



# Brn2 et Zic1 spécifient l'identité neuronale des cellules souches embryonnaires murines lors de la différenciation induite par l'acide rétinoïque

Sylvia Urban

## ► To cite this version:

Sylvia Urban. Brn2 et Zic1 spécifient l'identité neuronale des cellules souches embryonnaires murines lors de la différenciation induite par l'acide rétinoïque. Biologie du développement. Université de Strasbourg, 2014. Français. NNT : 2014STRAJ115 . tel-01241689

**HAL Id: tel-01241689**

**<https://theses.hal.science/tel-01241689>**

Submitted on 10 Dec 2015

**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.



École Doctorale  
des Sciences de la Vie  
et de la Santé  
S T R A S B O U R G

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

IGBMC

**THÈSE** présentée par :

**Sylvia URBAN**

soutenue le : 26 septembre 2014

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences de la Vie et de la Santé

Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

***Brn2 et Zic1 spécifient l'identité neuronale  
des cellules souches embryonnaires murines  
lors de la différenciation induite  
par l'acide rétinoïque***

**THÈSE dirigée par :**

**M. Davidson Irwin**

Docteur, IGBMC, Illkirch

**RAPPORTEURS :**

**Mme Nieto Laurence**

Professeur, IPBS, Toulouse

**Mme Pituello Fabienne**

Docteur, CBD, Toulouse

---

**AUTRE MEMBRE DU JURY :**

**Mme Giangrande Angela**

Docteur, IGBMC, Illkirch

## *Remerciements*

Je tiens avant tout à remercier les membres de mon jury, Laurence Delacroix, Angela Giangrande et Fabienne Pituello d'avoir accepté d'évaluer ce travail.

Je tiens également à remercier le Dr. Irwin Davidson qui m'a accueillie dans son laboratoire pour la première fois en juillet 2008. Stage volontaire de M1, stage de M2 et puis thèse, cela fait maintenant 6 ans que j'ai les pieds dans cette équipe. Merci de m'avoir fait confiance sur un projet « ES » parfois délicat, à force d'y croire parfois la roue tourne !

### *Merci à la Team Davidson et autres IGBMCiens*

Anas, Dana, Daniil, Diana, Fédé, Gab, Isa, Igor, Marie, Patrick, Shilpy (ou pour faire court, le 4011 et le 4039). Gab, est ce que j'ai le droit de mettre un « ON » dans les remerciements ?? (dit oui, s'il te plait !!!). Anas, mon nouveau voisin, et néanmoins ami...depuis ton arrivée je vis en quelque sorte une deuxième thèse. Que ce soit des discussions purement expérimentales, des conversations scientifiques générales, ou tout simplement des discussions « poissons » autour d'un café, j'ai découvert en toi un conseiller et un confident. Merci d'avoir pris le temps pour moi. Patrick, sans que je rentre dans les détails (ce n'est ni le moment, ni l'endroit...) entre thésards on se comprend non ? N'oublie pas, on a une partie de morpion à faire ! Isa, même si on a parfois parlé trop fort, ça a toujours été un plaisir de discuter avec toi. Merci pour tes petits coups de main et ta disponibilité.

Merci aux anciens, Céline, Raph, Manu, Dom, Tom, Amin

Merci aux « Anges de la culture », à Muriel, Bernard, Serge, Céline, Tao, Amandine et à toutes les personnes que j'ai pu croiser dans les services communs et plateformes de l'IGBMC. Stéphanie, je te rassure, je ne t'oublie pas.... Comment pourrais-je d'ailleurs !! Tu as commencé à te moquer de moi le jour où je suis arrivée avec un plâtre, et depuis ce jour là, je ne compte même plus le nombre de fois où nous nous sommes charriées mutuellement. En tout cas ça a toujours été un plaisir de discuter avec toi.

### *Merci à mes ami(e)s ...*

A celles qui sont là depuis un bout de temps, Stéphanie G, Stéphanie M, Sabrina, Laure, tout simplement merci de m'avoir donné votre amitié et votre confiance. J'ai vécu avec chacune de vous des moments que je n'oublierai pas, qu'ils soient bons (souvent) ou moins bons (rarement), c'est ça aussi l'amitié ! Lycée, prépa, fac, thèse, vous avez chacune été un soutien clé dans une période importante de ma vie. Stéphanie M, je compte bien venir enfin te rendre visite, que ce soit à côté du Lac de Constance ou à Dubaï. Laure, Barcelone et pleins d'autres aventures nous attendent... hâte d'y être !!

Ségolène & Co, un jour quelqu'un m'a dit que la vie n'était pas faite uniquement d'amitiés, mais aussi de « moments » d'amitiés. Je crois que ces moments d'amitiés je les ai découverts grâce à vous. Je n'oublierai jamais les restos, les barbecues et autres dé-crémaillères fêtées

avec vous. Sans oublier votre bonne humeur et nos (vos) discussions pendant les footings (j'avoue, je ne suis pas capable de faire deux choses à la fois...). Je n'aime pas courir, et pourtant vous l'avez fait, vous avez réussi à me faire prendre goût aux footings.

Les «3C » de la bibliothèque : Caro, Charlène, Camille, parfois le soutien vous arrive de la ou vous ne l'attendez pas. Après plusieurs soirées, rando, ou pique-niques, c'est finalement autour de notre rédaction de thèse que nous nous sommes vraiment retrouvées. Qu'est ce que ça aurait été sans nos « moments bibliothèques » ? J'ai réellement appris à vous découvrir durant ces dernières semaines, autour d'un café, d'une tablette de chocolat ou simplement au détour d'un passage à la bibliothèque à des horaires ou seul un thésard peut s'y perdre. Pareil pour toi Mélanie, on se connaissait (de la fac et du 4<sup>e</sup> étage de l'IGBMC) mais se retrouver les deux premières à la bibliothèque le matin ça permet de discuter plus amplement. On n'a peut être pas porté chance à l'équipe de France, mais au fond, on était pour l'Allemagne non ??

Les Chipster's, qu'aurai été ma thèse sans le hand ? Sûrement pas la même chose... Malgré une luxation du coude (Merci Aude, 3 mois sans hand, 3 semaines sans pipettes !), vous m'avez permis de me défouler 3x par semaine. Le surnom d'« Einstein » que vous m'avez donné n'est sûrement pas le plus approprié, mais il m'a bien fait rire. La thèse se termine mais une nouvelle saison est engagée à vos côtés, et j'espère bien qu'on battra couette-couette histoire de la commencer en beauté. En attendant, dès que j'ai fini ce paragraphe je fais 200 abdos, 30 pompes et 5 minutes de gainage, prépa physique oblige !

Les amis du volley (VBO), ça a été un plaisir de venir découvrir ce sport avec vous. Un beach volley sur la plage du K-Beach suivi d'un petit Mojito, rien de mieux pour retrouver la forme et la motivation. Pauline, j'espère que nos « moments d'amitiés » (volley, rando, concert) vont continuer d'exister, ça a été un vrai plaisir de les partager avec toi en tout cas.

Pour finir, Miriam, je crois que je vais « enfin » avoir le temps de venir à Barcelone. Il faut bien en profiter un peu après la thèse non ? et bien sur, je n'oublierai pas d'emmener mes « Running » pour un petit footing avec toi !

### Merci à ma famille

A mes parents, ma sœur, Seb, Titi et Mathis, vous m'avez toujours soutenue, et ce bien avant la thèse. Même si vous avez parfois trouvé que j'avais la belle vie (je ne vois pas du tout ce qui vous fait dire ça !!!), ça n'aurai pas été aussi « beau » sans votre soutien. Vous m'avez souvent demandé ce que je faisais réellement au labo les week-ends, je vous invite donc à lire les quelques pages qui suivent ces remerciements pour comprendre ce que c'est réellement le travail de thésard.... Ou plutôt, MON travail de thésarde !

|                                  |   |
|----------------------------------|---|
| Abréviations .....               | 1 |
| Liste des figures et tables..... | 4 |
| Contenu du CD .....              | 6 |

## INTRODUCTION

### 1 CHAPITRE I : LES CELLULES ES : ORIGINE ET UTILISATIONS ..... 9

#### 1.1 Origine des cellules souches embryonnaires ..... 9

#### 1.2 Capacité de prolifération *in vitro* et maintien de la pluripotence..... 11

##### 1.2.1 Implication de différentes voies de signalisation ..... 11

###### 1.2.1.1 La voie LIF-STAT ..... 12

###### 1.2.1.2 La voie MAPK-ERK ..... 13

###### 1.2.1.3 La voie Phosphoinositide 3-kinase ..... 14

###### 1.2.1.4 La voie TGFb ..... 14

###### 1.2.1.5 La voie Wnt ..... 15

##### 1.2.2 Cycle cellulaire ..... 15

##### 1.2.3 Des facteurs de transcription impliqués dans le maintien de la pluripotence ..... 16

###### 1.2.3.1 Rôle des facteurs de transcription OCT4, NANOG et SOX2 ..... 16

###### 1.2.3.2 Surexpression des facteurs de transcription et différenciation..... 17

#### 1.3 Potentiel de différenciation *in vitro* ..... 18

##### 1.3.1 Endoderme ..... 19

##### 1.3.2 Mésoderme ..... 19

##### 1.3.3 Ectoderme ..... 20

##### 1.3.4 Cellules de la lignée germinale ..... 21

#### 1.4 Des cellules ES dernière génération : les iPS..... 21

### 2 CHAPITRE II : LES CELLULES ES : UN SYSTEME EXPERIMENTAL POUR LA COMPREHENSION DES RESEAUX TRANSCRIPTIONNELS *IN VIVO*..... 24

#### 2.1 Différenciation neuronale *in vivo*..... 24

##### 2.1.1 Neurulation..... 24

##### 2.1.2 La différenciation du tube neural ..... 24

###### 2.1.2.1 Le tube neural, à l'origine du système nerveux central..... 24

###### 2.1.2.2 La voie Notch : au cœur de la différenciation neuronale ..... 25

##### 2.1.3 Les différents composants du système nerveux central ..... 26

##### 2.1.4 Biologie cellulaire de la neurogenèse..... 27

###### 2.1.4.1 Les cellules neuroépithéliales ..... 27

###### 2.1.4.2 Les cellules radiales gliales..... 27

###### 2.1.4.3 Les progéniteurs basaux ..... 28

###### 2.1.4.4 Cortex et neurones glutamatergiques ..... 29

#### 2.2 Modèle de différenciation neuronale des cellules ES *in vitro*..... 30

|       |                                                                          |    |
|-------|--------------------------------------------------------------------------|----|
| 2.2.1 | Importance de différentes voies de signalisation.....                    | 30 |
| 2.2.2 | Protocole de différenciation en neurones glutamatergiques in vitro ..... | 30 |

### **3 CHAPITRE III : LA TRANSCRIPTION ET SA REGULATION..... 33**

#### **3.1 Mécanismes généraux de l'initiation de la transcription ..... 33**

|         |                                                     |    |
|---------|-----------------------------------------------------|----|
| 3.1.1   | Les différents composants du promoteur minimum..... | 33 |
| 3.1.1.1 | La boîte TATA.....                                  | 33 |
| 3.1.1.2 | BRE .....                                           | 34 |
| 3.1.1.3 | Inr.....                                            | 34 |
| 3.1.1.4 | DPE.....                                            | 34 |
| 3.1.1.5 | MTE.....                                            | 35 |
| 3.1.2   | Les îlots CpG .....                                 | 35 |
| 3.1.3   | Régions régulatrices en dehors du promoteur.....    | 36 |
| 3.1.3.1 | Régions proximales .....                            | 37 |
| 3.1.3.2 | Régions distales .....                              | 37 |

