre les sites BstB1 et BamH1, contenant le site de recombinaison FRT et les sites de restriction AscI et Sbf1.

Pour introduire cette mutation dans le vecteur pL451, j'ai cloné deux amorces contenant une extrémité complémentaire à BstBI en 5' et une extrémité complémentaire à BamHI en 3' (Figure 53). Le plasmide obtenu a été nommé pL451AS (Figure 54).

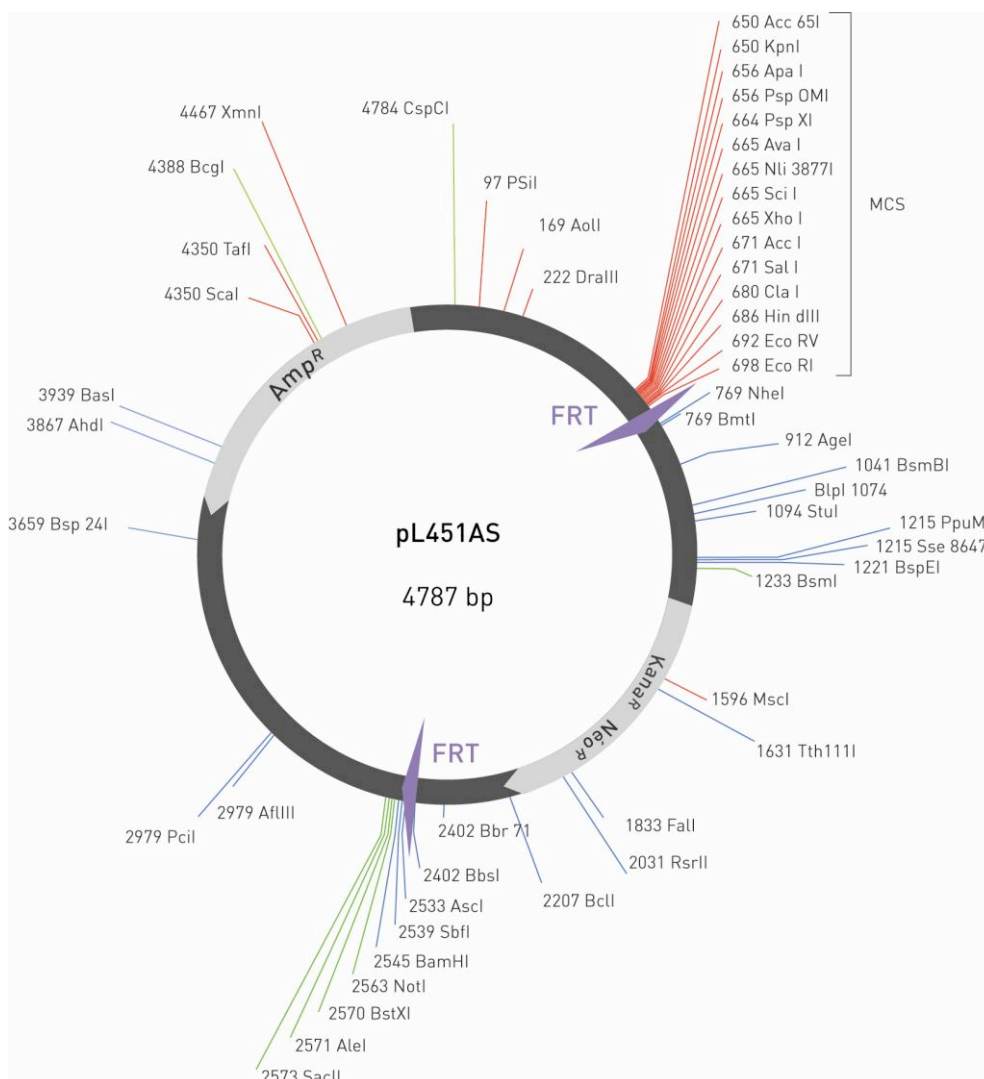


Figure 54: Cartographie du plasmide pL451

ii. Le vecteur comportant les bras d'homologie: pL451AS+BH

Pour réaliser la deuxième étape de la recombinaison homologue, deux fragments d'homologie sont sélectionnés de part et d'autre de la position d'introduction de la cassette de résistance à la néomycine. Les fragments sélectionnés doivent avoir une taille moyenne de 500bp. Le fragment qui portera la mutation pour le site catalytique doit comporter au moins 500bp après la mutation, de sorte que la recombinaison homologue se fasse préférentiellement à l'extérieur de la mutation (et ainsi minimiser le nombre de clones qui ne posséderont pas la mutation) (Figure 56).

Le bras d'homologie 5' est amplifié à l'aide d'amorces spécifiques à partir du BAC (PCR (i)). Deux amorces comprenant le site BsiWI et une partie de la séquence du bras d'homologie sont utilisées afin d'amplifier deux fragments l'un en 5' et l'autre en 3', de part et d'autre du site BsiWI (PCR (ii)). (Figure 51). Deux possibilités sont ensuite possibles pour générer le bras d'homologie 5' portant la mutation:

1. Une première possibilité consiste en une troisième PCR (iii) avec les deux amorces extrêmes regroupant les deux fragments issus de la deuxième PCR (ii). Nous avons privilégié cette étape mais nous avons remarqué que beaucoup de mutations sont introduites au cours de cette dernière étape de PCR. C'est un inconvénient majeur puisque le bras d'homologie muté porte la séquence codante du site catalytique. D'autres mutations surnuméraires produiront des effets secondaires non désirés.
2. Une alternative consiste à effectuer une double ligation avec les deux fragments d'intérêt amplifiés lors de la PCR (ii) et le plasmide digéré par les enzymes KpnI et XhoI. Nous n'avons pas testé cette méthode mais elle permettrait de palier à l'augmentation des mutations décrites dans l'alternative 1.

Le contrôle indispensable en aval de cette étape est un séquençage minutieux afin de vérifier la présence de la mutation KR et l'absence de mutations supplémentaires non désirées.

Nous avons réalisé ces constructions pour les gènes Chd3, Chd4 et Chd8. Cependant seule la construction Chd4 ne possédait pas de mutations indésirables dans la séquence codante.

iii. Deuxième recombinaison homologue

Cette deuxième recombinaison homologue permet, en ciblant la lysine du site catalytique dans le plasmide pL253R, d'introduire la mutation qui la remplace par une arginine. Cette étape s'accompagne aussi de l'introduction dans l'intron le plus proche, une cassette de résistance à la néomycine.

2. Génération de lignées ES mutantes

Recombinaison dans les cellules ES

Les cellules ES possédant l'allèle étiqueté sont électroporées dans les conditions standards décrites en annexe 2. Les clones sont sélectionnés avec le milieu D15 contenant 250ug/ml de G418 pendant 6 à 7 jours. Une sélection négative au ganciclovir de concentration 2pg/ul est appliquée pendant 2 jours. Les clones Néo^R et Ganc^R sont repiqués, amplifiés puis congelés. Une partie de la culture est néanmoins mise de côté afin d'extraire l'ADN et de procéder au génotypage des clones par Southern blot.

La construction Chd4 KR a été transfectée dans la lignée ES Hdac2-6HFH. Les résultats du génotypage montrent que plusieurs clones ont été obtenus grâce à la méthode décrite (Figure 55).

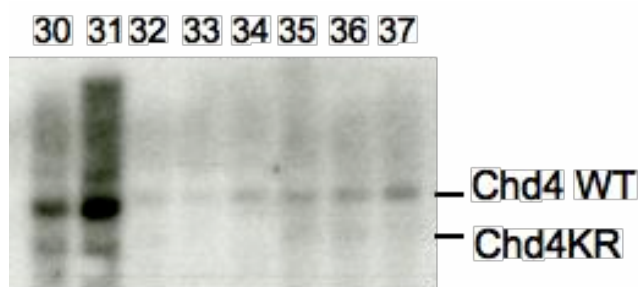


Figure 55: Génotypage par digestion EcoR1 des clones de la lignée Hdac2-6HFH électroporés avec le plasmide portant la mutation du site catalytique pour le gène Chd4. Les clones ayant correctement recombinés sont les numéros 30, 31, 32, 35 et 36.

Comme décrit plus tôt (page 155), un problème potentiel provient d'une courte séquence d'homologie (environ 75 paires de bases) entre la mutation et la cassette de résistance à la néomycine, au niveau de laquelle un événement de recombinaison

homologue peut théoriquement avoir lieu (Figure 56). Par conséquent, tous les clones résistants au G418 n'ont pas obligatoirement la mutation de la lysine en arginine.

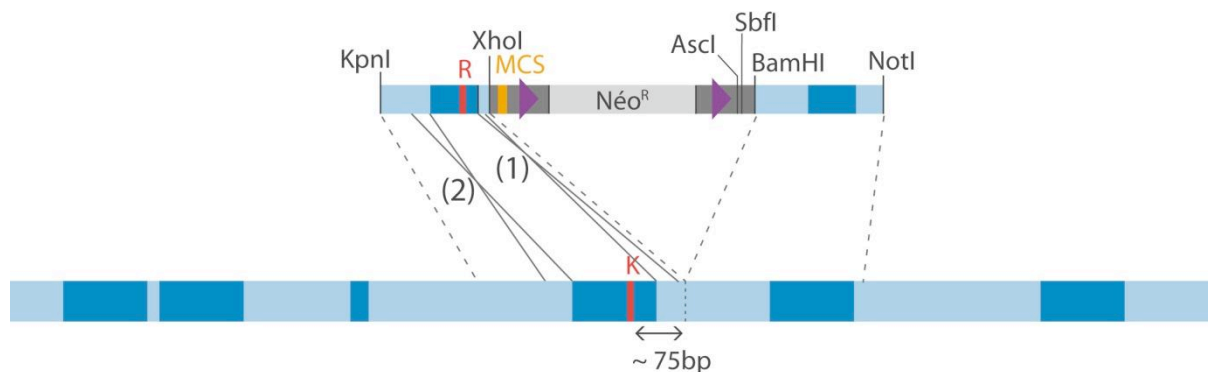


Figure 56: Deux possibilités de recombinaison. (1) recombinaison qui n'apporte pas la mutation (2) Recombinaison qui apporte la mutation

Le génotypage par Southern blot des clones de cellules ES recombinaison ont été vérifié par une seconde fois par cinq enzymes différentes. Afin d'identifier les clones qui ont effectivement introduit la mutation au niveau du site catalytique, le site BsiWI a été utilisé. Cependant ce site est rare dans le génome murin. C'est pourquoi il est nécessaire d'utiliser une double digestion. Une première enzyme (Nde1 pour l'allèle *Chd4*) détermine une bande spécifique pour l'allèle sauvage. L'enzyme BsiWI coupe au milieu de la bande sauvage pour créer une bande plus courte qui permet de distinguer les clones ayant incorporé la mutation au niveau du site catalytique (Figure 57).

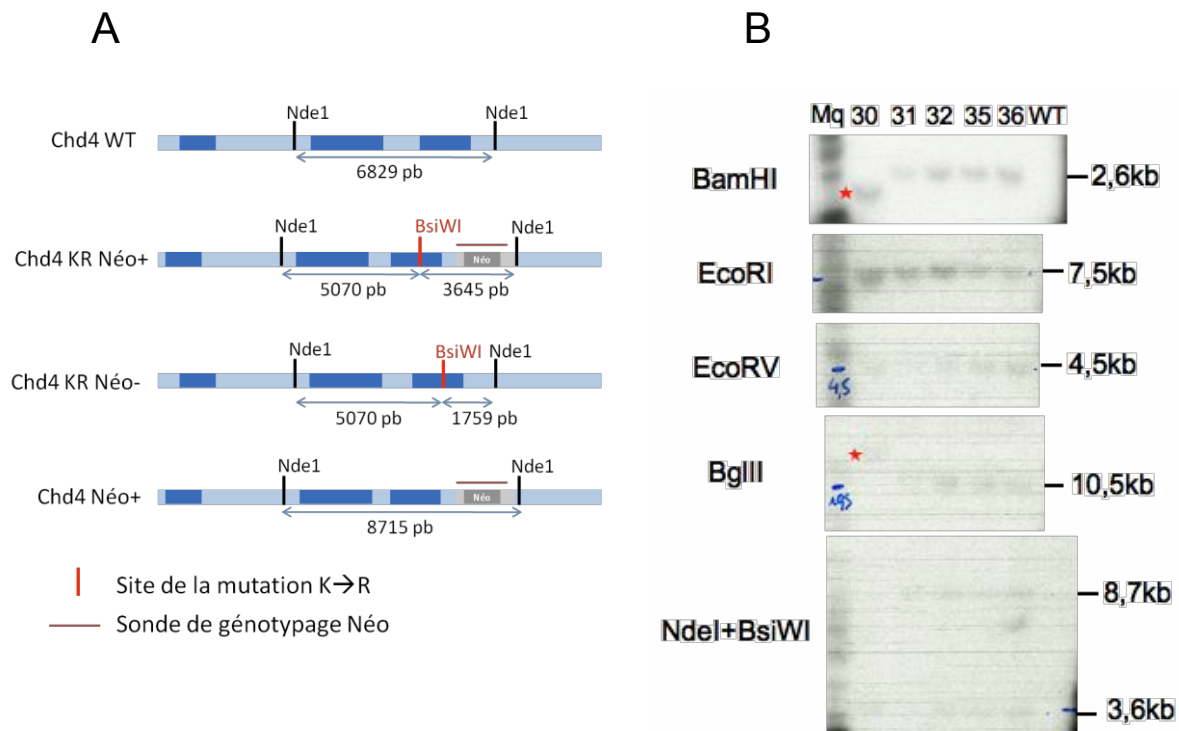


Figure 57: Génotypage des cellules ES Hdac2-6HFH électroporés avec la construction Chd4 KR afin de vérifier l'insertion correcte de la construction Chd4KR. A: Description des allèles possibles après électroporation. B: Analyse par Southern blot à l'aide d'une sonde ciblant la cassette de résistance à la néomycine. Le clone 30 montre des tailles aberrantes pour les digestions enzymatiques BglII et BamHI (identifiées par des étoiles rouges), ce dernier a donc été éliminé.

Les clones ayant incorporés la mutation avec succès sont appelés mutants KR.

Un dernier génotypage consiste à identifier l'allèle sur lequel la mutation s'est introduite. Ce type de vérification est nécessaire lorsque la mutation KR est réalisée dans la lignée de cellule ES étiquetée correspondant (ce qui n'est pas le cas dans la lignée présentée Hdac2-6HFH Chd4 KR). Pour ce faire, deux enzymes de restrictions peuvent être utilisées: SbfI et AscI. Leurs sites de coupure sont rares dans le génome murin. Le site SbfI est incorporé dans les constructions d'une part dans le site de clonage multiple juste après l'étiquette 3FTH (Chapitre 1), et d'autre part juste après le deuxième site FRT à proximité de la cassette Néo. Le site AscI est incorporé d'une part entre l'extrémité carboxyterminale de la protéine étiquetée et l'étiquette 3FTH (Chapitre 1) et d'autre part juste après le deuxième site FRT (Chapitre 5). A l'aide de ces deux enzymes, il est possible par simple digestion enzymatique et analyse par Southern blot d'identifier si la mutation KR est apparue sur l'allèle qui a été étiqueté. La digestion enzymatique fait apparaître une bande, dont la taille

est située à 20kb, lorsque l'allèle qui a été étiqueté possède aussi la mutation KR (Figure 58).

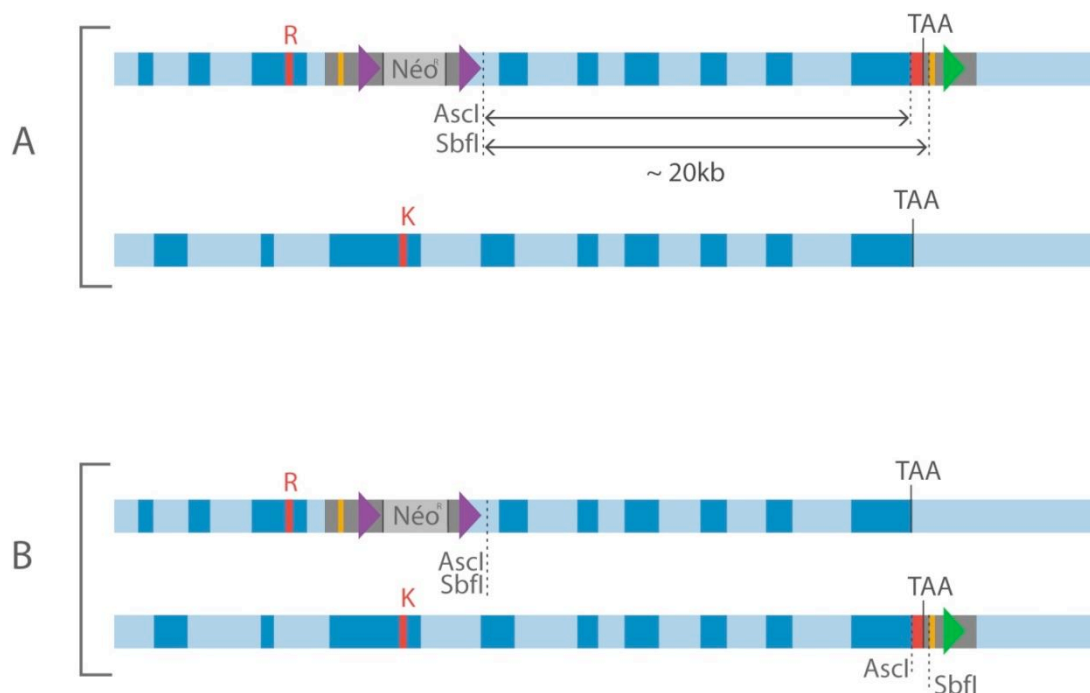


Figure 58: Stratégie de génotypage pour distinguer l'allèle *Chd4* muté non étiqueté et l'allèle *Chd4* muté possédant l'étiquette en C-ter. A: L'allèle qui a incorporé la mutation KR possède également l'étiquette en C-term. B: la recombinaison et l'introduction de la mutation KR a eu lieu sur l'allèle non étiqueté.

Retrait de la cassette de sélection de résistance à la néomycine

L'ajout de la cassette de résistance à la néomycine (taille de 2 kb) au milieu de l'intron précédent ou suivant l'exon comportant le cœur du site catalytique, peut perturber l'épissage du gène *Chd* ou réduire son taux de transcription. C'est pourquoi il est indispensable d'exciser la cassette avant d'utiliser la lignée cellulaire dans des expériences de ChIP-seq. L'excision se fait par la co-electroporation du plasmide codant pour la recombinase FLP et d'un plasmide portant un gène de résistance à l'hygromycine B. Le plasmide n'étant pas linéarité, l'expression de la Recombinase FLP est transitoire dans la cellule. La recombinase FLP permet aux deux sites FRT de s'apparier et par recombinaison d'exciser la séquence se trouvant à l'intérieur.

3. Projets futurs concernant les lignées KR

En aval des résultats obtenus en ChIP-seq sur les lignées sauvages étiquetées, il serait très intéressant d'obtenir des données de séquençage pour les lignées KR. Par exemple pour le complexe NuRD comprenant entre autre les protéines Chd4 et Hdac2, les pics détectés couvrent souvent une distance de plus d'un kb. Une des hypothèses possible postulerait que le complexe se déplace le long de l'ADN et que le protocole développé ne permet pas de fixer la protéine sur son site de liaison, mais la fixe à tous les endroits que le complexe scanne.

Dans le cas du complexe NuRD, une des protéines se fixant directement à l'ADN est Chd4. C'est pourquoi dans la lignée étiquetée pour *Hdac2*, la mutation KR sera introduite dans un allèle du gène *Chd4*. Ainsi, la protéine Chd4 stabilisée sur ses cibles immobilisera les protéines Hdac2 qui lui seront liées. Cependant, Chd4 n'est pas l'unique partenaire d'Hdac2 et les autres complexes ne seront quant à eux pas stabilisés.

Conclusion

Gelbart et al ont montré en 2005 que les gènes liés par protéine ISW2p (membre de la famille Swi/Snf chez la levure) ne reflétaient pas les gènes dérégulés lors d'un knock-out de cette protéine. Ils ont ainsi mis en évidence les propriétés dynamiques d'ISW2p qui scanne le génome et se lie de façon très transitoire sur ses cibles. La protéine ISW2p possède comme les protéines Chd un domaine ATPase capable d'utiliser l'énergie de l'ATP. A l'instar de ces auteurs, nous avons réalisé une mutation spécifique au niveau d'une lysine exceptionnellement conservée au niveau du site catalytique. Cette mutation permet d'immobiliser la protéine ISW2p sur ses sites de recrutement. Nous avons réalisé un travail similaire sur Chd4 dans les cellules souches de souris. Les lignées obtenues pourront être prochainement utilisées dans des expériences de ChIP-seq afin de déterminer plus précisément les sites de liaisons de Chd4. Les données ChIP-seq obtenues pour le complexe NuRD montrent beaucoup de régions enrichies pour Chd4 (et Hdac2) dont la taille est supérieure à 1kb. Ce résultat suggère qu'une fois recruté, le complexe NuRD se déplacerait sur les nucléosomes voisins (et peut-être ciblerait ainsi de son activité catalytique ces nucléosomes). Si le remodelage de chromatine par Chd4 fonctionne de manière similaire à celui de la protéine ISW2p de levure, on s'attend à ce que la mutation du site catalytique fige la protéine Chd4 au niveau de son site de recrutement initial. Si cette hypothèse est vraie, on s'attend à observer des régions d'enrichissement plus fines pour Chd4 (par exemple, des pics plus fins tels que ceux observés pour Chd8). L'introduction de

cette mutation catalytique dans Chd4 devrait donc nous permettre de mieux comprendre les mécanismes de recrutement de Chd4, et potentiellement donner des éléments nouveaux sur le mécanisme de remodelage de chromatine par cette enzyme.

B. Etude d'un variant d'histone impliqué dans la spermatogenèse de *Mus musculus*

1. Etude des variants d'histones chez la souris (travail initié en 2006)

La famille de l'histone H2a est celle qui possède le plus grand nombre de variants, suivie par l'histone H2b, qui comprend deux fois moins et H3 qui en compte seulement 5 (H3.1, H3.2, H3.3, H3t et CENPA). L'histone H4 ne possède pas de variants. Une étude à l'échelle du génome entier de la souris a permis d'identifier par homologie de séquence l'ensemble des variants de l'histone H2a et d'établir leur phylogénie ([Figure 59](#): Phylogénie des variants de l'histone H2A). Parmi eux, les variants les plus connus sont H2aX, H2aZ, H2aY, MacroH2a.1 et MacroH2a.2. Une sous-famille de l'histone H2a spécifiquement exprimée dans les testicules (Govin et al., 2007). De plus, la famille H2a.Bbd, identifiée en 2001 chez l'homme (Chadwick and Willard, 2001), a été retrouvée chez la souris. L'expression d'un variant H2A.Bbd murin a été confirmé par RT-QPCR, et par la suite, dans la bibliographie par Eirin-Lopez et al (Eirin-Lopez et al., 2008) et Ishibashi et al. (Ishibashi et al., 2010).

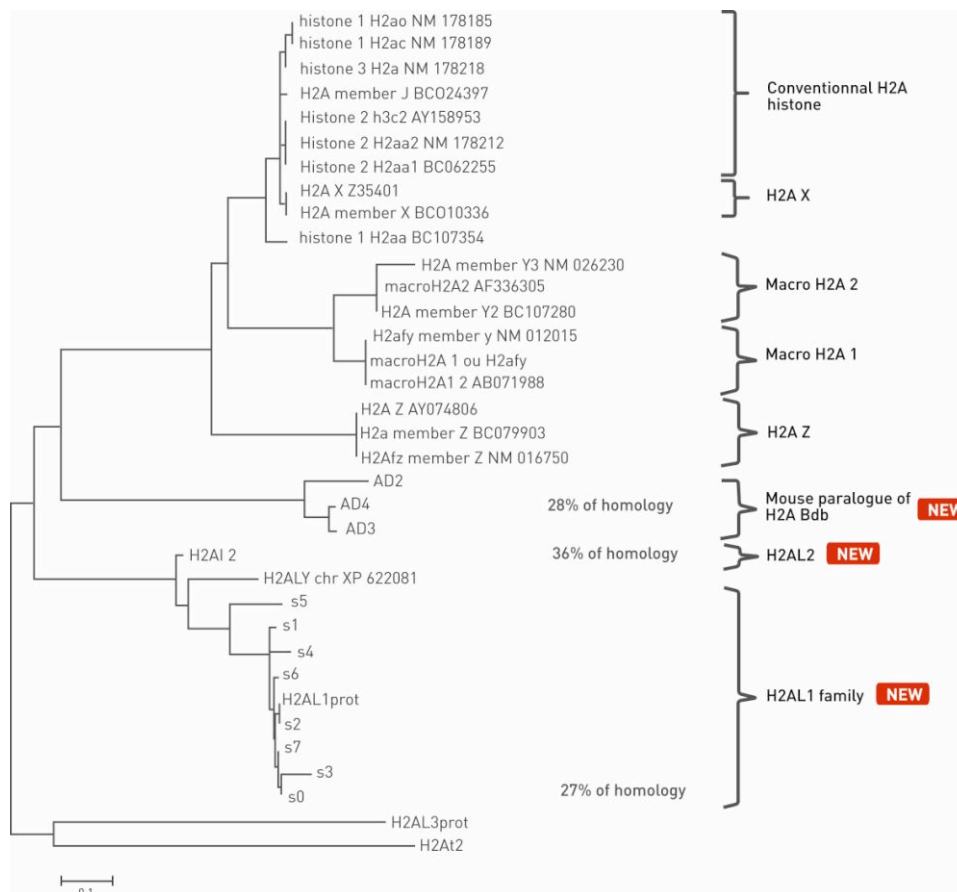


Figure 59: Phylogénie des variants de l'histone H2A. L'histone conventionnelle de référence est histone 1 H2ac NM_178189

Dans le cadre d'un projet en collaboration avec le laboratoire de S. Khochbin, nous sommes intéressés à la localisation génomique d'un de ces variants sur la chromatine. Notre objectif consiste plus particulièrement à définir des séquences de liaison sur le genome. Pour ce faire, nous avons choisi d'étiqueter le variant candidat pour d'identifier ses sites d'attachement sur le génome par CHIP-seq.

2. La famille H2al1/2

La famille H2AL1/2 possède 30% d'identité avec l'histone canonique et 50% d'homologie. Parmi les membres de cette sous-famille, le variant H2al2 a la particularité d'être localisé sur le chromosome 2 de la souris alors que les autres membres de la sous-famille sont tous localisés sur le chromosome X, le plus souvent autour de régions répétées. De plus la séquence de H2al2 étant légèrement divergente par rapport aux autres membres de la famille, nous avons pu spécifiquement (par RT-QPCR) détecter son expression

restreinte aux testicules. Les membres H2AL1 s0, s1, s2, s3, s4, s5, s6 et s7 sont très homologues entre eux. De plus, le gène s2 est dupliqué en 7 copies sur le chromosome X au milieu de fortes régions répétées. Il est impossible de définir précisément les gènes qui sont exprimés par RT-QPCR. Par conséquent, le variant H2a2 a été une cible de choix pour être étiqueté par la méthode du recombineering décrit dans le Chapitre 1. Génération de lignées ES étiquetées pour les facteurs d'intérêts (page 95).

Une analyse des génomes d'autres espèces séquencées à ce jour a montré que ce variant est présent chez certains mammifères. L'étude phylogénique de ce variant est illustrée dans la [Figure 60](#).

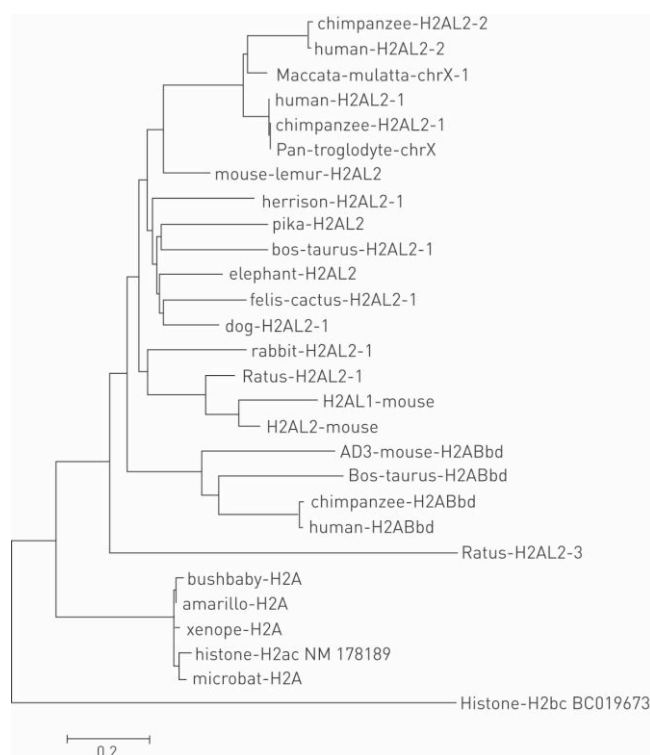


Figure 60: Phylogénie des différents variant de la famille H2A1/2. Ce variant est absent chez le tatou (Armadillo), le Galago (bushbaby), le Xenope ou la Chauve-souris (microbat)

3. Construction d'une lignée de souris possédant un allèle étiqueté pour le gène H2a2 (NM_026627)

La stratégie adoptée a été similaire à celle décrite dans le Chapitre 1. Génération de lignées ES étiquetées pour les facteurs d'intérêts. Brièvement, un BAC, contenant le gène H2AL2, a été sélectionné et inséré dans une souche de bactérie exprimant les protéines de la recombinaison homologe. 5000 paires de bases de part à d'autre du gène ont été incorporées par recombinaison homologe dans un plasmide bactérien. L'étiquette 6HFH suivit d'une cassette de résistance à la néomycine a été introduite en fusion avec le C-terminal du gène *H2A12*.

Le plasmide obtenu a été électroporé dans une lignée de cellules souches murine AT1. Les clones ayant correctement recombinés ont été sélectionnés en présence de G418. Enfin les clones ayant correctement recombinés ont été génotypés par Southern Blot avec l'enzyme de restriction BamHI selon la méthode décrite dans la [Figure 61](#).

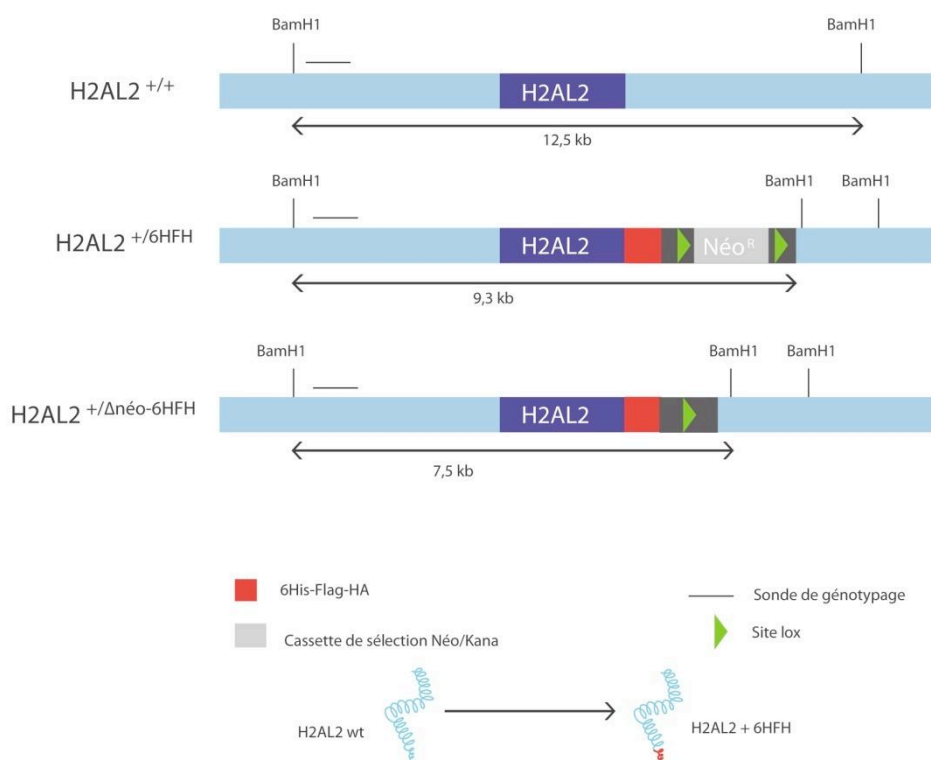


Figure 61: Stratégie de construction de la lignée étiquetée pour H2AL2. L'allèle + correspond à l'allèle sauvage.

Deux clones ayant correctement recombinés ont été injectés dans des blastocytes murins dans le but de dériver une lignée portant l'allèle H2al2 étiqueté. Nous avons obtenu des chimères (possédant des cellules du blastocyte originel et cellules injectées) fertiles, qui

ont permis la transmission de l'allèle étiqueté, donnant naissance à une nouvelle lignée: "H2a2 Néo+". Les souris qui possèdent l'allèle étiqueté possèdent aussi la cassette de résistance à la néomycine.

Les souris ne possèdent pas de phénotype particulier visible.

Visualisation de la protéine étiquetée

L'extraction des protéines des spermatides condensées des souris WT/WT et WT/6HFH a été réalisée grâce à un protocole fourni par l'équipe de Saadi Khochbin. Brièvement, les testicules ont été disséqués et l'albuginée (enveloppe protectrice) enlevée. Les spermatides ont été extraites grâce à une série de tampons de lyse et les protéines libérées par sonication. L'analyse des fractions purifiées par Western Blot a mis en évidence la protéine H2a2 WT (12,8 KDa) révélée par un anticorps spécifique et la protéine H2a2-6HFH (16,6 KDa) révélée par un anticorps dirigé contre l'épitope HA (Figure 62). La bande correspondant à H2a2-6HFH n'a pas été révélée par l'anticorps endogène car l'expression de la forme étiquetée est trop faible, et seul un anticorps hautement spécifique comme HA permet de la détecter. Cette hypothèse a été validée par RT-QPCR par F. Boussouar à Grenoble. La présence de la cassette de résistance à la néomycine inhibe donc l'expression de l'allèle étiqueté. Des observations similaires avaient déjà été faites pour d'autres lignées dans le laboratoire.

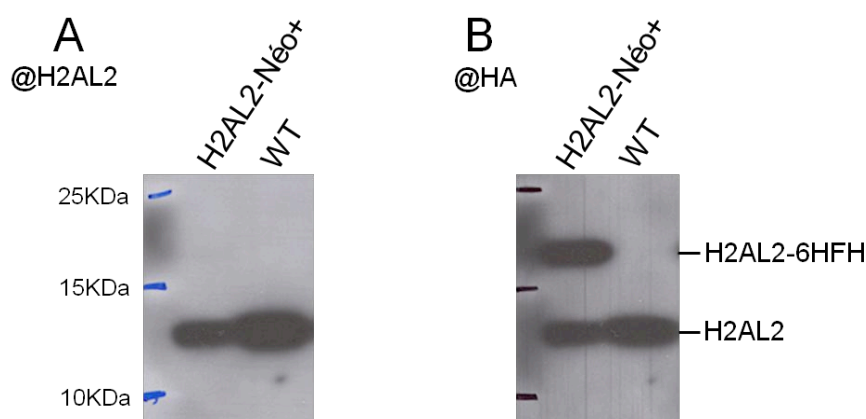


Figure 62: Expression du variant H2a2 étiqueté dans les spermatides condensées de souris contrôlée sur gel d'acrylamide 18%. Le changement de taille de la protéine correspond à la taille de l'étiquette. 43000 spermatozoïdes ont été déposés sur la piste H2a2 étiquetée, 70000 sur la piste contrôle. La même membrane a été révélée premièrement par un anticorps spécifique (Ref:SAB771; gentiment fourni par l'équipe de Saadi Khochbin à Grenoble) à H2a2 (Partie A) puis l'anticorps anti-HA a été ajouté (Partie B).

Test de fertilité des souris portant l'allèle étiqueté H2AL2-6HFH

Dans la lignée H2aI2 Néo+

Un test de fertilité sur des mâles WT/WT, WT/6HFH et 6HFH/6HFH sur la lignée H2aI2 Néo+ a montré que les mâles possédant les deux allèles étiquetés sont stériles. Cette observation est surprenante puisque il existe 15 gènes dans la sous-famille H2aI1/2 et l'ajout d'une étiquette au niveau d'un seul gène semble être dominant.

Les effets de la présence de l'étiquette sur la fertilité des mâles, est en cours d'étude à Grenoble.

Afin de pallier l'effet délétère de la cassette de résistance à la néomycine, cette dernière a été enlevée par croisement avec une lignée de souris exprimant la recombinaison Cre. La délétion de la cassette a été vérifiée par génotypage des individus. Une nouvelle lignée a donc été établie: "H2aI2 ΔNéo"

Dans la lignée H2aI2 ΔNéo

Dans cette lignée, la restauration de l'expression du variant H2aI2-6HFH a été validée par RT-QPCR à Grenoble.

Un test de fertilité ne semble pas restaurer la fertilité des mâles possédant deux allèles étiquetés.

Ces mâles seront ensuite sacrifiés afin d'isoler la chromatine des spermatides condensés. Les fragments d'ADN associés à la protéine H2aI2-6HFH seront isolés par ChIP-seq et séquencés grâce à la technologie Illumina.

L'analyse de ces fragments permettra de mieux comprendre l'incorporation de ces variants et leur rôles lors de la spermatogenèse et plus spécifiquement dans la compaction de la chromatine.

III. Discussion et conclusion

Mise en place d'un schéma robuste d'étude de protéines chromatinienne

Etude des cibles des protéines chromatinienne se liant à l'ADN

Nos travaux ont démontré la faisabilité d'un projet, à grande échelle, d'étude d'une famille de protéines couplée à une technologie puissante et innovante, le séquençage haut débit sur plateforme Illumina. De façon similaire à Kim et al., nous avons choisi d'étiqueter les protéines d'intérêt afin de nous affranchir de la production d'anticorps endogènes (tous n'étant pas commercialement disponibles). Nous avons standardisé un protocole d'immunoprécipitation assurant la reproductibilité et la comparaison des échantillons entre eux. De plus, à l'instar de Chen et al., nous avons choisi la technique du séquençage à haut débit permettant une cartographie complète (sauf les éléments répétés) des éléments liés par les protéines d'intérêt.

La technique de recombineering adapté de Liu et al., a permis de générer efficacement des lignées étiquetées pour chacune des protéines étudiées (**Chapitre 1**. Génération de lignées ES étiquetées pour les facteurs d'intérêts). Le succès de cette stratégie a été quasi-total pour l'ensemble des protéines étiquetées au laboratoire:

1. Ce projet de thèse: Chd1, Chd2, Chd3, Chd4, Chd5, Chd6, Chd8, Chd9
2. Autres projets au laboratoire: Hdac2, Chaf1a
3. Projet d'I. Hmitou: Smarca4, Smarca3, Smarca5, Smarcad1, Ep400, Atrx, Chd1L
4. Projets en collaboration: H2a2, tH2b, Atad2, Brdt (F. Boussouar) et de certaines sous-unités de l'ARN polymérase III et du complexe TFIIC (S. Graziani et L. Carrière)

Deux protéines n'ont pas pu être étiquetées, Chd7 et Hells (Smarca6). Premièrement, la disponibilité d'un BAC est requise pour effectuer la recombinaison dans les cellules ES. L'étape la plus limitative pour la construction du vecteur de recombinaison est la première recombinaison homologue (page 99) même si l'introduction du BAC dans la souche de recombinaison SW102 peut aussi se révéler une tâche très ardue.

Les étiquettes choisies ont toutes été utilisées avec succès dans des expériences d'immunoprécipitation grâce au protocole performant que notre équipe a mis au point. Parmi elles, l'étiquette 3FTH a montré la meilleure efficacité (**Chapitre 2**. Mise au point d'un protocole performant pour l'immunoprécipitation efficace des protéines étiquetées).

Les échantillons immunoprécipités ont été séquencés au CNG à Evry par nos collaborateurs Ivo et Martha Gut. L'analyse des séquences a consisté à déterminer des zones d'enrichissement des protéines d'intérêt ou pics grâce au logiciel SICER, puis à comparer les échantillons entre eux ainsi que les données publiées grâce au logiciel seqMINER (Ye et al., 2010). (**Chapitre 3. Analyse des données CHIP-seq**). La logique et la robustesse de cette étude nous ont permis d'obtenir des résultats solides sur un ensemble de 9 protéines (8 protéines Chd et Hdac2). L'approche expérimentale mise au point est d'autant plus robuste qu'elle a été validée dans un autre projet du laboratoire au cours des années 2009/2010 sur 7 autres protéines chromatinienne (I Hmitou).

Etude des gènes-cibles des protéines chromatinienne

Afin de déterminer les sites des facteurs Chd qui sont importants dans la régulation de la transcription du gène le plus proche, une analyse par perte de fonction shRNA à l'aide du vecteur pHYPER (Berlivet et al., 2007) a été réalisée pour chaque facteur dans la lignée étiquetée. L'identification des gènes dérégulés permet de déterminer les cibles directes et indirectes des protéines étudiées.

En comparant ces données transcriptomiques aux données de position des facteurs étudiés il est alors possible de déterminer les cibles directes qui associent la liaison du facteur et la dérégulation lors de la perte d'expression. L'identification des cibles directes et indirectes sont des données indispensables à la construction des cartes de réseaux de régulation.

Cependant, il est nécessaire d'attirer l'attention sur le grand nombre de sites liés par certaines protéines et le nombre plus faible de gènes qui sont dérégulés lors de leur perte d'expression. Une hypothèse expliquant cette observation est qu'il existe vraisemblablement un grand nombre de facteurs agissant de concert pour réguler un même gène cible et que l'absence d'un seul facteur peut ne pas suffire à induire une dérégulation détectable. C'est pourquoi l'analyse des puces d'expression après perte de fonction de nos remodeleurs reste délicate, car une fine dérégulation peut être significative. Par exemple, nous avons pu mettre en évidence que le seuil de dérégulation de 2, généralement choisi lors des études transcriptomiques, était inadapté pour la protéine Chd4. Il est donc nécessaire d'étudier attentivement les données, d'adapter les seuils en fonction des protéines étudiées et de se fier à des changements associés à des p-values significatives.

Etude de la fonctionnalité d'un domaine spécifique

Enfin, un projet annexe développé au cours de cette thèse propose une méthode d'étude des fonctions de remodelage des protéines Chd. Cette méthode, couplant l'efficacité du recombineering avec une mutation ponctuelle, peut s'appliquer directement à tous les protéines possédant un domaine ATPase cités dans la [Figure 20](#).

NuRD est un complexe impliqué dans la pluripotence des cellules ES

Les conclusions quant au complexe NuRD ont été détaillées dans les Chapitres 3 et 4. Brièvement, le complexe de remodelage NuRD co-localise et interagit de façon significative avec Oct4 et Nanog qui sont des facteurs majeurs de la pluripotence (nos propres données, (Liang et al., 2008; Pardo et al., 2010; van den Berg et al., 2010). De plus, une portion de ces cibles cellules ES est partagée avec d'autres facteurs importants : Esrrb, CTCF et Tcfcp2l1. Les données de localisation génomique du complexe NuRD ont montré que ce dernier se lie à 3823 gènes-cibles dont de nombreux gènes essentiels à la pluripotence. Inversement, la plupart des gènes ayant été reportés dans des études de maintien de la pluripotence ou impliqués dans la reprogrammation des iPS possèdent un enrichissement de NuRD sur leur promoteur.

Les cellules ES $Chd4^{hypo/KO}$ ont une instabilité phénotypique caractérisé par une tendance à la différenciation en cellules type du trophoctoderme (expression du marqueur trophoctodermique Hand1). *A contrario*, la lignée invalidée pour Hdac2 ne présente aucun signe de différenciation. Certains gènes importants pour la pluripotence sont dérégulés par la mutation de Chd4 ou l'absence d'Hdac2. Cependant, tous ne sont pas dérégulés dans la même direction ce qui peut refléter la complexité et le nombre important de facteurs régulateurs de la transcription qui contrôlent ces gènes.

Les données de localisation de Chd4 et Hdac2 ne sont pas totalement recouvrantes. Les protéines Chd4 et Hdac2 possèdent vraisemblablement chacune un ou plusieurs rôles indépendamment du complexe NuRD. Ceci a déjà été montré pour la protéine Hdac2 qui fait partie d'autres complexes régulateurs comme Sin3A. Cette hypothèse est appuyée par l'observation des dérégulations dans les lignées $Chd4^{hypo/hypo}$ ou $Hdac2^{-/-}$ qui peuvent être opposées et suggérerait une implication dans différents complexes de rôles opposés.

Les protéines Chd forment une famille de protéines chromatinienne très homologues entre elles mais possédant des fonctions uniques

Le rôle des protéines de la famille Chd restait globalement à ce jour inconnu. Les résultats obtenus au cours de cette étude permettent donc d'associer les protéines Chd à certaines fonctions cellulaires. La comparaison des données pour chaque Chd permettra de comprendre si ces dernières ont des fonctions redondantes ou si chacune possède un rôle bien particulier qui lui est unique.

Les premiers résultats sur les protéines Chd1, Chd4 et Chd8 montrent que ces trois protéines ont des sites de fixations sur la chromatine assez différents. Cette observation penche en faveur d'une divergence des fonctions des protéines Chd. Ceci serait en accord avec les observations faites lors des expériences d'inactivation génique (shRNA) ou de perte d'expression des différentes protéines, qui montrent que chaque Chd est indispensable à la survie ou au moins au maintien des cellules ES.

Chd1 et Chd8 sont caractérisées par des sites de liaisons particulièrement bien définis sur des régions sur le génome et, la plupart du temps, à proximité d'un promoteur. Il serait intéressant de voir si les pics qui ne sont pas à proximité d'un promoteur connu pourraient marquer certains gènes encore non identifiés. Chd1 est retrouvé plus préférentiellement associé avec les gènes activement transcrits. Chd8 n'est pas retrouvé massivement avec les gènes les plus transcrits mais sur certains gènes particuliers.

Les protéines Chd1, Chd4 et Chd8 font chacune partie d'une sous-famille des Chd. L'analyse des facteurs Chd2, Chd3, Chd5, Chd6 et Chd9 sera d'autant plus informative qu'elle fournira des données sur les différentes sous-familles Chd.

Enfin, une des tâches les plus ardues, lors de l'analyse des protéines Chd, sera de comprendre si les protéines Chd sont indispensables à la viabilité cellulaire indépendamment du contexte cellulaire ou si elles participent au maintien de la pluripotence des cellules ES.

Les réseaux de régulation de transcription importants dans le maintien de la pluripotence des cellules ES

Un trio infernal

Le cœur du réseau transcriptionnel important pour la pluripotence des cellules ES est formé par trois protéines qui interagissent fortement entre elles: Oct4, Sox2 et Nanog (Chambers and Tomlinson, 2009) (Figure M7, (Zhou et al., 2007). L'importance d'Oct4 dans le maintien de la pluripotence en cellules ES est l'illustration par excellence, démontrée en 2000 par Niwa (Niwa et al., 2000), Figure 7). De plus, Oct4 jouerait le rôle d'ancrage de la chromatine, autour duquel viendrait se greffer des complexes de régulation (van den Berg et al., 2010). La génération d'une lignée nulle pour la protéine Sox2 induit la différenciation des cellules ES en cellules du trophoctoderme (Masui et al., 2007). Loh et al. trouvent que Sox2 et Oct4 ont des sites de liaisons juxtaposés (Loh et al., 2006). Cependant, de façon très étonnante les protéines Sox2 et Oct4 n'ont pas une affinité très forte entre elles. La délétion de Nanog dans les cellules ES, n'induit pas la différenciation mais fragilise les cellules ES qui sont plus promptes à différencier. De même ce facteur n'est pas requis pour la reprogrammation des cellules ES, mais pourrait permettre de faire passer les cellules reprogrammées du stade indifférencié à pluripotent (Silva et al., 2009).

La reprogrammation des cellules ES se fait soit par fusion des cellules différenciées avec des cellules indifférenciées (créant des cellules polyploïdes) soit par expression ectopique des transcrits codants pour des facteurs permettant la reprogrammation. Il a été montré que des facteurs différents de Sox2 et Nanog sont capables eux aussi d'induire la reprogrammation: Esrrb, Klf4 et c-Myc. La reprogrammation est stimulée par l'inhibition des histones déacetylases (Chambers and Tomlinson, 2009). Il serait donc important de connaître la réelle contribution de ces facteurs par rapport au trio le plus étudié, qui pourraient avoir une importance aussi haute qu'Oct4, Sox2 et Nanog.

Cartographie du/des réseaux

La cartographie des réseaux qui maintiennent la pluripotence des cellules ES est nécessaire afin de comprendre les mécanismes mis en jeu dans ce phénotype si particulier. A l'aide des cellules souches de nombreuses maladies pourront être soignées par le remplacement des organes déficients (ex: maladie de Parkinson, Diabète, l'arthrose etc... source <http://stemcells.nih.gov/onfo/health.asp>).

Des études de plus en plus nombreuses décrivent le rôle de nouvelles protéines importantes pour les cellules ES. S'il est souvent difficile de déterminer si le rôle de ces nouveaux facteurs est spécifique aux cellules ES ou intervient dans des processus cellulaires plus généraux (indépendamment du contexte cellulaire), il n'en demeure pas moins qu'un nombre important de protéines est requis. L'identification de ces protéines et leurs interactions est un challenge majeur de notre décennie et fait l'objet de nombreuses études ([Tableau 2](#), page 53)

Perspectives de cette étude

Lignées de choix

Les lignées que nous avons générées sont un matériel de choix pour d'une part les études de localisation sur la chromatine par ChIP (ChIP-QPCR, ChIP-on-Chip, ChIP-seq) et d'autre part pour les études d'interaction par identification des partenaires protéiques. Plusieurs questions soulevées dans la bibliographie pourraient être résolues.

Par exemple, Liang et al. ont suggéré que le complexe NuRD pouvait être présent dans les cellules ES sous la forme d'un complexe appelé NODE qui exclurait la protéine Chd4 et inclurait la protéine Oct4 (Liang et al., 2008). Cette hypothèse a été démentie un an plus tard par Pardo et al. ainsi que van den Berg et al. qui ont isolé le complexe NuRD dans son intégralité avec la protéine Oct4 (Pardo et al., 2010; van den Berg et al., 2010). Une mauvaise qualité d'immunoprécipitation et des conditions trop astringentes de purification n'avaient pas permis à Liang et al. de détecter l'interaction avec Chd4. L'immunoprécipitation du complexe NuRD dans les cellules ES grâce à la double purification de la protéine Chd4 d'une part et de la protéine Hdac2 d'autre part pourrait permettre de répondre à ces questions. L'étude quantitative molécule/molécule pourrait aussi permettre de répondre à la question de l'existence d'un complexe possédant la sous-unité Chd4 mais pas Hdac2 dans les cellules ES (il est d'ores et déjà bien établi qu'il existe de nombreux complexes Hdac2 sans Chd4 comme le complexe Sin3A) et si le complexe NuRD peut être présent sans la protéine Oct4.

De même, nous avons pu montrer grâce à nos résultats de ChIP-seq que Chd1 et Chd8 co-localisaient à un certain nombre de sites plus particulièrement aux promoteurs de certains gènes les mieux exprimés en cellules ES. Une première hypothèse consisterait à tester par immunoprécipitation la présence d'interaction physique de ces deux protéines (même si une étude plus ciblée montre une co-localisation plus limitée sur certains gènes). Si

tel est le cas, la mise en évidence d'un nouveau complexe de régulation et d'activation pourrait être testée. De même, l'interaction de ces deux protéines avec le facteur de transcription E2f1, qui partage le plus de sites de co-localisation avec Chd1 et Chd8, doit être testée. Chd1 est souvent retrouvée liée sur les promoteurs des gènes les plus exprimés. De plus, comparativement à Chd8, la protéine Chd1 est liée un peu plus en aval du promoteur (certaines fois quasiment au début de l'ORF). Elle pourrait donc avoir un lien avec l'ARN polymérase II. Cette hypothèse peut être validée par immunoprécipitation des protéines associées avec Chd1 ou par co-immunoprécipitation.

En ce qui concerne Chd8, les résultats de ChIP-seq n'ont pas permis de valider son interaction avec CTCF. Une étude avait montré l'interaction de Chd8 avec l'ARN polymérase III (Yuan et al., 2007). De nombreuses études de ChIP-seq ont été très récemment publiées sur la polymérase III et nos collaborateurs du CEA (L. Carrière, S. Graziani et M. Werner) ont généré des données ChIP-seq concernant l'ARN polymérase 3 mais aussi le complexe TFIIIC. Il serait donc particulièrement intéressant de vérifier l'hypothèse de cette interaction par co-localisation des protéines sur le génome murin mais aussi par immunoprécipitation des complexes associés dans le contexte ES. De surcroît, cette même expérience pourrait confirmer l'appartenance de Chd8 à un complexe fonctionnel de 900-800KDa, donnée avancée par Thompson et al. (Thompson et al., 2008). D'autres articles lient la protéine Chd8 au régulateur du cycle cellulaire, la cycline E2 (Caldon et al., 2009; Rodriguez-Paredes et al., 2009) ou à la voie Wnt (Kobayashi et al., 2002; Yamashina et al., 2006) ou encore à p53 et l'histone H1 (Nishiyama et al., 2009). De la même manière, il serait intéressant de vérifier ces liens dans les cellules ES murines.

Au moment où ce manuscrit est rédigé (juillet 2010), une publication concernant le ChIP-seq de la protéine Chd7 est sur le point d'être publiée. Les auteurs ont réalisé l'immunoprécipitation de Chd7 avec un anticorps endogène dans les cellules ES Chd7 WT et dans les cellules Chd7^{-/-} qui servent de contrôle (communication personnelle de M. Schnetz, (Schnetz et al., 2010). Ces données seront particulièrement précieuses car elles complètent parfaitement notre étude sur la famille Chd puisque nous n'avons pas réussi à étiqueter cette protéine. Plusieurs études se dessinent grâce à ces données. Tout d'abord la comparaison de Chd7 avec les 8 autres Chd permettra d'avoir une vision globale et complète des sites de liaison de la famille Chd dans les cellules ES murines. Cette information pourra fournir des hypothèses de travail quant à la question d'une complémentarité des fonctions des Chd. De façon plus ciblée, l'hypothèse de la collaboration de Chd7 avec le complexe BAF (ici Brg1/Smarca4) avancée par Bajpai et al., pourra être testée dans les cellules ES (Bajpai et al., 2010). De même, l'interaction de Chd8 et Chd7 pourra être confirmée ou infirmée dans les cellules ES par immunoprécipitation des protéines (Batsukh et al., 2010). Leur possible

co-localisation sur la chromatine permettra de proposer des hypothèses quant à leur fonction mutuelle. Une hypothèse intéressante à tester serait leur possible complémentarité: en effet d'après Nishiyama et al., Chd8 aurait une activité anti-apoptotique au cours des stades précoces du développement (Nishiyama et al., 2009). Au fur et à mesure de l'évolution de l'embryon, l'apoptose est nécessaire à la mise en place de certaines fonction (ex: le système immunitaire, le système nerveux central) et Chd7 (qui a une expression préférentielle dans les cellules différenciées) pourrait prendre le relais des fonctions assurées par Chd8. Afin de valiser cette hypothèse, une expérience pilote nécessaire vise à comparer les sites de fixation de Chd8 et Chd7. Si ceux-ci sont similaires ou suggèrent des fonctions proches, il serait intéressant de suivre la différenciation des cellules ES et d'observer l'extinction de Chd8 et l'apparition de Chd7 à l'aide des étiquettes par immunofluorescence.

Une des rares publications sur Chd6 a montré que cette dernière co-localisait avec l'ARN polymérase II. Des données de position sur l'ARN polymérase II de haute qualité sont disponibles (Goldberg et al., 2010). Il est alors aisé de comparer la localisation de Chd6 avec l'ARN polymérase II et pourquoi pas la marque H3K36me3 qui marque spécifiquement les gènes transcrits. A l'instar de la plupart des autres Chd, Chd6 a été identifiée comme appartenant à un groupe fonctionnel d'une taille d'environ 2-3 MDa (Lutz et al., 2006). Il serait intéressant d'identifier ce complexe dans les cellules ES et de comparer la composition des complexes immunoprécipités par les différents Chd.

Le rôle de Chd9 reste plus mystérieux. Cette protéine, que l'on pensait réduite à être uniquement exprimée dans les os et le cartilage, est retrouvée comme une cible certaine d'Oct4 mais aussi d'autres protéines impliquées dans le maintien de la pluripotence (Kim et al., 2008a; Orkin et al., 2008; Sharov et al., 2008). Les données de positions de la protéine Chd9 seront d'une importance certaine afin d'éclaircir ce point. En effet, si le gène *Chd9* était fortement réprimé dans les cellules ES, il porterait les marques d'hétérochromatine et selon toute vraisemblance il serait étonnant d'observer un site de liaison d'Oct4 dans l'hétérochromatine constitutive. Deux hypothèses peuvent être formulées à ce jour: soit la protéine Chd9 est importante dans la pluripotence et son rôle reste à déterminer, soit le gène *Chd9* est partiellement réprimé (il pourrait faire partie des gènes bivalent) et la présence d'Oct4 permet une levée de cette inhibition immédiate lors de l'initiation de la différenciation.

L'ensemble des données de position, une fois analysées, pourra permettre d'identifier en profondeur les réseaux contrôlant la pluripotence ainsi que leurs liens (Figure 64) et de comprendre le rôle des Chd ainsi que les différences entre ces différents membres (Figure 63).

L'étude des Chd pourra ensuite être poursuivie lors de la différenciation neuronale grâce aux propriétés des 46C (page 101). De plus, par simple injection des cellules ES dans les blastocytes, les souris exprimant chaque Chd étiqueté pourront être générées. L'étude comparative des Chd pourra alors s'étendre à n'importe quel tissu de souris adulte.

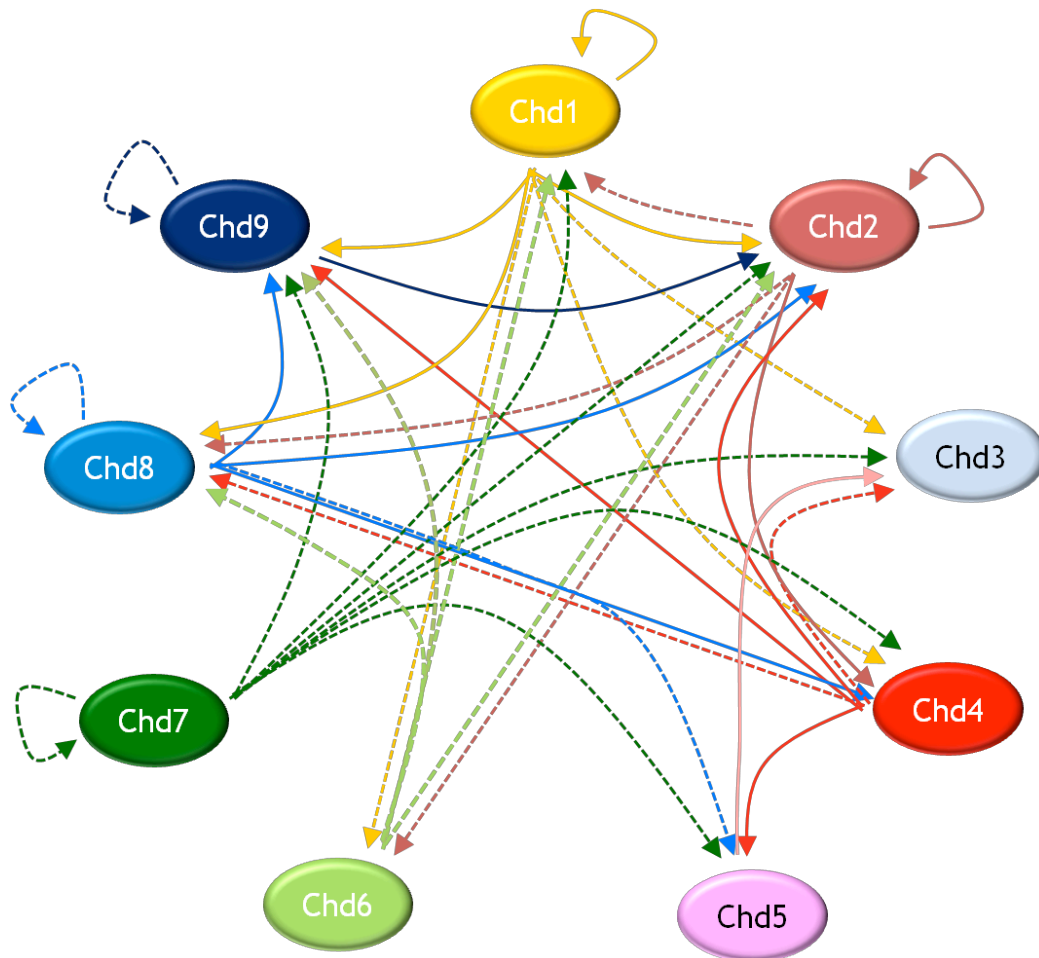


Figure 63: Réseau d'occupation des Chd sur leurs promoteurs dans les cellules ES établi à l'aide des données de séquençage Chd1 F, Chd4 F, Chd4 DSG, Chd8 F Chd1 DSG, Chd2 F, Chd2 DSG, Chd3 F, Chd5 F, Chd6 F, Chd6 DSG, Chd8 DSG, Chd9 F, Chd9 DSG. Cette figure est à compléter avec les résultats de séquençage suivants: Chd3 DSG et Chd5 DSG. Les interactions visualisées avec une seule expérience Chip-seq sont indiquées en pointillées et avec deux expériences (F et DSG) en trait plein.

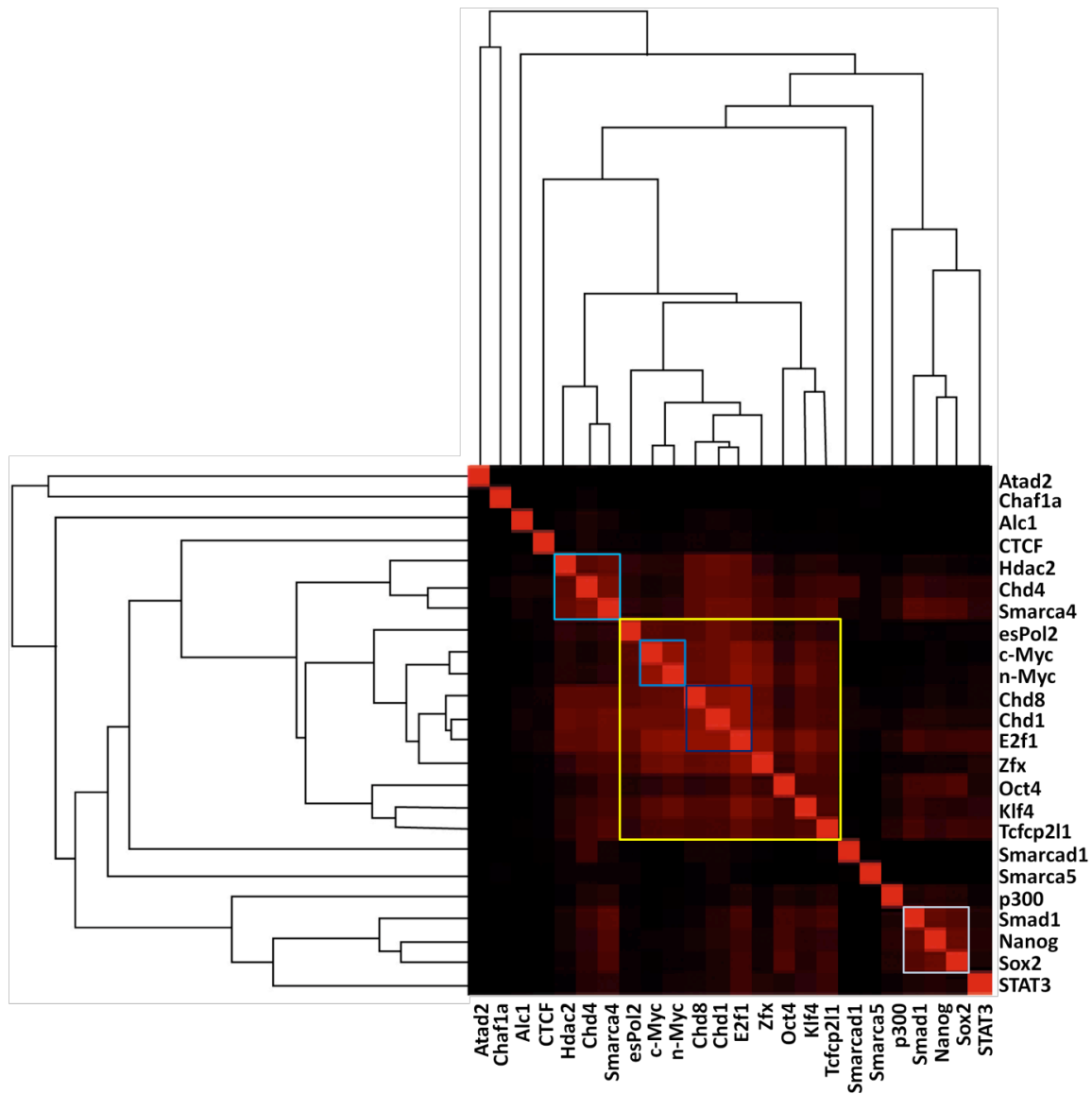


Figure 64: Exemple de résultat de co-occupation des protéines chromatinienne obtenu pour l'analyse de nos propres données et les données publiées grâce au programme Pirp, Cluster et Java Treeview. Ce sont des données préliminaires car les méthodes de calcul des corrélations et du clustering doivent être améliorées. Cependant, les données les plus robustes sont indiquées en bleu, et un large réseau de protéines partageant les mêmes cibles cellules Est indiqué en jaune. Cette figure est à compléter avec nos propres données, Chd7, le facteur Esrrb, et des protéines polycomb (Suz12, Ezh2, Ring1B).

Conclusion

L'étude d'une famille entière de remodeleurs sur le génome entier de la souris, était très ambitieuse et semblait impossible il y quelques années. Nous avons pu montrer que grâce aux nouvelles technologies couplées à une approche systématique, il est possible d'étudier la localisation d'un large nombre de protéines par ChIP-seq et sans anticorps spécifique de chaque protéine. En effet, la génération d'anticorps immunoprécipitants reste une étape limitative pour un certains nombre de protéines. Dans cette étude, nous proposons une méthode alternative combinant l'étiquetage des protéines d'intérêt à un protocole d'immunoprécipitation performant et robuste. Le développement des outils informatiques permettant d'intégrer les nombreuses données publiées est un outil indispensable à cette analyse.

On peut anticiper que les technologies haut débit permettront d'obtenir d'ici quelques années une vision globale de l'ensemble des réseaux transcriptionnels gouvernants la pluripotence des cellules ES. Les enjeux sont de taille puisque les retombées potentielles au niveau médical pourraient être de première importance, avec par exemple de nouvelles méthodes de reprogrammation et de différenciation plus efficaces.