#### **3.2 Les acteurs protéiques de la transcription..... 38**

|         |                                                    |    |
|---------|----------------------------------------------------|----|
| 3.2.1   | L'ARN polymérase II et la machinerie de base ..... | 38 |
| 3.2.1.1 | L'ARN polymérase II.....                           | 38 |
| 3.2.1.2 | Les facteurs généraux de la transcription.....     | 39 |
| 3.2.2   | Modèles d'assemblage du PIC.....                   | 39 |
| 3.2.2.1 | Assemblage séquentiel .....                        | 39 |
| 3.2.2.2 | Modèle de l'holoenzyme .....                       | 40 |
| 3.2.3   | Les cofacteurs de la transcription.....            | 40 |

#### **3.3 Régulation de la transcription..... 41**

#### **3.4 Transcription et épigénétique..... 42**

|         |                                                        |    |
|---------|--------------------------------------------------------|----|
| 3.4.1   | Vue d'ensemble des modifications d'histones .....      | 42 |
| 3.4.2   | Mise en place des principales marques d'histones ..... | 43 |
| 3.4.2.1 | Acétylation.....                                       | 43 |
| 3.4.2.2 | Phosphorylation.....                                   | 43 |
| 3.4.2.3 | Méthylation .....                                      | 44 |

### **4 CHAPITRE IV : LA REGULATION DE LA TRANSCRIPTION PAR LES RECEPTEURS DE L'AR ..... 46**

#### **4.1 La voie de l'acide rétinoïque ..... 46**

#### **4.2 Les récepteurs de l'acide rétinoïque ..... 47**

|       |                                                     |    |
|-------|-----------------------------------------------------|----|
| 4.2.1 | Structure des récepteurs à l'acide rétinoïque ..... | 48 |
| 4.2.2 | Le LBD .....                                        | 49 |
| 4.2.3 | Le DBD .....                                        | 49 |

#### **4.3 Reconnaissance de l'ADN ..... 50**

#### **4.4 Modèle d'activation des gènes cibles de l'AR..... 50**

#### **4.5 L'acide rétinoïque dans la différenciation neuronale..... 51**

## **5 CHAPITRE V : ZIC1 ET BRN2, DEUX FACTEURS CLES DE LA DIFFERENCIATION NEURONALE ? ..... 54**

|            |                                                                                |           |
|------------|--------------------------------------------------------------------------------|-----------|
| <b>5.1</b> | <b>Zic1.....</b>                                                               | <b>54</b> |
| 5.1.1      | Structure des gènes Zic.....                                                   | 54        |
| 5.1.2      | Expression et régulation des gènes Zic.....                                    | 55        |
| 5.1.3      | Fonctions biologiques.....                                                     | 55        |
| 5.1.4      | Les particularités de ZIC1.....                                                | 56        |
| <b>5.2</b> | <b>Brn2.....</b>                                                               | <b>57</b> |
| 5.2.1      | Fonction du facteur Brn2.....                                                  | 58        |
| 5.2.1.1    | Brn2 dans la différenciation neuronale in vivo.....                            | 58        |
| 5.2.1.2    | Brn2 dans la différenciation neuronale in vitro.....                           | 59        |
| 5.2.2      | Structure de Brn2.....                                                         | 59        |
| 5.2.3      | Reconnaissance de l'ADN.....                                                   | 60        |
| 5.2.4      | Mécanismes de régulation de la liaison et de l'activité des protéines POU..... | 60        |
| 5.2.4.1    | Séquence du site de fixation.....                                              | 60        |
| 5.2.4.2    | Modifications post-traductionnelles.....                                       | 61        |
| 5.2.4.3    | Localisation.....                                                              | 62        |
| 5.2.5      | Partenaires protéiques des protéines à domaine POU.....                        | 63        |
| 5.2.5.1    | Facteurs généraux de la transcription.....                                     | 63        |
| 5.2.5.2    | POU-SOX.....                                                                   | 63        |
| 5.2.5.3    | Famille des facteur PAX.....                                                   | 64        |
| 5.2.5.4    | ASCL1.....                                                                     | 65        |

## **RÉSULTATS**

### **PROJET DE THESE..... 68**

### **A BRN2-ZIC1 AXIS SPECIFIES THE NEURONAL FATE OF RETINOIC ACID TREATED EMBRYONIC STEM CELLS. .... 72**

### **RESULTATS COMPLEMENTAIRES.....125**

|           |                                                                              |            |
|-----------|------------------------------------------------------------------------------|------------|
| <b>1.</b> | <b>Matériel et méthodes.....</b>                                             | <b>125</b> |
| <b>2.</b> | <b>Hes3.....</b>                                                             | <b>127</b> |
| 2.1.      | Description du facteur de transcription HES3.....                            | 127        |
| 2.2.      | Résultats.....                                                               | 128        |
| 2.2.1.    | Réduction de l'expression du gène Hes3 au niveau ARN.....                    | 128        |
| 2.2.2.    | La perte d'expression de HES3 n'abolit pas la différenciation neuronale..... | 128        |
| 2.2.3.    | L'expression de facteurs neuronaux est altérée par la perte de Hes3.....     | 128        |
| 2.3.      | Discussion et perspectives.....                                              | 129        |
| <b>3.</b> | <b>Brn2 : optimisation du protocole de différenciation.....</b>              | <b>131</b> |
| 3.1.      | Objectif de l'optimisation.....                                              | 131        |
| 3.2.      | Résultats comparatifs des deux protocoles de différenciation.....            | 131        |

|        |                                        |     |
|--------|----------------------------------------|-----|
| 3.2.1. | Réduction du niveau d'expression ..... | 131 |
| 3.2.2. | Observation du phénotype .....         | 132 |
| 3.2.3. | Impact sur certains gènes cibles ..... | 132 |
| 3.3.   | Conclusion .....                       | 132 |

## DISCUSSION

|                                                                                                          |            |
|----------------------------------------------------------------------------------------------------------|------------|
| <b>DISCUSSION.....</b>                                                                                   | <b>134</b> |
| 1. Brn2 spécifie l'identité neuronale des cellules ES. ....                                              | 134        |
| 2. Un faible nombre de site de fixation hautement conservés.....                                         | 134        |
| 3. Absence de redondance fonctionnelle entre les protéines Brn1 et Brn2 .....                            | 136        |
| 4. Quel rôle pour Brn5 ? .....                                                                           | 137        |
| 5. La phosphorylation de Brn2 est-elle importante dans le processus de différenciation neuronale ? ..... | 137        |
| 6. Zic1, une cible directe de Brn2 ? .....                                                               | 138        |
| 7. Une alternative à l'approche lentivirale : le système Crispr-Cas9 .....                               | 139        |

## ANNEXES

|                                                                                                                                          |            |
|------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>HOW THE RAR-RXR HETERODIMER RECOGNIZES THE GENOME .....</b>                                                                           | <b>143</b> |
| <b>RETINOIC ACID RECEPTORS RECOGNIZE THE MOUSE GENOME THROUGH BINDING ELEMENTS WITH DIVERSE SPACING AND TOPOLOGY .....</b>               | <b>169</b> |
| <b>PHOSPHORYLATION OF THE RETINOIC ACID RECEPTOR RAR2 IS CRUCIAL FOR THE NEURONAL DIFFERENTIATION OF MOUSE EMBRYONIC STEM CELLS.....</b> | <b>196</b> |

## REFERENCES

|                            |            |
|----------------------------|------------|
| <b>BIBLIOGRAPHIE .....</b> | <b>209</b> |
|----------------------------|------------|

## Abréviations

|             |                                                                                 |
|-------------|---------------------------------------------------------------------------------|
| 5caC.....   | 5-carboxycytosine                                                               |
| 5fC.....    | 5-formylcytosine                                                                |
| AADC.....   | L-Amino Acid Decarboxylase                                                      |
| ADH7 .....  | Alcool Dehydrogénase 7                                                          |
| ADN .....   | Acide désoxyribonucléique                                                       |
| AF.....     | Activation Function                                                             |
| Akt.....    | voir PKB (synonyme)                                                             |
| AMPc .....  | Adénosine Monophosphate Cyclique                                                |
| Apc .....   | Adenomatosis Polyposis Coli                                                     |
| AR .....    | Acide Rétinoïque                                                                |
| ARN .....   | Acide ribonucléique                                                             |
| ARNm .....  | ARN messenger                                                                   |
| ASCL1.....  | Achaete-scute homolog 1                                                         |
| Atoh1 ..... | Atonal Homolog 1                                                                |
| bFGF.....   | Basic Fibroblast Growth Factor                                                  |
| bHLH.....   | basic Helix-Loop-Helix                                                          |
| BLBP .....  | Brain Lipid-Binding Protein                                                     |
| BMP .....   | Bone Morphogenic Protein                                                        |
| Bp .....    | Base pairs                                                                      |
| BRE.....    | TFIIB Response Element                                                          |
| CBP .....   | CREB Binding Protein                                                            |
| CDK .....   | Cyclin Dependant Kinase                                                         |
| ChIP.....   | Chromatin ImmunoPrecipitation                                                   |
| CRABP.....  | Cellular Retinoic-Acid Binding Protein                                          |
| CRH.....    | Corticotropin releasing Hormone                                                 |
| CRISPR..... | Clustered Regularly Interspaced Short Palindromic Repeats                       |
| CRM1 .....  | Chromosome Region Maintenance 1                                                 |
| CTD.....    | C-Terminal Domain                                                               |
| CUX .....   | Cut-like homeobox                                                               |
| DAX1 .....  | Dosage sensitive sex-reversal, Adrenal hypoplasia congenital on the X chromosom |
| DBD .....   | DNA Binding Domain                                                              |
| Dll.....    | Delta like                                                                      |
| DNMT .....  | DNA Methyl Transferase                                                          |
| DPE .....   | Downstream Promoter Element                                                     |
| DR .....    | Direct Repeat                                                                   |
| EB.....     | Embryoid Body                                                                   |
| EC.....     | Embryonic Carcinoma                                                             |
| ERE .....   | Estrogen Response Element                                                       |
| ERK.....    | Extracellular signal-Regulated Kinase                                           |
| ES .....    | Embryonic Stem                                                                  |
| ESRRB ..... | EStrogen-Related Receptor Beta                                                  |
| FGF.....    | Fibroblast Growth Factor                                                        |
| Foxd3.....  | Forkhead box protein D3                                                         |
| Gab .....   | Grb2-Associated binding protein                                                 |
| GCN5 .....  | General Control of amino acid synthesis protein 5                               |
| GFAP.....   | Glial Fibrillary Acidic Protein                                                 |
| GLAST ..... | GLutamate ASpartate Transporter                                                 |
| GNAT.....   | Gcn5-related N-acetyltransferases                                               |
| Gp130 ..... | Glycoprotein 130                                                                |
| GR .....    | Glucocorticoid Receptor                                                         |