IV. Bibliographie

- Ambrosetti, D. C., Scholer, H. R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* *275*, 23387-23397.
- Anderson, K. V., Jurgens, G., and Nusslein-Volhard, C. (1985). Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. *Cell* *42*, 779-789.
- Andersson, T., Sodersten, E., Duckworth, J. K., Cascante, A., Fritz, N., Sacchetti, P., Cervenka, I., Bryja, V., and Hermanson, O. (2009). CXXC5 is a novel BMP4-regulated modulator of Wnt signaling in neural stem cells. *J Biol Chem* *284*, 3672-3681.
- Anton, R., Kestler, H. A., and Kuhl, M. (2007). Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. *FEBS Lett* *581*, 5247-5254.
- Armstrong, J. A., Papoulas, O., Daubresse, G., Sperling, A. S., Lis, J. T., Scott, M. P., and Tamkun, J. W. (2002). The *Drosophila* BRM complex facilitates global transcription by RNA polymerase II. *Embo J* *21*, 5245-5254.
- Arnosti, D. N., and Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* *94*, 890-898.
- Aubert, J., Dunstan, H., Chambers, I., and Smith, A. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* *20*, 1240-1245.
- Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* *17*, 126-140.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M., and Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* *8*, 532-538.
- Badenhorst, P., Voas, M., Rebay, I., and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev* *16*, 3186-3198.
- Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S. Y., Voas, M., Rebay, I., Cherbas, P., and Wu, C. (2005). The *Drosophila* nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes Dev* *19*, 2540-2545.
- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* *34*, W369-373.
- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C. P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* *463*, 958-962.
- Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M., and Cavalli, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev* *17*, 2406-2420.
- Batsukh, T., Pieper, L., Koszucka, A. M., von Velsen, N., Hoyer-Fender, S., Elbracht, M., Bergman, J. E., Hoefsloot, L. H., and Pauli, S. (2010). CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Hum Mol Genet* *19*, 2858-2866.
- Baumann, C., and De La Fuente, R. (2009). ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. *Chromosoma* *118*, 209-222.
- Beddington, R., and Robertson, E. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* *105*, 733-737.
- Bender, W., and Fitzgerald, D. P. (2002). Transcription activates repressed domains in the *Drosophila* bithorax complex. *Development* *129*, 4923-4930.

- Berlivet, S., Guiraud, V., Houllard, M., and Gerard, M. (2007). pHYPER, a shRNA vector for high-efficiency RNA interference in embryonic stem cells. *Biotechniques* 42, 738, 740-733.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. *Nat Genet* 35, 215-217.
- Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B., and Young, R. A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 23, 2484-2489.
- Biswas, D., Takahata, S., Xin, H., Dutta-Biswas, R., Yu, Y., Formosa, T., and Stillman, D. J. (2008). A role for Chd1 and Set2 in negatively regulating DNA replication in *Saccharomyces cerevisiae*. *Genetics* 178, 649-659.
- Boiani, M., and Scholer, H. R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol* 6, 872-884.
- Bosman, E. A., Penn, A. C., Ambrose, J. C., Kettleborough, R., Stemple, D. L., and Steel, K. P. (2005). Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet* 14, 3463-3476.
- Bourillot, P. Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N., and Savatier, P. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. *Stem Cells* 27, 1760-1771.
- Bowen, N. J., Fujita, N., Kajita, M., and Wade, P. A. (2004). Mi-2/NuRD: multiple complexes for many purposes. *Biochim Biophys Acta* 1677, 52-57.
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947-956.
- Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349-353.
- Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255-256.
- Brehm, A., Langst, G., Kehle, J., Clapier, C. R., Imhof, A., Eberharter, A., Muller, J., and Becker, P. B. (2000). dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *Embo J* 19, 4332-4341.
- Brinster, R. L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exp Med* 140, 1049-1056.
- Brook, F. A., and Gardner, R. L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A* 94, 5709-5712.
- Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., and Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* 6, 1287-1295.
- Burdon, T., Smith, A., and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* 12, 432-438.
- Buszczak, M., and Spradling, A. C. (2006). Searching chromatin for stem cell identity. *Cell* 125, 233-236.

- Cai, S., Lee, C. C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* *38*, 1278-1288.
- Caldon, C. E., Sergio, C. M., Schutte, J., Boersma, M. N., Sutherland, R. L., Carroll, J. S., and Musgrove, E. A. (2009). Estrogen regulation of cyclin E2 requires cyclin D1 but not c-Myc. *Mol Cell Biol* *29*, 4623-4639.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* *298*, 1039-1043.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* *14*, 155-164.
- Carter, S. L., Eklund, A. C., Mecham, B. H., Kohane, I. S., and Szallasi, Z. (2005). Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics* *6*, 107.
- Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* *132*, 885-896.
- Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* *116*, 499-509.
- Chadwick, B. P., and Willard, H. F. (2001). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* *152*, 375-384.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* *113*, 643-655.
- Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* *23*, 7150-7160.
- Chambers, I., and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. *Development* *136*, 2311-2322.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* *133*, 1106-1117.
- Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P., and Ng, H. H. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* *25*, 6031-6046.
- Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R. E., Hoeijmakers, J. H., and Vermeulen, W. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol* *20*, 7643-7653.
- Clapier, C. R., and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* *78*, 273-304.
- Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. *Cell* *127*, 469-480.
- Creyghton, M. P., Markoulaki, S., Levine, S. S., Hanna, J., Lodato, M. A., Sha, K., Young, R. A., Jaenisch, R., and Boyer, L. A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell* *135*, 649-661.

- Dang, D. T., Chen, X., Feng, J., Torbenson, M., Dang, L. H., and Yang, V. W. (2003). Overexpression of Kruppel-like factor 4 in the human colon cancer cell line RKO leads to reduced tumorigenicity. *Oncogene* 22, 3424-3430.
- Dang, D. T., Pevsner, J., and Yang, V. W. (2000). The biology of the mammalian Kruppel-like family of transcription factors. *Int J Biochem Cell Biol* 32, 1103-1121.
- Darr, H., Mayshar, Y., and Benvenisty, N. (2006). Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* 133, 1193-1201.
- de Celis, J. F., Barrio, R., and Kafatos, F. C. (1996). A gene complex acting downstream of *dpp* in *Drosophila* wing morphogenesis. *Nature* 381, 421-424.
- de Hoon, M. J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453-1454.
- De La Fuente, R., Viveiros, M. M., Wigglesworth, K., and Eppig, J. J. (2004). ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev Biol* 272, 1-14.
- de la Serna, I. L., Ohkawa, Y., and Imbalzano, A. N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet* 7, 461-473.
- de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., *et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* 7, 663-676.
- Dejosez, M., Krumenacker, J. S., Zitur, L. J., Passeri, M., Chu, L. F., Songyang, Z., Thomson, J. A., and Zwaka, T. P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* 133, 1162-1174.
- Denslow, S. A., and Wade, P. A. (2007). The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26, 5433-5438.
- Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M., *et al.* (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* 5, 355-365.
- Dhillon, N., and Kamakaka, R. T. (2000). A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR. *Mol Cell* 6, 769-780.
- Dobrev, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Farinas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. *Cell* 125, 971-986.
- Dodge, J. E., Kang, Y. K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. *Mol Cell Biol* 24, 2478-2486.
- Downs, J. A., Allard, S., Jobin-Robitaille, O., Javaheri, A., Auger, A., Bouchard, N., Kron, S. J., Jackson, S. P., and Cote, J. (2004). Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol Cell* 16, 979-990.
- David, G., Ye, Z., Hammond, H., Chen, G., Pyle, A., Donovan, P., Yu, X., and Cheng, L. (2005). Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. *Stem Cells* 23, 1489-1501.
- Dupont, S., Mamidi, A., Cordenonsi, M., Montagner, M., Zacchigna, L., Adorno, M., Martello, G., Stinchfield, M. J., Soligo, S., Morsut, L., *et al.* (2009). FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell* 136, 123-135.

- Eberharter, A., Ferrari, S., Langst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., and Becker, P. B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *Embo J* 20, 3781-3788.
- Eberharter, A., Vetter, I., Ferreira, R., and Becker, P. B. (2004). ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *Embo J* 23, 4029-4039.
- Eirin-Lopez, J. M., Ishibashi, T., and Ausio, J. (2008). H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *Faseb J* 22, 316-326.
- Elfring, L. K., Daniel, C., Papoulas, O., Deuring, R., Sarte, M., Moseley, S., Beek, S. J., Waldrip, W. R., Daubresse, G., DePace, A., *et al.* (1998). Genetic analysis of brahma: the Drosophila homolog of the yeast chromatin remodeling factor SWI2/SNF2. *Genetics* 148, 251-265.
- Endoh, M., Endo, T. A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A. P., Okano, M., Brockdorff, N., Vidal, M., and Koseki, H. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* 135, 1513-1524.
- Euskirchen, G. M., Rozowsky, J. S., Wei, C. L., Lee, W. H., Zhang, Z. D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M. B., *et al.* (2007). Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res* 17, 898-909.
- Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-156.
- Fazio, T. G., Huff, J. T., and Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134, 162-174.
- Feng, B., Jiang, J., Kraus, P., Ng, J. H., Heng, J. C., Chan, Y. S., Yaw, L. P., Zhang, W., Loh, Y. H., Han, J., *et al.* (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* 11, 197-203.
- Fernandez, P. C., Frank, S. R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. *Genes Dev* 17, 1115-1129.
- Ficz, G., Heintzmann, R., and Arndt-Jovin, D. J. (2005). Polycomb group protein complexes exchange rapidly in living Drosophila. *Development* 132, 3963-3976.
- Fitzgerald, D. P., and Bender, W. (2001). Polycomb group repression reduces DNA accessibility. *Mol Cell Biol* 21, 6585-6597.
- Flanagan, J. F., Blus, B. J., Kim, D., Clines, K. L., Rastinejad, F., and Khorasanizadeh, S. (2007). Molecular implications of evolutionary differences in CHD double chromodomains. *J Mol Biol* 369, 334-342.
- Flaus, A., Martin, D. M., Barton, G. J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res* 34, 2887-2905.
- Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2, 160-169.
- Francis, N. J., Kingston, R. E., and Woodcock, C. L. (2004). Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574-1577.
- Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki, J., and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* 16, 784-789.

- Fujita, N., Jaye, D. L., Geigerman, C., Akyildiz, A., Mooney, M. R., Boss, J. M., and Wade, P. A. (2004). MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell* *119*, 75-86.
- Fujita, N., and Wade, P. A. (2004). Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods* *33*, 81-85.
- Fujita, T., Igarashi, J., Okawa, E. R., Gotoh, T., Manne, J., Kolla, V., Kim, J., Zhao, H., Pawel, B. R., London, W. B., *et al.* (2008). CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. *J Natl Cancer Inst* *100*, 940-949.
- Futatsumori-Sugai, M., Abe, R., Watanabe, M., Kudou, M., Yamamoto, T., Ejima, D., Arakawa, T., and Tsumoto, K. (2009). Utilization of Arg-elution method for FLAG-tag based chromatography. *Protein Expr Purif* *67*, 148-155.
- Fyodorov, D. V., Blower, M. D., Karpen, G. H., and Kadonaga, J. T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev* *18*, 170-183.
- Galan-Caridad, J. M., Harel, S., Arenzana, T. L., Hou, Z. E., Doetsch, F. K., Mirny, L. A., and Reizis, B. (2007). Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* *129*, 345-357.
- Gardner, R. L. (1983). Origin and differentiation of extra-embryonic tissues in mouse. *Int Rev Exp Pathol*, 63-133.
- Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., Ramalho-Santos, J., McManus, M. T., Plath, K., Meshorer, E., and Ramalho-Santos, M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* *460*, 863-868.
- Gelbart, M. E., Bachman, N., Delrow, J., Boeke, J. D., and Tsukiyama, T. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev* *19*, 942-954.
- Germain, D. (2009). [Induced pluripotent stem cells (iPS cells)]. *Pathol Biol (Paris)* *57*, 555-559.
- Ghaleb, A. M., Nandan, M. O., Chanchevalap, S., Dalton, W. B., Hisamuddin, I. M., and Yang, V. W. (2005). Kruppel-like factors 4 and 5: the yin and yang regulators of cellular proliferation. *Cell Res* *15*, 92-96.
- Goldberg, A. D., Banaszynski, L. A., Noh, K. M., Lewis, P. W., Elsaesser, S. J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., *et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* *140*, 678-691.
- Goodfellow, P. N., and Camerino, G. (1999). DAX-1, an 'antitestis' gene. *Cell Mol Life Sci* *55*, 857-863.
- Gostissa, M., Yan, C. T., Bianco, J. M., Cogne, M., Pinaud, E., and Alt, F. W. (2009). Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. *Nature* *462*, 803-807.
- Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thevenon, J., Catena, R., Davidson, I., Garin, J., Khochbin, S., and Caron, C. (2007). Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* *176*, 283-294.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* *124*, 957-971.
- Gu, H., Bock, C., Mikkelsen, T. S., Jager, N., Smith, Z. D., Tomazou, E., Gnirke, A., Lander, E. S., and Meissner, A. (2010). Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nat Methods* *7*, 133-136.
- Guo, J., Li, Z. C., and Feng, Y. H. (2009). Expression and activation of the reprogramming transcription factors. *Biochem Biophys Res Commun* *390*, 1081-1086.

- Guo, Y., Costa, R., Ramsey, H., Starnes, T., Vance, G., Robertson, K., Kelley, M., Reinbold, R., Scholer, H., and Hromas, R. (2002). The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. *Proc Natl Acad Sci U S A* *99*, 3663-3667.
- Hall, B. D. (1979). Choanal atresia and associated multiple anomalies. *J Pediatr* *95*, 395-398.
- Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., *et al.* (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* *5*, 597-609.
- Hanna, L., Foreman, R., Tarasenko, I., Kessler, D., and Labosky, P. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* *16*, 2650-2661.
- Hao, J., Li, T. G., Qi, X., Zhao, D. F., and Zhao, G. Q. (2006). WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol* *290*, 81-91.
- Hartlepp, K. F., Fernandez-Tornero, C., Eberharter, A., Grune, T., Muller, C. W., and Becker, P. B. (2005). The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol Cell Biol* *25*, 9886-9896.
- Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Lleonart, M., Jackson, S., and Wang, Z. Q. (2001). Disruption of Trp49 causes early embryonic lethality and defects in cell cycle progression. *Nat Genet* *29*, 206-211.
- Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., and *et al.* (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev* *2*, 1513-1516.
- Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., *et al.* (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* *6*, 283-289.
- Hester, M. E., Song, S., Miranda, C. J., Eagle, A., Schwartz, P. H., and Kaspar, B. K. (2009). Two factor reprogramming of human neural stem cells into pluripotency. *PLoS One* *4*, e7044.
- Ho, L., and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature* *463*, 474-484.
- Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., and Crabtree, G. R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* *106*, 5187-5191.
- Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A. I., Ranish, J., and Crabtree, G. R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* *106*, 5181-5186.
- Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* *121*, 465-477.
- Hooker, C. W., and Hurlin, P. J. (2006). Of Myc and Mnt. *J Cell Sci* *119*, 208-216.
- Houbaviy, H. B., Murray, M. F., and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. *Dev Cell* *5*, 351-358.
- Huarte, M., Sanz-Ezquerro, J. J., Roncal, F., Ortin, J., and Nieto, A. (2001). PA subunit from influenza virus polymerase complex interacts with a cellular protein with homology to a family of transcriptional activators. *J Virol* *75*, 8597-8604.

- Hurd, E. A., Capers, P. L., Blauwkamp, M. N., Adams, M. E., Raphael, Y., Poucher, H. K., and Martin, D. M. (2007). Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mamm Genome* *18*, 94-104.
- Ishibashi, T., Li, A., Eirin-Lopez, J. M., Zhao, M., Missiaen, K., Abbott, D. W., Meistrich, M., Hendzel, M. J., and Ausio, J. (2010). H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res* *38*, 1780-1789.
- Ishihara, K., Oshimura, M., and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* *23*, 733-742.
- Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., and Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* *90*, 145-155.
- Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* *442*, 533-538.
- Iwai, N., Kitajima, K., Sakai, K., Kimura, T., and Nakano, T. (2001). Alteration of cell adhesion and cell cycle properties of ES cells by an inducible dominant interfering Myb mutant. *Oncogene* *20*, 1425-1434.
- Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* *41*, 941-945.
- Jin, J., Cai, Y., Yao, T., Gottschalk, A. J., Florens, L., Swanson, S. K., Gutierrez, J. L., Coleman, M. K., Workman, J. L., Mushegian, A., *et al.* (2005). A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J Biol Chem* *280*, 41207-41212.
- Joaquin, M., and Watson, R. J. (2003). Cell cycle regulation by the B-Myb transcription factor. *Cell Mol Life Sci* *60*, 2389-2401.
- Jonsson, Z. O., Jha, S., Wohlschlegel, J. A., and Dutta, A. (2004). Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. *Mol Cell* *16*, 465-477.
- Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008). Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* *36*, 5221-5231.
- Jung, M., Peterson, H., Chavez, L., Kahlem, P., Lehrach, H., Vilo, J., and Adjaye, J. (2010). A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. *PLoS One* *5*, e10709.
- Kahn, T. G., Schwartz, Y. B., Dellino, G. I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the *Drosophila* *ubx* gene. *J Biol Chem* *281*, 29064-29075.
- Kaji, K., Caballero, I. M., MacLeod, R., Nichols, J., Wilson, V. A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* *8*, 285-292.
- Kalmar, T., Lim, C., Hayward, P., Munoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* *7*, e1000149.
- Kashyap, V., Rezende, N. C., Scotland, K. B., Shaffer, S. M., Persson, J. L., Gudas, L. J., and Mongan, N. P. (2009). Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev* *18*, 1093-1108.

- Kidder, B. L., Palmer, S., and Knott, J. G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* 27, 317-328.
- Kim, D. H., Villeneuve, L. M., Morris, K. V., and Rossi, J. J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* 13, 793-797.
- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008a). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049-1061.
- Kim, J. B., Greber, B., Arauzo-Bravo, M. J., Meyer, J., Park, K. I., Zaehres, H., and Scholer, H. R. (2009a). Direct reprogramming of human neural stem cells by OCT4. *Nature* 461, 649-643.
- Kim, J. B., Sebastiano, V., Wu, G., Arauzo-Bravo, M. J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., *et al.* (2009b). Oct4-induced pluripotency in adult neural stem cells. *Cell* 136, 411-419.
- Kim, J. B., Zaehres, H., Arauzo-Bravo, M. J., and Scholer, H. R. (2009c). Generation of induced pluripotent stem cells from neural stem cells. *Nat Protoc* 4, 1464-1470.
- Kim, J. B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Arauzo-Bravo, M. J., Ruau, D., Han, D. W., Zenke, M., and Scholer, H. R. (2008b). Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* 454, 646-650.
- King, I. F., Francis, N. J., and Kingston, R. E. (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol Cell Biol* 22, 7919-7928.
- Knoepfler, P. S., Zhang, X. Y., Cheng, P. F., Gafken, P. R., McMahon, S. B., and Eisenman, R. N. (2006). Myc influences global chromatin structure. *Embo J* 25, 2723-2734.
- Kobayashi, M., Kishida, S., Fukui, A., Michiue, T., Miyamoto, Y., Okamoto, T., Yoneda, Y., Asashima, M., and Kikuchi, A. (2002). Nuclear localization of Duplin, a beta-catenin-binding protein, is essential for its inhibitory activity on the Wnt signaling pathway. *J Biol Chem* 277, 5816-5822.
- Konev, A. Y., Tribus, M., Park, S. Y., Podhraski, V., Lim, C. Y., Emelyanov, A. V., Vershilova, E., Pirrotta, V., Kadonaga, J. T., Lusser, A., and Fyodorov, D. V. (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* 317, 1087-1090.
- Krebs, A., Frontini, M., and Tora, L. (2008). GPAT: retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* 9, 533.
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., *et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4, e1000242.
- Kurdistani, S. K., Robyr, D., Tavazoie, S., and Grunstein, M. (2002). Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet* 31, 248-254.
- Laajala, T. D., Raghav, S., Tuomela, S., Lahesmaa, R., Aittokallio, T., and Elo, L. L. (2009). A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments. *BMC Genomics* 10, 618.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H. F., Pereira, C. F., Leleu, M., Piccolo, F. M., Spivakov, M., *et al.* (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* 12, 618-624.
- Langst, G., Bonte, E. J., Corona, D. F., and Becker, P. B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell* 97, 843-852.

- Lee, J. H., Hart, S. R., and Skalnik, D. G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* **38**, 32-38.
- Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313.
- Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S., Gokhale, S., Scholer, H. R., Tomilin, A., and Jaenisch, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1**, 403-415.
- Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., Wu, H., Aebbersold, R., Graef, I. A., and Crabtree, G. R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201-215.
- Levine, S. S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R. E. (2002). The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E., and Reinberg, D. (2010a). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* **24**, 368-380.
- Li, L., Sun, L., Gao, F., Jiang, J., Yang, Y., Li, C., Gu, J., Wei, Z., Yang, A., Lu, R., *et al.* (2010b). Stk40 links the pluripotency factor Oct4 to the Erk/MAPK pathway and controls extraembryonic endoderm differentiation. *Proc Natl Acad Sci U S A* **107**, 1402-1407.
- Li, Y., McClintick, J., Zhong, L., Edenberg, H.J., Yoder, M.C., and Chan, R.J. (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood*, 635-637.
- Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S. Y., Qin, J., Wong, J., Cooney, A. J., Liu, D., and Songyang, Z. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* **10**, 731-739.
- Liu, P., Jenkins, N. A., and Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-484.
- Liu, Y., and Labosky, P. A. (2008). Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. *Stem Cells* **26**, 2475-2484.
- Lluis, F., Pedone, E., Pepe, S., and Cosma, M. P. (2008). Periodic activation of Wnt/beta-catenin signaling enhances somatic cell reprogramming mediated by cell fusion. *Cell Stem Cell* **3**, 493-507.
- Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-440.
- Loh, Y. H., Zhang, W., Chen, X., George, J., and Ng, H. H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev* **21**, 2545-2557.
- Looijenga, L. H., Stoop, H., de Leeuw, H. P., de Gouveia Brazao, C. A., Gillis, A. J., van Roozendaal, K. E., van Zoelen, E. J., Weber, R. F., Wolffenbuttel, K. P., van Dekken, H., *et al.* (2003). POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Res* **63**, 2244-2250.
- Lorch, Y., Zhang, M., and Kornberg, R. D. (1999). Histone octamer transfer by a chromatin-remodeling complex. *Cell* **96**, 389-392.
- Lu, P. Y., Levesque, N., and Kobor, M. S. (2009). NuA4 and SWR1-C: two chromatin-modifying complexes with overlapping functions and components. *Biochem Cell Biol* **87**, 799-815.

- Lund, A. H., and van Lohuizen, M. (2004). Epigenetics and cancer. *Genes Dev* 18, 2315-2335.
- Lusser, A., Urwin, D. L., and Kadonaga, J. T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol* 12, 160-166.
- Lutz, T., Stoger, R., and Nieto, A. (2006). CHD6 is a DNA-dependent ATPase and localizes at nuclear sites of mRNA synthesis. *FEBS Lett* 580, 5851-5857.
- MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 17, 9-26.
- Mallanna, S. K., and Rizzino, A. (2010). Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells. *Dev Biol* 344, 16-25.
- Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R. A., and Jaenisch, R. (2008a). Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell* 3, 132-135.
- Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., *et al.* (2008b). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134, 521-533.
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78, 7634-7638.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., *et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* 9, 625-635.
- Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M. G., Sharov, A. A., and Ko, M. S. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS One* 1, e26.
- Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo J* 18, 4261-4269.
- Maurange, C., Lee, N., and Paro, R. (2006). Signaling meets chromatin during tissue regeneration in *Drosophila*. *Curr Opin Genet Dev* 16, 485-489.
- Mazin, A. V., Mazina, O. M., Bugreev, D. V., and Rossi, M. J. (2010). Rad54, the motor of homologous recombination. *DNA Repair (Amst)* 9, 286-302.
- McDaniel, I. E., Lee, J. M., Berger, M. S., Hanagami, C. K., and Armstrong, J. A. (2008). Investigations of CHD1 function in transcription and development of *Drosophila melanogaster*. *Genetics* 178, 583-587.
- Medina, P. P., and Sanchez-Cespedes, M. (2008). Involvement of the chromatin-remodeling factor BRG1/SMARCA4 in human cancer. *Epigenetics* 3, 64-68.
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33, 5868-5877.
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.
- Meneghini, M. D., Wu, M., and Madhani, H. D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* 112, 725-736.

- Menon, T., Yates, J. A., and Bochar, D. A. (2010). Regulation of androgen-responsive transcription by the chromatin remodeling factor CHD8. *Mol Endocrinol* *24*, 1165-1174.
- Merrill, B. J. (2008). Develop-WNTs in somatic cell reprogramming. *Cell Stem Cell* *3*, 465-466.
- Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* *7*, 540-546.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* *10*, 105-116.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* *448*, 553-560.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* *113*, 631-642.
- Mizuguchi, G., Shen, X., Landry, J., Wu, W. H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* *303*, 343-348.
- Mizuguchi, G., Vassilev, A., Tsukiyama, T., Nakatani, Y., and Wu, C. (2001). ATP-dependent nucleosome remodeling and histone hyperacetylation synergistically facilitate transcription of chromatin. *J Biol Chem* *276*, 14773-14783.
- Mizuguchi, G., and Wu, C. (1999). Nucleosome remodeling factor NURF and in vitro transcription of chromatin. *Methods Mol Biol* *119*, 333-342.
- Mohrmann, L., and Verrijzer, C. P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta* *1681*, 59-73.
- Montgomery, N. D., Yee, D., Montgomery, S. A., and Magnuson, T. (2007). Molecular and functional mapping of EED motifs required for PRC2-dependent histone methylation. *J Mol Biol* *374*, 1145-1157.
- Motegi, A., Sood, R., Moinova, H., Markowitz, S. D., Liu, P. P., and Myung, K. (2006). Human SHPRH suppresses genomic instability through proliferating cell nuclear antigen polyubiquitination. *J Cell Biol* *175*, 703-708.
- Muller, J., and Kassis, J. A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr Opin Genet Dev* *16*, 476-484.
- Nagarajan, P., Onami, T. M., Rajagopalan, S., Kania, S., Donnell, R., and Venkatachalam, S. (2009). Role of chromodomain helicase DNA-binding protein 2 in DNA damage response signaling and tumorigenesis. *Oncogene* *28*, 1053-1062.
- Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M. S., and Niwa, H. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* *26*, 7772-7782.
- Narducci, M. G., Fiorenza, M. T., Kang, S. M., Bevilacqua, A., Di Giacomo, M., Remotti, D., Picchio, M. C., Fidanza, V., Cooper, M. D., Croce, C. M., *et al.* (2002). TCL1 participates in early embryonic development and is overexpressed in human seminomas. *Proc Natl Acad Sci U S A* *99*, 11712-11717.
- Narducci, M. G., Virgilio, L., Engiles, J. B., Buchberg, A. M., Billips, L., Facchiano, A., Croce, C. M., Russo, G., and Rothstein, J. L. (1997). The murine Tcl1 oncogene: embryonic and lymphoid cell expression. *Oncogene* *15*, 919-926.

- Niakan, K. K., Davis, E. C., Clipsham, R. C., Jiang, M., Dehart, D. B., Sulik, K. K., and McCabe, E. R. (2006). Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol Genet Metab* *88*, 261-271.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* *95*, 379-391.
- Nioi, P., Nguyen, T., Sherratt, P. J., and Pickett, C. B. (2005). The carboxy-terminal Neh3 domain of Nrf2 is required for transcriptional activation. *Mol Cell Biol* *25*, 10895-10906.
- Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* *19*, 5453-5465.
- Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol Cell Biol* *25*, 5084-5094.
- Nishiyama, M., Nakayama, K., Tsunematsu, R., Tsukiyama, T., Kikuchi, A., and Nakayama, K. I. (2004). Early embryonic death in mice lacking the beta-catenin-binding protein Duplin. *Mol Cell Biol* *24*, 8386-8394.
- Nishiyama, M., Oshikawa, K., Tsukada, Y., Nakagawa, T., Iemura, S., Natsume, T., Fan, Y., Kikuchi, A., Skoultchi, A. I., and Nakayama, K. I. (2009). CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat Cell Biol* *11*, 172-182.
- Niwa, H. (2001). Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct Funct* *26*, 137-148.
- Niwa, H. (2007a). How is pluripotency determined and maintained? *Development* *134*, 635-646.
- Niwa, H. (2007b). Open conformation chromatin and pluripotency. *Genes Dev* *21*, 2671-2676.
- Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* *12*, 2048-2060.
- Niwa, H., Masui, S., Chambers, I., Smith, A. G., and Miyazaki, J. (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol Cell Biol* *22*, 1526-1536.
- Niwa, H., Miyazaki, J., and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* *24*, 372-376.
- Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* *460*, 118-122.
- Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* *123*, 917-929.
- Noma, K., Cam, H. P., Maraia, R. J., and Grewal, S. I. (2006). A role for TFIIC transcription factor complex in genome organization. *Cell* *125*, 859-872.
- Nowak, D. E., Tian, B., and Brasier, A. R. (2005). Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* *39*, 715-725.
- Ohta, K., Nobukuni, Y., Mitsubuchi, H., Fujimoto, S., Matsuo, N., Inagaki, H., Endo, F., and Matsuda, I. (1992). Mutations in the Pit-1 gene in children with combined pituitary hormone deficiency. *Biochem Biophys Res Commun* *189*, 851-855.

- Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* *448*, 313-317.
- Okita, K., and Yamanaka, S. (2006). Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Curr Stem Cell Res Ther* *1*, 103-111.
- Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y., Boon, K., Keaveney, M., *et al.* (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *Embo J* *17*, 2019-2032.
- Orkin, S. H., Wang, J., Kim, J., Chu, J., Rao, S., Theunissen, T. W., Shen, X., and Lévassieur, D. N. (2008). The Transcriptional Network Controlling Pluripotency in ES Cells. *Cold Spring Harb Symp Quant Biol*.
- Pan, G., and Thomson, J. A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res* *17*, 42-49.
- Papapetrou, E. P., Tomishima, M. J., Chambers, S. M., Mica, Y., Reed, E., Menon, J., Tabar, V., Mo, Q., Studer, L., and Sadelain, M. (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc Natl Acad Sci U S A* *106*, 12759-12764.
- Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* *20*, 2041-2054.
- Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M. M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* *6*, 382-395.
- Park, I. H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., Lerou, P. H., Lensch, M. W., and Daley, G. Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* *451*, 141-146.
- Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* *27*, 3769-3779.
- Pasini, D., Cloos, P. A., Walfridsson, J., Olsson, L., Bukowski, J. P., Johansen, J. V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* *464*, 306-310.
- Pazin, M. J., Bhargava, P., Geiduschek, E. P., and Kadonaga, J. T. (1997). Nucleosome mobility and the maintenance of nucleosome positioning. *Science* *276*, 809-812.
- Pekarsky, Y., Koval, A., Hallas, C., Bichi, R., Tresini, M., Malstrom, S., Russo, G., Tschlis, P., and Croce, C. M. (2000). Tcl1 enhances Akt kinase activity and mediates its nuclear translocation. *Proc Natl Acad Sci U S A* *97*, 3028-3033.
- Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* *139*, 1290-1302.
- Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. *Nat Methods* *6*, S22-32.
- Pereira, L., Yi, F., and Merrill, B. J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* *26*, 7479-7491.
- Persson, J., and Ekwall, K. (2010). Chd1 remodelers maintain open chromatin and regulate the epigenetics of differentiation. *Exp Cell Res* *316*, 1316-1323.
- Petruk, S., Sedkov, Y., Riley, K. M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J. B., Brock, H. W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* *127*, 1209-1221.
- Pfaffle, R. W., DiMattia, G. E., Parks, J. S., Brown, M. R., Wit, J. M., Jansen, M., Van der Nat, H., Van den Brande, J. L., Rosenfeld, M. G., and Ingraham, H. A. (1992).

- Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science* 257, 1118-1121.
- Phelan, M. L., Schnitzler, G. R., and Kingston, R. E. (2000). Octamer transfer and creation of stably remodeled nucleosomes by human SWI-SNF and its isolated ATPases. *Mol Cell Biol* 20, 6380-6389.
- Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd, and Grant, P. A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433, 434-438.
- Quan, T. K., and Hartzog, G. A. Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of *Saccharomyces cerevisiae* Spt4-Spt5 in transcription. *Genetics* 184, 321-334.
- Radovick, S., Nations, M., Du, Y., Berg, L. A., Weintraub, B. D., and Wondisford, F. E. (1992). A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science* 257, 1115-1118.
- Radziwill, G., Erdmann, R. A., Margelisch, U., and Moelling, K. (2003). The Bcr kinase downregulates Ras signaling by phosphorylating AF-6 and binding to its PDZ domain. *Mol Cell Biol* 23, 4663-4672.
- Rahl, P. B., Lin, C. Y., Seila A. C., Flynn R. A., McCuine S., Burge C. B., Sharp P. A. and Young R. A. (2010). c-Myc regulates transcriptional pause release. *Cell* 141, 432-445.
- Rao, M. (2004). Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev Biol* 275, 269-286.
- Reisman, D., Glaros, S., and Thompson, E. A. (2009). The SWI/SNF complex and cancer. *Oncogene* 28, 1653-1668.
- Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., and Dynlacht, B. D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* 16, 245-256.
- Reya, T., and Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nature* 434, 843-850.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17, 1030-1032.
- Ringrose, L. (2006). Polycomb, trithorax and the decision to differentiate. *Bioessays* 28, 330-334.
- Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* 280, 24731-24737.
- Rodriguez-Paredes, M., Ceballos-Chavez, M., Esteller, M., Garcia-Dominguez, M., and Reyes, J. C. (2009). The chromatin remodeling factor CHD8 interacts with elongating RNA polymerase II and controls expression of the cyclin E2 gene. *Nucleic Acids Res* 37, 2449-2460.
- Roeder, R. G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 579, 909-915.
- Rouleau, N., Domans'kyi, A., Reeben, M., Moilanen, A. M., Havas, K., Kang, Z., Owen-Hughes, T., Palvimo, J. J., and Janne, O. A. (2002). Novel ATPase of SNF2-like protein family interacts with androgen receptor and modulates androgen-dependent transcription. *Mol Biol Cell* 13, 2106-2119.
- Rowland, B. D., and Peeper, D. S. (2006). KLF4, p21 and context-dependent opposing forces in cancer. *Nat Rev Cancer* 6, 11-23.
- Saha, A., Wittmeyer, J., and Cairns, B. R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* 7, 437-447.

- Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., and Nishinakamura, R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* *133*, 3005-3013.
- Sakamoto, I., Kishida, S., Fukui, A., Kishida, M., Yamamoto, H., Hino, S., Michiue, T., Takada, S., Asashima, M., and Kikuchi, A. (2000). A novel beta-catenin-binding protein inhibits beta-catenin-dependent Tcf activation and axis formation. *J Biol Chem* *275*, 32871-32878.
- Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* *20*, 3246-3248.
- Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. *Science* *311*, 1118-1123.
- Sato, A., Kishida, S., Tanaka, T., Kikuchi, A., Kodama, T., Asashima, M., and Nishinakamura, R. (2004). Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt signaling by localizing to heterochromatin. *Biochem Biophys Res Commun* *319*, 103-113.
- Savarese, F., Davila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C. F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A. G., and Grosschedl, R. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev* *23*, 2625-2638.
- Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., Zhang, X., Song, L., Wang, Z., Laframboise, T., *et al.* (2009). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res* *19*, 590-601.
- Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., Adams, D. J., Flicek, P., Crawford, G. E., Laframboise, T., *et al.* (2010). CHD7 Targets Active Gene Enhancer Elements to Modulate ES Cell-Specific Gene Expression. *PLoS Genet* *6*, e1001023.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* *128*, 735-745.
- Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X. Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* *38*, 700-705.
- Schwartz, Y. B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* *8*, 9-22.
- Sharov, A. A., Masui, S., Sharova, L. V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M. S. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* *9*, 269.
- Shen, X., Kim, W., Fujiwara, Y., Simon, M. D., Liu, Y., Mysliwiec, M. R., Yuan, G. C., Lee, Y., and Orkin, S. H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* *139*, 1303-1314.
- Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature* *406*, 541-544.
- Shen, X., Ranallo, R., Choi, E., and Wu, C. (2003). Involvement of actin-related proteins in ATP-dependent chromatin remodeling. *Mol Cell* *12*, 147-155.
- Shindo, T., Manabe, I., Fukushima, Y., Tobe, K., Aizawa, K., Miyamoto, S., Kawai-Kowase, K., Moriyama, N., Imai, Y., Kawakami, H., *et al.* (2002). Kruppel-like zinc-finger transcription factor KLF5/BTEB2 is a target for angiotensin II signaling and an essential regulator of cardiovascular remodeling. *Nat Med* *8*, 856-863.

- Shur, I., and Benayahu, D. (2005). Characterization and functional analysis of CReMM, a novel chromodomain helicase DNA-binding protein. *J Mol Biol* *352*, 646-655.
- Shur, I., Socher, R., and Benayahu, D. (2006). In vivo association of CReMM/CHD9 with promoters in osteogenic cells. *J Cell Physiol* *207*, 374-378.
- Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* *138*, 722-737.
- Sims, R. J., 3rd, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* *280*, 41789-41792.
- Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H. D., and Cordes, S. P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* *138*, 885-897.
- Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol* *15*, 259-267.
- Smith, A. G. (2001). Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* *17*, 435-462.
- Song, J. J., and Kingston, R. E. (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* *283*, 35258-35264.
- Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* *6*, 846-856.
- Srinivasan, S., Armstrong, J. A., Deuring, R., Dahlsveen, I. K., McNeill, H., and Tamkun, J. W. (2005). The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. *Development* *132*, 1623-1635.
- Stephens, R. M., and Schneider, T. D. (1992). Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J Mol Biol* *228*, 1124-1136.
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.
- Suda, Y., Suzuki, M., Ikawa, Y., and Aizawa, S. (1987). Mouse embryonic stem cells exhibit indefinite proliferative potential. *J Cell Physiol*, 197-201.
- Sun, C., Nakatake, Y., Akagi, T., Ura, H., Matsuda, T., Nishiyama, A., Koide, H., Ko, M. S., Niwa, H., and Yokota, T. (2009). Dax1 binds to Oct3/4 and inhibits its transcriptional activity in embryonic stem cells. *Mol Cell Biol* *29*, 4574-4583.
- Sun, C., Nakatake, Y., Ura, H., Akagi, T., Niwa, H., Koide, H., and Yokota, T. (2008). Stem cell-specific expression of Dax1 is conferred by STAT3 and Oct3/4 in embryonic stem cells. *Biochem Biophys Res Commun* *372*, 91-96.
- Sun, H., Lesche, R., Li, D. M., Liliental, J., Zhang, H., Gao, J., Gavrilova, N., Mueller, B., Liu, X., and Wu, H. (1999). PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proc Natl Acad Sci U S A* *96*, 6199-6204.
- Sun, Y., Jiang, X., Chen, S., Fernandes, N., and Price, B. D. (2005). A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc Natl Acad Sci U S A* *102*, 13182-13187.
- Surapureddi, S., Yu, S., Bu, H., Hashimoto, T., Yeldandi, A. V., Kashireddy, P., Cherkaoui-Malki, M., Qi, C., Zhu, Y. J., Rao, M. S., and Reddy, J. K. (2002). Identification of a transcriptionally active peroxisome proliferator-activated receptor alpha -interacting

- cofactor complex in rat liver and characterization of PRIC285 as a coactivator. *Proc Natl Acad Sci U S A* *99*, 11836-11841.
- Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C., and Izpisua Belmonte, J. C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proc Natl Acad Sci U S A* *103*, 10294-10299.
- Szutorisz, H., and Dillon, N. (2005). The epigenetic basis for embryonic stem cell pluripotency. *Bioessays* *27*, 1286-1293.
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* *11*, 1553-1558.
- Tai, H. H., Geisterfer, M., Bell, J. C., Moniwa, M., Davie, J. R., Boucher, L., and McBurney, M. W. (2003). CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins. *Biochem Biophys Res Commun* *308*, 170-176.
- Takahashi, K., Mitsui, K., and Yamanaka, S. (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* *423*, 541-545.
- Takahashi, K., Murakami, M., and Yamanaka, S. (2005). Role of the phosphoinositide 3-kinase pathway in mouse embryonic stem (ES) cells. *Biochem Soc Trans* *33*, 1522-1525.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* *131*, 861-872.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* *126*, 663-676.
- Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev* *9*, 1211-1222.
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* *14*, 1025-1040.
- Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* *455*, 1124-1128.
- Taya, S., Yamamoto, T., Kanai-Azuma, M., Wood, S. A., and Kaibuchi, K. (1999). The deubiquitinating enzyme Fam interacts with and stabilizes beta-catenin. *Genes Cells* *4*, 757-767.
- Teitell, M. A. (2005). The TCL1 family of oncoproteins: co-activators of transformation. *Nat Rev Cancer* *5*, 640-648.
- Thompson, B. A., Tremblay, V., Lin, G., and Bochar, D. A. (2008). CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* *28*, 3894-3904.
- Thompson, P. M., Gotoh, T., Kok, M., White, P. S., and Brodeur, G. M. (2003). CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. *Oncogene* *22*, 1002-1011.
- Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D., *et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* *464*, 1082-1086.
- Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* *83*, 1021-1026.

- Umlauf, D., Goto, Y., and Feil, R. (2004). Site-specific analysis of histone methylation and acetylation. *Methods Mol Biol* 287, 99-120.
- Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M., and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods* 5, 829-834.
- van Attikum, H., and Gasser, S. M. (2005). The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol* 6, 757-765.
- van den Berg, D. L., Snoek, T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R. A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 6, 369-381.
- van den Berg, D. L., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R. A. (2008). Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol* 28, 5986-5995.
- van der Weyden, L., Adams, D. J., and Bradley, A. (2002). Tools for targeted manipulation of the mouse genome. *Physiol Genomics* 11, 133-164.
- Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P. B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* 388, 598-602.
- Vert, J. P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* 7, 520.
- Vissers, L. E., van Ravenswaaij, C. M., Admiraal, R., Hurst, J. A., de Vries, B. B., Janssen, I. M., van der Vliet, W. A., Huys, E. H., de Jong, P. J., Hamel, B. C., *et al.* (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 36, 955-957.
- Walfridsson, J., Khorosjutina, O., Matikainen, P., Gustafsson, C. M., and Ekwall, K. (2007). A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly. *Embo J* 26, 2868-2879.
- Wang, H. B., and Zhang, Y. (2001). Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor. *Nucleic Acids Res* 29, 2517-2521.
- Wang, J., Rao, S., Chu, J., Shen, X., Lévassieur, D. N., Theunissen, T. W., and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444, 364-368.
- Wang, K., Sengupta, S., Magnani, L., Wilson, C. A., Henry, R. W., and Knott, J. G. (2010). Brg1 is required for Cdx2-mediated repression of Oct4 expression in mouse blastocysts. *PLoS One* 5, e10622.
- Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., *et al.* (2009a). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* 138, 660-672.
- Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., and Zhao, K. (2009b). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138, 1019-1031.
- Watanabe, S., Umehara, H., Murayama, K., Okabe, M., Kimura, T., and Nakano, T. (2006). Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* 25, 2697-2707.
- Wei, C. L., Miura, T., Robson, P., Lim, S. K., Xu, X. Q., Lee, M. Y., Gupta, S., Stanton, L., Luo, Y., Schmitt, J., *et al.* (2005). Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells* 23, 166-185.
- Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science* 193, 848-856.

- Wernig, M., Meissner, A., Cassady, J. P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* 2, 10-12.
- Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448, 318-324.
- Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolnikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., and Shi, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* 125, 467-481.
- Wong, L. H., McGhie, J. D., Sim, M., Anderson, M. A., Ahn, S., Hannan, R. D., George, A. J., Morgan, K. A., Mann, J. R., and Choo, K. H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* 20, 351-360.
- Wu, Q., Chen, X., Zhang, J., Loh, Y. H., Low, T. Y., Zhang, W., Zhang, W., Sze, S. K., Lim, B., and Ng, H. H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 281, 24090-24094.
- Yamanaka, S., Li, J., Kania, G., Elliott, S., Wersto, R. P., Van Eyk, J., Wobus, A. M., and Boheler, K. R. (2008). Pluripotency of embryonic stem cells. *Cell Tissue Res* 331, 5-22.
- Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmJc-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* 125, 483-495.
- Yamashina, K., Yamamoto, H., Chayama, K., Nakajima, K., and Kikuchi, A. (2006). Suppression of STAT3 activity by Duplin, which is a negative regulator of the Wnt signal. *J Biochem* 139, 305-314.
- Yang, J., Gao, C., Chai, L., and Ma, Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS One* 5, e10766.
- Yates, J. A., Menon, T., Thompson, B. A., and Bochar, D. A. (2010). Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8. *FEBS Lett* 584, 689-693.
- Ye, T., Krebs, A. R., Choukrallah, A. M., Plewniak, F., Davidson, I., and Tora, L. (2010). SeqMINER: A genome wide mapping data interpretation platform. To be published.
- Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003a). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 115, 281-292.
- Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* 21, 183-186.
- Yu, H. B., Kunarso, G., Hong, F. H., and Stanton, L. W. (2009). Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *J Biol Chem* 284, 31327-31335.
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917-1920.
- Yuan, C. C., Zhao, X., Florens, L., Swanson, S. K., Washburn, M. P., and Hernandez, N. (2007). CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription. *Mol Cell Biol* 27, 8729-8738.
- Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y., Aburatani, H., Ura, K., Koseki, H., Niwa, H., and Nishinakamura, R. (2009). Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression. *Stem Cells* 27, 796-805.

- Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952-1958.
- Zeng, P. Y., Vakoc, C. R., Chen, Z. C., Blobel, G. A., and Berger, S. L. (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* 41, 694, 696, 698.
- Zhang, J. Z., Gao, W., Yang, H. B., Zhang, B., Zhu, Z. Y., and Xue, Y. F. (2006). Screening for genes essential for mouse embryonic stem cell self-renewal using a subtractive RNA interference library. *Stem Cells* 24, 2661-2668.
- Zhang, L., Schroeder, S., Fong, N., and Bentley, D. L. (2005). Altered nucleosome occupancy and histone H3K4 methylation in response to 'transcriptional stress'. *Embo J* 24, 2379-2390.
- Zhang, X., Zhang, J., Wang, T., Esteban, M. A., and Pei, D. (2008). Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* 283, 35825-35833.
- Zhang, Z., Jones, S., Hagood, J. S., Fuentes, N. L., and Fuller, G. M. (1997). STAT3 acts as a co-activator of glucocorticoid receptor signaling. *J Biol Chem* 272, 30607-30610.
- Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W. K., Shahab, A., Kuznetsov, V. A., *et al.* (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* 1, 286-298.
- Zhou, Q., Chipperfield, H., Melton, D. A., and Wong, W. H. (2007). A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 104, 16438-16443.
- Zindy, F., Eischen, C. M., Randle, D. H., Kamijo, T., Cleveland, J. L., Sherr, C. J., and Roussel, M. F. (1998). Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes Dev* 12, 2424-2433.
- Zipori, D. (2004). The nature of stem cells: state rather than entity. *Nat Rev Genet* 5, 873-878.

Annexes

I. Figures supplémentaires

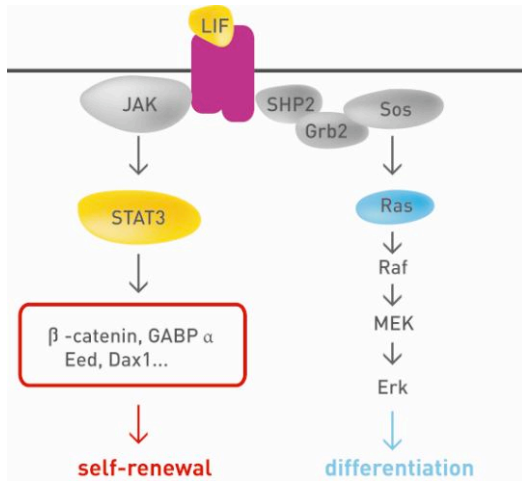


Figure S1: Voie de signalisation de LIF, largement inspiré de YOKOTA, T. <http://www.m.kanazawa-u.ac.jp/eng/gaiyou/dep02/dep2-08.html>

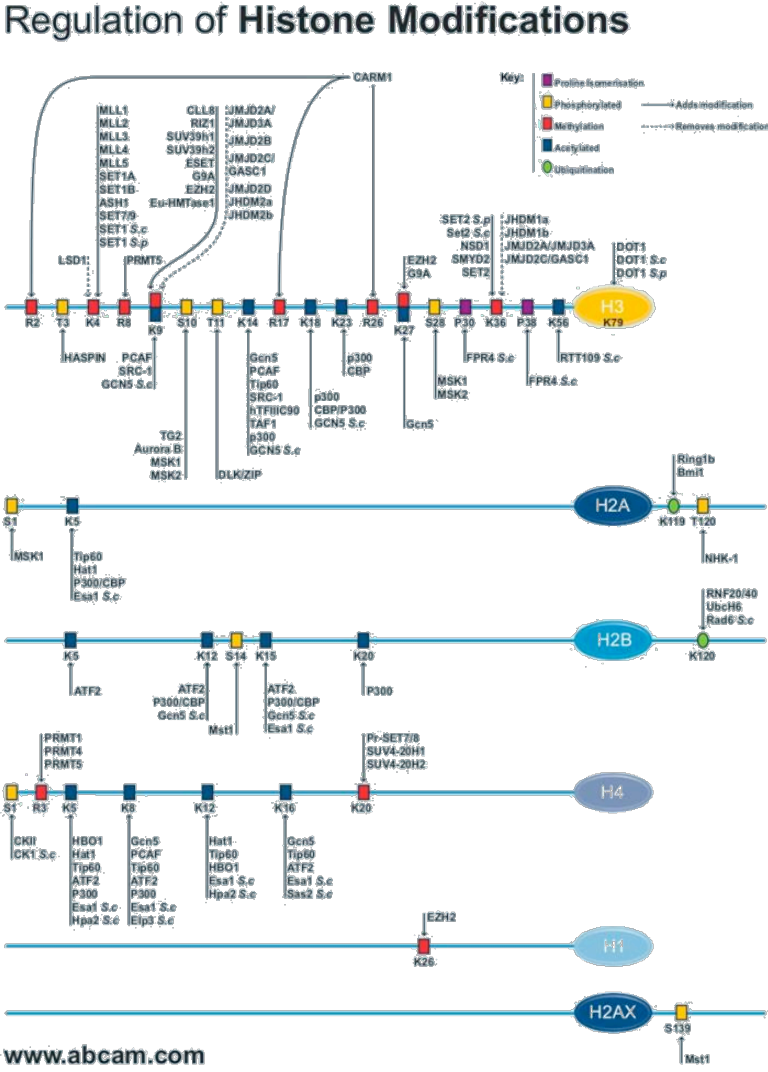


Figure S2: Modifications d'histones et enzymes modificateuses d'après Abcam

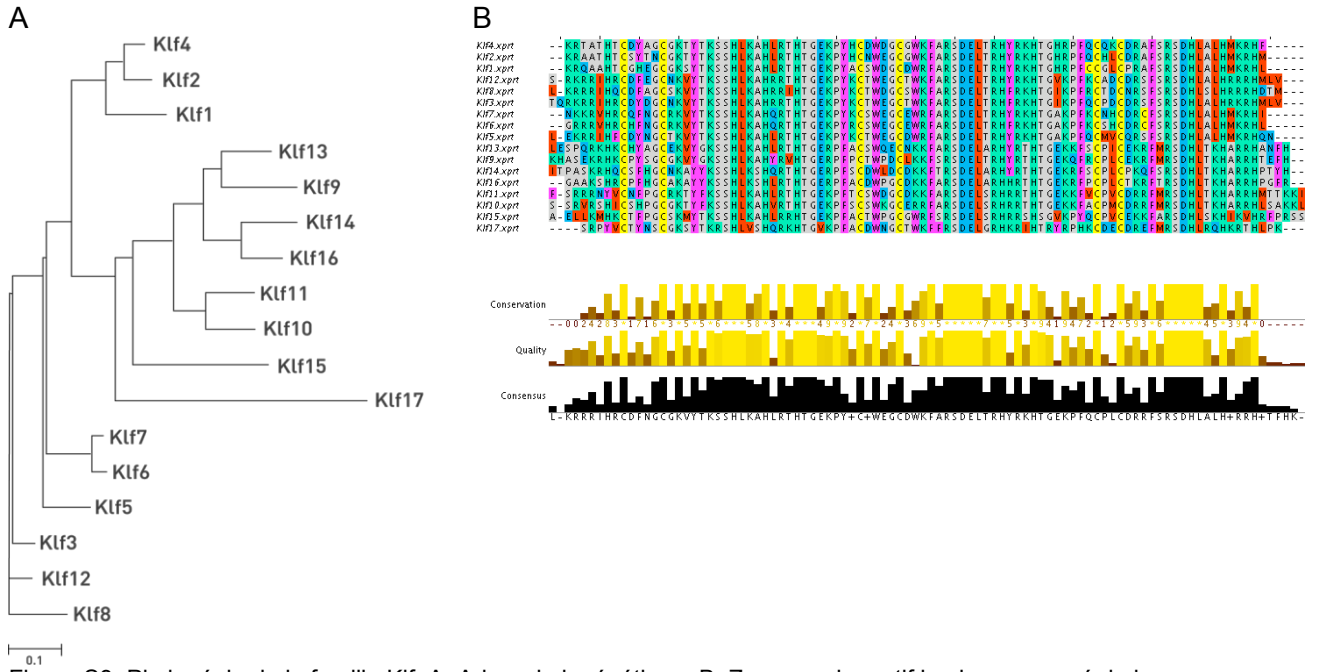


Figure S3: Phylogénie de la famille Klf. A: Arbre phylogénétique. B: Zoom sur le motif le plus conservé de la famille.

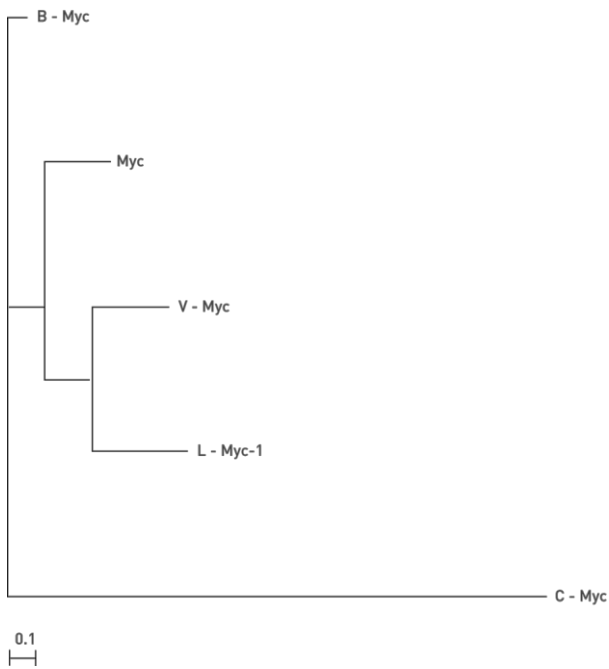


Figure S4: Phylogénie de la famille Myc. Le gène V-Myc correspond à n-Myc

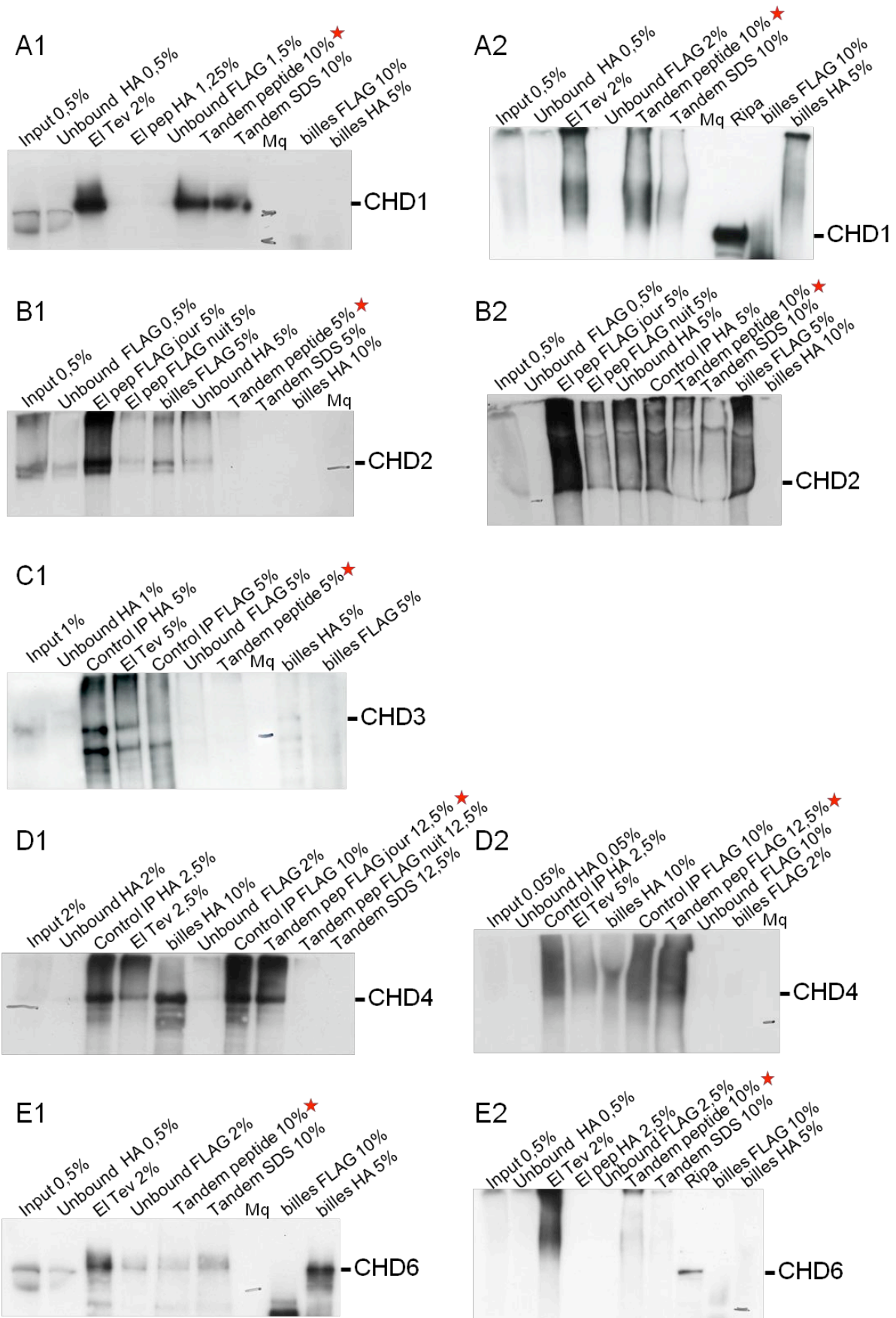


Figure S6: Vérification par Western Blot de la qualité des préparations de chromatine immunoprécipités dont l'ADN a été séquencé sur plateforme Illumina. EI: Elution; Unbound: fraction non liée. Pep: peptide; Tandem: dernière fraction éluée après deux étapes d'immunoprécipitation; Mq: marqueur de taille moléculaire (la seule

bande visible sur tous les gels est la bande à 260KDa). La fraction qui a été envoyée au séquençage est indiquée par une étoile rouge.

Protéine	Poids (Kda)
Chd1	196,38
Chd2	210,8
Chd3	228,97
Chd4	217,55
Chd5	222,68
Chd6	305,39
Chd7	334
Chd8	280,8
Chd9	321,89
Hdac2	55,33
Atad2	155,28
Chaf1a	101,93

Tableau S3: Poids en KDa des différentes protéines qui ont été testées avec le protocole final

Données ChIP-seq publiées dans les ES murines

Auteur	Facteurs	Numéro d'accèsion	génomme	Année
Mikkelsen et al.	H3K4me3, H3K27me3, H3K36me3, H3K9me3, H4K20me3, RPol2, H3	GSE12241	mm8	2007
Chen et al.	Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1, CTCF	GSE11431	mm8	2008
Ku et al.	Suz12, Ezh2, Ring1B	GSE13084	mm8	2008
Marson et al.	Oct4, Nanog, Sox2, Tcf3	GSE11724	mm8	2008
Meissner et al.	H3K4me1, H3K4me2	GSE11172	mm8	2008
Bilodeau et al.	SetDB1, H3K9me3	GSE18371		2009
Bilodeau et al.	TBP	GSE22303	mm8	2010
Blackledge et al.	Kdm2a	GSE21202		2010
de Dieuleveult et al.	Chd1, Chd2, Chd3, Chd6, Chd8, Chd9		mm9	2010
Golberg et al.	H3.3, Pol2	GSE16893	mm9	2010
Hmitou et al.	Smarca4, Smarca5, Alc1, Atrx, Smarca3, Smarcd1		mm9	2010
Jounier et al.	Hdac2, Chd4		mm9	2010
Krebs et al.	Gcn5, Luzp1	GSE21717		2010
Li et al.	Jarid2, MTF2	GSE19708		2010
Peter et al	RNAPII_ser7P	GSE21917	mm8	2010
Rahl et al	RNAPII_ser2P, RNAPII_ser5P	GSE15519	mm8	2010
Rugg-Gunn et al	H3K4me3, H3K27me3	GSE15519	mm8	2010
Schnetz et al.	Chd7, p300	GSE22341		2010
Seila et al	RNAPII_hypoP	GSE12680	mm8	2010

Tableau S4: Vue globale de toutes les données ChIP-seq disponibles en juillet 2010

Protéine	Fixation	Nombre de pics
Pol2	F	15741
Esrrb	F	7066
Hdac2 DSG	DSG	12361
Hdac2	F	1697
Klf4	F	14263
Nanog	F	14087
nMyc	F	9130
Oct4	F	5454
p300	F	470
Smad1	F	878
smarca4	F	13556
smarca5	F	15115
smarcad1	F	2829
Sox2	F	5080
STAT3	F	3117
Tcfcp2l1	F	34220
Zfx	F	17534
Suz12 Chen	F	3657
Suz12 Ku	F	2340
Ezh2	F	5681
Ring1B	F	1300

Tableau S5: Tableau récapitulatif du nombre de pics détectés pour chaque protéine avec le logiciel SICER

	paramètres	nb de	
QUEST	1.2	1250	régions
	2.2	320	
	3.2	111	
	3.3	2901	
SICER	200 600 FDR-3	20121	fenêtres
	200 600 FDR-9	13265	
SISSR		4660	pics

Tableau 6: Comparaison des nombres de pics obtenus avec différents algorithmes et paramètres

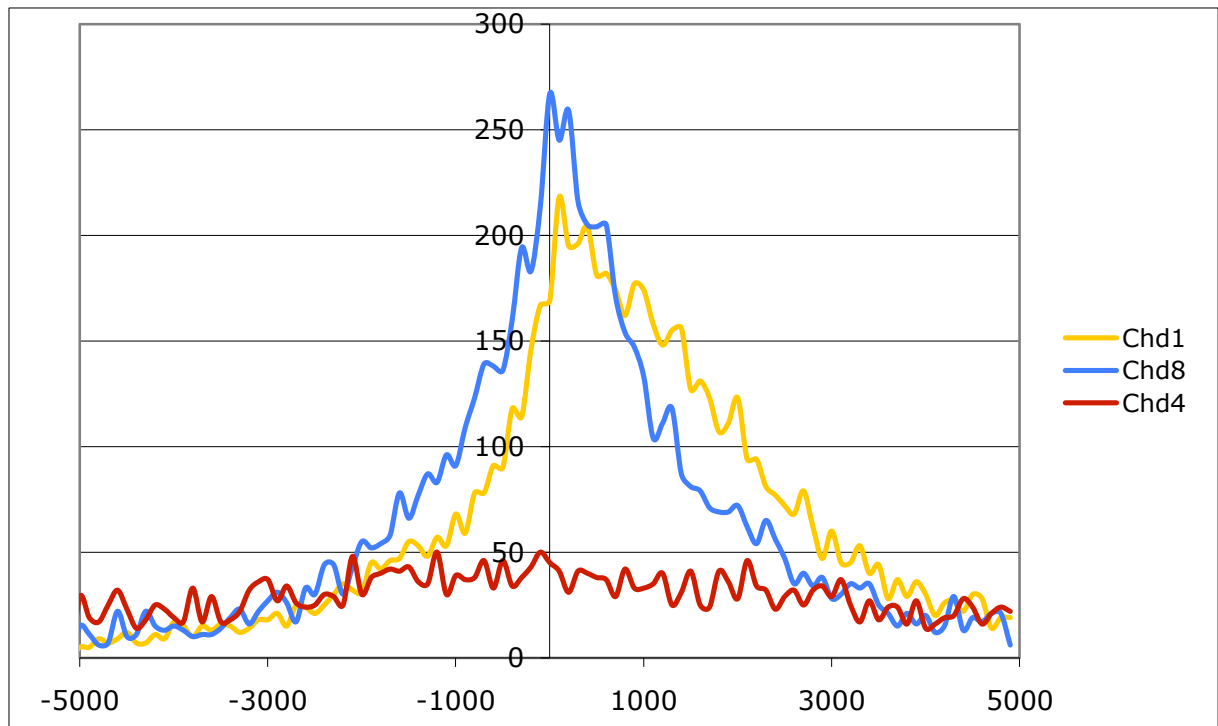


Figure 7: Occupation de CHD1, CHD4 ou CHD8 sur leurs propres promoteurs cibles. L'enrichissement donné est le nombre gènes dont le milieu du pic tombe dans un intervalle de 100 paires de bases.