|                        |                                                               |
|------------------------|---------------------------------------------------------------|
| GRB2.....              | Growth factor Receptor Bound protein 2                        |
| GSK3b.....             | Glycogen Synthase Kinase 3b                                   |
| GST .....              | Glutathione S Transferase                                     |
| GTF .....              | General Transcription Factor                                  |
| HAT.....               | Histone Acétyltransferase                                     |
| HDM.....               | Histones DéMéthylases                                         |
| HLH.....               | Helix-Loop-Helix                                              |
| HMT .....              | Histones Méthyl-Transférase                                   |
| ICM .....              | Inner Cell Mass                                               |
| Id.....                | Inhibitor of differentiation                                  |
| Inr .....              | initiateur                                                    |
| iPS .....              | Induced Pluripotent Stem                                      |
| IR.....                | Inverted Repeat                                               |
| Jag.....               | Jagged                                                        |
| JAK.....               | Janus-Associated tyrosine Kinase                              |
| Klf4.....              | Krueppel-like factor 4                                        |
| KO .....               | Knock Out                                                     |
| LBD.....               | Ligand Binding Domain                                         |
| LCOR .....             | Ligand-dependant Co-Repressor                                 |
| LIF .....              | Leukemia Inhibitory Factor                                    |
| LIFR .....             | LIF Receptor                                                  |
| Maml .....             | Mastermind like                                               |
| MAPK .....             | Mitogen-Activated Protein Kinase                              |
| MEF.....               | Mouse Embryonic Fibroblast                                    |
| MEK.....               | MAPK/ERK Kinase                                               |
| MORE .....             | More Palindromic Oct factor Recognition Element               |
| MTE .....              | Motif Ten Element                                             |
| MYTL1.....             | Myelin transcription factor 1-like protein                    |
| NC .....               | Negative Cofactor                                             |
| NCID .....             | Notch Intracellular Cytoplasmic Domain                        |
| NCoR1.....             | Nuclear-Receptor Corepressor 1                                |
| SMRT .....             | Silencing Mediator for Retinoid and Thyroid hormone receptors |
| NE.....                | NeuroEpithelial                                               |
| Neurog.....            | Neurogénine                                                   |
| NORE.....              | N-Oct3 Responsive Element                                     |
| NR .....               | Nuclear Receptor                                              |
| NTD.....               | N-terminal Domain                                             |
| OCT.....               | OCTamer binding protein                                       |
| PAX.....               | Paired box protein                                            |
| PI3K .....             | PhosphoInositide 3-Kinase                                     |
| PIC.....               | Preinitiation Complex                                         |
| PIT1 .....             | Pituitary-specific positive transcription factor 1            |
| PKA.....               | Protein Kinase A                                              |
| PKB .....              | Protein kinase B                                              |
| PolII.....             | ARN Polymérase II                                             |
| PORE.....              | Palindromic Oct factor Recognition Element                    |
| POU <sub>H</sub> ..... | POU Homeo domain                                              |
| POU <sub>S</sub> ..... | POU Specific domain                                           |
| PRAME .....            | Preferentially Expressed Antigen in Melanoma                  |
| PRMT .....             | Protein arginine MethylTransferase                            |
| P-TEFb .....           | Positive Transcription Elongation Factor b                    |
| RALDH .....            | RetinAldehyde DeHydrogenase                                   |
| RAR.....               | Retinoic Acid Receptor                                        |

|             |                                                        |
|-------------|--------------------------------------------------------|
| RARE .....  | Retinoic Acid Response Element                         |
| RBP4 .....  | Retinol-Binding Protein                                |
| RDH10 ..... | Retinol Dehydrogenase 10                               |
| RG .....    | Radial Glial                                           |
| RIP140..... | 140 kDa Receptor Interacting Protein                   |
| RMN.....    | Résonance Magnétique Nucléaire                         |
| RPB .....   | RNA Polymerase B                                       |
| RXR.....    | Retinoid X Receptor                                    |
| S100B ..... | S100 calcium binding protein B                         |
| SALL4.....  | Sal-like protein 4                                     |
| SET .....   | Su(var)3-9, Enhancer of zest, Trithorax                |
| Shh.....    | Sonic HedgeHog                                         |
| SMAD .....  | Small Mothers Against Decapentaplegic                  |
| SNC .....   | Système Nerveux Central                                |
| SNP.....    | Système Nerveux Périphérique                           |
| SOX.....    | SRY-related HMG bOX                                    |
| SRB .....   | Suppressor of RNA Polymerase B                         |
| SSEA .....  | Stage-Specific Embryonic Antigen                       |
| STAT .....  | Signal Transducer and Activator of Transcription       |
| STRA6.....  | Stimulated by Retinoic Acid gene 6)                    |
| TAF .....   | TBP Associated Factor                                  |
| TALEN.....  | Transcription Activator-Like Effector Nuclease         |
| TBP.....    | TATA Binding Protein                                   |
| TBR2 .....  | T-box brain protein 2                                  |
| TBX3.....   | T-box transcription factor 3                           |
| TDG.....    | Thymine DNA Glycosylase                                |
| TET.....    | Ten-Eleven Translocation                               |
| TGF .....   | Transforming Growth Factor                             |
| TSS .....   | Transcription Start Site                               |
| UAS .....   | Upstream Activating Sequence                           |
| URS .....   | Upstream Repressing Sequence                           |
| USA .....   | Upstream Stimulatory Activity                          |
| UTF1 .....  | Undifferentiated embryonic cell transcription factor 1 |
| ZFN .....   | Zinc Finger Nuclease                                   |
| ZF-NC ..... | Zinc finger N-flanking region                          |
| ZIC .....   | Zinc Finger Of Cerebellum                              |
| ZOC.....    | Zic1-3 odd-paired conserved                            |

## Liste des figures et tables

Figure 1 : Isolation des lignées pluripotentes au cours du développement de l'embryon murin

Figure 2 : Voie de signalisation LIF/STAT

Figure 3 : Voie de signalisation MAPK-ERK

Figure 4 : Voie de signalisation PI3K

Figure 5 : Voie de signalisation BMP

Figure 6 : Relations conflictuelles entre les différents facteurs impliqués dans la pluripotence

Figure 7 : Trois principales méthodes de différenciation des cellules ES

Figure 8 : Exemples de lignées différenciées à partir de corps embryonnaires

Figure 9 : Mise en place du tube neural

Figure 10 : Modèle de différenciation neuronale dans le cortex cérébral

Figure 11 : Vue simplifiée des différents intermédiaires de la différenciation neuronale

Figure 12 : Méthodes de différenciation des cellules ES *in vitro*

Figure 13 : Différenciation neuronale *in vivo* et *in vitro*

Figure 14 : Protocole de différenciation neuronale des cellules ES

Figure 15 : Comparaison des promoteurs de type compact et de type dispersé

Figure 16 : Les séquences de régulation des gènes de classe II

Table 1 : Composition des différents GTF et leurs fonctions chez l'homme

Figure 17 : Deux modèles d'assemblage du PIC

Figure 18 : Régulation du cycle transcriptionnel

Table 2 : Exemples de modifications d'histones et de leur rôle dans la transcription

Figure 19 : Récapitulatif des différentes marques d'histones

Figure 20 : La voie de signalisation de l'acide rétinoïque

Figure 21 : Représentation schématique de la structure d'un récepteur nucléaire

Figure 22 : Représentation du LBD du récepteur RXR

Figure 23 : Organisation différentielle du dimère RAR-RXR en fonction de l'espacement des hémisites

Figure 24 : Mécanisme de régulation de la transcription par les récepteurs nucléaires de classe II

Figure 25 : Organisation chromosomale des cinq membres de la famille *Zic* chez la souris

Table 3 : Les différents membres de la famille POU et leur domaine d'expression

Figure 26 : Organisation structurale des protéines à domaine POU

Figure 27 : Organisation des domaines POU en fonction de la séquence ADN reconnue

## Contenu du CD

Un CD a été joint à ce manuscrit. Il se trouve en 3<sup>ème</sup> page de couverture. Sur ce CD figurent :

- 1) Un tableau excel « Supplemental Dataset 1 » associé au Manuscrit n°1, qui comprend l'ensemble des résultats de RNAseq pour les différentes lignées décrites dans le manuscrit, ainsi que les analyses d'ontologie associées.
- 2) Un tableau excel « Supplemental Dataset 2 » associé au Manuscrit n°1 dans lequel figurent les résultats d'analyse de ChIP-seq (Coordonnées de pics détectés par MACS et gènes associés).
- 3) Un tableau excel « Supplemental Dataset 3 » associé au Manuscrit n°1 comprenant la liste des sites de fixation pour Brn2 dans les différents jeux de données analysées, ainsi que la liste des gènes qui sont associés à ces pics.
- 4) Un tableau excel « Supplemental Table 1 » associé à la publication annexe n°2 comprenant la liste des sites fixés par RAR ainsi que les gènes associés à ces sites.
- 5) Un tableau excel « Supplemental Table 2 » associé à la publication annexe n°2 comprenant la liste des gènes dérégulés suite au traitement des corps embryonnaires par l'acide rétinoïque (après 24h de traitement)

# ***Introduction***

# ***Chapitre 1***

*Les cellules ES :*

*Origine et utilisations*

# 1 Chapitre I : Les cellules ES : origine et utilisations

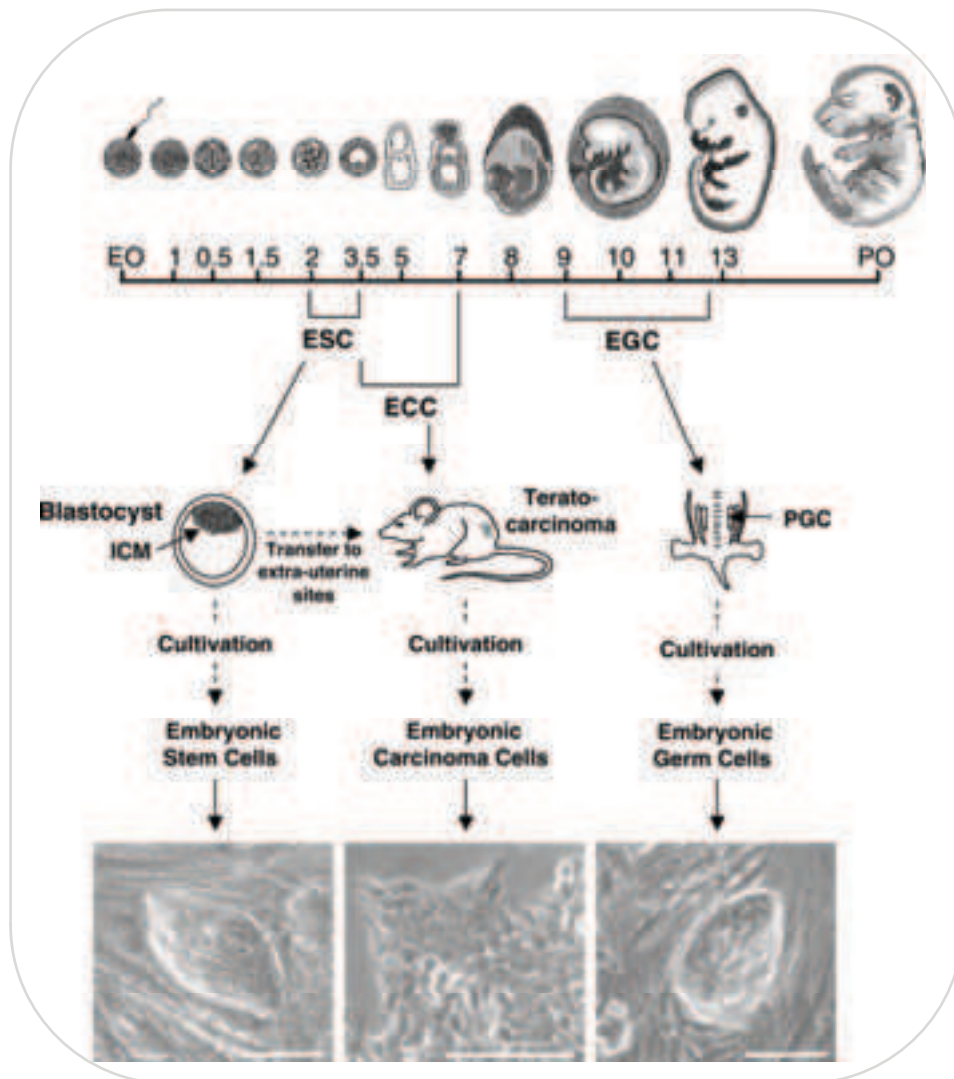
Les cellules souches embryonnaires (ES) sont des cellules pluripotentes issues de l'embryon. Ces cellules peuvent soit proliférer afin d'accroître leur nombre, soit se différencier en un tissu spécifique. Ces caractéristiques font d'elles un outil de choix pour la recherche en biologie.