Gene	Chd1	Chd4	Chd8	Name
Polycomb subunit genes				
NM_199196	1	0	1	Suz12
NM_001042623	1	0	1	Phc1
NM_013627	0	1	0	Pax6
NM_023699	0	0	1	nfatc4
NM_011277	1	0	0	Ring1b
NM_009066	0	0	1	Ring1a
NM_001163307	0	0	1	Mel18
NM_021876	0	1	1	Eed
NM_144811	1	0	1	Cbx7
NM_007971	1	0	1	Ezh2
Differentiation genes				
NM_010202	1	1	1	Fgf4
NM_010144	1	1	0	Ephb4.b
NM_010111	1	0	0	Efnb2
NM_001110266	0	1	0	Vegfa4
NM_001025257	1	0	1	Vegfa3
Core pluripotency circuit				
NM_028610	1	1	1	Dppa4.1
NM_028615	1	1	1	Dppa2
NM_010849	1	1	1	Myc
NM_001013745	0	0	1	Zfp57
NM_009559	1	0	0	Zfp57
NM_028016	0	0	1	Nanog
NM_013633	1	1	1	Oct4
NM_013611	1	1	1	Nodal
NM_175303	1	1	1	Sall4a
NM_010637	1	0	0	Klf4
NM_008452	1	0	1	Klf2
NM_023844	1	1	1	Jam2
NM_001164152	1	0	1	Tcf3.6
NM_021390	1	0	0	Sall1

Tableau S7: Présence (1) et absence (0) des protéines Chd1, Chd4 et Chd8 aux promoteurs de certains gènes dans les ES murines

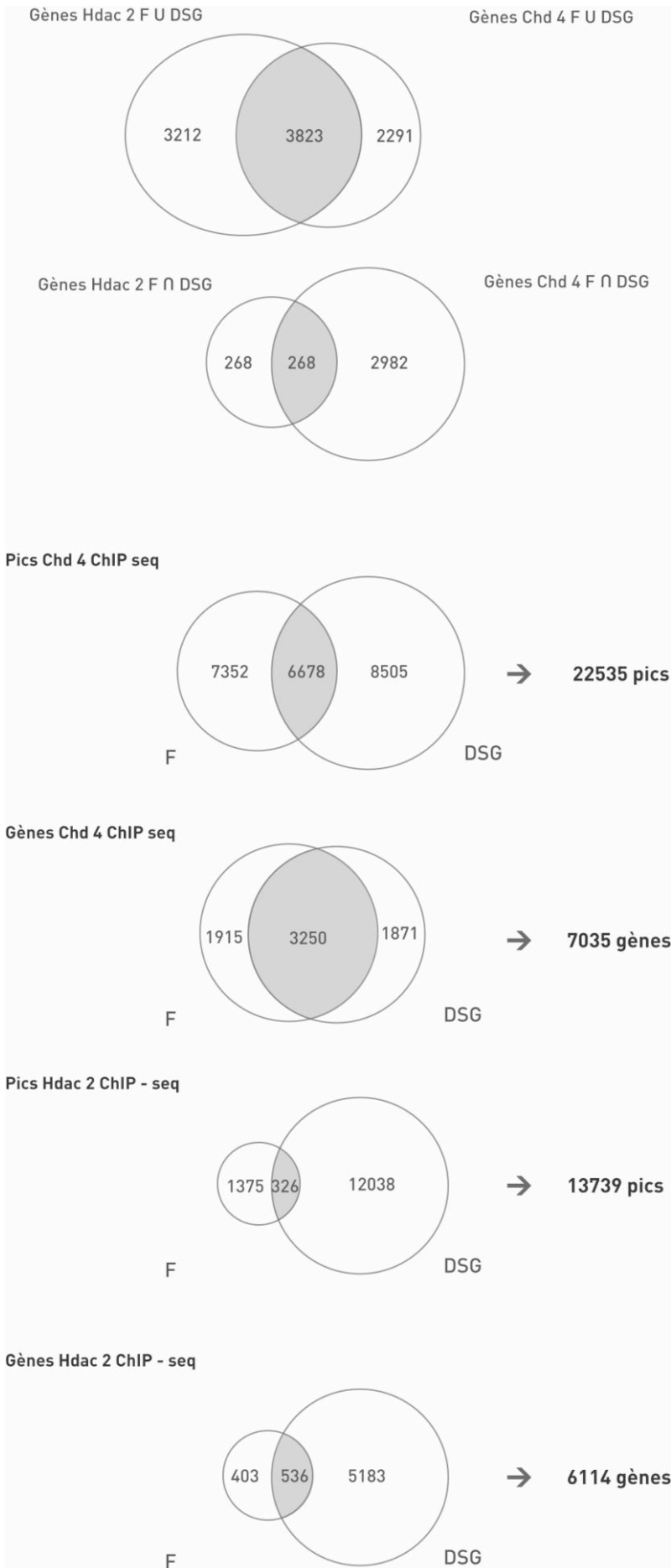


Figure S8: Comparaison des données de positions de la protéine Chd4 fixation DSG ou formaldéhyde et Hdac2 DSG ou formaldéhyde. Les pics correspondent aux régions liées alors que les gènes sont le(s) gène(s) le(s) plus proche(s) associé au pic détecté. De nombreux pics se trouvent hors d'une région génique (enhanceur-promoteur-ORF-terminateur) ou plusieurs pics peuvent correspondre à un même gène c'est pourquoi ce nombre est plus restreint.

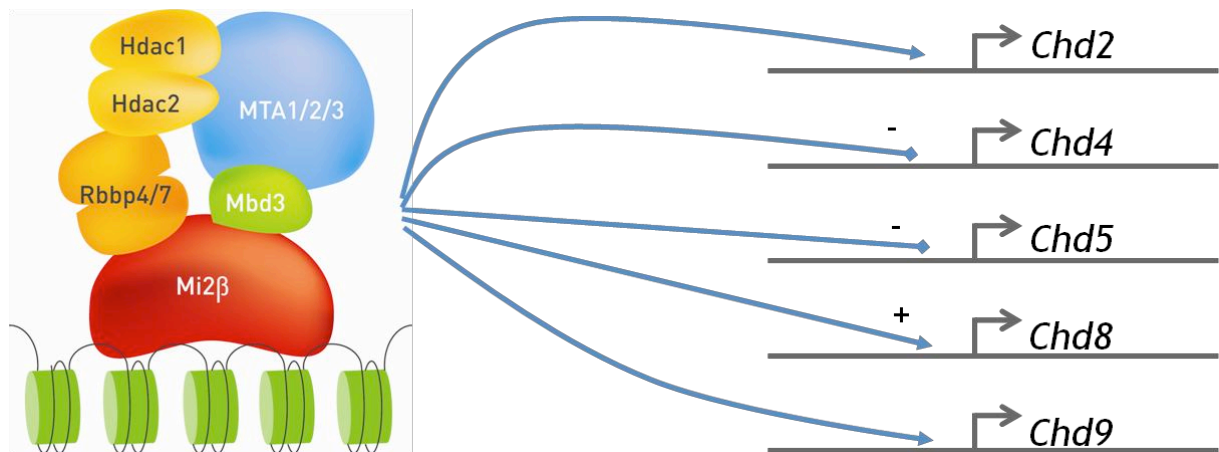


Figure S9: NuRD régule l'expression de certaines protéines Chd. Les expériences de RNA-seq ont permis de montrer que pour Chd5 et Chd4 il s'agissait d'une inhibition et pour Chd8 une stimulation de l'expression, mais pour les gènes Chd2 et Chd9, les données sont contradictoires entre Chd4^{hypo/hypo} et Hdac2^{-/-}.

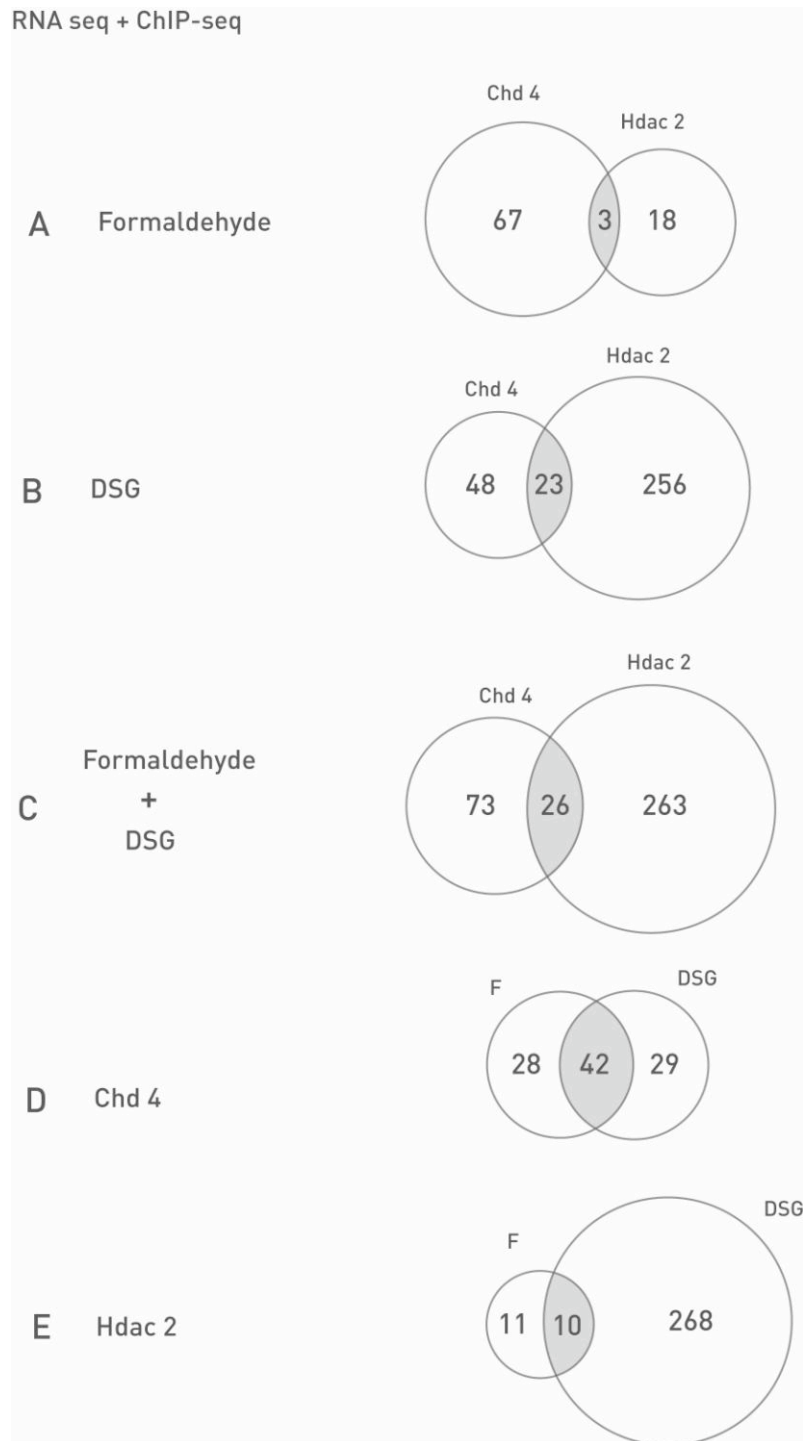


Figure 10: Représentation schématique des données ChIP-seq et RNA-seq sur le complexe NuRD. A

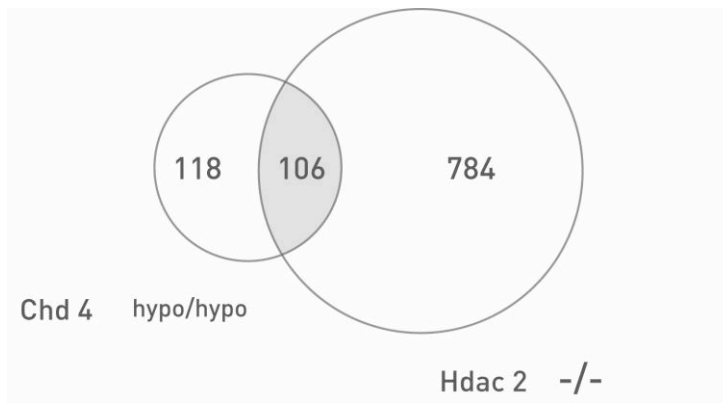


Figure 11: Analyse des données RNA-seq seules. Chd4 et Hdac2 régulent 106 en commun (Attention: mais pas uniquement sans le même sens).

II. Protocole de recombineering

Par Sébastien Graziani

Protocole Recombineering (Etapas de A à Z)

But : Insérer par recombineering un TAG en C-ter d'un gène dans un locus de souris.

I) Constructions in silico

Localiser le gène d'intérêt avec ENSEMBL.

Dans les options, sélectionner dans DAS source l'option [129S7/AB2.2 clones](#) pour commander le BAC. Comme le TAG doit être inséré en C-ter choisir 2 BAC qui pourront contenir 10 kb de part et d'autre du C-ter. Les commander (Geneservice).

Extraire sous forme FASTA la séquence -10 kb - « C-ter (codon stop) »- +10 kb et la mettre dans strider.

Repérer précisément le codon stop et adopter une stratégie pour le génotypage des cellules ES. Ce génotypage se fera par Southern blot. Il faut choisir une enzyme de restriction qui coupe idéalement 2 fois (une à l'extérieur des miniarms, et l'autre à l'intérieur). Cette enzyme doit être une de celles utilisées par le labo pour les southern (ne doit pas être inhibée par CpG méthylation).

BamHI, NcoI, EcoRI, XhoI, et XbaI sont apportés par la construction. En plus de ces sites le plasmide pL452-FHH (fayçal) comprend, après le stop du TAG, les sites de restriction pour EcoRV et HindIII. Le plasmide pL452-FHHseb (Sébastien) contient les sites KpnI et XbaI :

Placer la sonde à l'extérieur des miniarms 5' ou 3'. Cette sonde ne doit pas être disposée dans des régions répétées. Pour le vérifier utiliser le programme Repeat Masker et sélectionner le génome de la souris !!!

Choisir les amorces PCR adéquates avec Primer3 Input pour générer une sonde de 600-700 pb (inclure de part et d'autre un site EcoRI pour le clonage futur dans bluescript : le but est de fabriquer la sonde en purifiant le plasmide la possédant en insert).

IMPORTANT : Si possible choisir 2 stratégies une avec une sonde en 5' et l'autre avec une sonde en 3' pour pouvoir voir un shift de 2 kb quand la K7 Néo sera excisée.

Tenir compte du placement des miniarms (la sonde doit être à l'extérieur des miniarms.

En fonction de l'enzyme choisie, définir le plasmide pL452 qui sera utilisé (Fayçal ou sébastien).

Le profil entre le WT et le mutant en southern devra être assez différent :

Ex : Si pour le WT on obtient une bande à 10 kb, dans l'idéal on doit avoir une bande à 5 kb pour le mutant.

Une fois la sonde placée, disposer les miniarms 5' et 3'. Ils doivent être idéalement placés à respectivement -5 kb et +5 kb du Stop et comprendre entre 250-500 pb. Dans tous les cas s'assurer qu'il y a 10 kb entre les miniarms. On peut avoir un bras court à 3kb et un bras long à 7kb.

Choisir les amorces adéquates avec Primer3 input.

Comme les miniarms seront amplifiés par PCR et clonés dans le retrieval plasmid (en Not I / Spe I pour le miniarm 5' et Spe I / Bam HI pour le miniarm 3') vérifier qu'il n'y a pas de sites de restriction Not I / Spe I dans le 5' et Spe I / Bam HI dans le 3'.

Ajouter dans l'amorce FW du miniarm 5' le site Not I, et dans l'amorce Rev le site SpeI.

Ajouter dans l'amorce FW du miniarm 3' le site Spe I, et dans l'amorce Rev le site Bam HI.

Les bras d'homologie doivent être compris entre 250-500 pb. L'amorce Rev du bras d'homologie 5' comprendra les 20 nucléotides avant le codon stop. L'amorce FW du bras d'homologie 3' comprendra les 20 nucléotides suivants codon stop compris.

Ajouter dans l'amorce FW du miniarm 5' le site Sal I, et dans l'amorce Rev le site Asc I.

Ajouter dans l'amorce FW du miniarm 3' le site Bam HI, et dans l'amorce Rev le site Not I.

IMPORTANT : Le bras d'homologie 5' ne devra pas contenir de sites Sal I et Asc I. Le 3' ne devra pas posséder de sites Bam HI et Not I.

Rappatrier dans strider les séquences des miniarms, des bras d'homologie, des sondes pour toutes les constructions, et les séquences de 20 kb flanquant le stop du gène d'intérêt.

Construire dans strider la carte du Gap repaired plasmid, du retrieval plasmid, du minitargeting vector, du retrieval plasmid et du neotargeted plasmid.

II Insertion du BAC dans la souche SW102 et Vérifications

Préculture des bactéries renfermant les BAC :

5 mL LB/Chloramphénicol 12,5µg/ml final.

Préparer un stock glycérol

2 échantillons par BAC. Mettre 1,2 mL de préculture avec 0,6 mL de glycérol 75% stérile.

Purifier les 2 BAC sélectionnés.

Miniprep d'ADN plasmidique:

1) Culture ON à 37°C sur 3 ml LB liquide + antibio ici : Chloramphénicol.

2) Récolter les cellules par centrifugation 2 min dans eppendorfs 2 ml.

- 3) Resuspension dans 100µl de solution I puis vortexer.
- 4) Ajouter 200µl de solution II puis inverser (5-6 fois).
- 5) Ajouter 150µl d'Acétate de sodium 3M pH 5,2 puis inverser (5-6 fois).
- 6) Centrifuger 10 min puis récupérer le surnageant dans un eppendorf 1,5 ml.
- 7) Ajouter 1ml d'Ethanol 100% (-20°C) puis laisser précipiter 10 min à -20°C.
- 8) Centrifuger 10 min à 13000 rpm et enlever le surnageant.
- 9) Laver le culot avec 1ml d'éthanol 70% puis centrifuger 10 min à 13000 rpm.
- 10) Enlever le surnageant et sécher le culot 15 min dans l'étuve à 37°C.
- 11) Resuspendre le culot dans 50µl d'eau stérile.

On doit obtenir environ 10µg pour une culture en LB Amp (digestion sur 2µl)

Solution II à préparer extemporanément :

NaOH 200mM

SDS 1%

Doser l'ADN pour vérifier sa pureté et la quantité. (faire une dilution au 1/100 et doser à 260 nm).

Vérifier sur gel le profil des BAC préalablement digérés par une, deux voire 3 enzymes de restriction (une, deux ou 3 digestions indépendantes). PS : j'ai vérifié mes BAC avec Kpn I et Eco RI.

Réaction :

10µl ADN BAC (miniprep)

3µL Tampon 10X

16µl eau_{dd}

1µl Enzyme

Incubation à 37°C pdt 3hrs au bain marie.

Migration sur gel agarose 1% ON à 40 V. (de 18h30 à 9h00).

Transformation par électroporation des cellules **SW102**

a) Pré-culture de 5 ml LB ON à **30-32 °C**.

b) Mettre de l'eau, des falcons 50 mL des eppendorfs et des cuvettes à 4°C.

c) Diluer au 50^{ème} (500µL dans 25 mL) et incuber 3-5h jusqu'à atteindre une DO₆₀₀ =0,6.

25 mL de culture correspond à 2 tubes de cellules compétentes.

d) Incuber les cellules dans la glace pdt 2 min puis les transférer (12,5 ml) dans 2 falcons 50 mL préalablement à 4°C.

e) Centrifuger à 4°C, 5000 rpm pdt 5 min.

Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le

falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.

f) Centrifuger à 4°C, 5000 rpm pdt 5 min.

g) Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois).

Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot). Resuspendre le culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 200µL par falcon).

h) Transférer 50µL de cellules electrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger 2 µL d'ADN. Pour un BAC utiliser 1-5µg d'ADN (ici environ 10 µg).

Transformer par électroporation (1750 V). Vérifier le time constant. Il doit être aux alentours de 5-6mv.

i) Transférer les cellules dans 1 mL de LB puis incubé à 30°C pdt 1h.

j) Étaler les bactéries (30 µL 150µL et le reste) sur le milieu de sélection ici : Chloramphénicol. Centrifuger à 5000 rpm et enlever le surnageant pour le reste. Ressuspendre dans 150µl et étaler.

k) Incuber à 30-32°C pour 16-24h.

NB : Inclure un contrôle sans BAC et l'étaler sur une boîte LB et une LB/Chloramphénicol.

Le lendemain :

Les colonies sont très petites. Il est préférable de les laisser une bonne partie de la journée.

Repiquer 5 clones dans 3 ml de LB/Chlor ON à 30°C.

Préparer un stock glycérol

1 échantillon par BAC. Mettre 1,2 mL de préculture avec 0,6 mL de glycérol 75% stérile.

Purifier les BAC à partir de la souche SW102

Miniprep d'ADN plasmidique Cf protocole.

Vérifier sur gel le profil des BAC préalablement digérés par une, deux voire 3 enzymes de restriction (une, deux ou 3 digestions indépendantes). PS : j'ai vérifié mes BAC avec Kpn I et EcoR I.

Digérer le BAC original (issu de la souche E. coli initiale) et 2 BAC issus de la souche SW102.

Comparer les profils et valider le BAC.

Vérifier les BAC par PCR. cf 19).

III) Vérification des couples d'amorces pour PCR (Sonde, miniarms et bras d'homologie).

Mix réactionnel :

1µl ADN BAC
 1µl Amorces FW 10µM
 1µl Amorces Rev 10µM
 2,5µl Tampon 10X
 0,2µl Taq
 1ul MgCl₂ 50mM
 17,3 µl ea_{dd}

Programme PCR :
 PCR Bac 60 avec
 ' à 94°C
 40" à 72°C
 Go To 29

Faire migrer 5-10 µl de la PCR dans un gel 1% agarose. Stocker à 4°C le reste de la PCR (important pour la suite) !!!!!

Amplification par PCR des miniarms, bras d'homologie et sonde pour clonage.

Refaire les PCR mais cette fois faire 4 tubes pour chaque produit de PCR (pour générer suffisamment de matériel pour le clonage). Faire un contrôle négatif pour chaque PCR (chaque couple d'amorces). Cf p32 cahier 1 SG.

Faire migrer 4µl PCR avec a coté 5-10µl de la PCR ayant servi à contrôler les amorces (pour vérifier la taille de l'amplicon et ainsi voir s'il n'y a pas eu d'inversions).

Purifier les produits PCR (96µl restant) sur colonne Nucleospin (Kit Nucléospin extract II macherey-nagel). Reprendre les échantillons dans 30µl d'eau_{dd}.

Vérification des BmQ dans la souche SW102 :

a) Par PCR

Prendre par exemple les amorces pour amplifier la sonde et tester 3 BmQ par construction.

Mettre en contrôle positif le reste de la PCR de vérification des amorces.

b) Par digestion

Choisir 5 enzymes de digestions différentes

IV Construction du Retrieval Plasmid

A) Clonage des miniarms 5' et 3' dans pL253

1) Digestion du miniarm 5' par Not I / Spel

Mix réactionnel :

30µL PCR

1µl Not I

1µl Spe I

3,5µl NEB2

2) Digestion du miniarm 3' par Spe I / Bam HI

Mix réactionnel :

30µL PCR

1µl Spe I

1µl Bam HI

3,5µl React 4

3) Digestion de pL253

2 options : si on veut cloner les miniarms 5' et 3' en même temps, il faut digérer pL253 par Not I / Bam HI. Si on veut procéder par étape, il faut commencer par digérer pL253 avec Not I / Spe I et cloner le miniarm 5'. Ensuite digérer pL253-miniarm 5' par Spe I / Bam HI et y cloner ensuite le miniarm 3'.

Dans tous les cas digérer 20µg de plasmide dans 30µl final.

Utiliser le tampon React 3 pour Not I / Bam HI.

En contrôle pour vérifier que les enzymes de restriction sont fonctionnelles digérer séparément 2µg de pL253 avec Spe I, Bam HI et Not I.

2µg pL253

1µl Enz

2µl Tampon 10X

qsp eaudd 20µl

Toutes les digestions sont déposées au bain marie à 37°C pendant 2h30

4) Purification sur gel des digestions (Kit Nucléospin extract II macherey-nagel)

Dépôt sur gel d'agarose 1% des digestions.

Utiliser des peignes grand format.

Laisser un puit vide entre chaque échantillon.

Migration à 80V (préconisé par le kit)

Placer le gel sur le transluminateur protégé avec du saran. Le gel est morcelé. Exposer un morceau de gel et découper le plus rapidement possible les bandes avec un scalpel.

Extraire chaque bande puis la découper en petits bouts pour faciliter l'extraction. Nettoyer la lame du scalpel entre chaque bande !!!!

Suivre le protocole d'extraction nucleospin extract II.

Vérifier sur gel de la bonne fonctionnalité des enzymes (cf controles de digestion).

5) Quantification des produits digérés et purifiés sur gel

a) Préparation du marqueur

Concentration initiale du marqueur II (λ HindIII) = 0,5 μ g/ μ l

Concentration initiale du marqueur IX (ϕ Hae III) = 1 μ g/ μ l

Concentration finale : 20 μ g/ml.

Mélanger 40 μ l de marqueur II et 20 μ l de marqueur IX avec 940 μ l de TAE 1X

TAE 1X = 94 μ l TAE bleu 10X + 846 μ l eau_{dd}.

Diluer au demi un aliquot de cette solution et déposer 20 μ l pour avoir le marqueur 200ng, 10 μ l pour le 100 ng et 5 μ l pour le 50ng.

b) Préparation des échantillons.

Charger sur gel l'échantillon non dilué et dilué au 1/5 pour les fragments PCR et non dilué, au 1/5 ou 1/10 pour les plasmides.

Faire migrer les échantillons sur un gel d'agarose 1%.

Estimer la quantité des échantillons sur le gel

La qualité de la photo est très importante pour l'estimation. Ne pas saturer le signal.

Formule : $\text{ng}/\mu\text{l} * 10^3 / (635 * \text{nb bases (de l'échantillon)}) = \text{pmoles}/\mu\text{l}$.

6) Ligation des miniarms dans pL253

2 options cf IV.c.

1^{ère} option : double ligation

Mix réactionnel :

Tampon 5X 4 μ l

pL253 (Not I / Bam HI) 0,1 pmole

miniarm 5' (Not I / Spe I) 0,5 pmole

miniarm 3' (Spe I / Bam HI) 0,5 pmole

Taq DNA Ligase 1 μ l

Qsp 20 μ l Eau_{dd}

Incuber ON à 16°C ou 1hr à TA. (PS : j'ai toujours ligué ON).

2^{ème} option : simple ligation

Mix réactionnel :

Tampon 5X 4 μ l
 pL253 (Not I / Spe I) 0,1 pmole
 miniarm 5' (Not I / Spe I) 0,5 pmole
 Taq DNA Ligase 1 μ l
 Qsp 20 μ l Eau_{dd}

NB : Dans les 2 cas faire un contrôle sans insert.

7) Transformation des produits de ligation

Préparer des boîtes LB agar / Amp 100 μ g/ml final (au labo ils utilisent 50 μ g/ml final).

100 μ l Topo 10 ou XL1-blue ou TBS chimiocompétentes avec 5 μ l de la ligation (j'ai travaillé en grande partie avec les TBS).

Incuber 30' sur glace

Choc thermique à 42°C pdt 45 sec.

Laisser 1 à 2' sur glace.

Ajouter 1 ml de LB puis incuber 1h à 37°C.

Étaler 100 μ l et le reste.

Inclure en contrôle des cellules sans ADN et faire le même traitement puis étaler 100 μ L sur LB et 100 μ l sur LB/Amp.

8) Purification de l'ADN des clones qui ont poussé

Piquer 8 clones par construction et lesensemencer dans 3 ml de LB/Amp. Incuber ON à 37°C.

Purification ADN plasmidique

Miniprep cf protocole précédent.

9) Vérification du clonage du retrieval plasmid

Digestion pour sortir l'insert

4 μ l ADN

0,3 μ l Enz I

0,3 μ l Enz II

2 μ l Tampon

13,4 μ l Eau_{dd}

Digérer 2-3hrs à 37°C dans un bain marie.

Migrer sur un gel agarose 1%.

Faire un stock glycérol de 2 clones positifs pour chaque construction.

NB : on peut également digérer pL253 et le retrieval plasmid par Spe I et les faire migrer sur gel.

Si la première option a été privilégiée, passer directement au j).

Pour la 2^{ème} option :

Après avoir vérifié le clonage du miniarm 5' dans pL253, il faut cloner le miniarm 3'.

Faire une midiprep du plasmide pL253-miniarm5' :

- 1) Préculture de 100-200 ml dans du LB/Amp ON à 37°C.
- 2) Centrifugation dans falcon 50 mL à 5000 rpm pour récupérer le culot pdt 5'.
- 3) Ajouter 4 ml de Solution de lyse S1 et resuspendre le culot.
- 4) Ajouter 4 ml de tampon S2 et insérer le tube 6-8 fois.
- 5) Ajouter 4 ml de tampon S3 frais ! et inverser 6-8 fois.
- 6) Equilibrer la colonne Nucleobond AX100 avec 2,5 ml de tampon N2.
- 7) Filtrer le lysat avec un erlen et un filtre puis déposer l'éluat dans la colonne.
- 8) Laver la colonne avec 10 ml de tampon N3.
- 9) Eluer le plasmide avec 5 ml de tampon N5.
- 10) Précipiter l'ADN avec 3,5 ml d'isopropanol.
- 11) Vortexer et centrifuger pendant 30' à 4°C et 14 000 rpm.
- 12) Ajouter 2 ml d'éthanol 70%.
- 13) Sécher le culot et le reprendre dans un volume approprié d'eau_{dd} (en général 100µl).
- 14) Déterminer la concentration de l'ADN au spectro.

Digérer 20µg de pL253-miniarm5' :

Mix réactionnel :

20µg pL253-miniarm5'

2µl Spe I

2µl Bam HI

3µl React 4

Eau_{dd} qsp 30µl

Digérer 2-3hrs à 37°C dans un bain marie.

Faire la ligation entre pL253-miniarm5' (Spe I / Bam HI) et PCR miniarm 3' (Spe I / Bam HI).

Transformer la ligation.

Faire une préculture de 6 clones qui ont poussé.

Analyser l'incorporation du bras d'homologie par digestion.

10) Séquençage des produits de ligation

On séquence deux clones de chaque construction. On séquence le brin sens et le brin anti-sens. L'amorce FW est l'amorce T7 car pL253 est un dérivé de pBSK. L'amorce Rev est celle qui a servi à amplifier le miniarm 3'.

Mix réactionnel :

1µl Big Dye

3,5 µl de diluant

0,5µl T7 FW 10µM ou Rev du miniarm 3'

14µl eau_{dd}

Programme PCR :

30 cycles : 95°C 20 sec

55°C 30 sec

60°C 4 min

Purification des produits PCR :

Déposer dans chaque tube 50µl d'éthanol absolu et 6µl d'acétate d'ammonium 10M.

Ajouter 5µg de glycogène (pour voir le culot)

Vortexer puis refroidir 10' sur glace.

Centrifuger à 13 rpm pdt 20' à 4°C.

Laver avec 100µl d'éthanol glacé puis sécher le culot à 37°C.

Resuspendre le culot dans 10µl de HiDi Formamide.

11) Analyse des séquences

Utiliser le programme sequencher.

III) Construction du Gap repaired Plasmid (GRP)

A) Digestion du retrieval plasmid

Faire soit une midiprep des retrieval plasmid ou purifier sur nucléospin les minipreps et les reprendre dans 30µl eau_{dd}. (PS : j'ai purifié les miniprep. PB on n'a pas beaucoup de matériel à transformer entre 10 et 50 ng). Cependant je n'ai pas eu de problèmes de transformation).

Mix réactionnel :

2µg ADN miniprep

3µl Tampon React 4

1µl Spe I
Eaudd Qsp 30µl

Digestion 3h30 à 37°C dans le bain marie.

Vérifier sur gel la digestion en chargeant 3µl sur un gel d'agarose 1%.

Purifier sur colonne Nucleospin les plasmides digérés par Spe I (il faut que la digestion du plasmide soit totale).

Quantifier sur gel en chargeant 5µl.

B) Transformation des SW102 renfermant les BmQ des constructions avec les retrieval plasmid correspondants.

- a) Pré-culture de 5 ml LB/chor ON à 30-32 °C.
- b) Mettre de l'eau, des falcons 50 mL des eppendorfs et des cuvettes à 4°C.
- c) Diluer au 50^{ème} (500µL dans 25 mL) et incuber 3-5h jusqu'à atteindre une $DO_{600} = 0,6$.
- d) Incuber les cellules dans la glace pdt 2 min puis les transférer (12,5 mL) dans 2 falcons 50 mL préalablement à 4°C.
- e) Traiter pendant 15' un échantillon à 30°C et un autre à 42°C pour induire la production des protéines de recombinaison. Agiter manuellement !!!!
- f) Refroidir les cellules dans de l'eau glacée pendant 5 min en agitant.
- g) Centrifuger à 4°C, 5000 rpm pdt 5 min.
- h) Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.
- i) Centrifuger à 4°C, 5000 rpm pdt 5 min.
- j) Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois).
- k) Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot).
- l) Resuspendre le culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 100µL par falcon).
- m) Transférer les cellules electrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger idéalement 50 ng de Retrieval plasmid digéré par Spe I.
- n) Transformer par électroporation à 1750 V. Transférer les cellules dans 1 mL de LB puis incuber à 30-32°C pdt 1h.
- o) Etaler les bactéries (1 µL 100µL et le reste) sur le milieu de sélection (LB/Amp pour sélectionner le retrieval plasmid).

Incuber à 32°C pour 16-24h.

NB : j'ai transformé de 5ng à 30ng faute d'avoir pu avoir plus. Dans tous les cas les transfo ont marché.

Le lendemain :

Préculture de 6 clones dans 3mL de LB/Amp ON à 30-32°C.

Purification des minipreps

Cf protocole habituel

C) Vérification du Gap repaired Plasmid : GRP (vérification de l'insertion des 10 kb)

Définir une stratégie de digestion entre le GRP et le Retrieval Plasmid.

Essayer d'avoir au max 3 bandes. Utiliser soit une simple ou une double digestion.

Idéalement on linéarise le RP et on obtient 2 bandes pour le GRP.

Digérer 5µl d'ADN dans 20µl final.

NB : 3µl devrait être suffisants.

Faire un stock glycérol de la souche contenant le Gap repaired plasmid validé.

IV) Construction du mini-targeting vector (MTV)

Si vous utilisez les plasmides déjà construits par fayçal ou sébastien passez directement au B).

A) Sous-clonage du Tag His-Flag-HA dans pL452 (Sal I / EcoRI)

Il faut utiliser le plasmide pL452 qui contient les séquences LoxP-Neo-LoxP (marqueur de sélection). On doit insérer dans le plasmide : - le TAG que l'on veut insérer en phase avec le C-ter du gène d'intérêt.

- les sites de restriction utilisés pour le génotypage. Ces sites doivent se trouver en aval du TAG.

Le clonage des bras d'homologie se fera en Sal I / Asc I pour le bras 5' et Bam HI / Not I pour le 3'. Remarque : le bras 3' est à l'extérieur du TAG.

Il faut pour cela ajouter dans la construction le site Asc I qui n'est pas présent dans pL452.

1) Sous-Colange du TAG His-Flag-HA dans pL452 (Sal I / EcoRI).

a) Amplification par PCR du TAG :

Le plasmide contenant le Tag 6*His-Flag-HA est le pOZFHH (fournit par fayçal)

On introduit à l'amorce FW le site Sal I et Asc I. Entre ces 2 sites est ajouté Xho I pour optimiser le clonage ultérieur du bras d'homologie 5' en Sal I / Asc I.

On introduit à l'amorce Rev les sites des enzymes utilisées pour le génotypage et pour finir le site Eco RI servant au clonage.

IMPORTANT : Comme Asc I coupe 8 nucléotides il faut en ajouter un pour conserver la phase de lecture. Ici un C est ajouté et par conséquent une proline sera créée.

NB : il n'est pas nécessaire d'ajouter les sites de restriction pour BamHI, NcoI, Eco RI, Xho I, et Xba I car ces sites sont respectivement présents dans la construction finale ainsi ou dans la séquence codant le TAG.

b) Cloner le TAG en Sal I / Eco RI dans pL452.

Réaliser la PCR en utilisant le programme PCR Bac60.

Faire un gel 2% pour vérifier la taille du fragment PCR (151 pb).

Purifier sur nucléospin le produit de PCR cf protocole habituel

Digérer 20µg de pL452 et les 30µl de PCR avec Sal I et Eco RI dans du NEB3.

Digérer en contrôle 2µg de pL452 avec Eco RI et Sal I.

Cf.

Cf. pour la suite du clonage. Le produit PCR est digéré sur un gel 1%.

Repiquer ici 12 à 24 voire 48 clones.

J'ai rencontré des problèmes pour ce clonage. J'ai du recommencer 3 fois de A à Z.

c) Séquencer le TAG pour voir s'il est bien en phase avec le C-ter.

J'ai obtenu 3 clones positifs sur 48 et 2 étaient ok pour le séquençage. Le troisième avait le Tag inséré à l'envers.

B) Clonage des bras d'homologie 5' et 3' dans pL452-FHH

IMPORTANT : En fonction des sites de restriction présents dans les bras d'homologie on doit choisir par quel bras d'homologie on doit cloner dans pL452-FHH.

Ex : pour une construction le bras d'homologie 5' (cloné en Sal I / Asc I) avait un site Bam HI. J'ai du cloner en premier le bras d'homologie 3' (cloné en Bam HI / Not I) puis en second le bras d'homologie 5'.

1) Clonage du miniarm 5' dans pL452-FHH

a) Digestion du miniarm 5' par Sal I / Asc I

Les tampons de Sal I et Asc I ne sont pas compatibles. Il faut faire une digestion séquentielle.

Mix réactionnel :

30µL PCR homologie 5'

1µl Asc I

3,5µl NEB4

b) Digestion de pL452-FHH

Mix réactionnel :

20µg pL452-TAG

2µl Asc I

3µl NEB4

qsp eau_{dd} 30µl

Vérifier sur gel en contrôle la digestion d'Asc I sur pL452-FHH.

c) Purification Nucléospin des produits digérés par Asc I.

les échantillons sont repris dans 30µl d'eau_{dd}.

d) Digestion de pL452-FHH (Asc I) et de PCR homologie 5' (Asc I) par Sal I

Mix réactionnel :

30µL

1µl Asc I (PCR homologie 5') ou 2µl (pL452-FHH)

3,5µl NEB4

e) Purifier sur gel les produits de digestion.

Suivre le protocole habituel de clonage.

f) Vérification de l'insertion du bras d'homologie 5' :

Digestion en Sal I / Eco RI.

Faire un stock glycérol du plasmide pL452-homologie5'-FHH vérifié.

g) Midiprep de pL452-homologie5'-FHH :

1) Préculture de 100-200 ml LB/amp ON à 37°C.

2) Centrifugation dans falcon 50 mL à 5000 rpm pour récupérer le culot pdt 5'.

3) Ajouter 4 ml de Solution de lyse S1 et resuspendre le culot.

4) Ajouter 4 ml de tampon S2 et insérer le tube 6-8 fois.

5) Ajouter 4 ml de tampon S3 frais ! et inverser 6-8 fois.

6) Equilibrer la colonne Nucleobond AX100 avec 2,5 ml de tampon N2.

- 7) Filtrer le lysat avec un erlen et un filtre puis déposer l'éluat dans la colonne.
- 8) Laver la colonne avec 10 ml de tampon N3.
- 9) Eluer le plasmide avec 5 ml de tampon N5.
- 10) Précipiter l'ADN avec 3,5 ml d'isopropanol.
- 11) Vortexer et centrifuger pendant 30' à 4°C et 14 000 rpm.
- 12) Ajouter 2 ml d'éthanol 70%.
- 13) Sécher le culot et le reprendre dans un volume approprié d'eau_{dd} (en général 100µl).

2) Clonage du miniarm 3' dans pL452-homologie5'-FHH

a) Digestion du miniarm 3' par Not I / Bam HI

30µL PCR homologie 3'

1µl Not I

1µl Bam HI

3,5µl React 3

b) Digestion du pL452-homologie5'-FHH par Not I / Bam HI

20µg pL452-homologie5'-TAG

2µl Not I

2µl Bam HI

3µl React 3

eau_{dd} qsp 30µl

Inclure en contrôle une digestion de pL452-homologie5'-FHH par Bam HI et Not I.

Digestion 2h30-3hr au bain marie à 37°C.

Suivre le protocole habituel de clonage.

c) Vérification de l'insertion du bras d'homologie 5' :

Digestion en Not I / Bam HI.

Faire un stock glycérol du plasmide pL452-homologie5'-FHH vérifié.

d) Vérification de la construction de la construction finale : pL452-homologie5'-FHH-homologie3'

- Séquençage des bras d'homologie 5' et 3'.

Le 5' est séquencé en FW avec l'amorce FB13 et en Rev avec l'amorce F14 (se fixe dans la cassette Néo). Avec F14 on pourra vérifier si le TAG est inséré en phase.

Pour le bras d'homologie 3' aucune amorce n'a été désignée. On peut utiliser les amorces utilisées pour la PCR homologie 3' ou alors synthétiser des amorces en amont et en aval du clonage.

- Digestion par NotI / Sal I (facultatif).

e) Faire un stock glycérol de la souche contenant le mini-targeting vector (MTV) vérifié.

IMPORTANT : vérifier dans le séquençage que le TAG est en phase et que les sites de restriction pour le génotypage sont bien présents.

V) Construction du neo-targeted plasmid (NTP)

A) Midiprep du MTV

B) Extraction de la mini-targeting cassette

1) Digestion MTV par Not I / Sal I

Mix réactionnel :

20µg MTV

2µl Not I

2µl Sal I

3µl React 3

Digestion 2h30-3h ou ON à 37°C

2) Purification sur gel

Il n'est souvent pas possible de purifier les 2 bandes issues de la digestion. Cette purification sur gel est faite pour éliminer le MTV qui n'a pas été digéré.

3) Quantification des produits.

C) Transformation de la mini-targeting cassette dans la souche SW102 contenant le **Gap Repaired Plasmid**.

1) Pré-culture de 5 ml LB ON à 30-32 °C.

2) Mettre de l'eau, des falcons 50 mL des eppendorfs et des cuvettes à 4°C.

3) Diluer au 50^{ème} (500µL dans 25 mL) et incuber 3-5h jusqu'à atteindre une $DO_{600} = 0,6$.

4) Incuber les cellules dans la glace pendant 2 min puis les transférer (12,5 mL) dans 2 falcons 50 mL préalablement à 4°C.

- 5) Traiter pendant 15' un échantillon à 30°C et un autre à 42°C pour induire la production des protéines de recombinaison. Agiter manuellement !!!!
- 6) Refroidir les cellules dans de l'eau glacée pendant 5 min.
- 7) Centrifuger à 4°C, 5000 rpm pdt 5 min.
- 8) Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.
- 9) Centrifuger à 4°C, 5000 rpm pdt 5 min.
- 10) Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois).
- 11) Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot). Resuspendre le culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 100µL par falcon).
- 12) Transférer les cellules electrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger 50ng de mini-targetting cassette.
- 13) Transformer par électroporation. Transférer les cellules dans 1 mL de LB puis incuber à 30-32°C pdt 1h.
- 14) Etaler les bactéries (1 µL 100µL et le reste) sur le milieu de sélection LB/Kanamycine de l'électroporation. En se mettant sur Kanamycine on selectionne les clones qui auront recombiné. La sélection Kanamycine est apportée par la mini-targetting cassette.
- 15) Incuber à 32°C pour 16-24h.

Le lendemain :

Préculture de 6 clones dans 3mL de LB/Kan ON à 30-32°C.

Purification des minipreps

Cf protocole habituel

Vérification du Minitargeting Vector MTV

Définir une stratégie de digestion entre le GRP et le MTV.

Essayer d'avoir au max 3 bandes. Utiliser soit une simple ou une double digestion.

Idéalement on linéarise le GRP et on obtient 2 bandes pour le NTP.

Digérer 5µl d'ADN dans 20µl final.

NB : 3µl devrait être suffisant.

Faire un stock glycérol de la souche contenant le Gap repaired plasmid validé.

Séquencage du plasmide final:

Il suffit d'utiliser le protocole de séquençage habituel, en utilisant 10 fois moins d'ADN soit 50ng d'ADN pour un plasmide de 17kb. En effet l'ADN inhibe la réaction la réaction de séquençage.

Les amorces utilisées sont une amorce spécifique placée dans le C terminal de la protéine d'intérêt par exemple le bras d'homologie 5' Forward (même si sa séquence de plusieurs dizaines de nucléotides n'en fait pas une amorce de choix pour le sequencage) et une amorce commune appelée communseqR se situant dans le polylinker au niveau du site EcoRI.

III. Protocoles de préparation de chromatine

Protocole A

PREPARATION DE CHROMATINE

1. Cultiver les cellules ES jusqu'à l'obtention de 8 boîtes de 10 cm de diamètre
2. Le jour de la préparation de chromatine, lorsque les cellules sont arrivées à confluence 70%, fixer en ajoutant 0.4% final de formaldéhyde (ajouter 110ul à 36,5%)
3. Incuber 10min à RT
4. Neutraliser l'action de la formaldéhyde grâce à 0,125M de glycine final. (ajouter 1ml de glycine 1,25M préparée extemporanément)
5. Incuber à RT 10min
6. Aspirer le milieu des boîtes et laver 2 fois avec 10ml de PBS 1X froid
7. Ajouter 2ml de PBS 1X avec des inhibiteurs: Complete (Roche) et PMSF 1mM final.
8. Gratter les boîtes, pooler dans un Falcon 15 ou 50ml
9. Centrifuger 5min à 2000rpm 4°C
10. Reprendre les culots avec du PBS1X froid et repartir dans des tubes eppendorf l'équivalent de $2 \cdot 10^7$ cellules
11. Centrifuger 5min 2000rpm 4°C
12. Reprendre le culot dans 400ul de Swelling Buffer
13. Incuber 20min dans la glace pour faire gonfler les cellules
14. Centrifuger 10min à 1500rpm 4°C
15. Reprendre le culot dans 400ul de tampon de lyse noyaux
16. Incuber 10min dans la glace
17. Soniquer afin d'obtenir des fragments d'environ 300 paires de bases. (8 cycles 30"ON 30"OFF avec la Bioruptor de chez Diagenode)
18. Conserver un aliquot de chromatine afin de reverser le crosslink, préparer l'ADN et le doser pour connaître la quantité d'ADN présente.

IMMUNOPRECIPITATION DE CHROMATINE

19. Préparer les billes : 50ul de slurry/IP (mélange 50% billes et 50% tampon de conservation), équilibrer avec 3 lavages de tampon de lyse+BSA 0,5ug/ul final (2 heures). Laver avec le tampon de lyse.
20. Possibilité de faire un pre-clearing (à 4°C pendant 4 heures) de la chromatine avec les billes non couplées aux anticorps.
21. Incuber les billes couplées aux anticorps avec la chromatine sur la nuit à 4°C.
22. Centrifuger la chromatine avec les billes
23. Garder le surnageant comme contrôle à déposer sur WB
24. Effectuer un lavage de 1ml de TSE
25. Centrifuger après chaque lavage à 2000rpm, 1min, 4°C
26. Effectuer un deuxième lavage dans du TSE + 150mM NaCl
27. Effectuer un troisième lavage dans 500mM NaCl
28. Effectuer un quatrième lavage avec le tampon de lavage
29. Effectuer un dernier lavage avec du TE

30. Eluer les billes avec le tampon d'élution pendant 15min à 65°C puis 15min à RT sous agitation
31. Centrifuger et conserver le surnageant qui contient la chromatine immunoprécipitée
32. Incuber 6h ou sur la nuit à 65°C pour reverser la fixation
33. Traiter les échantillons à la protéinase K pendant 2 heures à 56°C
34. Purifier l'ADN à l'aide d'une extraction phénol-chloroforme suivie d'une précipitation à l'EtOH.
35. Quantifier l'ADN ainsi isolée par immunoprecipitation

TAMPONS

Swelling Buffer

5mM PIPES pH8.0
85mM KCl
0,5% NP-40
inhibiteurs de protéases

Tampon de lyse Noyaux

0,1% SDS
1 % Triton X-100
3mM EDTA
25mM Tris-HCl pH8
150mM NaCl
inhibiteurs de protéases

TSE

25 mM Tris pH7,5
2,5 mM EDTA
0,3M Sucrose

Tampon de lavage

10mM Tris pH8
0,25M LiCl
0,5% NP-40
0,5% Na-déoxycholate
1 mM EDTA

Tampon d'élution

50mM Tris pH8
10mM EDTA
1 % SDS

Protocole B

Ce protocole nous a gentiment été fourni par l'équipe d'Irwin Davidson (IGBMC)

Day 1:

CHROMATIN PREPARATION

Use 1P10 dish per IP reaction (per antibody per timepoint/condition). That roughly corresponds to 150-300 μg of chromatin.

Cross-link by adding 0.4% formaldehyde to the growth medium {100 μl formaldehyde 35% to 10 ml culture medium} or replace medium by PBS with 0.4% formaldehyde (freshly prepared). Incubate 10 min at RT.

Stop the cross-linking reaction by adding glycine to a final concentration of 0.125 M . Incubate 10 min at RT.

Rinse cells twice in chilled PBS {8 ml}.

Scrape cells with 1-2 ml PBS containing proteases inhibitors and transfer to a falcon50.

Pellet cells 5-10 min at 2500 rpm (4°C).

Remove supernatant and rinse pellet in chilled PBS and centrifuge at 2500 rpm for 10 min at 4°C.

Remove supernatant and add 1 ml/10⁷ cells of Cell Lysis Buffer (use +/-150 μl of lysis Buffer per dish). Incubate on ice 10 min. Caution, we find it better to incubate at room temperature in order to avoid SDS precipitation and poor lysis (sometimes even incubate at 37°C). Be sure that the lysis is complete before going to sonication. If not, insist by pipetting up and down and/or add more lysis buffer.

Proceed directly to sonication (previously optimised to generate fragments of 200-600bp). In Diagenode sonicator water-bath, sonicate 15x10 sec on and 30 sec off in a maximal volume of 2 ml in Falcon15.

Transfer sonicated chromatin to small tubes, centrifuge 10 min at 14000 rpm and recuperate supernatant. Caution: the pellet should be small and dense (sometimes a bit brown). If the pellet is large, the cells were not properly lysed and you will have a mix of very large (unfragmented) DNA and small DNA fragments. Keep a 50 μl aliquot (to check sonication and estimate chromatin concentration) and stock the rest of the chromatin at -80°C

Check sonication and concentration

Reverse cross-link the 50 μl aliquot. Add 450 μl water and 20 μl NaCl 5M and incubate for 4 hours to overnight at 65°C.

Add 10 μl EDTA 0.5M pH 8.0, 20 μl Tris 1M pH 8.0, 2 μl proteinase K 20 mg/ml and incubate 1 hour at 50°C.

Phenol-chloroform extraction

Precipitate DNA with 1/10 vol NaOAc 3M, 2-2.5 vol EtOH 100% and 10 μg Glycogen (1 μl).

Incubate 30 min at -20°C then, centrifuge 15 min at maximal speed. Wash the pellet in EtOH 70% and centrifuge 10 min at maximal speed. Air-dry the pellet and resuspend in 50 μl ddH₂O.

Estimate DNA quantity and purity by D.O. measurement on 1, 2 and 4 μl .

Electrophorese a 5 μl sample on a 1.5% agarose gel.

If sonication is incomplete, perform extra cycles of sonication on the prepared chromatin.

Day 2:

Blocked beads

To block the beads, use 0.5mg/ml of BSA and 0.2mg/ml sonicated salmon sperm DNA or yeast tRNA and rotate for 4 hours at 4°C.

(50 µl BSA 10 mg/ml and 20 µl sssDNA 10 mg/ml to 1 ml of the beads preparation).

The blocked Protein G Sepharose preparation is stable for a month at 4°C (but not the M2 resin that has to be blocked freshly).

Pre-clear

For each IP condition prepare a 1.5 ml microtube with 300 µg of sonicated chromatin (defreeze it gently by RT incubation) and diluted 8X with protease inhibitors-containing CHIP Dilution Buffer. DON'T FORGET to take a 20 µl Input sample from one of these tubes (Input control).

Add 50 µl blocked beads (PAS for rabbit IP and PGS for mouse IP) and incubate for 2 hours at 4°C under rotation.

Centrifuge at 2500 rpm at 4°C and recuperate supernatant.

Chromatin-antibody Incubation

Add 2-5 µg of antibody to the pre-cleared chromatin and incubate overnight at 4°C under rotation.

For a Flag ChIP, incubate the pre-cleared chromatin with 50 µl blocked M2 resin (Caution: don't add blocked PAS or PGS the next morning!!)

Day 3:

Immunoprecipitation

For IP with specific antibody, add 50 µl blocked beads (PAS or PGS) and further incubate for 2-3 hours at 4°C under rotation.

Centrifuge at 2500 rpm for 1 min at 4°C and remove supernatant.

Wash the beads by a 5 min incubation at 4°C (under rotation) in 1 ml of the following buffers:

Low Salt Buffer, once

High Salt Buffer, once

Low LiCl, once

High LiCl, once (facultative)

TE, twice

Centrifuge at 2500 rpm for 1 min at 4°C and remove supernatant.

Elute the DNA by adding 250 µl fresh Elution Buffer. Incubate by rotating 15 min at RT, mix by medium power vortexing.

Centrifuge at 2500 rpm for 1 min at 4°C and transfer supernatant to a fresh tube.

Repeat the elution step and combine the eluates in the same tube. Centrifuge once more the tube to get rid of remaining beads and collect eluate.

Reversion of cross-link

Add 20 µl NaCl 5M to the 500 µl eluates and incubate for 4 hours to overnight at 65°C.

Don't forget to reverse cross-link the input sample (20 µl input sample + 480 µl Elution Buffer + 20 µl NaCl 5M => ON at 65°C).

Day4:

Proteinase K digestion

Add 10 μ l EDTA 0.5M pH8, 20 μ l Tris 1M pH8 and 2 μ l proteinase K to the 500 μ l reversed IP chromatin and incubate for 1 hour at 50°C.

DNA purification

phenol-chloroform extraction

EtOH precipitation with 1 μ l Glycogen and resuspend in 50 μ l ddH₂O.

Solutions

Cell Lysis Buffer

10 mM EDTA pH 8.0
50 mM Tris pH 8
1% SDS

ChIP dilution Buffer

0.01 % SDS
1.1 % Triton X-100
1.2 mM EDTA, pH 8
16.7 mM Tris, pH 8
167 mM NaCl
Protease inhibitors, added freshly

(50 ml)
50 μ l SDS 10 %
550 μ l Triton X-100
240 μ l EDTA 250 mM, pH 8
835 μ l Tris 1 M, pH 8
1.67 ml NaCl 5M
Proteases inhibitors, added freshly

Low Salt Buffer

0.1% SDS
1 % Triton X-100
2 mM EDTA pH 8
20 mM Tris, pH 8
150 mM NaCl

(50 ml)
500 μ l SDS 10 %
500 μ l Triton X-100
400 μ l EDTA 250mM, pH 8
1 ml Tris 1 M, pH 8
1.5 ml NaCl 5 M

High Salt Buffer

0.1% SDS
1 % Triton X-100
2 mM EDTA pH 8
20 mM Tris, pH 8
500 mM NaCl

(50 ml)
500 μ l SDS 10 %
500 μ l Triton X-100
400 μ l EDTA 250mM, pH 8
1 ml Tris 1 M
5 ml NaCl 5 M

Low LiCl Buffer

0.25 M LiCl
1 % NP-40/Igepal
1 % deoxycholate
10 mM Tris, pH 8
1 mM EDTA, pH 8

(50 ml)
12.5 ml LiCl 1 M
0.5 ml NP-40/Igepal
5 ml deoxycholate 10 %
0.5 ml Tris 1M, pH 8
200 μ l EDTA 250 mM, pH 8

High LiCl Buffer

0.5 M LiCl
1 % NP-40/Igepal
1 % deoxycholate
10 mM Tris, pH 8
1 mM EDTA, pH 8

(50 ml)
25 ml LiCl 1 M
0.5 ml NP-40/Igepal
5 ml deoxycholate 10 %
0.5 ml Tris 1M, pH 8
200 μ l EDTA 250 mM, pH 8

IP elution Buffer

1 % SDS
 0.1 M NaHCO₃
 (50 ml)
 5 ml SDS 10 %
 5 ml NaHCO₃ 1 M

Comments

-before you start, sterilize everything !! Especially bench and pipettes. If the pipettes have been around the gene you plan to ChIP, you must ethanol, bleach and UV them. Use only filter tips throughout the entire procedure.

The working solutions can be exposed 3 min to UV light, just before use.

-sonication needs to be optimized for each cell lines by varying the power and/or the number of pulses. The sample should be kept on ice at all time (place tube in ice-containing glass beaker)

-the antibody is critical to obtain good ChIP results. First, check if ChIP has already been performed for your target protein. If not, prefer polyclonal antibodies that give good results in EMSA (supershifted the targeted complex). The antibody should be affinity purified. You can first try to ChIP your protein with the specific antibody and use an aliquot of eluted material to check by western blot that your protein is precipitable in its chromatin state.

- protein A Sepharose (PAS) is used when IP reaction is performed with a rabbit antibody, whereas protein G Sepharose (PGS) is used with mice antibodies.

Protocole C

Ce protocole a été adapté d'un protocole qui nous a gentiment été fourni par l'équipe de Michel Werner au CEA

Immuno-précipitation de la Chromatine (Protocole de Laurent Kuras modifié), 01/2006

Culture et crosslink

- Cultiver les cellules ES jusqu'à l'obtention de la quantité requise.
- Ajouter sous sorbonne de la formaldéhyde pour une concentration à 1% final (protocole levure initial) ou 0.4% final (ce que nous avons testé). Laisser 10 min en agitant toutes les 2 min.
- Cross-link arrêté avec 0.5M de glycine final (protocole de levure initial) ou 0.125M final (ce que nous avons testé) pendant 5 min.

- Retirer le milieu avec la formaldéhyde et faire 2 lavages au PBS 1X.
- Reprendre chaque boîte de 10 cm dans 2ml de PBS1X plus inhibiteurs. Gratter les cellules et les collecter dans des Falcon.
- Centrifuger 3 min à 4000 rpm.
- Reprendre les cellules dans 5 ml de FA/SDS/PMSF. **Toujours travailler dans la glace.**
- Transférer dans tubes Falcon de 14 ml en évitant de faire de bulles, centrifuger idem.
- **A cette étape on peut congeler dans N2I puis à -80°C**

Broyage et préparation de la chromatine

- Reprendre les cellules avec 1,3 ml de FA/SDS/PMSF et transvaser dans un tube Eppendorf 2 ml. Ajouter 0,3 ml de billes de verre (425-600 μm) par tube.

Sonication de la chromatine

- Utiliser la boîte et le portoir spéciaux ainsi que le boy. La sonde doit atteindre le niveau de 0,5 ml dans un tube eppendorf de 2 ml sans toucher la paroi. Les tubes sont dans un mélange eau/glace /NaCl pendant la sonication.

-**Réglage de l'appareil: pulse, 60%, puissance 4.** Soniquer 3 fois 40 secondes avec un intervalle de 20 secondes (taille moyenne des fragments 400 pb)

-transférer dans un tube Greiner de 12 ml et ajouter 3 ml de FA/SDS/PMSF. Mettre sur roue rotative à 4°C 30 min (à 1h) avec un angle d'inclinaison faible. **Pour ChIP-on-chip : diluer moins la chromatine, ajouter 0,5 ml de FA/SDS/PMSF.**

- Centrifuger JA20 10000 rpm 30 min 4°C. **La chromatine est maintenant soluble, reprendre le surnageant**

- Aliquoter (650 μl) et congeler dans N2I si l'on veut.

Contrôle de la taille des fragments d'ADN

- Prendre un aliquot de 200 μl de chromatine, ajouter 50 μl de tampon Pronase X5 et 12,5 μl Pronase (20 mg/ml ds H₂O), incuber 1h à 37°C (facultatif) puis à 65°C O/N dans une **étuve** ce qui permet de digérer les protéines et de réverser le cross-link. Ajouter 0,7 μl de RNase à 10 mg/ml et incuber 1h à 37°C.

- Purifier l'ADN sur colonne Qiagen (PCR purification kit) en suivant le protocole. Charger en 2 fois sur la colonne (2X750 μl). Eluer avec 50 μl de Tris 10 mM pH8,5. Analyse de la taille des fragments par électrophorèse sur gel d'agarose (1,5%, TAE x 0,5 100V) sur 12 μl + 2,7 μl tampon d'échantillon Xylène Cyanol.

Immunoprécipitation

Préparation des billes couplées aux anticorps

- Prélever 50 µl de la solution de stock de Dynabeads IgG (souris pour 12CA5, 8WG16, 9E10).
- Laver avec 3X 500 µl d'eau stérile
- 2X 500 µl de PBS BSA 0,1% (solution gardée à 4°C).
- (Pour plusieurs IP – 1 tube par 2 IP, lavage avec 1 ml).
- Reprendre dans 100 µl de PBS BSA 0,1%. (Plusieurs IP – regrouper).
- Ajouter 1 µl de 12CA5 (FA 3,6 µg/µl) (9E10 – 2 µl à 0,7 µg/µl, 8WG16 – 2-5 µl purifié à 1 µg/µl).
- Incuber 30 min à 1300 rpm dans l'agitateur Eppendorf à 30°C.
- Laver 2X 500 µl de PBS BSA 0,1 %
- Laver 1X 500 µl de PBS BSA 0,1 % sous agitation 1300 rpm 10 min à 30°C.
- Laver 1X 500 µl de PBS BSA 0,1 %.
- (Plusieurs IP : 1 ml pour 2 IP, regrouper et distribuer).

Réaction d'immunoprécipitation

- Décongeler un tube de chromatine, centrifuger 15 min 12000rpm à 4°C.
 - Ajouter aux billes/anticorps 500 µl de la chromatine avec 50 µl de PBS BSA 10 mg/ml.
 - Incuber 2 h 1300 rpm à 21°C.
 - Reprendre avec 500 µl de FA/SDS, changer de tube.
 - Laver avec 3X 1 ml de FA/SDS + NaCl (concentration finale de 500 mM).
 - Le dernier lavage 10 min 1300 rpm 21°C.
 - Laver avec 500 µl de tampon IP.
 - Laver avec 500 µl de TE.
 - Eluer pendant 20 min à 65°C 600 rpm avec 125 µl de tampon pronase. Refroidir à température ambiante.
 - Reprendre le surnageant, ajouter 6,25 µl de pronase (20 mg/ml dans H₂O).
 - Incuber 1h à 37°C (facultatif) puis O/N à 65°C dans **étuve**.
 - Sur les 100µl de chromatine restante non immunoprécipitée (ADN total), rajouter 25µl de tampon Pronase X5 et 6,25µl de pronase. Incuber 1h à 37°C (facultatif) puis O/N à 65°C dans **étuve**.
 - Traitement RNase après IP et de l'ADN total: ajouter 3,3µl de RNase diluée 10 fois dans l'eau et incuber 1h à 37°C.
 - Purification des fragments d'ADN (PCR kit Qiagen). Eluer avec 100µl de Tris 10mM pH8,5 pendant 5min. Centrifuger. Faire une seconde élution idem. Même volume d'élution pour l'ADN total.
- Pour CHIP-on-chip : élution avec 50 µl de tampon.

Tampons :

Hepes KOH 1M pH 7,5 : 11,915g (Rnase et Dnase free)+ H₂O stérile après, tamponner avec KOH concentrée et ajuster à 50 ml final

Tris 20mM pH 8: stock 1 M stérilisé et filtré + H₂O stérile

Glycine 2,5 M: filtré et stérilisé

FA/SDS/PMSF: solutions stocks stériles. Tampon final filtré (sans le PMSF...)

- Hepes KOH pH 7,5 50 mM 25 ml de 1M
- NaCl 150 mM 15ml de 5M
- EDTA 1 mM 1ml de 0,5M + qsp 500ml H₂O
- Triton X100 1% 5ml de 100%
- Déoxycholate de Na 0,1% 5ml de 10% (solution à 10% : peser + H₂O stérile et filtrer)
- SDS 0,1% 5ml de 10%
- PMSF 1 mM à ajouter extemporanément (T_{1/2}=30 min en solution aqueuse): solution stock 20 mg/ml (112 mM, x100) ou 200 mM (x200) dans Isopropanol (propanol 2)

Tampon Pronase X5: solutions stocks stériles.

- Tris pH7,5 125 mM: 250µl de 1M
- EDTA 25 mM 100µl de 0,5M + 1150µl H₂O
- SDS 2,5% 500µl de 10%

Tampon IP : solutions stériles et Tampon final filtré

- Tris 10mM pH8 0,5ml de 1M
- LiCl 0,25M 12,5ml de 1M
- EDTA 1mM 0,1ml de 0,5M
- NP40 0,5% 0,25ml de 100%
- Déoxycholate de Na 0,5% 2,5ml de 10%
- H₂O stérile qsp 50ml

Tampon d'échantillon:

- 0,0625% Xylène cyanol
- 50% Glycérol
- 10 mM EDTA

PBS BSA 0,1%, 50 – 100 ml, filtrer et garder à 4°C

PBA BSA 10 mg/ml, 10 ml, filtrer et garder à 4°C

Protocole D

Ce protocole a été développé par l'équipe de Matthieu Gérard en collaboration avec S. Graziani et L. Carrière et rédigé par S. Jounier et H. Humbertclaude

But: Immuno-précipiter la chromatine liée à une protéine d'intérêt par cross-link in vivo dans des cultures cellulaires de cellules souches embryonnaires.

Préliminaire : La veille préparer du 2-3 litres de PBS1x et stocker les bouteilles à 4°C.

1er JOUR

PREPARATION DE LA CHROMATINE

Temps de préparation: 1 jour à partir du moment où l'on a les cellules (3 ou 10 boîtes pour le tandem).

- Changer le milieu des cellules 2h avant la manip. (9H)
- Les compter (il doit y avoir env. $2 \cdot 10^7$ cellules par boîte): 1 boîte est réservée pour le comptage et pour faire des extraits totaux pour contrôle en WB.
- Préparer la Glycine 20-30min avant la fin des 2h.

CROSS-LINK AU FORMALDEHYDE

Remarque : On traite les boîtes 12 par 12 maximum. Donc si on a plusieurs séries de boîtes, il est indispensable de les décaler de 15 min (20 min si DTBP).

Option: Si un cross-link DTBP est nécessaire, suivre le protocole suivant:

Pour cette partie, mettre de la glace dans un ou deux grands bacs de rétention. On disposera les boîtes sur la glace.

Faire 2 lavages au PBS 1x (froid)

Mettre 10 ml de PBS 1x dans chaque boîte et ajouter 100µl de DTBP 100x (500mM, préparé extemporanément).

Incuber 30 min sur la glace.

Faire 2 lavages au PBS 1x (froid).

Stopper la réaction avec du Quenching Buffer, 10 min sur la glace.

Faire 3 lavages au PBS 1x (froid).

Mettre du milieu D15 (37°C) dans chaque boîte et continuer la suite de la manipe avec le Formaldéhyde.

Ajouter 110µl de formaldéhyde 36,5% (Formaldéhyde frais réservé pour les IP) directement dans le milieu de culture (= 0.4% FINAL) et agiter en douceur pour bien mélanger.

Incuber 10' à Temp. amb.

STOPPER LE CROSS-LINK.

Ajouter 1ml de Glycine 1,25M préparée extemporanément.

(soit 4,69g pour 50ml d'H₂O, chauffer à 37°C pendant 5')

Laisser 10' en contact.

Remarque: Jeter dans la bouteille déchets formaldéhyde

LAVAGES et RECUPERATION DES CELLULES

Faire 2 lavages avec 10ml de PBS glacé.

Puis, mettre 2ml de PBS glacé avec inhibiteurs de protéases (PMSF dilué 100× et COMPLETE dilué 50×). Gratter les cellules avec un scrappeur.

Récupérer dans 1 ou 2 Falcon 15ml.

Mettre dans la glace et continuer la manip. au labo.

RECUPERATION DU CULOT CELLULAIRE (12H)

Remarque: Travailler avec des pointes à filtres (protection contre des ADN contaminants, substances toxiques, tampons moussants ...) et des solutions stériles.

Centrifuger 5' à 2500rpm à 4°C.

Aspirer le surnageant à la pompe à vide.

LYSE DES CELLULES

Ajouter sur le culot cellulaire 1ml de FA/SDS/90mM KCl (contenant les inhibiteurs de protéases) pour $2 \cdot 10^7$ cellules. Resuspendre le culot d'abord avec 1ml de FA/SDS/90mM KCl et compléter en fonction du nombre de boîte.

Bien remettre en suspension avec la pipette Gilson;

Transférer 1ml de suspension cellulaire dans des tubes Eppendorfs 1,5ml.

Laisser sur la glace pendant 15'.

RECUPERATION DE LA CHROMATINE (ADN ET PROTEINES LIEES PAR LA FORMALDEHYDE)

Centrifuger 20' à 12000 rpm à 4°C

En principe on n'a pas besoin du surnageant, on peut en garder un pour un contrôle en WB (protéines solubles non fixées). → *Fraction soluble*

Aspirer le surnageant avec la pompe à vide.

LAVAGE DE LA CHROMATINE PURIFIEE (à lancer avant le déjeuner)

Reprendre le culot dans 1ml de FA/SDS/90mM KCl avec inhibiteurs de protéases pour $2 \cdot 10^7$ cellules.

Bien remettre en suspension à la pipette Gilson et laver 1H sur la roue à 4°C, 11 rpm.

CHROMATINE POUR SONICATION

Centrifuger 10' 14000 rpm à 4°C

Aspirer les surnageants sauf 1 qui servira pour un contrôle en WB. → *Lavage*

Reprendre le culot dans 400ul de FA/SDS/90mM KCl +inhib.

Remettre en suspension à la pipette.

Garder dans la glace.