Les cellules ES de plusieurs organismes modèles (hamster, lapin, rat, poulet...) ont été dérivées, mais les cellules ES murines restent la référence dans ce domaine d'étude. Ces cellules ES murines peuvent en effet être réimplantées et former un embryon viable. Concernant les cellules souches embryonnaires humaines, des problèmes éthiques ont pendant longtemps empêché leur utilisation. Une lignée de cellules souches embryonnaires humaines a bien été dérivée par une équipe américaine (Thomson, Itskovitz-Eldor et al. 1998), mais il a fallu attendre 2004 en France pour obtenir les dérogations permettant de travailler sur ce type de cellules.

Les cellules ES ont la particularité de pouvoir se différencier dans de nombreux types cellulaires de l'organisme. Ces cellules sont caractérisées par leur capacité d'autorenouvellement et par leur pluripotence. En effet, exceptés les tissus extra-embryonnaires, les cellules ES peuvent se différencier dans tous les tissus de l'organisme. Leur capacité d'autorenouvellement leur permet de proliférer et de garder leurs propriétés pendant de longues périodes et en font un outil attractif notamment pour la médecine régénérative.

## 1.1 Origine des cellules souches embryonnaires

Au cours de l'embryogenèse, la première entité est le zygote, qui correspond à l'ovule après fécondation. Le zygote pourra donner naissance à un organisme entier, capacité définie sous le terme de totipotence et qui est maintenue jusqu'au stade 8 cellules. Les divisions cellulaires suivantes vont donner naissance au blastocyste, composé du trophoblaste et d'une masse cellulaire interne (ICM : Inner Cell Mass), composée elle-même de cellules non différenciées (*Figure 1*). Les cellules composant cette ICM ne sont plus totipotentes, mais sont définies sous le terme de pluripotentes (pour revue (Wobus and Boheler 2005)).



Wobus et Boheler 2005

**Figure 1 : Isolation des lignées pluripotentes au cours du développement de l'embryon murin.** Les cellules souches embryonnaires (ES), les cellules de tératocarcinomes (EC) et les cellules de la lignée germinale sont isolées à différents stades du développement de l'embryon de souris. Les cellules ES sont dérivées de l'ICM entre le jour embryonnaire 2 et 3. Les cellules de tératocarcinomes embryonnaires sont issues de la transplantation d'embryon de stade E3,5-E7 au niveau de sites extra-utérin. Les cellules de la lignée germinale sont dérivées à partir de cellules germinales primaires isolées elles mêmes des crêtes génitales entre E9 et E12,5

Les cellules ES sont dérivées des embryons au stade préimplantatoire. Elles ont été isolées chez la souris par deux groupes en 1981 (Evans and Kaufman 1981; Martin 1981), alors que les cellules ES humaines ont été isolées en 1998 (Thomson, Itskovitz-Eldor et al. 1998). Avant cela, des études sur des cellules de carcinomes embryonnaires (EC : Embryonic Carcinoma) avaient ouvert la voie de l'utilisation des cellules souches, mais ces cellules issues de tératocarcinomes présentaient finalement de nombreux défauts et perdaient leur potentiel de différenciation *in vitro* (pour revue (Andrews 2002)).

Les recherches sur les cellules souches de tératocarcinomes ont tout de même permis l'avancée des connaissances et les techniques mises au point pour leur culture ont été utilisées pour la culture des cellules ES. Evans and Kaufmann ont ainsi cultivé les cellules ES sur une couche de fibroblastes embryonnaires murins inactivés (cellules viables mais division mitotique impossible par traitement à la mitomycine, appelées cellules nourricières ou « feeders »), technique qui avait été utilisée précédemment pour les cellules EC. Le groupe de Martin a quant à lui utilisé pour les cellules ES un milieu spécialement conditionné précédemment pour la culture des cellules EC.

Les cellules ES dérivées par ces deux groupes proviennent de l'ICM et peuvent être maintenues *in vitro*, sans perte apparente de leur capacité de pluripotence. Ceci a notamment été vérifié *in vivo* en introduisant des cellules ES directement dans les blastocystes. La souris chimère résultant de cette manipulation est la preuve que les cellules ES sont capables de produire l'ensemble des lignées cellulaires de l'organisme, y compris la lignée germinale (Bradley, Evans et al. 1984). *In vitro*, la capacité de ces cellules à se différencier dans l'ensemble des cellules somatiques de l'organisme a été rapidement démontrée (Evans and Kaufman 1981; Wobus, Holzhausen et al. 1984; Doetschman, Eistetter et al. 1985), alors qu'il a fallu un peu plus de temps pour étudier leur capacité à se différencier en cellules germinales (Hubner, Fuhrmann et al. 2003; Geijsen, Horoschak et al. 2004).

Après cette première dérivation de cellules ES, d'autres expériences ont suivi et ont montré que la réussite de la dérivation dépendait entre autre de la souche de souris. Ainsi, il est simple de dériver des cellules ES de la souche sv129, alors que les protocoles ont du être adaptés pour d'autres souches (Ledermann and Burki 1991; Schoonjans, Kreemers et al. 2003). Dans l'ensemble des cas, les cellules ES issues de ces dérivations peuvent être maintenues *in vitro* pendant de nombreux passages, tout en conservant un caryotype inchangé.

Ces cellules ont également été caractérisées pour leur temps de génération relativement court (12-15h) (Rohwedel, Sehlmeier et al. 1996).

## 1.2 Capacité de prolifération *in vitro* et maintien de la pluripotence

Les cellules ES sont dérivées sans l'ajout d'agent d'immortalisation. Elles conservent un caryotype diploïde et n'entrent pas en sénescence au cours de leur dérivation. Les cellules ES peuvent proliférer sans limites apparentes et ne sont pas sujettes aux inhibitions de contact. La capacité des cellules ES de se diviser de manière symétrique afin de générer deux cellules filles identiques est définie sous le terme d'autorenouvellement. Cette capacité d'autorenouvellement passe en quelque sorte par la suppression de la différenciation au cours de la prolifération. En fait, la croissance d'une population pluripotente de cellules ES nécessite un équilibre entre des signaux de survie, de prolifération et d'autorenouvellement. Des études faites *in vitro* et *in vivo* ont permis d'identifier différents régulateurs qui semblent jouer un rôle important dans le maintien des capacités d'autorenouvellement et de pluripotence des cellules ES. Parmi ces régulateurs se trouvent notamment des facteurs de signalisation extracellulaire ainsi que des facteurs de transcription (pour revue (Liu, Lu et al. 2007)).

### 1.2.1 Implication de différentes voies de signalisation

Les cellules souches embryonnaires murines sont caractérisées par l'expression de glycoprotéines de surface spécifiques, telles que SSEA-1 (Stage-specific embryonic antigen 1) (Solter and Knowles 1978) ainsi que par l'expression de facteurs de transcription tels que OCT4 (Scholer, Balling et al. 1989; Scholer 1991) et Nanog (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). De plus, une autre caractéristique des cellules ES est leur forte activité télomérase (Thomson, Itskovitz-Eldor et al. 1998).

C'est une combinaison de différentes voies de signalisation qui va permettre d'orchestrer le maintien des cellules ES dans un état pluripotent. Par exemple, il est connu que la protéine BMP (Bone Morphogenic Protein) favorise l'autorenouvellement et la pluripotence des cellules ES en présence de LIF (Leukemia Inhibitory Factor) (Ying, Nichols et al. 2003). De plus, l'activation de la voie canonique Wnt permet de conserver un phénotype non différencié, à la fois pour les cellules mES et hES, en favorisant le maintien de l'expression d'OCT4 et de Nanog en absence de LIF (Sato, Meijer et al. 2004). D'autres éléments, tels que les acides aminés (Tang, Hornstein et al. 2001) ou les inositols (Meyuhas 2000) peuvent également être

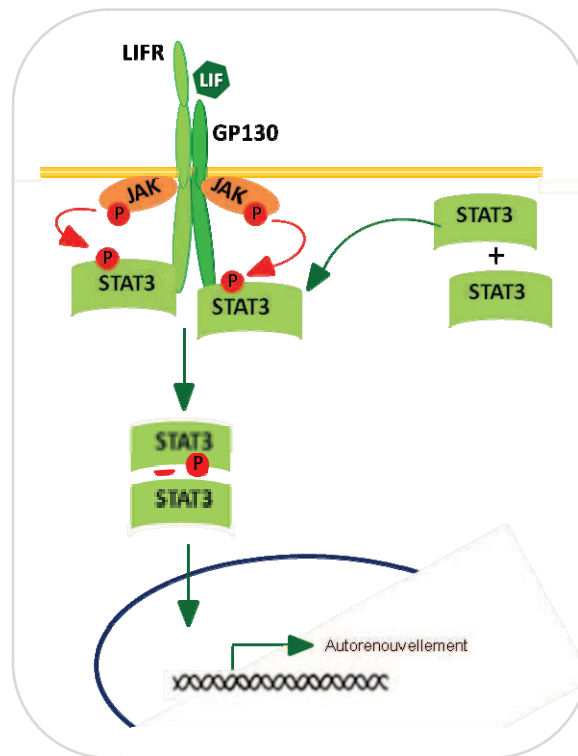
impliqués dans la survie des cellules ES. Dans cette partie, je vais détailler brièvement les différentes voies de signalisation qui ont une importance dans le choix du maintien des cellules ES dans un état pluripotent ou de leur engagement dans un processus de différenciation.

#### *1.2.1.1 La voie LIF-STAT*

La génération des cellules ES nécessitait initialement la présence d'une couche de cellules nourricières et la culture dans un milieu contenant du sérum de veau fœtal. Le fait que ces cellules nourricières soient indispensables a conduit à l'hypothèse qu'elles apportaient un ou plusieurs éléments/facteurs critiques essentiels à l'autorenouvellement, ou bloquant la différenciation.

Deux groupes indépendants ont identifié le LIF comme étant l'agent tropique nécessaire au maintien des deux caractéristiques clés des cellules ES que sont la pluripotence et la capacité d'autorenouvellement (Smith, Heath et al. 1988; Williams, Hilton et al. 1988). Le LIF est une glycoprotéine soluble de la famille des interleukines 6. Il se fixe à son récepteur LIFR, qui dimérise avec le récepteur membranaire gp130 (Davis, Aldrich et al. 1993). L'activation de gp130 provoque la phosphorylation de STAT (Signal Transducer and Activator of Transcription) via l'activation de JAK (Janus-Associated tyrosine Kinase). Cette phosphorylation de STAT entraîne sa dimérisation, puis sa translocation dans le noyau et la fixation à ses cibles au niveau de l'ADN. Il a été démontré que la capacité du LIF à maintenir la fonction d'autorenouvellement des cellules ES dépend plus précisément de l'activation d'un membre de la famille STAT, en l'occurrence STAT3 (*Figure 2*) (Niwa, Burdon et al. 1998). Plus précisément, la répression de STAT3 entraîne la différenciation de cellules ES (Niwa, Ogawa et al. 2009). L'utilisation d'une forme modifiée de STAT3, où la région codante de STAT est fusionnée au domaine de liaison du ligand du récepteur à l'œstradiol, permet de vérifier l'effet inverse. Ainsi, l'activation de l'expression de STAT3 par le traitement à l'œstradiol est suffisante au maintien de la pluripotence en l'absence de LIF dans le milieu (Matsuda, Nakamura et al. 1999).

La voie de signalisation LIF-STAT3 n'est pas la seule impliquée dans le maintien de la pluripotence des cellules ES. En effet, si l'activation de STAT3 suffit pour maintenir la capacité d'autorenouvellement des cellules ES, en absence de sérum, LIF et STAT3, à eux seuls, n'arrivent pas à maintenir ces cellules. Ceci signifie qu'il y a, dans le sérum, de



**Figure 2 : Voie de signalisation LIF/STAT.** Le LIF appartient à la famille de cytokines à interleukines de type 6 (IL6). L'action de cette cytokine se fait via un récepteur commun (gp130) couplé à un récepteur ligand spécifique (LIFR). La formation du complexe LIF/LIFR/gp130 résulte en l'activation de JAK qui est associé au récepteur, puis au recrutement d'une protéine à domaine SH2 (Src homology-2) telle que STAT3. La fixation de STAT3 au récepteur est suivie de sa phosphorylation au niveau de la tyrosine 705, ce qui va entraîner sa dimérisation avec une autre protéine STAT3 phosphorylée. Les dimères sont ensuite transloqués dans le noyau où ils pourront se fixer aux régions promotrices / enhancer cibles.