Remarque: on peut stocker l'échantillon à -80°C et continuer le protocole le lendemain.

Le fait de congeler et de décongeler influence le résultat de la sonication.

SONICATION (diagenode 144)

Si le lot de formaldéhyde est ouvert récemment (moins d'1 mois):

17 cycles de 20'' ON/ 40'' OFF.

Remarque: on peut ajouter 100µl de Glass Beads pour casser la chromatine

A chaque série, mettre de la glace avec l'eau de la cuve jusqu'au niveau adéquat.
 Ne pas rajouter de glace dans le sonicateur entre les cycles.
 Si le lot de formaldéhyde est vieux, il faut diminuer le nombre de cycles ou le temps de sonication:
 (ex: 15 cycles de 15'' on et 40'' off)

RECUPERATION DE LA CHROMATINE SONIQUEE

Centrifuger 10' à 14000 rpm.
 Faire attention car le culot est peu visible.
 Prélever les 90% du surnageant. → *Input*
 Garder une aliquote de 20µl pour vérifier la sonication. Mettre le reste a -80°C.

REVERSION DU CROSS-LINK (vers 17h)

Réverser le cross -link des 20µl de chromatine aliquotée :
 Ajouter 480µl d'H₂O, 20µl de NaCl 5M et 15µg de RNase A (=0,5 µl). Incuber de 4 H à toute la nuit à 65°C.

2^{ème} JOUR

Récupérer les échantillons à 65°C et ajouter:
 -10µl EDTA 0,5M pH8,0
 -2µl de PROTEINASE K à 20mg/ml
 Incuber 1H à 50°C.

EXTRACTION PHENOL/CHLOROFORME

Ajouter 2 volumes de Phénol Chloroforme (réservé CHIP)
 Vortexer 1'
 Centrifuger 5' à 13000rpm à Temp. amb.
 Récupérer la phase aqueuse.
 Précipiter l'ADN en ajoutant:
 -1/10ème de volume de Na-Acétate 3M pH5,2
 -2 volumes d'Ethanol 100%.
 -10ug de Glycogène(1µl).
 Laisser précipiter 30' à -20°C.
 Centrifuger 15' à 14000 rpm à 4°C.
 Laver le culot avec 500 µl d'Ethanol 70%.
 Centrifuger 10' à 14000 rpm.
 Laisser sécher le culot et le resuspendre dans 20µl d'H₂O.

Estimer la quantité d'ADN et la pureté au Nanodrop par DO.
 Faire migrer 2 et 5 µl sur un gel d'agarose 1,5% pour voir la taille des fragments d'ADN.
 (Utiliser du «bleu light» 10×concentré, boîte des marqueurs de taille, frigo couloir).
 Si la sonication est insuffisante, décongeler l'échantillon et refaire 2cycles (ou plus si besoin) de sonication.

3^{ème} JOUR

IMMUNOPRECIPITATION HA

Le protocole est donné pour 1 seule IP ou pour 4 IP (si on veut faire du tandem).

On utilise un portoir aimanté pour faire tous les changements de solution des billes.

PREPARATION DES BILLES COUPLEES AVEC L'ANTICORPS HA

Dans un tube Eppendorf mettre 50µl (+5µl de marge) ou 200µl (+10µl de marge) de Dynabead Pan mouse IgG (Invitrogen ; Réf 110-42).

Faire 3 lavages avec 1ml d' H₂O ultrapure (filtrée 0,2 µm).

Faire 2 lavages avec 1ml de PBS1x-BSA 0,1%.

Reprendre les billes dans 100µl ou 400µl de PBS1x-BSA 0,1%.

Ajouter 3µg ou 12µg d'anticorps HA-7 (1µg/µl ; SIGMA-ALDRICH ; H3663).

Incuber 1h à 30°C sur un thermomixeur, 1300rpm.

BLOCAGE DES BILLES HA

Préparer le *tampon de blocage*: 1ml FA/SDS/90mM KCl+ 25µl BSA 20mg/ml+20 µl tRNA 10mg/ml

Faire 1 lavage avec 1ml de FA/SDS/90mM KCl.

Bloquer les billes HA avec 1 ml (par tube) de *tampon de blocage*.

Incuber sur la roue 1h à température ambiante (ou 4h à 4°C).

IMMUNOPRECIPITATION HA

Dans 1 ou 4 tubes Eppendorf mettre 250µg ou 4x250µg de chromatine soniquée.

Compléter si besoin à 500µl avec du tampon de lyse FA/SDS/90mM KCl+ Complete 1/50° +PMSF 1/100°.

Retirer le surnageant des billes HA.

Reprendre les billes dans 50µl ou 200µl de FA/SDS/90mM KCl+ Complete 1/50° +PMSF 1/100°.

Ajouter les billes HA bloquées dans les tubes de chromatine.

Incuber 2h à température ambiante sur la roue.

Récupérer 1 surnageant pour contrôle. → *FT HA(Flow Through)*

Laver les billes HA avec les tampons suivants :

1ml Low salt Buffer	X2
1ml High salt buffer	X3

S'il s'agit du tandem faire un dernier lavage avec 1 ml de FA/SDS/ KCl 90mM en poolant les 4 IP.

Remarque : A cette étape, on peut si on le souhaite, récupérer une fraction des billes avant élution pour contrôle (1/4 ou 1/2 IP Par exemple) → 1/4 ou 1/2 IP HA

ELUTION AVEC PEPTIDE HA

Pour l'IP simple on passe directement à l'étape « réversion du cross-link, 4^{ème} jour » si on veut récupérer l'ADN ou on met directement 20µl de Lämmli 3x sur les billes si on veut vérifier l'IP en Western Blot.

Enlever le FA/ SDS/ KCl 90mM

Eluer les complexes Protéines/Chromatine avec le tampon d'élution :

220µl FA/SDS/90 mM KCl+ Complete 1/50° +PMSF 1/100°+ 40 µl de peptide HA (4 µg/µl)

- *OPTION A* :Incuber à 4°C sur la roue pendant la nuit.

- *OPTION B* : Incuber à 15°C pendant 4h sur le thermomixeur

Récupérer les 260µl d'éluat

Faire une 2^{ème} élution à 4°C sur la roue pendant la nuit

On peut récupérer les éluats une fois que les billes Flag sont prêtes :

Pour *l'OPTION A* : Récupérer les 250µl de surnageant (étape ELUTION PEPTIDE HA) des billes HA.

Ajouter 250µl de FA/SDS/90mM KCl+ Complete 1/50° +PMSF 1/100°.

Pour *l'OPTION B* : récupérer les 260µl de la 2^{ème} élution et les pooler avec la 1^{ère} élution

Prélever 1/4 ou 1/2 de l'IP HA post-élution pour contrôle → 1/4 ou 1/2 IP HA post-élution
 Récupérer les billes et mettre 20µl de Lämmli 3x pour contrôle en western blot → Billes HA

4ème JOUR

IMMUNOPRECIPITATION FLAG

PREPARATION DES BILLES COUPLEES AVEC L'ANTICORPS FLAG

Dans un tube Eppendorf mettre 50µl (prendre 5µl de marge) de Dynabead Pan mouse IgG (Invitrogen ; Réf 110-42).

Faire 3 lavages avec 1ml d' H₂O ultrapure (filtrée 0,2µm).

Faire 2 lavages avec 1ml de PBS1x-BSA 0,1%.

Reprendre les billes dans 100µl de PBS1x-BSA 0,1%.

Ajouter 3µg d'anticorps Flag M2 (1µg/µl ; SIGMA-ALDRICH ; F1804).

Incuber 1h à 30°C sur un thermomixeur, 1300rpm.

BLOCAGE DES BILLES FLAG

Préparer le Tampon de Blocage: 1ml FA/SDS/90mM KCl+ 25µl BSA 20mg/ml+20 µl tRNA 10mg/ml

Faire 1 lavage avec 1ml de FA/SDS/90mM KCl.

Bloquer les billes FLAG avec 1 ml (par tube) de Tampon de Blocage.

Incuber sur la roue 1h à température ambiante (ou 4h à 4°C).

IMMUNOPRECIPITATION FLAG

Retirer le surnageant des billes FLAG.

Reprendre les billes dans 50 µl de FA/SDS/90mM KCl+ Complete 1/50^e +PMSF 1/100^e.

Ajouter les billes FLAG bloquées à la chromatine.

Incuber 2h à température ambiante sur la roue.

Récupérer 1 surnageant pour contrôle → FT Flag (Flow Through)

Laver les billes FLAG avec les tampons suivants :

1ml Low salt Buffer	X2
1ml High salt buffer	X3

OPTION 1 : ELUTION SDS

Sur les 50 µl de billes Flag, éluer avec 250µl de tampon d'élution pré-chauffé à 65°C : 1% SDS-0,1M NaHCO₃

Incuber 15 min sur la roue à température ambiante.

Vortexer doucement.

Collecter l'éluat (250µl)

Eluer une seconde fois avec 250 µl de tampon d'élution

Pooler les deux éluats (500µl) → Tandem

Récupérer les billes et mettre 20µl de Lämmli 3x pour contrôle en western blot → Billes Flag

OPTION 2 : ELUTION PEPTIDE FLAG

Sur les 50 µl de billes Flag, éluer les complexes Protéines/Chromatine avec le tampon d'élution : 220µl FA/SDS/90 mM KCl+ Complete 1/50^e +PMSF 1/100^e+ 40 µl de peptide Flag (4 µg/µl)

- *OPTION A* : Incuber à 4°C sur la roue pendant la nuit.

- *OPTION B* : Incuber à 15°C pendant 4h sur le thermomixeur

Récupérer les 260µl d'éluat

Faire une 2^{ème} élution à 4°C sur la roue pendant la nuit

Pooler les 2 éluats de 260µl → *Tandem*
Récupérer les billes et mettre 20µl de Lämmli 3x pour contrôle en western blot → *Billes Flag*

OPTION 3 : ELUTION GLYCINE

Sur les 50 µl de bille Flag, éluer 10 min à température ambiante avec 260µl de Glycine 0,1M pH 2,5 sur le thermomixeur

Récupérer le surnageant

Éluer une 2^{ème} fois 10 min à température ambiante avec 260µl de Glycine 0,1M pH 2,5 sur le thermomixeur

Pooler les 2 éluats → *Tandem*

Récupérer les billes et mettre 20µl de Lämmli 3x pour contrôle en western blot → *Billes Flag*

Remarque : le pH est très acide, on tamponne avec 3-4 gouttes de tampon Tris 1M pH8 et on vérifie le pH en prenant 5 µl d'éluat auxquels on ajoute 5µl de Laemli 3x. Celui ci vire au jaune si le pH est trop acide. On rajoute du tampon TRIS jusqu'à ce que le Lämmli vire au bleu.

REVERSION DU CROSS-LINK

A cette étape traiter en parallèle les 500µl de l'éluat du tandem HA-FLAG (étape précédente) et 20µl d'Input.

- Compléter à 500 µl avec du tampon d'élution.
- Ajouter 20 µl de NaCl 5 M et 0,5 µl (15µg) de RNase A.
- Incuber sur la nuit à 65°C.

Remarque : on peut faire la réversion du cross-link que sur la moitié de l'éluat du tandem et l'autre moitié peut être contrôlé en Western Blot

5ème JOUR

PURIFICATION DE LA CHROMATINE IMMUNOPRECIPITEE

DIGESTION PROTEINASE K

Ajouter 10 µl EDTA 0,5 M pH 8.
20µl TRIS 1M pH 8.
2 µl Protéinase K 20µg/µl.
Incuber 1h à 50°C au bain marie.

PURIFICATION D'ADN

Manipuler sous la hotte car le Phénol/ Chloroforme est un produit dangereux.
Ajouter 530 µl de phénol/ Chloroforme 4°C.
Vortexer 30 sec.
Centrifuger 15 min à 4°C 13000 rpm.
Prélever le surnageant (phase aqueuse contenant l'ADN).
Ajouter 50µl de NaOAc 3M, 1 ml d'Ethanol 100% glacé et 1 µl de glycogène 10µg/µl.
Vortexer.
Incuber à -20°C pendant 20 min.
Centrifuger 10 min 13000 rpm, 4°C (le culot doit être visible).
Aspirer le surnageant.
Laver le culot avec 400 µl d'Ethanol 70%.

Centrifuger 5 min, 13000 rpm, 4°C.
Aspirer le surnageant.
Laisser sécher le culot et reprendre dans 20µl d'H₂O
Quantifier au nanodrop.

SOLUTIONS : PREPARATIONS ET REFERENCES

FA/SDS/90mM KCl (se garde au frigo)

50mM TRIS pH 7.5

90mM KCl

1mM EDTA

1% TRITON 100×

0.1% Na DEOXYCHOLATE

0.1% SDS

	POUR 1 LITRE	POUR 250ML
TRIS 1M pH 7.5	50ML	12.5ML
KCl 1M	90ML	22.5ML
EDTA 0.5M	2ML	500µL
TRITON 100×	10ML	2.5ML
Na DEOXYCHOLATE 10%	10ML	2.5ML
SDS 10%	10ML	2.5ML
QSP H ₂ O		
FILTRER 0.2µ AVEC UNE UNITE DE FILTRATION.		

FORMALDEHYDE 36,5% (sous la hotte)

SIGMA

REF: F8775

QTE: 25ML

GLYCINE (sur étagère pesée)

SIGMA

REF: G7126

75G/MOLE

Préparer à 1,25M (10×concentré): peser 4,7g pour 50ml d'H₂O

Chauffer à 37°C pendant 5' pour aider la dissolution

Se prépare extemporanément

DTBP 100X (500mM) (au frigo couloir)

Se prépare extemporanément et 100×concentré.

Précipite à Temp amb. Sortir la bouteille 15' avant la pesée car très hydrophile

PIERCE

Ref: 20665

309,28g/mole

S'utilise à 5mM final

Peser 77,32mg pour 500µl d'H₂O

QUENCHING BUFFER

100mM Tris.HCl pH8

150mM NaCl

PMSF (avec produits toxiques sur étagère pesée)

ATTENTION ! produit avec demi vie en milieux aqueux de seulement 30 minutes. Donc à diluer au dernier moment.

Phenylmethylsulfonylfluoride

SIGMA

Ref: P-7626

174,2g/mole

S'utilise à 1mM final(1×concentré)

Se prépare 100× concentré (0,1M) dans l'Ethanol absolu: Peser 0,1742g pour 10ml d'Ethanol.

A aliquoter et stocker a -20°C (congélateur pièce 9)

COMPLETE (au frigo pièce 9)

ROCHE

REF: 11836145001

3×20 TABLETTES

Se prépare 50×concentré

S'utilise 1×concentré

Diluer 20 tablettes dans 40ml d'H₂O. Aliquoter et stocker a -20°C. (congélateur pièce 9)

CELL SCRAPER

Total Length: 38cm, Blade Length: 25mm

TPP (Tissu Cell culture products)

REF : TP99004

RNase A (congélateur pièce 9)

SIGMA

REF: R 4612-50MG

30 mg/ml

PROTEINASE K(frigo couloir)

ROCHE

4×250MG

Se prépare à 10mg/ml dans H₂O. Aliquoter et stocker à -20°C (congélateur pièce 9)

REF: 03115852001

GLASS BEADS, ACID-WASHED

SIGMA

REF :G8772-500G

PHENOL-CHLOROFORME (congélateur pièce 9 réservé CHIP))

FLUKA

REF: 77618

QTE: 100ML

GLYCOGENE (frigo couloir)

ROCHE

REF:10901393001

Se prépare dans l'H₂O à 10mg/ml

S'aliquote et se stocke à -20°C.

BLUE LIGHT

0,01g Xylène Cyanol

2g de Ficoll

Qsp 10ml TAE 10x

FICOLL 400

SIGMA ALDRICH

Réf : F4375

PBS1x-BSA 2% (20mg/ml) (frigo pièce 9)

Albumin Bovine-Fraction V

EUROMEDEX

500g

REF : 04-100-810^E

Se prépare dans le PBS1x.

Peser 1g et le diluer dans 50ml de PBS1x.

Filtrer dans une seringue avec un filtre 0,2 μ .

PBS1x-BSA 0,1% (1mg/ml) (frigo pièce 9)

Se prépare à partir de la solution mère PBS1x-BSA 2%.

Pou 250 ml : 12,5ml de PBS1x-BSA 2%

qsp 250ml PBS1x

Filtrer avec une unité de filtration de 250 ml (0,2 μ)

Dynabeads® Pan Mouse IgG

INVITROGEN

Réf : 110-42

ANTICORPS MONOCLONAL ANTI-HA PRODUCT IN MOUSE CLONE HA-7

(-20°C pièce 9)

SIGMA ALDRICH

200 μ l

Réf : H3663

Aliquoter par 50 μ l et stocker à -20°C.

ANTICORPS ANTI-Flag M2 PRODUCT IN MOUSE

(-20°C couloir)

SIGMA ALDRICH

1mg

Réf : F1804

Se prépare à 1 μ g/ μ l dans l'H₂O : reprendre les 1 mg dans 1ml d'H₂O

Aliquoter par 50 μ l et stocker à -20°C.

YEAST tRNA 10mg/ml

INVITROGEN

2 x 25mg

REF : 15401-029

Reprendre les 25 mg dans 2,5ml d'H₂O

Aliquoter par 50 μ l et stocker à -20°C.

PEPTIDE HA 4 μ g/ μ l

ANSYNTH SERVICES

Sequence : YPYDV PDYA
N-term : Free
C-term : Amide
Purity : >90% with HPLC >100 mg

Pos	Description	Price
01	Synthesis 9 amino acids >100mg after purification	Euro 325,--
02	Purification > 100 mg > 90% with HPLC	Euro 240,--
	Subtotal	Euro 565,--

La quantité est indiquée en mg sur le flacon.

Dissoudre dans du TRIS 50mM pH7,3

PEPTIDE FLAG

ANSYNTH SERVICES

Sequence : DYKDD DDK
 N-term : Free
 C-term : Amide
 Purity : >90% with HPLC >100 mg

Pos	Description	Price
03	Synthesis 8 amino acids >100mg after purification	Euro 300,--
04	Purification > 100 mg > 90% with HPLC	Euro 230,--
	Subtotal	Euro 530,--

La quantité est indiquée en mg sur le flacon.
 Dissoudre dans du TRIS 500mM pH7,3

SOLUTION D'ELUTION

1% SDS

0,1M NaHCO₃

Préparer la solution extemporanément

Pour 50ml

SDS 10% 5 ml

NaHCO₃ 1M 5 ml

LiCl

SIGMA

REF : L9650

100g

Igepal CA-630

SIGMA

REF : I3021-100ML

100ml

LOW SALT BUFFER

0,1% SDS

1% Triton

2mM EDTA pH8

20 mM Tris pH8

150mM NaCl

Pour 500ml

SDS 10% 5 ml

Triton X-100 5 ml

EDTA 250mM pH8 4 ml

Tris 1M pH8 10 ml

NaCl 5M 15 ml

GLYCINE 0,1 M pH2,5

SIGMA

Réf : G7126

peser 375 mg de glycine et dissoudre dans 50 ml d'eau ultrapure

Ajuster le pH à 2,5 avec quelques gouttes de HCl fumant

HIGH SALT BUFFER

0,1% SDS

1% Triton

2mM EDTA pH8

20 mM Tris pH8

500mM NaCl

Pour 500ml

SDS 10% 5 ml

Triton X-100	5 ml
EDTA 250mM pH8	4 ml
Tris 1M pH8	10 ml
NaCl 5M	50 ml

LOW LiCl BUFFER

0,25 M LiCl	
1% Igepal	
1% deoxycholate	
10 mM Tris pH8	
1 mM EDTA pH8	
	Pour 500 ml
LiCl 1M	125ml
Igepal	5 ml
deoxycholate 10%	50 ml
Tris 1 M pH8	5 ml
EDTA 250 mM pH8	2 ml

Triton X-100

SIGMA
REF : T9284-500ML
500ml

NaHCO₃

Sodium bicarbonate
SIGMA
REF : S5761
1kg

Protocole E

Ce protocole a été rédigé par Matthieu Gérard

Principe:

Les cellules (obtenues à partir de culture ou de tissu) sont placées dans la solution I (hypotonique), qui les fait gonfler. On ajoute ensuite de l'Igepal (aussi appelé NP40) pour lyser les membranes plasmiques, tout en préservant les noyaux. On sépare ensuite les noyaux du cytoplasme et autres organelles par centrifugation à travers un coussin de sucrose. On remet les noyaux en suspension, et on ajoute la MNase (Micrococcal Nuclease), qui va pénétrer à l'intérieur des noyaux. La digestion de la chromatine se fait donc in situ, dans les noyaux. La MNase va couper le plus souvent l'ADN entre 2 nucléosomes (car il est plus accessible à l'enzyme). On arrête la digestion en ajoutant une solution EDTA, et on élimine l'excès d'enzyme en centrifugeant les noyaux. Le surnageant obtenu après cette étape de centrifugation est appelée fraction S1. Le S1 contient des mononucléosomes qui ont été relargués très rapidement lors de la digestion. On reprend les noyaux dans le tampon N-ChIP dans laquelle les fragments de chromatine vont s'éluer. Après incubation, on centrifuge les noyaux pour récupérer la chromatine soluble dans le surnageant (fraction S2).

A noter qu'une fraction de chromatine insoluble reste associée aux membranes nucléaires, au niveau du culot (ou pellet). En soniquant, on libère une partie de cette fraction insoluble.

Préparation de l'expérience / des solutions

La veille, décongeler les tubes de solution I, II, III, tampon MNase, TEEP20 en transférant les tubes congelés sur un portoir en chambre froide.

Le matin, bien homogénéiser les solutions, prélever les volumes requis et ajouter les composés suivants, selon les besoins de l'expérience:

- Complete 25x (Roche, 1 tablette pour 2 ml H₂O; ajouter 40 µl de cette dilution / ml de tampon)

- PMSF 0,1M – Attention toxique ! Il s'agit d'un inhibiteur des serine-protéases. Attention, ½ vie du PMSF = 30 minutes. A ajouter au dernier moment aux suspensions de noyaux, et à la solution II juste avant emploi. Inutile d'en ajouter aux tampons à l'avance car ce composé réagit avec l'eau et s'inactive. Le PMSF est dissous dans EtOH, gardé à -20°C et re-préparé fréquemment (2 fois/mois).

Juste avant de démarrer les préparations de noyaux, régler la centrifugeuse (Ultracentrifugeuse Beckman) à 4°C et mettre le vide en route (nécessaire pour que la machine réfrigère). Vérifier que le rotor SW41 et ses godets sont à 4°C. De même, régler la centrifugeuse eppendorf 5810R à 4°C. Vérifier que les rotors requis (à tubes eppendorfs et A-4-62) sont à 4°C.

Exemple : Solutions pour préparer des noyaux à partir de cellules ES (4 boîtes 10 cm):

9 ml de sol I : 360 µl complete

3 ml sol II 0,4% Igepal: 120 µl complete

17 ml sol III: (pour 2 coussins de sucrose)

3 ml de solution MNase (pas de complete)

Protocole

1^{er} JOUR

Collecte des ES

A partir de 4 boîtes de diamètre 10 cm (cellules ES 80% confluentes)

- 1) Trypsiner les cellules avec 2 ml de trypsine/boîte, puis neutraliser avec 5 ml de D10/boîte. Dissocier les cellules et les placer dans 2 tubes 14 ml à fond conique. Centrifuger 5 minute à 1000 rpm. Eliminer le surnageant.
- 2) Reprendre les cellules avec 2 x 6 ml de PBS 1x. Pooler les 2 tubes et centrifuger 5 minute à 1000 rpm.
- 3) Reprendre les cellules dans 3 ml de solution I froide et placer sur la glace.
- 4) Centrifuger à 1000 rpm 5 min à 4°C: utiliser le rotor A-4-62 et la centrifugeuse eppendorf 5810R.
- 5) Sur la glace : jeter le surnageant, puis reprendre les cellules dans 1 ml de solution I. Remettre les cellules en suspension avec une p1000 et poursuivre en IV.

Préparation des noyaux (sur la glace ou à 4°C):

- 6) Ajouter un volume de solution II (généralement 1,1 ml) – homogénéiser avec la p1000 et placer sur la glace. Il faut incubé pendant 10 min exactement (pour que les

membranes plasmiques soient lysées), puis démarrer immédiatement la centrifugation sur coussin de sucrose. Pour respecter ces 10 minutes, il convient de déposer la suspension de cellules dans {soll +soll} sans tarder sur les tubes contenant la solution III: avec une p1000, déposer en douceur 2 ml sur chaque coussin de sucrose.

- 7) Equilibrer les tubes 2 à 2, à 0,01 g près sur la balance. Placer chaque tube sur la glace, puis dans un godet de rotor SW41. Fermer les godets avec le tournevis adapté, et les accrocher sur le rotor.
- 8) Placer le rotor sur son axe, mettre en route le vide et centrifuger à 10000 x g pendant 30 min à 4°C (SW41: 9000 rpm).
- 9) A la fin de la centrifugation, les noyaux sont collés au fond du tube (petit culot plus ou moins visible selon la quantité de matériel et le type cellulaire).
- 10) Pour éliminer le liquide au dessus, utiliser une trompe à vide connecté à 3tips de p1000 empilés. Eliminer tout le liquide sauf les derniers 500ul (environ). Enlever ce dernier volume avec une p1000.
- 11) Reprendre les noyaux dans 500 ul de tampon MNase (sans complete). Pipeter environ 15-20 fois sans faire de bulles pour que les noyaux soient remis en suspension. Il est important de ne plus avoir d'amas avant de passer à l'étape de digestion MNase. Attention, les noyaux décanter assez rapidement, il faudra donc les resuspendre en douceur avant d'en pipeter pour les digestions MNase. Placer les tubes sur la glace.

Comptage des noyaux: Prélever une aliquot de 10 ul de noyaux. Ajouter 20 ul de tampon MNase (sans enzyme). A proximité du microscope, ajouter 30 ul de Bleu Trypan, homogénéiser en pipettant, puis déposer 20ul sur la cellule de comptage. Laisser les noyaux décanter 30 s sur la cellule avant de commencer à compter.

Rendements typiques: on obtient environ $2 \cdot 10^7$ noyaux à partir de 2 boîtes de cellules ES de diamètre 10 cm, 60% confluentes.

Préparation des mono-nucléosomes par digestion MNase sur les noyaux.

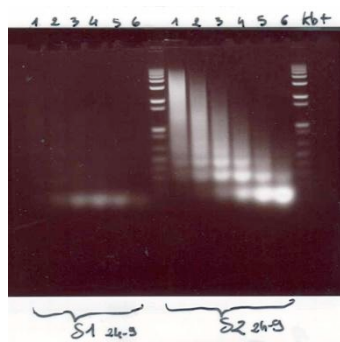
Choisir la quantité de MNase selon l'objectif de l'expérience (mononucléosomes ou oligonucléosomes, voir la figure ci-dessous). La MNase (stock à 200 u/ul) est diluée à 10 unité/ul dans le tampon MNase juste avant l'emploi :

- 12) Ajouter la quantité choisie de MNase aux 500ul de noyaux et homogénéiser avec la p1000.
- 13) Incuber exactement 10 minutes à 37°C dans un bain-marie.
- 14) Placer les tubes sur la glace et ajouter 30 ul de solution "STOP MNase". Homogénéiser en retournant le tube fermé 3 fois. Centrifuger immédiatement 5 min à 2000g, à 4°C.
- 15) Prélever le surnageant; il s'agit de la fraction S1, qui comprend des nucléosomes rapidement relargués par la digestion et des protéines solubles qui ont diffusé à partir des noyaux. Le S1 contient aussi la MNase.
- 16) Au culot, ajouter 500 ul de Tampon N-ChIP froid (avec inhibiteurs de protéases). Remettre les noyaux en suspension en douceur en pipetant avec une p1000. Mettre sur la roue à 4°C pendant 2h (rotation 10 rpm).
- 17) Centrifuger 5 min à 2000g, à 4°C. Le surnageant = S2, comprenant les mono ou oligonucléosomes relargués par la digestion MNase. Le culot (pellet) comprend les membranes nucléaires et la chromatine insoluble. Garder S1, S2 et pellet sur la glace.
- 18) Analyser 3 ul de fraction S1 et S2 sur gel agarose 1,3%. Ajouter 17 ul de bleu-SDS, et incuber 15 min à 65°C avant dépôt sur gel et migration à 100 volts.

Dans la figure ci-dessous sont indiquées les quantités de MNase à utiliser pour obtenir des oligonucléosomes de différentes tailles ou des mononucléosomes. Attention: le nombre de

noyaux dans la suspension et le volume réactionnel influencent fortement l'efficacité de digestion par la MNase. Garder ces paramètres constants pour obtenir des résultats reproductibles. **Les conditions de digestion utilisées pour la figure: 11 millions de noyaux, dans un volume de 500ul.**

Tube	MNase (unités)	chromatine obtenue dans S2:
1	3u	polynucléosomes de grandes tailles
2	6u	tri/tetra-nucléosomes
3	12u	di/tri-nucléosomes
4	24u	di-nucléosomes majoritaires
5	48u	70% mononucléosomes
6	96u	95% mononucléosomes



Mesurer la concentration de chromatine: DO 260nm

19) Prélever 3 ul de S1 ou S2 + 97 ul H₂O.

Pour 2 boîtes (diamètre 10 cm) de cellules ES, on attend environ 400ug de chromatine au total.

Immunoprecipitation de chromatine native

20) Dans un tube eppendorf, sur le glace, combiner:

- 700 ul de ChIP buffer 1x
- **30 ug** de chromatine (30 ug = généralement autour de 35 ul de fraction S2) pour les ChIP sur histones et jusqu'à **250 ug** de chromatine pour les ChIP sur chromatine de cellules ES-TAP-tag
- 10 ug d'anticorps (10 à 30 ul).

Attention: Lorsqu'on utilise de grandes quantités de chromatine (>30ug), il convient d'ajuster les concentrations en NaCl, Tris, etc (voir dans la partie II) de l'échantillon pour obtenir la composition du tampon N-ChIP 1x.

Important: Préparer un tube "Input" qui comprend 30 ug (ou la quantité choisie pour 1 point de l'expérience) de chromatine. Ajouter un volume équivalent de 2 x elution buffer, et compléter à 500 ul avec le 1 x elution buffer. Congeler ce tube jusqu'au moment de l'extraction phenol de l'ensemble des échantillons.

21) Agiter sur la roue (10 rpm) pendant la nuit.

2eme JOUR

Penser à préchauffer la solution d'éluion à 65°C.

- 22) Prélever le volume requis de protein-G sepharose et l'équilibrer dans le tampon ChIP buffer 1x en lavant 3 fois le slurry (avec 10 ml de tampon), dans un falcon 14ml à fond conique. Centrifuger à 1000 x g 1 min (rotor A-4-62, centrif 5810R) pour sédimenter les billes. Eliminer les surnageants grâce à la trompe à vide connectée à 3 tips 1ml emboîtés (couper les extrémités des tips pour une bonne aspiration).
- 23) Ajouter 50 ul de slurry / tube IP. Utiliser une p200 avec un tip coupé pour pipeter les billes de manière homogène Remettre en rotation à 4°C pendant 2-4h.
- 24) Placer le rotor à tubes eppendorf dans la centrif 5810R, à 4°C. Centrifuger les tubes à 1000xg pendant 1 min. Collecter le surnageant, qui correspond à la fraction non liée par l'anticorps (Unbound). Reprendre le slurry dans 1 ml de tampon de lavage A.
- 25) Transférer le contenu dans un tube de 14 ml comportant 11 ml de tampon de lavage A. Inverser le tube 3 fois. Installer le rotor 4-4-62 dans la 5810R, et centrifuger 1 min à 1000 x g.
- 26) Aspirer le surnageant avec la trompe à vide (attention à ne pas aspirer de billes; laisser un petit volume de solution au dessus des billes).
- 27) Ajouter 12 ml de tampon de lavage B, et inverser le tube 3 fois. Centrifuger 1 min à 1000 x g.
- 28) Aspirer le surnageant avec la trompe à vide. Ajouter 12 ml de tampon de lavage C. Inverser 3 fois, puis centrifuger 1 min à 1000 x g. Eliminer le surnageant avec la trompe à vide, puis avec une p200 pour éliminer tout le liquide.
- 29) Ajouter 500 ul de solution d'éluion (préchauffée à 65°C). Transférer dans des tubes eppendorf.
- 30) Agiter sur la roue à température ambiante pendant 30 minutes.
- 31) Centrifuger à 2000 rpm 30 s et collecter le surnageant.

Extraction phénol/Chloroforme des échantillons

- 32) Penser à décongeler le tube "input" pour le préparer en parallèle
- 33) Ajouter 500 ul de phenol/CHCl₃ – vortexer
- 34) Centrifuger 10 min à 13 000 rpm
- 35) Collecter le surnageant (380 ul, avec une p1000) dans un nouveau tube
- 36) Ajouter 3 ul de glycogène (10ug/ul) et 22 ul de NaCl 5M. Homogénéiser avec la p1000.
- 37) Ajouter 400 ul d'isopropanol. Agiter en retournant les tubes.
- 38) Placer à 4°C pendant quelques heures (ou O/N)
- 39) Centrifuger à 4°C à 13 000 rpm
- 40) Eliminer le surnageant – le culot doit être bien visible. Centrifuger 15 s les tubes vidés pour collecter et éliminer les derniers ul de solution isopropanol avec une p200.
- 41) Ajouter 750 ul d'ethanol 70%. Retourner les tubes 2-3 fois
- 42) Centrifuger 5 min à 13 000 rpm, puis éliminer le surnageant. Centrifuger 15 s les tubes vidés pour collecter et éliminer les derniers ul de solution ethanol 70% avec une p200.
- 43) Reprendre les culot avec 50 ul de TE. Laisser 15 minutes à RT, homogénéiser par vortex et congeler à -20°C.