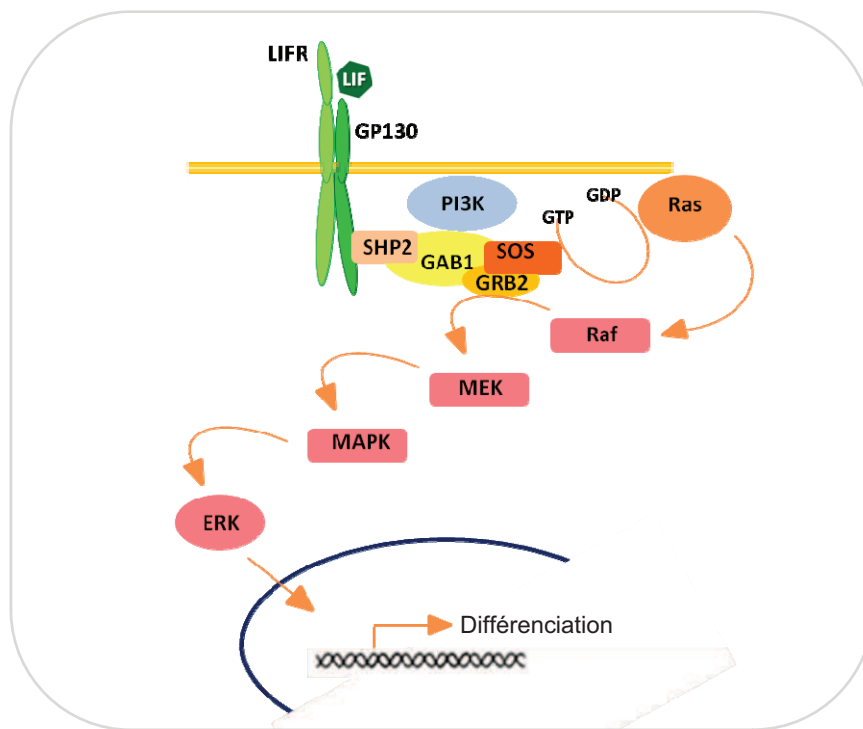
nombreux facteurs qui de manière simultanée avec LIF-STAT3, permettent le maintien des capacités de pluripotence et d'autorenouvellement. Qui plus est, en parallèle de l'activation de la voie STAT3, la fixation du LIF au récepteur LIFR/gp130 permet également l'activation des voies MAPK (Mitogen-Activated Protein Kinase) et PI3K (PhosphoInositide 3-Kinase) que je vais décrire ci-dessous (pour revue (Graf, Casanova et al. 2011))

#### *1.2.1.2 La voie MAPK-ERK*

La voie de signalisation en aval de gp130 ne se limite pas à l'activation de STAT3, mais comprend également la stimulation de la voie MAPK. Les cellules ES présentent une forte activité de ERK (Extracellular signal-Regulated Kinase) lorsqu'elles sont stimulées pour entrer dans un processus de différenciation. De ce fait, la suppression de la voie de signalisation ERK permet aux cellules ES de conserver leurs propriétés d'autorenouvellement.

La voie MAPK-ERK permet de réguler de nombreuses réponses cellulaires dans les cellules somatiques ainsi que dans la prolifération et la différenciation. De manière simplifiée, la voie ERK est engagée suite au recrutement d'un complexe contenant l'adaptateur GRB2 (Growth factor Receptor Bound protein 2) et le facteur Sos guanine-nucleotide-exchange. Ce recrutement permet d'activer Ras et d'initier une cascade de transphosphorylations impliquant Raf, MEK (MAPK/ERK kinase) et finissant par l'activation d'ERK (Kolch 2000). La version active d'ERK permet de phosphoryler des cibles cytoplasmiques, mais peut également être transloquée dans le noyau et moduler l'activité transcriptionnelle (*Figure 3*). Le recrutement du complexe GRB-Sos au niveau du récepteur peut se faire de manière indirecte. Dans le cas du récepteur LIFR-gp130, l'intermédiaire dans le recrutement de GRB-Sos est une tyrosine phosphatase appelée SHP-2. La phosphorylation de SHP-2 va permettre de générer des sites de fixation pour GRB2 et GAB1 (Grb2-Associated Binding protein 1). La protéine GAB1 va permettre de recruter PI3K et stabiliser le complexe au niveau de la membrane, ce qui va favoriser son « couplage » avec Ras. Dans les cellules ES, le fait d'utiliser une version mutée de gp130 abolit l'initiation de la voie au niveau de Ras et entraîne un autorenouvellement des cellules. De plus, l'atténuation spécifique de la voie ERK, soit par l'utilisation d'inhibiteur de MEK, soit par l'expression forcée de phosphatases pour ERK, réduit la différenciation et favorise ainsi l'autorenouvellement (Burdon, Stracey et al. 1999).

Cette voie MAPK-ERK est directement liée à la voie LIF-STAT par des boucles de régulations. En effet, de manière intéressante, il semblerait que la voie ERK-MAPK soit



**Figure 3 : Voie de signalisation MAPK-ERK.** Une fois activée suite à l'action du LIF, gp130 peut s'associer avec la protéine SHP-2. Cette protéine représente un « effecteur » positif de la voie MAPK et entraîne le recrutement de Gab1. Le complexe formé notamment de gp130/SHP-2/Gab1 va permettre d'initier une cascade de phosphorylation dont le point culminant est l'activation des kinases ERK.

capable d'inhiber la voie JAK/STAT3 en agissant sur STAT3. Il a en effet été montré que dans les cellules ES présentant un knock-out (KO) de la phosphatase SHP2, la phosphorylation de STAT3 est plus importante après la stimulation au LIF comparé aux cellules sauvages (Chan, Johnson et al. 2003). Ces données confirment l'interaction entre ces deux voies de signalisation et indiquent que c'est en réalité un équilibre entre les voies STAT3 et ERK qui va déterminer l'efficacité de l'autorenouvellement des cellules ES (Burdon, Chambers et al. 1999).

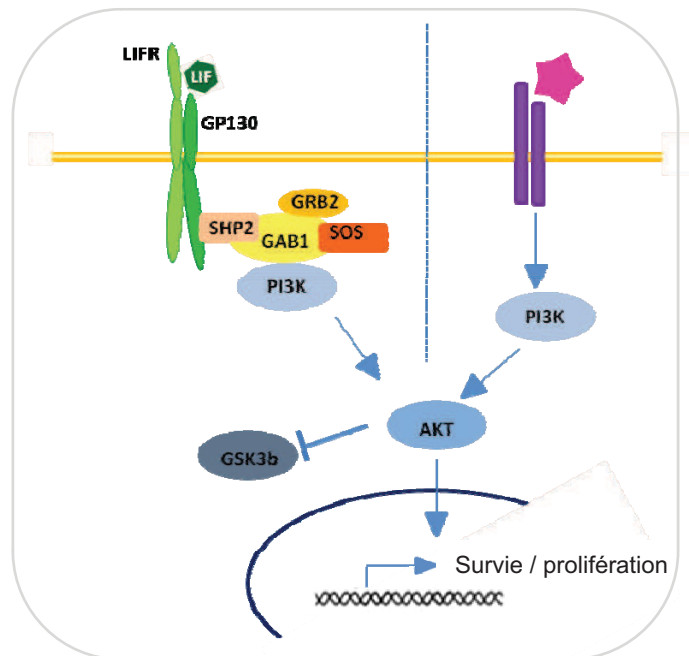
#### *1.2.1.3 La voie PhosphoInositide 3-kinase*

La voie PI3K est importante pour la prolifération, la survie et le maintien des cellules ES. Cette voie peut être activée par des facteurs qui participent au maintien de la pluripotence des cellules ES, tel que le LIF ou bFGF (basic Fibroblast Growth Factor) (*Figure 4*). Il a été démontré que l'inhibition de PI3K et de Akt entraîne une différenciation des cellules ES murines, suggérant que la voie PI3K/Akt est nécessaire pour le maintien de la pluripotence. Akt permet de contrôler divers processus cellulaires et ce via différentes protéines cibles. Plusieurs expériences ont permis de montrer que dans la voie PI3K/Akt impliquée dans la pluripotence des ES, ni l'action d'Akt sur le cycle cellulaire, ni l'activation de mTor ne sont impliqués. En revanche, il semblerait que la voie PI3K/Akt agisse via l'activation de GSK3b (Glycogen Synthase Kinase 3b). Une interaction avec la voie ERK décrite ci-dessus est également envisagée, puisque le fait de bloquer la voie PI3K entraîne une augmentation de la signalisation MAPK-ERK.

#### *1.2.1.4 La voie TGFb*

La capacité de pluripotence des ES est également sous le contrôle de la superfamille des TGF $\beta$  (Transforming Growth Factor  $\beta$ ). Cette famille comprend plus de 40 membres, parmi lesquels TGF $\beta$ , l'Activine, BMP et Nodal sont des ligands associés aux cellules ES. Dans les cellules ES murines, BMP4 serait important pour le maintien d'un état non différencié, alors que dans les cellules ES humaines, TGFb/Activin/Nodal seraient impliqués (pour revue (Valdimarsdottir and Mummery 2005)).

Dans les cellules ES murines, en présence de sérum, LIF/STAT3 sont suffisants pour maintenir l'autorenouvellement des cellules. Ceci n'est pas vrai en l'absence de sérum, laissant supposer que le sérum apporte un facteur supplémentaire nécessaire à cet autorenouvellement. Ce facteur pourrait être BMP4, puisque l'addition de BMP4 au milieu permet au LIF de maintenir les cellules ES en culture en l'absence de sérum (Ying, Nichols et



**Figure 4 : Voie de signalisation PI3K.** L'association du LIF au dimère LIFR/gp130 induit l'activation des membres de la famille PI3K. L'un des effecteurs de cette voie n'est autre que PKB, également connue sous le nom d'Akt. L'une des cibles de cette voie n'est autre que GSK3b (également ciblé par la voie Wnt). Ainsi, la phosphorylation de GSK3b va le rendre inactif (voir le descriptif de la voie Wnt). La voie de signalisation PI3K peut également être stimulée par des facteurs de croissance tel que bFGF.

al. 2003). Plus précisément, BMP4 est capable de phosphoryler SMAD1/5 (*Figure 5*). Cette activation de SMAD a pour conséquence l'expression de la protéine Id (Inhibitor of differentiation), qui va contrecarrer la fonction des facteurs neurogéniques de type bHLH (basic Helix-Loop-Helix) et par conséquent bloquer la différenciation neuronale (Ying, Nichols et al. 2003; Gerrard, Rodgers et al. 2005). Il a également été supposé que BMP4 était capable de maintenir les cellules ES en bloquant la voie de signalisation MAPK (Qi, Li et al. 2004). Contrairement à son action dans les cellules ES murines (décrit ci-dessus), BMP4 n'est pas capable de maintenir les propriétés d'autorenouvellement des cellules ES humaines et induirait même leur différenciation en endoderme primitif.

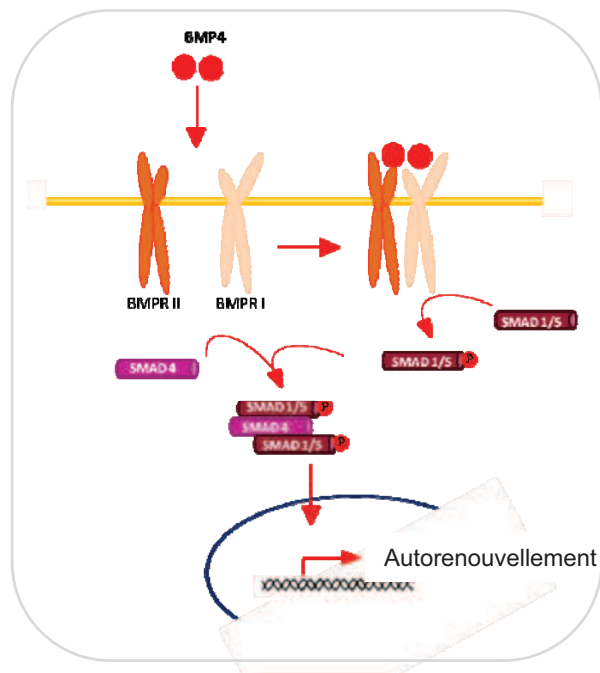
#### *1.2.1.5 La voie Wnt*

Outre la voie LIF-STAT3 et la voie TGF $\beta$  (BMP4), la voie Wnt est également impliquée dans le maintien de la pluripotence des cellules ES murines. Via son récepteur Frizzled, Wnt va permettre au complexe cytoplasmique formé de GSK3 $\beta$  et d'Apc (Adenomatosis Polyposis Coli) de ne plus phosphoryler la bêta-caténine. Cette bêta-caténine non phosphorylée est transloquée dans le noyau, où elle formera un complexe avec les facteurs TCF et activera l'expression de gènes cibles (pour revue (Willert and Jones 2006)). Suite à l'activation de la voie Wnt, le facteur de transcription TCF3 active des gènes clés dans le maintien de la pluripotence, collaborant notamment avec les facteurs centraux de ce maintien que sont OCT4, SOX2 et NANOG (Masui 2010).

#### *1.2.2 Cycle cellulaire*

Les voies de signalisation citées ci-dessus, notamment les voies MAPK-ERK et PI3K, peuvent également agir directement sur le cycle cellulaire.

Comparées aux cellules somatiques, les cellules ES présentent des caractéristiques spécifiques dans leur cycle cellulaire. L'une des plus importantes est la durée de ce cycle cellulaire, qui n'est que de 10h pour les cellules mES (Stead, White et al. 2002) et de 15-16h pour les cellules hES (Becker, Ghule et al. 2006). A titre comparatif, les cellules somatiques telles que les fibroblastes se divisent en environ 24h. Ce raccourcissement du cycle cellulaire semble être dû à une phase G1 raccourcie comparée aux cellules somatiques. Il a été montré en utilisant pour modèle des cellules de carcinome embryonnaire que ces cellules pluripotentes en phase G1 sont plus vulnérables à la différenciation par traitement à l'acide rétinoïque (AR). En effet, des cellules synchronisées traitées à l'AR en phase G1 vont se



**Figure 5 : Voie de signalisation BMP.** Le récepteur de type II contient un domaine kinase constitutivement actif qui suite à la fixation de BMP4 va phosphoryler le récepteur de type I. Suit ensuite l'activation de SMAD1/5 par le biais de sa phosphorylation. SMAD1/5 une fois phosphorylée va pouvoir s'associer à SMAD4. Ce complexe hétérotrimérique pourra être transloqué dans le noyau et stimuler l'expression de ses gènes cibles.

différencier beaucoup plus rapidement que si ce sont des cellules en phase S qui sont traitées (Mummery, van den Brink et al. 1987). Le raccourcissement de cette étape dans le cas des cellules ES pourrait implicitement être important pour le maintien de l'autorenouvellement des cellules.

Les voies de signalisation MAPK-ERK et PI3K peuvent réguler l'expression de cyclines importantes pour le maintien de la phase G1 ou la transition G1/S. Par exemple, les cellules ES expriment un très faible niveau de cyclines de type D. L'expression de ces cyclines augmente en revanche lors de la différenciation, que ce soit *in vitro* ou *in vivo* et il a été montré que la voie PI3K régulait l'expression de la cycline D1.

L'autre particularité des cellules ES ayant un lien avec le cycle cellulaire est leur activité télomérase. Les chromosomes eucaryotes présentent à leur extrémité des régions appelées télomères, qui jouent un rôle de protection des chromosomes lors de la phase de réplication. La synthèse et le maintien de ces télomères sont sous le contrôle de l'enzyme appelé télomérase. Des études dans les ES ont montré que le niveau d'expression de la télomérase est élevé dans ces cellules (Thomson, Itskovitz-Eldor et al. 1998). L'activité télomérase va ensuite diminuer au cours de la différenciation (Miura, Mattson et al. 2004). Contrairement à ce qu'il se passe dans les cellules somatiques où le nombre de divisions est limité, les cellules souches ou cancéreuses ont une forte activité télomérase.

### *1.2.3 Des facteurs de transcription impliqués dans le maintien de la pluripotence*

#### *1.2.3.1 Rôle des facteurs de transcription OCT4, NANOG et SOX2*

OCT4 (POU5F1), SOX2 et NANOG sont des facteurs de transcription clés dans la formation ou le maintien de la masse cellulaire interne au stade préimplantatoire de l'embryon, ainsi que pour la pluripotence et la capacité d'autorenouvellement des cellules ES.

Le niveau d'OCT4 est critique pour le maintien des cellules ES dans un état indifférencié. Alors qu'une baisse d'expression d'OCT4 entraîne la différenciation en trophoctoderme, qu'un niveau normal d'expression permet de maintenir la pluripotence, sa surexpression dans les cellules ES entraîne leur différenciation en endoderme et mésoderme primitif (Niwa, Miyazaki et al. 2000). Néanmoins, OCT4 à lui seul n'est pas capable de maintenir les cellules ES dans un état indifférencié. NANOG, un autre facteur de transcription, a également été identifié comme étant impliqué dans le maintien de la pluripotence des cellules ES (Chambers, Colby et al. 2003; Mitsui, Tokuzawa et al. 2003). Les facteurs OCT4 et NANOG

sont tous deux exprimés dans les cellules de l'ICM et de l'épiblaste. Leur expression disparaît au cours du développement précoce de l'embryon. Tout comme pour OCT4, le niveau d'expression de NANOG définit le devenir de la cellule. Ainsi, *in vitro*, la surexpression du facteur NANOG permet le maintien des cellules ES dans un état non différencié et ce en absence de LIF (Chambers, Colby et al. 2003). En revanche, une réduction de l'expression de NANOG induit la différenciation des cellules ES (Hatano, Tada et al. 2005).

A ces deux facteurs peut être associé le facteur SOX2. Comme pour les deux facteurs précédents, la perte d'expression de SOX2 dans les cellules ES entraîne leur différenciation prématurée, dans le cas de SOX2 en trophoctoderme. De plus, SOX2 est connu pour son action coopérative avec OCT4 au niveau de nombreux promoteurs. Ce binôme intervient entre autres pour la régulation de l'expression du gène *Nanog* (Kuroda, Tada et al. 2005). Des analyses plus larges ont été réalisées et démontrent que NANOG, OCT4 et SOX2 corégulent de nombreux gènes (Ivanova, Dobrin et al. 2006) et occupent conjointement de nombreux loci (Loh, Wu et al. 2006)

Le trio OCT4, SOX2 et NANOG représente le réseau central de la pluripotence (« core pluripotency network ») qui joue un rôle essentiel dans le maintien de la pluripotence, mais le réseau transcriptionnel qui régule cet état est en réalité bien plus complexe (Loh and Lim 2011).

#### 1.2.3.2 *Surexpression des facteurs de transcription et différenciation*

L'importance du trio de facteur OCT4, SOX2 et NANOG dans le maintien de la pluripotence mais également dans la différenciation est démontrée par les expériences de surexpression de ces facteurs. En effet, la surexpression d'OCT4 provoque une différenciation mésodermale (Niwa, Miyazaki et al. 2000), celle de SOX2 a pour conséquence une spécification neuroectodermale (Kopp, Ormsbee et al. 2008) alors que la surexpression de NANOG dans les cellules ES humaines les dirige vers une différenciation en mésendoderme (Yu, Pan et al. 2011).

Si la surexpression de ces trois facteurs entraîne une différenciation spécifique en différents tissus, des résultats similaires sont obtenus pour d'autres facteurs de transcription. Ainsi, la surexpression de ESRRB, SALL4 ou TBX3 dirige les cellules vers une différenciation endodermale (Ivanova, Dobrin et al. 2006; Zhang, Tam et al. 2006; Lu, Yang

et al. 2011), alors qu'une surexpression de DAX1 entraîne une différenciation en trophectoderme (Sun, Nakatake et al. 2009).

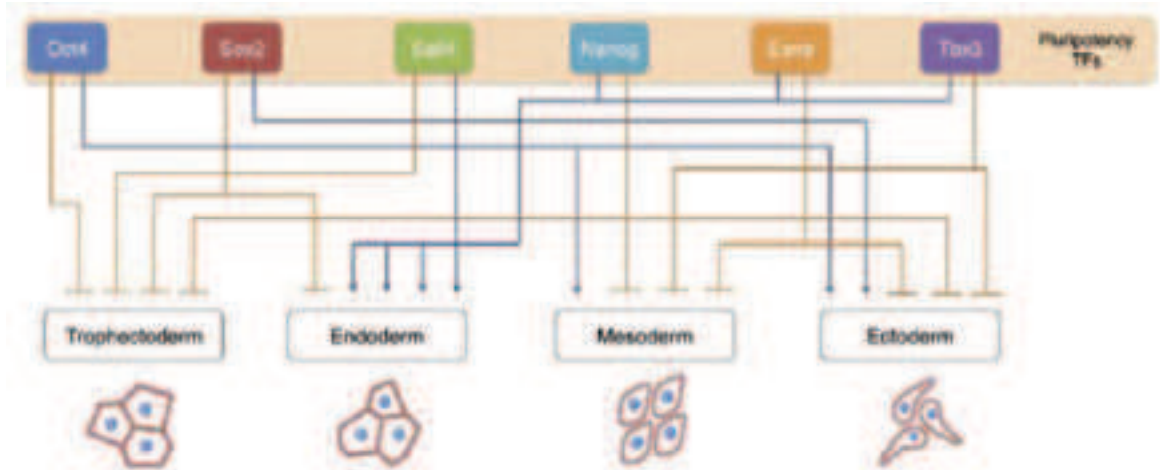
Ces études montrent qu'un important nombre de facteurs de transcription sont ainsi impliqués dans le maintien de la pluripotence et la différenciation. De plus, des boucles de régulations entre ces différents facteurs existent. En guise d'exemple, il a récemment été montré l'existence d'une régulation spécifique entre DAX1, ESRRB et OCT4 (Uranishi, Akagi et al. 2013).

L'ensemble des facteurs de transcription impliqués dans le maintien de la pluripotence et le fait que leur surexpression engendre dans de nombreux cas la différenciation, font que (Loh and Lim 2011) émettent l'hypothèse que l'état pluripotent des cellules ES ne serait pas un état basal (« ground state »). Il faudrait plutôt parler d'un état précaire/instable au cours duquel de nombreux facteurs de transcriptions entrent en compétition pour spécifier la lignée dans laquelle les cellules vont se différencier (*Figure 6*).