Solutions

NaCl 5M (MW: 58,44)
Pour 1 l, peser 292,2 g

KCl 2M (MW: 74,55)
Pour 1 l, peser 150 g

Solution I pour la préparation des noyaux (500 ml)

Volume ou pesée	Concentration finale
51,3 g Sucrose	0,3M
15 ml KCl 2M	60 mM
1,5 ml NaCl 5M	15mM
2,5 ml MgCl ₂ 1M	5mM
0,5 ml EGTA 0,1M	0,1mM
7,5 ml Tris-HCl 1M PH7,5	15 mM

se conserve à -20°C, par aliquots de 30 ml

Solution III pour le coussin de sucrose (500 ml)

Tous les composés de la solution I → mêmes quantités sauf pour le sucrose:

205,2 g Sucrose 1,2M final

se conserve à -20°C, par aliquots de 34 ml. Avec 34 ml, on fait 4 tube de 8 ml de SW41 pour la centrifugation des cellules lysées.

Solution II pour la lyse des membranes plasmiques et la libération des noyaux.

Solution II = Solution I + Igepal (aussi appelé NP40) :

- 0,35 % (volume/volume) pour le cerveau, le foie fœtal
- 0,4% (v/v) pour les cellules ES
- 0,8% (v/v) pour les MEFs, dont la membrane plasmique est très résistante à l'Igepal.

Igepal CA-630: SIGMA I3021-100ML

Tampon MNase (200 ml):

Volume ou pesée	Concentration finale
21,9 g Sucrose	0,32 M
0,8 ml MgCl ₂ 1M	4 mM
0,2 ml CaCl ₂ 1M	1mM
0,8 ml NaCl 5M	20 mM
10 ml Tris HCl 1M PH 7,5	50 mM

Solution STOP MNase:

20 mM EDTA dans H₂O : 2 ml EDTA 0,5M PH 8 + 48 ml H₂O

Ajouter 30 ul de cette solution pour 500 ul de noyaux digérés MNase, soit 1,2 mM EDTA final.

Note: des concentrations plus fortes (par ex 10 mM) en EDTA sont plus efficaces pour arrêter l'enzyme mais perturbent le relarguage des oligonucléosomes en dehors des noyaux.

MNase 10 unités / ul

MNase NEB #M0247S (environ 8000 Kunitz units/40 ul, soit 200 units/ul)

Tampon Bleu/SDS

A préparer en tubes eppendorf de 2 ml:

200 ul de Bleu TAE 10X

200 ul de SDS 10%

1300 ul H₂O

Pour l'analyse de 3 ul de chromatine + 17 ul de Tampon Bleu/SDS, sur gel agarose 1,3%

Attention: chauffer l'échantillon à 65°C pendant 15 min pour dissocier les nucléosomes avant de loader le gel.

Tube à centrifuger pour SW41: Beckman Centrifuge tubes 14x89 mm, ref 331372

ChIP buffer 1 x

50 mM TrisHCl pH 7,5
5 mM EDTA
50 mM NaCl

1 litre

50 ml de solution 1M
10 ml de solution 0,5M
10 ml de solution 1M

Conversion du TEEP20 (fractions S2) en tampon ChIP (volumes indiqués pour 500ul)

+40 mM TrisHCl + 40ul Tris 1M
+4mM EDTA + 8ul EDTA 0,5 M
+30mM NaCl +6ul NaCl 5M

Attention: homogénéiser immédiatement après avoir ajouté un composé – de haute concentrations en sel destabilisent les nucléosomes.

Conversion du Tampon MNase (fractions S1) en tampon ChIP (volumes indiqués pour 500ul)

+4mM EDTA + 8ul EDTA 0,5 M
+30mM NaCl +6ul NaCl 5M

Attention: homogénéiser immédiatement après avoir ajouté un composé.

ChIP buffer 2

100mM TrisHCl pH 7,5
10 mM EDTA
100mM NaCl

Pour 50 ml

5 ml 1M
1 ml 0,5M
1 ml 5M

ChIP washing buffer A, B and C

Les 3 tampons comprennent:

50 mM TrisHCl pH 7,5
10 mM EDTA

La concentration en NaCl varie selon le tampon A, B, C

A: 75 mM NaCl

B: 125 mM NaCl

C: 175 mM NaCl

1 litre

50 ml de solution 1M
20 ml de solution 0,5M

15 ml de solution 5M /l

25 ml de solution 5M /l

35 ml de solution 5M /l

Elution Buffer

50mM Tris-HCl (PH 7,5)
50mM NaCl
5mM EDTA
1%SDS

200 ml

10 ml de solution 1M
2 ml de solution 5M
2 ml de solution 0,5M
20 ml 10% SDS

2 x Elution Buffer (pour traiter l'input)

100mM Tris-HCl (PH 7,5)
100mM NaCl
10mM EDTA
2%SDS
Qsp H2O

50 ml

5 ml de solution 1M
1 ml de solution 5M
1 ml de solution 0,5M
10 ml 10% SDS
33 ml H2O

Na Butyrate 200 x (soit 1M)

La concentration finale (1x) dans les tampons est de 5mM

Sigma B5887 (MW= 110)
9,1 ml H₂O + 1g NaButyrate

Glycogène: 10 ug/ul

TE buffer

10 mM Tris-HCl (PH 7,5)
1mM EDTA

50 ml

0,5 ml de solution 1M
100 ul de solution 0,5M

Protein-G Sepharose 4 Fast Flow

GE Healthcare
Ref 17-0618-01

IV. Protocole de bioinformatique

Analyse non exhaustive des données ChIP-seq

Abréviations utilisées :

RDL : remodeleurs de chromatine

A. Données d'entrées

Deux types de fichiers sont à considérer

1. Les fichiers BED

Sont les fichiers contenant pour chaque ligne une séquence lue dans la piste. Le nombre de lignes est égal aux nombres de lectures faites.

On parle de fichier BED pour un fichier texte qui respecte le format BED. On ajoute par précaution .bed a la fin de chaque fichier mais un .txt pourrait tout autant convenir du moment que le format est respecté.

Voilà le format attendu (d'après UCSC):

BED format

BED format provides a flexible way to define the data lines that are displayed in an annotation track. BED lines have three required fields and nine additional optional fields. The number of fields per line must be consistent throughout any single set of data in an annotation track. The order of the optional fields is binding: lower-numbered fields must always be populated if higher-numbered fields are used.

The first three required BED fields are:

1. chrom - The name of the chromosome (e.g. chr3, chrY, chr2_random) or scaffold (e.g. scaffold10671).
2. chromStart - The starting position of the feature in the chromosome or scaffold. The first base in a chromosome is numbered 0.
3. chromEnd - The ending position of the feature in the chromosome or scaffold. The chromEnd base is not included in the display of the feature. For example, the first 100 bases of a chromosome are defined as chromStart=0, chromEnd=100, and span the bases numbered 0-99.

The 9 additional optional BED fields are:

4. name - Defines the name of the BED line. This label is displayed to the left of the BED line in the Genome Browser window when the track is open to full display mode or directly to the left of the item in pack mode.

5. score - A score between 0 and 1000. If the track line useScore attribute is set to 1 for this annotation data set, the score value will determine the level of gray in which this feature is displayed (higher numbers = darker gray). This table shows the Genome Browser's translation of BED score values into shades of gray:

shade

score in range ≤ 166 167-277 278-388 389-499 500-611 612-722 723-833 834-944 ≥ 945

6. strand - Defines the strand - either '+' or '-'.

7. thickStart - The starting position at which the feature is drawn thickly (for example, the start codon in gene displays).

8. thickEnd - The ending position at which the feature is drawn thickly (for example, the stop codon in gene displays).

9. itemRgb - An RGB value of the form R,G,B (e.g. 255,0,0). If the track line itemRgb attribute is set to "On", this RGB value will determine the display color of the data contained in this BED line. NOTE: It is recommended that a simple color scheme (eight colors or less) be used with this attribute to avoid overwhelming the color resources of the Genome Browser and your Internet browser.

10. blockCount - The number of blocks (exons) in the BED line.

11. blockSizes - A comma-separated list of the block sizes. The number of items in this list should correspond to blockCount.

12. blockStarts - A comma-separated list of block starts. All of the blockStart positions should be calculated relative to chromStart. The number of items in this list should correspond to blockCount.

Example:

Here's an example of an annotation track that uses a complete BED definition:

```
track name=pairedReads description="Clone Paired Reads" useScore=1
```

```
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
```

```
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

Dans notre cas, le fichier BED adéquat est celui qui est en input de SICER. Le fichier SISSR est inutilisable.

2. Le fichier de détection de pics

Ce fichier contient la liste des pics, avec principalement une colonne avec le chromosome, une autre avec le début du pic, une autre avec la fin du pic puis autres colonnes supplémentaires peuvent être ajoutées (scores, p-values)

chr1	4773800	4777799	103	28	3.34509781716e-14	2.2972280567	8.18307518003e-14
chr1	4790800	4802199	193	51	3.80477631247e-26	2.36326297704	2.28784509443e-25
chr1	4840200	4842799	76	10	5.34198570464e-28	4.74611777151	3.61579292365e-27
chr1	4843600	4849999	115	27	4.45600630667e-20	2.65986132419	1.72221005747e-19
chr1	5002600	5015599	172	48	3.40503338457e-21	2.23775289666	1.43147017045e-20
chr1	7076600	7081999	112	27	8.51338131204e-19	2.59047363747	2.98197405508e-18
chr1	7085200	7093199	114	36	1.91606602482e-11	1.97754907146	3.80147109978e-11
chr1	7387200	7389199	68	4	4.75849672166e-46	10.6163160679	9.5060403818e-45
chr1	8571400	8573599	47	6	9.85821921416e-19	4.89183191362	3.42995769629e-18
chr1	8923800	8926999	51	10	8.44183903746e-13	3.18489482036	1.85413454087e-12
chr1	9533600	9539999	211	69	6.9107557217e-18	1.90966981272	2.25762914856e-17
chr1	9690800	9693599	75	17	1.39371763945e-14	2.75509932557	3.51777094412e-14
chr1	9758400	9758799	22	1	4.21593298318e-19	13.7387619702	1.51824124166e-18
chr1	9928800	9934799	131	33	4.51904475684e-20	2.47903280729	1.74484718713e-19

3. Les données publiées

i. Les trouver

Chercher l'article et dans le mat et meth trouver le numéro d'accension

<http://www.ncbi.nlm.nih.gov/geo/>

Télécharger les données au format.bed

Deziper (avec gunzip)

```
>GUNZIP GSM165892.GZ
```

Renseigner l'alignement qui a été utilisé : mm8 ou mm9

Exemple :

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12241>

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16893>

ii. Les convertir

Se placer dans le bon dossier avec un terminal Linux.

Utiliser la commande liftOver selon la ligne de commande suivante

```
>./LIFTOVER GSM165892.BED CHAINE-DE-CONVERSION  
RDL.BED UNMAPPEDRDL.TXT
```

Exemple pour passer de mm8 à mm9 : chaine de conversion mm8ToMm9

B. GPAT

1. Input

Le fichier avec les pics détectés

Pour les fichiers d'ouput de :

Sicer

Il faut juste remplacer l'extension par .bed

SISSR

Il faut simplement ouvrir le fichier et enlever les 57 premières lignes.

Enregistrer le fichier en texte avec séparateur tab, retour chariot de windows

Quest

Ouvrir le fichier.

Enlever toutes les lignes qui commencent par un P, ne garder que celles qui commencent par R

Scinder la colonne 2 en deux colonnes distinctes avec le séparateur ' – '

Sauver le fichier en texte séparateur tab, retour chariot de windows

2. Run de GPAT

Options de fenêtres de recherche

Choisir le genome de référence

Gene annotation : Refseq ID

Research mode :

1. Direct search

2. Window search

3. Promoteur search

Autres options

Chromosomic position annotation : option

Expression data : ES celle ou nos propres données

Display option : ?

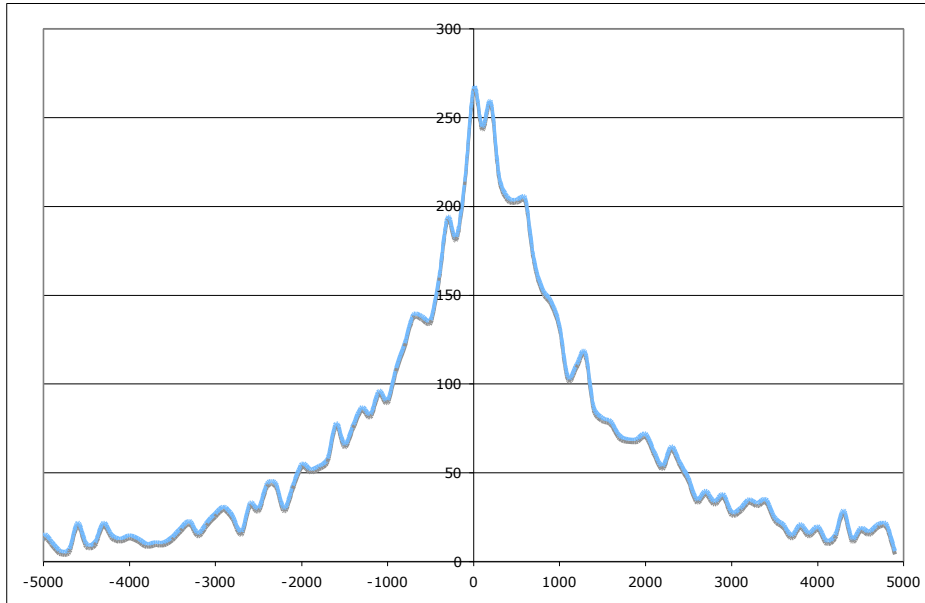
Enregistrer le résultat sous forme de tableau

3. Autre résultat pour visualisation

Sauver le fichier

Partie haute

Ce tableau donne pour une fenêtre de 100 bp, le nombre de pics dont le centre se trouve dans cette fenêtre. On peut extraire ses données et faire un tableau excel et visualiser en faisant un nuage de points, une ligne ou un diagramme.



Partie basse

On retrouve le nom du gène, la distance par rapport au TSS et une valeur correspondant à l'expression du gène (ces données peuvent aussi être rentrées par l'utilisateur au début)

C. Comparaison de pics

1. Input

Les deux fichiers d'entrée sont deux fichiers type excel avec trois colonnes.

La première colonne contient le chromosome (1)

Il faut reprendre les fichiers d'input de GPAT, insérer une colonne supplémentaire qui comprend le milieu du pic puis ajouter deux autres colonnes contenant chacune d'elle -500pb (2) et +500pb (3) de part et d'autre du milieu du pic.

Il faut ensuite créer un fichier contenant les colonnes (1), (2) et (3) dans l'ordre.

2. Comparaison avec Galaxy

A trouver

D. Promoteur profile

1. Input

L'input de ce programme est le fichier BED de la protéine étudiée comprenant tous les tag séquencés.

2. Le programme

L'algorithme

L'algorithme à utiliser est promoter-profile

La chaîne de caractère pour exécuter le programme

```
PYTHON PROMOTER_PROFILE.PY FICHER-ENTRE.BED
```

A vérifier

L'algorithme va utiliser un fichier (refseq_mm9) qui doit être dans le même dossier

De même le chemin écrit « en dur » dans l'algorithme en python doit correspondre à l'emplacement du dossier pour que les calculs soient faits

3. Output

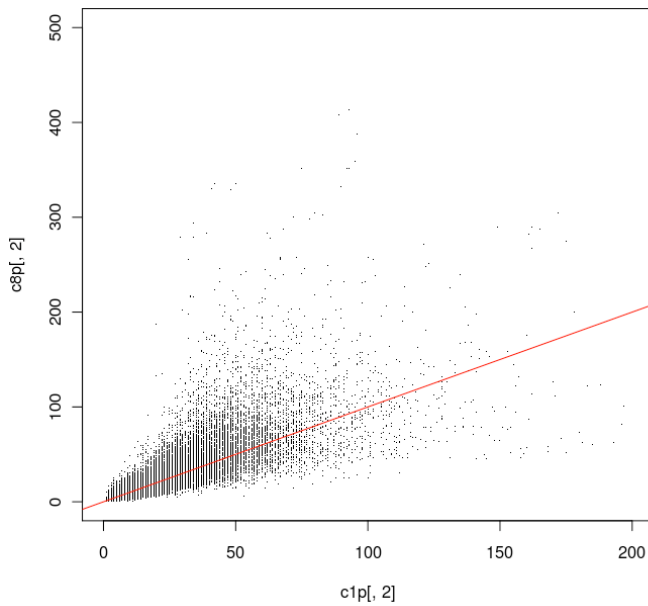
Le fichier de sortie est un tableau à deux colonnes. La première indique le nom du gène et la deuxième le nombre de tag trouvés dans une fenêtre de xxxpb autour du promoteur.

4. Un exemple de résultat

Comparaison de fichiers

On peut faire la même opération pour deux fichiers et les comparer entre eux

Comparaison par un blot sur R



Comparaison en mettant des couleurs pour un gradient

Deuxième exemple

On peut ajouter en parallèle les données d'expression.

Trier les gènes et faire un code couleur

E. Multiple profile

Identique à promoter profile sauf que l'on donne en input une liste bien définie (ex : une liste de pics)

F. Python Distribution

5. Input

Il faut en input donner une liste de références (ex : une liste de pics)

Ajouter à cela les fichiers Wig = fichiers de densité

Les fichiers liste de référence

Les fichiers de référence doivent contenir une colonne chromosome, deux colonnes pour le début et la fin de la zone d'intérêt (ex : un pic = une zone d'intérêt)

Généralement on prendra un fichier de détection de pics Sicer paramètres 200 200 FDR10⁻⁹

Pour cela faire une copie du fichier de détection des pics et la renommer

```
EX : SAMPLES/SICER/SICERRDL/
EDITION > DUPLIQUER
EDITION > RENOMMER RDL200-9.BED
```

Puis transférer la copie dans le dossier python Arnaud

Les fichiers Wig

Il y a déjà des fichiers Wiggle faits. Vérifier que cela n'a pas déjà été fait.

Pour faire tourner le Wig Maker (en java) il faut d'abord nettoyer les fichiers BED qui sont en input de sicer. C'est-à-dire enlever toutes les lignes qui correspondent à une séquence qui n'a pas pu être alignée sur le génome mm9 à la sortie du séquenceur.

Cette étape n'est pas nécessaire pour les fichiers qui ont été convertis du mm8 au mm9.

Cette étape n'est généralement pas nécessaire pour les fichiers téléchargés. En revanche, il faut vérifier au préalable que le format bed soit respecté. Colonnes décrite au début de ce document et séparées par des tabs "\t"

Nettoyer un fichier BED

Mettre le fichier à nettoyer dans le dossier python/Bed

Entrer la commande suivante :

```
PYTHON BED_CLEANER.PY RDL_SICER.BED
```

Une fois l'opération terminée, déplacer le fichier de sortie RDL_sicer.bed_clean.bed dans le dossier python.

Générer un dossier de fichier WIG

Pour générer le dossier fichiers de densités, une application java : Wig_Maker.jar va être utilisée. Entrer la commande suivante :

```
JAVA -Xmx3000M -JAR WIGMAKER3.JAR
RDL_SICER.BED_CLEAN.BED 150 25 MOUSE
```

Xmx3000m : alloue 3Gigas de mémoire vive maximum

150 : taille de la découpe des fragments séquencés (pour les données de la littérature, utiliser plutôt 200)

25 : pas

mouse : pour le génome de la souris. Moins de mémoire vive nécessaire

Sortie : un dossier de fichiers (un par chromosome) wig

6. Faire tourner le script Distribution

Distribution permet avec un fichier de référence (par exemple pics de RDL) de comparer les densités d'autres facteurs, c'est-à-dire l'enrichissement des tag séquencés aux mêmes emplacements sur le génome.

Bon à savoir : une liste de plus de 15000lignes demande une analyse très laborieuse (longue et gourmande en mémoire).

La ligne de commande à utiliser est :

```
PYTHON DISTRIBUTIONV5.PY RDL200-9.BED WIG_FOR_RDL
```

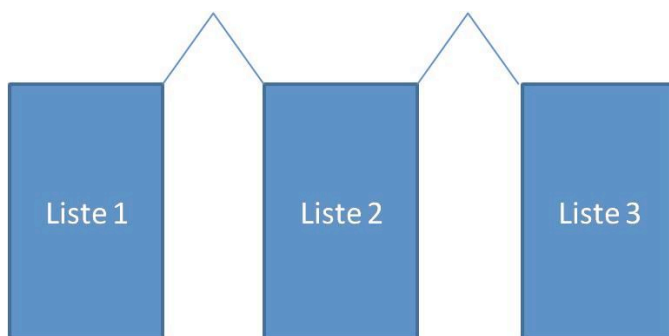
Sortie: un fichier correspondant à une matrice nombre de lignes = longueur du fichier RDL200-9.bed, et 400 colonnes

Répéter cette opération jusqu'à obtenir tous les fichiers voulus.

Ceci prend au maximum une dizaine de minutes.

7. Concatener des fichiers de sortie

Pour faire la figure visualisée par Java cluster et faut rassembler les fichiers d'intérêt.



```
PASTE OUTPUT_RDL200-9_WIGS_FOR_RDL OUTPUT_RDL200-
9_WIGS_FOR_ESH3K4ME3 OUTPUT_RDL200-
9_WIGS_FOR_ESH3K36ME3 > RDL_K43_K36_OVERRDL.TXT
```

fichier de sortie: RDL_K43_K36_overRDL.txt pour : densité de RDL+H3K4me3+H3K36me3 sur les pics de RDL

8. Gene cluster 3.0

Input

Pour ouvrir l'interface graphique qui commande l'application double cliquer sur l'exécutable dans le dossier LREGE.

Pour entrer le fichier : Open file > RDL_K43_K36_overRDL.txt

Les options

K-mean

C'est celui qui nécessite le moins de temps donc on privilégiera cette analyse pour les fichiers qui comportent plus de 1000 lignes et 3(*400) colonnes.

Les options à activer sont :

- ✓ Organise gènes
- ✓ nombre de sous-groupes ou clusters : par défaut 10
- ✓ k-mean
=>Execute

Correlation (uncentered)

Sortie standart

En sortie on obtient 2 fichiers.

9. Java tree view

Ouvrir Java Tree view

Sur un Mac

Dans Application, ouvrir l'application JavaTreeview.

On pourra vérifier que cette application se lance avec une mémoire suffisante.

Sur linux

i. Avec le raccourci (solution préférée)

Dans le dossier LREGGE, lancer dans un terminal, lancer le .bat

ii. En ligne de commande

Se placer dans le bon repertoire :

```
JAVA -XMX3000M -JAR TREEVIEW.JAR
```

Ouvrir un fichier

Open

Ouvrir le fichier généré par Cluster3.0 en .cdt

Les options

Une fois que l'image s'est affichée.

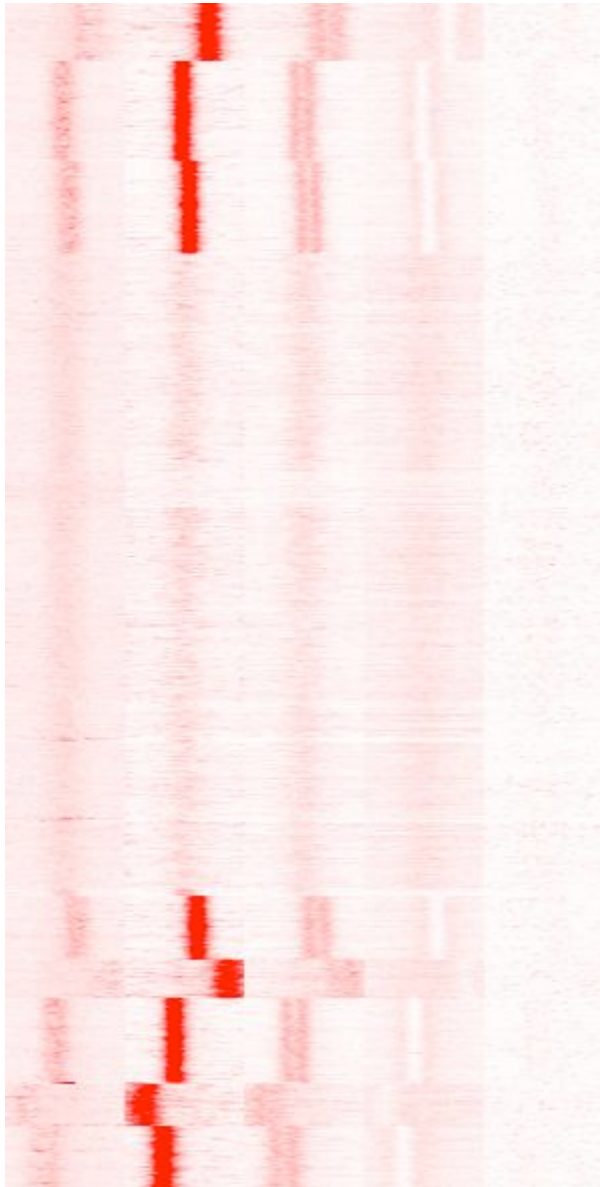
Option>pixels settings

Mettre full visualisation en x et y

Ne pas toucher à global (pour l'instant)

Régler le seuil d'intensité. La valeur affichée correspond à valeur pour laquelle le seuil d'intensité est atteint.

Généralement on pensera aussi à mettre le blanc comme couleur de zéro. La couleur pour les valeurs d'intensité sont de préférence plutôt foncées (rouge, bleu, vert..etc..)



Les résultats

Export de l'image

i. Le plus simple

Une copie d'écran sur mac: Alt+Maj+3

Une sélection d'une partie d'écran sur mac: Alt+Maj+4

Sur PC : Imprécran, coller sur word ou power point

ii. Le plus mieux

Export>Export image

Décocher save selection

Nommer l'image et ajouter l'extension .png

Export d'une liste simple

Si on veut savoir à quelle région correspond les pics que l'on regarde. Ex : un sous groupe.

Export>export list

Sauver en nommant exactement ce que l'on extrait pour retrouver les données plus tard.

Utilisation de cette liste pour trouver les gènes correspondants (GPAT)

Cliquer sur le fichier et l'ouvrir (Fichier>ouvrir avec) avec une feuille de calcul (Open office)

Mettre comme autre séparateur de texte : deux points : , tiret haut - , et tiret bas _

Ouvrir le fichier et vérifier que l'on a bien 4 colonnes.

Enregistrer

Renommer le fichier et ajouter l'extension .bed

Ouvrir une fenêtre Explorateur et sélectionner l'onglet GPAT

Choix du format de données : BED

Browse>Choisir le fichier d'intérêt

Choix du génome mouse mm9

Entrer

Vérifier que les données ont bien été prises en compte et valider

Sauver le tableau.txt

Les gènes qui correspondent sont identifiés.

Pour le reste de l'analyse se reporter au paragraphe ci-dessus.

Utilisation de R pour visualiser la moyenne de l'intensité des pics

Exporter les données sous R pour pouvoir visualiser un profil moyen :

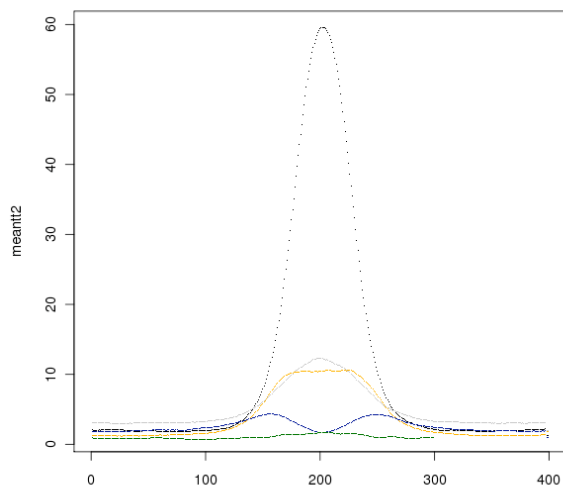
Ouvrir un terminal et taper la commande R

```
> table <- read.table('Chd8-K4me3me2me1-CTCF_densities_K_G10_data_categories2et3.cdt')
```

```

> mean_t=mean(table)
> length(mean_t)
> meantt <- as.matrix(mean_t)
> class(meantt)
[1] "matrix"
> matplot(meantt)
> matplot(meantt, pch = '.')
> meantt1 <- meantt[4: 402]
> length(meantt1)
[1] 399
> meantt2 <- meantt[403: 801]
> meantt3 <- meantt[802: 1200]
> meantt4 <- meantt[1201: 1599]
> meantt5 <- meantt[1600: 1898]
> matplot(meantt2, pch = '.')
> matpoints(meantt1, pch = '.', col = 280)
> matpoints(meantt3, pch = '.', col = 'orange')
> matpoints(meantt4, pch = '.', col = 'darkblue')
> matpoints(meantt5, pch = '.', col = 'darkgreen')
> savePlot('profil_moyen_sur_deux_categories_Ch8.png')

```



```

> jpeg('rplot.jpg')
> plot(x, y)
> dev.off()

```

G. Analyse Gene Ontology

<http://www.geneontology.org/GO.tools.microarray.shtml>

<http://llama.med.harvard.edu/funcassociate/>

Tutorial

```
> x = data.frame(a=c(1,2,3),b=c("GO:0004526","GO:0031570","GO:0051053"))
>
> x
  a      b
1 1 GO:0004526
2 2 GO:0031570
3 3 GO:0051053
> x$b
[1] GO:0004526 GO:0031570 GO:0051053
Levels: GO:0004526 GO:0031570 GO:0051053
> as.character(x$b)
[1] "GO:0004526" "GO:0031570" "GO:0051053"
> lapply(as.character(x$b),Term)
[[1]]
      GO:0004526
"ribonuclease P activity"

[[2]]
      GO:0031570
"DNA integrity checkpoint"

[[3]]
      GO:0051053
"negative regulation of DNA metabolic process"

> lapply(as.character(x$b),Ontology)
[[1]]
GO:0004526
      "MF"

[[2]]
```

```
GO:0031570
  "BP"

[[3]]
GO:0051053
  "BP"

> y = lapply(as.character(x$b),Ontology)
> unlist(y)
GO:0004526 GO:0031570 GO:0051053
  "MF"      "BP"      "BP"
> unlist(y,use.names=FALSE)
[1] "MF" "BP" "BP"
> cbind(x,unlist(y,use.names=FALSE))
  a          b unlist(y, use.names = FALSE)
1 1 GO:0004526          MF
2 2 GO:0031570          BP
3 3 GO:0051053          BP
> cbind(x,Ont=unlist(y,use.names=FALSE))
  a          b Ont
1 1 GO:0004526  MF
2 2 GO:0031570  BP
3 3 GO:0051053  BP
> x=cbind(x,Ont=unlist(y,use.names=FALSE))
> x
  a          b Ont
1 1 GO:0004526  MF
2 2 GO:0031570  BP
3 3 GO:0051053  BP
>
```


Résumé

Les cellules souches embryonnaires (cellules ES) ont la capacité unique de se diviser indéfiniment et de pouvoir se différencier en de multiples types cellulaires. Elles apparaissent donc très prometteuses comme agents thérapeutiques dans les traitements médicaux du futur. Un enjeu majeur de la recherche actuelle consiste à comprendre la contribution des protéines régulatrices de la chromatine à la plasticité et au contrôle de l'expression du génome des cellules. La famille des remodeleurs Chd, qui fait partie de la super famille SNF2, comprend neuf membres, soit le tiers des remodeleurs exprimés dans les cellules ES murines. L'objectif principal de ce projet de thèse a consisté à identifier de manière exhaustive les gènes cibles de chaque facteur pour comprendre comment ils participent à la régulation du génome et se partagent le remodelage de la chromatine.

Nous avons entrepris un projet à grande échelle dans lequel chaque gène codant chaque Chd a été fusionné, à son extrémité carboxy-terminale, à une séquence codant une étiquette, par recombinaison homologe en cellules ES. Les cellules ES étiquetées ont ensuite été utilisées pour des expériences d'immunoprécipitation de chromatine (ChIP-seq). La présence de l'étiquette a permis de standardiser et d'optimiser les méthodes d'immunoprécipitation des protéines. Les fragments d'ADN isolés ont ensuite été séquencés dans le laboratoire d'Ivo Gut (CEA/CNG -Evry- et CNAG -Barcelone-). Nous avons également analysé les transcriptomes des cellules ES où la déplétion de chaque protéine Chd a été réalisée, par hybridation sur puce et RNA-seq.

Ces données ont permis de montrer le rôle de NuRD (Chd4, Hdac2) au sein des réseaux de la régulation transcriptionnelle des ES. Les données obtenues pour les facteurs Chd1, Chd8 et Chd4 montrent des rôles différents mais interconnectés pour chaque protéine. Enfin, ces données nous ont permis de proposer des hypothèses pour expliquer comment ces protéines contribuent à la régulation du génome.

Summary

Embryonic stem cells have the unique capacity to divide indefinitely and to differentiate in all cell types. They are promising for therapeutic applications for future treatments. Understanding the contribution of chromatin remodeler to the control of gene expression is critical to reach this goal. The Chd family which belongs to the SNF2 superfamily, contains nine members and represent about one third of the chromatin remodelers expressed in ES cells. Our aim was to identify the target genes of each remodeler, to understand how they are involved in transcription regulation and how they share chromatin regulation in ES.

This large-scale study started with the tagging of each Chd. Using homologous recombination, we generated 9 ES cell lines each carrying a tagged allele. The C-terminus of each gene was fused with a FLAG-HA tag. This tag allowed efficient chromatin immunoprecipitation for all the proteins. We performed tandem affinity chromatin IP for each protein. Immunoprecipitated DNA was sequenced onto an Illumina platform in the lab of I. Gut (CEA/CNG and CNAG). We next analysed the transcriptome of Chd-depleted ES cells, using microarray hybridization and RNA-seq strategy.

Our data show that NuRD (Chd4, Hdac2) is extensively involved in ES transcriptional pluripotency core network. Data obtained for Chd1, Chd8 and Chd4 show that each Chd has a particular binding pattern onto the mouse ES cell genome, suggesting different roles. However they regulated each other and belong to the same complex core circuit. Finally, our results lead to new hypothesis to explain how Chd proteins contribute to genome regulation.