### 1.3 Potentiel de différenciation *in vitro*

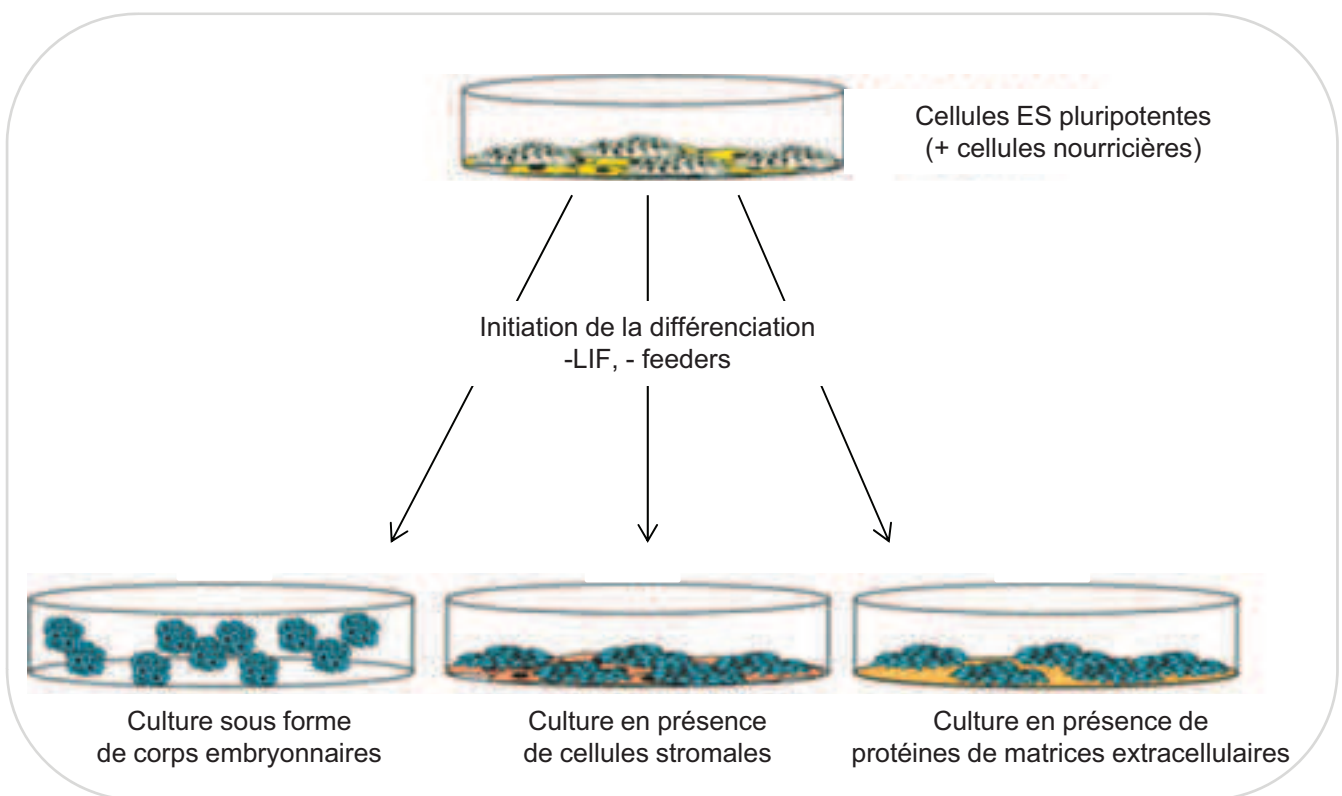
En l'absence de cellules nourricières et de facteurs tels que le LIF, les cellules ES murines vont spontanément se différencier et, en fonction des conditions, générer des progéniteurs des trois feuillets embryonnaires que sont le mésoderme, l'endoderme et l'ectoderme (pour revue (Smith 2001)).

Différentes approches ont été utilisées pour initier la différenciation des cellules ES (*Figure 7*) (pour revue (Keller 2005)). Une première méthode consiste à cultiver les cellules ES dans des gouttes suspendues, conditions dans lesquelles ces cellules vont s'agréger et former des colonies tridimensionnelles appelées corps embryonnaires (EB : Embryoid Bodies) (Doetschman, Eistetter et al. 1985). Lors de la formation de ces EB, des cellules de type endodermes vont former la couche externe de ces agrégats cellulaires. Après quelques temps, une cavité interne, tapissée de cellules ectodermes apparaît. Une spécification ultérieure de cellules mésodermes va finalement permettre d'obtenir des EB contenant les cellules des trois feuillets embryonnaires. La structure tridimensionnelle de ces EB permet d'avoir des interactions entre les cellules souvent importantes dans les programmes de développement.



Loh et Lim 2011

**Figure 6 : Relations conflictuelles entre les différents facteurs impliqués dans la pluripotence.** Le maintien de la pluripotence des cellules ES est le résultat d'une compétition entre les différents facteurs de pluripotence. L'ensemble des facteurs de pluripotence sont impliqués dans la répression de certaines voies de différenciation, tout en favorisant d'autres voies. L'expression de l'ensemble de ces facteurs dans les cellules pluripotentes crée en quelque sorte un équilibre ce qui permet de garder ces cellules dans un état non différencié.



D'après Keller 2005

**Figure 7 : Trois principales méthodes de différenciation des cellules ES.** Différentes méthodes ont été utilisées afin de différencier les cellules ES. Dans l'ensemble des méthodes, les cellules ES sont dans un premier temps déplétées des cellules nourricières sur lesquelles elles étaient cultivées. Différentes propositions de culture sont ensuite possibles ; selon les protocoles utilisés et les cellules finales attendues, la différenciation peut se faire sous forme de corps embryonnaires, sous forme de monocouche en présence de cellules stromales, ou en présence de protéines de matrices extracellulaires.

Une autre technique consiste en la différenciation des ES sous forme de monocouche, la croissance se faisant sur une matrice de protéines extracellulaires (Nishikawa, Nishikawa et al. 1998) ou en présence d'un milieu de différenciation spécifique (Takahashi, Lord et al. 2003; Fico, Manganelli et al. 2008). La culture de cellules ES peut également se faire en présence de cellules stromales (Nakano, Kodama et al. 1994).

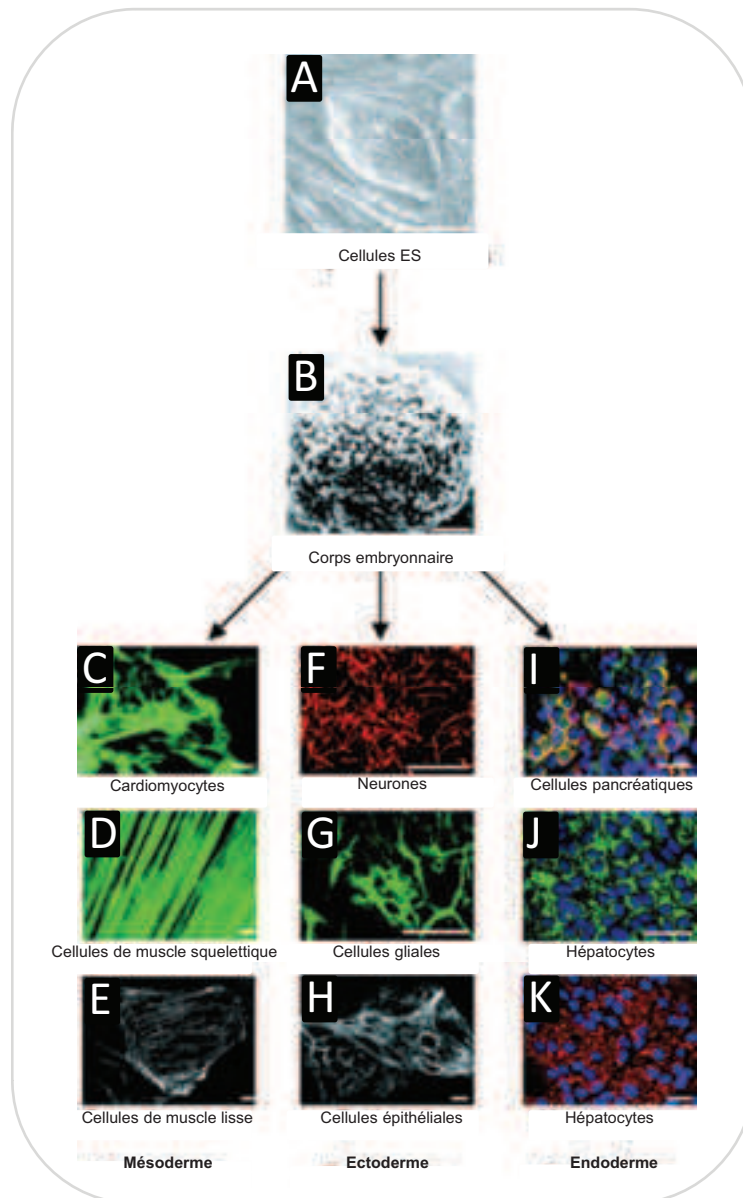
L'étude de la différenciation spécifique en une lignée cellulaire donnée passe par deux principes fondamentaux. Dans un premier temps, cette différenciation doit pouvoir se faire de manière efficace et reproductible. Dans un deuxième temps, les cellules matures obtenues suite à la différenciation doivent présenter les fonctions appropriées. Les protocoles de différenciation décrit ci-dessus ont permis de générer un large spectre de lignées spécialisées à partir des cellules ES (*Figure 8*).

### 1.3.1 Endoderme

La génération de tissus dérivant de l'endoderme, plus précisément de cellules B pancréatiques et de cellules hépatiques, a été très largement étudiée du fait de leur potentiel en thérapie génique dans le traitement respectif du diabète de type I (Soria 2001) et de la cirrhose du foie (Alison, Islam et al. 2009). Deux problèmes majeurs sont néanmoins rencontrés pour la différenciation en cellules endodermes ce qui ralentit les progrès dans ce domaine. Dans un premier temps, les marqueurs utilisés pour l'endoderme définitif ainsi que pour les précurseurs hépatiques et pancréatiques sont des marqueurs également présents dans l'endoderme viscéral, qui est extra-embryonnaire. Le deuxième problème rencontré est l'absence d'inducteur spécifique pour faire entrer les cellules dans la voie de différenciation endodermale. Malgré ces difficultés, des protocoles ont finalement été mis en place pour la différenciation en cellules pancréatiques (Stoffel, Vallier et al. 2004), en hépatocytes (Hamazaki, Iiboshi et al. 2001) ou encore en cellules intestinales (Yamada, Yoshikawa et al. 2002).

### 1.3.2 Mésoderme

Le feuillet embryonnaire correspondant au mésoderme donnera plus tard naissance aux muscles, aux os, au cartilage ou encore au système vasculaire. De nombreux protocoles ont émergés pour la différenciation en cellules mésodermes *in vitro* à partir des cellules ES. Ainsi, les cellules ES ont été différenciées *in vitro* avec succès en cellules du muscle lisse (en présence d'une matrice de collagène) (Xiao, Zeng et al. 2007), muscle strié (Rohwedel,



*D'après Wobus et Boheler 2005*

**Fig. 8 : Exemples de lignées différenciées à partir de corps embryonnaires.** Les cellules souches embryonnaires cultivées en présence LIF sur une couche de cellules nourricières forment une colonie de cellules pluripotentes (A). L'une des techniques permettant de différencier ces cellules consiste à les cultiver sur forme de structure tridimensionnelle appelée corps embryonnaire. La différenciation via ces corps embryonnaires permet d'obtenir des lignées mésodermes (C, D, E), ectodermes (F, G, H) et endodermes (I, J, K). Echelle = 0.5  $\mu\text{m}$  (H), 20  $\mu\text{m}$  (I), 25  $\mu\text{m}$  (C, D, E), 30  $\mu\text{m}$  (J, K), 50  $\mu\text{m}$  (B, G), and 100  $\mu\text{m}$  (A, F).

Maltsev et al. 1994), ostéoblastes (Buttery, Bourne et al. 2001), cardiomyocytes (Boheler, Czyz et al. 2002). Cette liste est non exhaustive et malgré des protocoles très variés pour ces différents types de différenciation, tous nécessitent le traitement avec des facteurs de croissance spécifiques et/ou la présence d'une matrice extracellulaire tel que le collagène lors de la différenciation en muscle lisse, suivi de différentes étapes de différenciation.

### 1.3.3 Ectoderme

Parmi les différentes lignées qui peuvent être produites par l'ectoderme au cours du développement embryonnaire, la lignée neuroectodermale va être à l'origine de la formation du système nerveux central (SNC) et périphérique (SNP). La lignée épithéliale permettra de former les tissus de l'épiderme.

La différenciation épithéliale des cellules ES peut être mise en évidence par la présence de filaments intermédiaires de cytokératine. Différentes techniques de différenciation ont permis d'obtenir des tissus avec une morphologie similaire à la peau embryonnaire normale. Les cellules obtenues expriment des marqueurs tardifs de la différenciation en épiderme et des marqueurs de fibroblastes similaires à ceux trouvés dans la peau native (Aberdam, Gambaro et al. 2007). Ces données de différenciation suggèrent que les cellules ES *in vitro* ont la capacité de reproduire / reformer de la peau parfaitement différenciée (Coraux, Hilmi et al. 2003).

La différenciation en neurones et en cellules gliales présente un grand intérêt dans le cadre des thérapies de traitement des maladies neurodégénératives. La différenciation neuronale des cellules ES a été publiée de manière indépendante par trois groupes dès 1995 (Bain, Kitchens et al. 1995; Fraichard, Chassande et al. 1995; Strubing, Ahnert-Hilger et al. 1995). Même si la différenciation spontanée en neurones est limitée, différentes stratégies ont été développées pour améliorer cette différenciation, telles que l'utilisation d'acide rétinoïque (Rohwedel, Guan et al. 1999). Au final, les trois principaux types cellulaires du système nerveux central que sont les neurones, les astrocytes et les oligodendrocytes peuvent être générés *in vitro*, en population relativement pure à partir de cellules ES. La différenciation de ces cellules a été hautement documentée, que ce soit pour les cellules gliales (Fraichard, Chassande et al. 1995; Tang, Shang et al. 2002; Chen, Daugherty et al. 2012) ou pour les différents types de neurones : dopaminergiques, sérotonergiques, GABAergiques, glutamatergiques et

cholinergiques (Bain, Kitchens et al. 1995; Lee, Lumelsky et al. 2000; Bibel, Richter et al. 2007; Chuang, Tung et al. 2011; Chung, Moon et al. 2011).

#### **1.3.4 Cellules de la lignée germinale**

Les cellules germinales représentent la lignée de cellules permettant la transmission génétique d'une génération à la suivante. Ces cellules sont catégoriquement différentes des cellules somatiques et présentent des caractéristiques uniques notamment une haploïdie chromosomale suite à une méiose. Ces dernières années, les biologistes du développement portent un grand intérêt à la recherche sur la dérivation des gamètes à partir des cellules ES. De nombreuses études montrent que les cellules ES murines peuvent être dérivées en précurseurs de la lignée germinale *in vitro*. Plus précisément, deux méthodes ont été développées : la première consiste en une différenciation spontanée des cellules en monocouche après avoir retiré du milieu les facteurs promouvant la pluripotence tels que le FGF ou le LIF et en l'absence de feeders (Hubner, Fuhrmann et al. 2003; Nayernia, Nolte et al. 2006). La deuxième méthode passe par la formation d'EB (Toyooka, Tsunekawa et al. 2003; Geijsen, Horoschak et al. 2004). Ces techniques ont permis de former des oocytes et gamètes mâles à partir de cellules ES murines.

#### **1.4 Des cellules ES dernière génération : les iPS**

Les cellules iPS (Induced Pluripotent Stem) sont un type de cellules pluripotentes dérivées à partir de cellules somatiques adultes. Elles sont en de nombreux points similaires aux cellules ES et peuvent se différencier en cellules des différents feuillets embryonnaires. En 2006, l'équipe de Takanashi et Yamanaka a montré que la conversion d'un fibroblaste en cellule de type cellule souche embryonnaire (« embryonic-like stem cell ») était possible. Cette reprogrammation nécessitait uniquement l'expression d'une combinaison de facteurs que sont Oct4, Sox2, Klf4 et c-myc (Takahashi and Yamanaka 2006).

La technologie des iPS a ouvert de nouvelles voies dans la modélisation et la compréhension des maladies humaines. Avant cela, l'obtention de cellules ES humaines présentant une certaine mutation était bloquée à cause des problèmes éthiques, puisqu'il fallait isoler des cellules ES à partir d'un embryon, et de ce fait détruire ce dernier (Revazova, Turovets et al. 2007). La reprogrammation des cellules à partir d'une simple biopsie de peau permet maintenant aux scientifiques de générer des cellules iPS de n'importe quel patient, atteint de n'importe quelle maladie. L'utilisation de ces cellules permet également de

contourner l'obstacle du rejet immunitaire qui peut avoir lieu après une transplantation thérapeutique. Cette nouvelle approche a également permis d'avoir un nouveau point de vue sur des maladies telles qu'Alzheimer (Yagi, Ito et al. 2011) ou Parkinson (Park, Arora et al. 2008).

Plus récemment, des expériences ont permis de convertir directement des fibroblastes en un autre type de cellules, notamment en neurones, sans repasser par le stade pluripotent (Vierbuchen, Ostermeier et al. 2010). Cette conversion directe est appelée transdifférenciation. Dans le cas de la reprogrammation en iPS, la compréhension des réseaux transcriptionnels liés à la pluripotence était importante. Dans le cas de la transdifférenciation, c'est la compréhension des réseaux transcriptionnels spécifiques à chaque voie de différenciation qui est un élément clé.

# ***Chapitre 2***

*Les cellules ES :*

*Un système expérimental  
pour la compréhension des  
réseaux transcriptionnels  
in vivo*

## 2 Chapitre II : Les cellules ES : un système expérimental pour la compréhension des réseaux transcriptionnels *in vivo*

L'embryogenèse *in vivo* comprend différentes étapes, toutes étant régulées par des signaux moléculaires bien spécifiques. Au stade zygote succède le stade 2 cellules, 4 cellules, 8 cellules, puis le stade blastocyste, stade auquel les cellules ES sont dérivées. Au stade blastocyste succède la gastrulation, étape pendant laquelle les trois feuillets de l'embryon vont se mettre en place. L'étape qui nous intéresse est celle postérieure à la gastrulation. A ce stade, l'ectoderme peut donner naissance soit à l'épiderme, soit au système nerveux. Dans la partie suivante je décrirai plus particulièrement le développement du système nerveux.

### 2.1 Différenciation neuronale *in vivo*

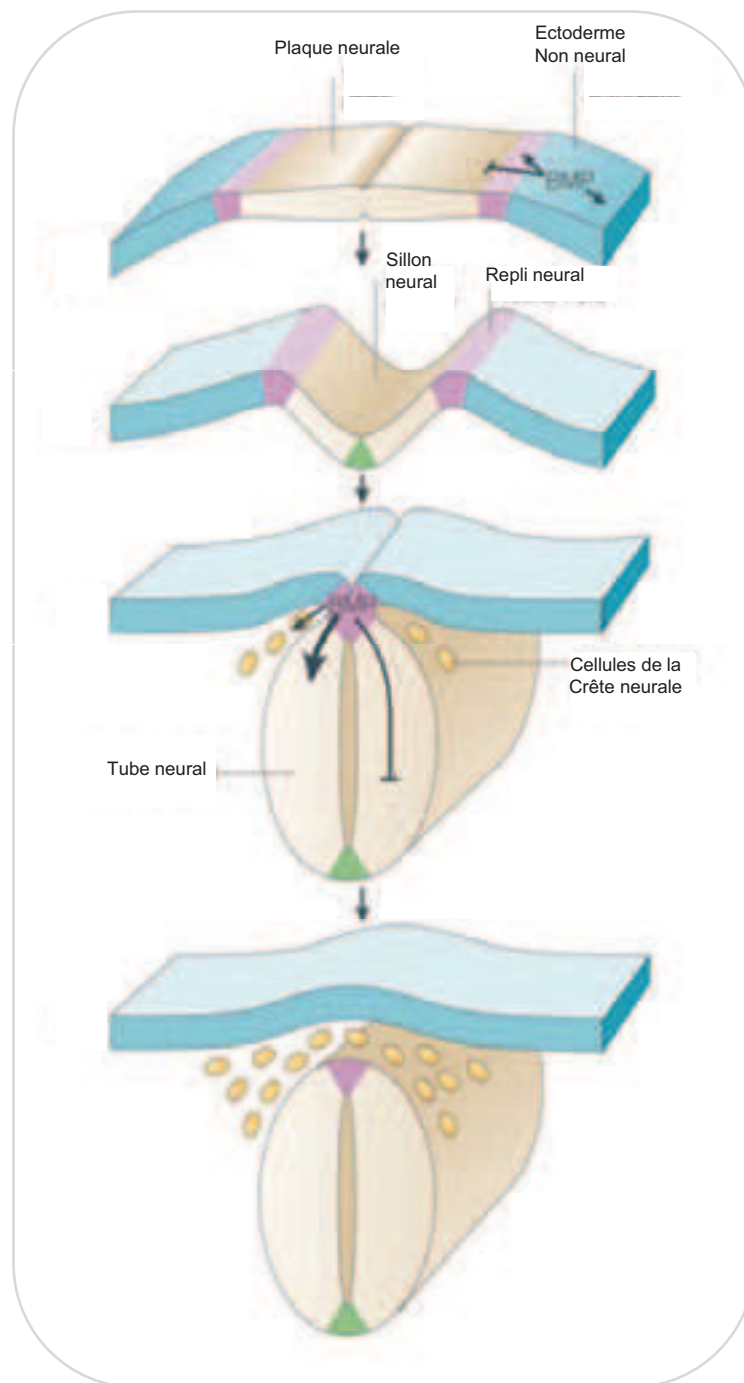
#### 2.1.1 Neurulation

La partie de l'ectoderme destinée à former le cerveau est identifiable dès les premiers stades de la spécification. L'épithélium va dans un premier temps s'épaissir et former la plaque neurale. Cette plaque neurale va ensuite se transformer en tube neural. Cette étape porte le nom de neurulation et à ce stade l'embryon est appelé neurula. Les parties antérieures du tube neural vont permettre la formation du cerveau, alors que les régions postérieures vont former la moelle épinière. Plus précisément, le tube neural constitue le futur système nerveux central (SNC) alors que parallèlement, dans la région dorsale du tube neural, les cellules de la crête neurale vont former le futur système nerveux périphérique (SNP) (pour revue (Liu and Niswander 2005)) (*Figure 9*). Les facteurs BMP jouent un rôle important dans cette étape de neurulation. En réalité, BMP4 contrôle la différenciation neuronale précoce selon un modèle par défaut. Ceci signifie que BMP4 induit la différenciation épidermique de l'épiblaste et inhibe la spécification neuronale. En revanche, tout événement inhibant l'action de BMP4 induit la différenciation neuroectodermique.

#### 2.1.2 La différenciation du tube neural

##### 2.1.2.1 Le tube neural, à l'origine du système nerveux central

Au cours du développement embryonnaire, le tube neural sera segmenté dans le sens antéropostérieur (Lumsden and Krumlauf 1996). Dans un premier temps, 3 vésicules principales peuvent être observées : le prosencéphale (forebrain), le mésencéphale (midbrain)



*D'après Liu et Niswander 2005*

**Figure 9 : Mise en place du tube neural.** Le système nerveux central est dérivé du tube neural, qui lui-même provient initialement de l'ectoderme. Après 7,5 jours de gestation, l'épithélium va s'épaissir pour former ce qui est appelé la plaque neurale. Cette plaque neurale va ensuite se replier pour former une gouttière, qualifiée de sillon neural. Après 8,5 jour, cette gouttière va se refermer le long de la ligne médiane de l'embryon. Le tube neural va petit à petit s'isoler, et les cellules de la crête neurale migrer le long du tube. Ces cellules seront à l'origine du système nerveux périphérique, des mélanocytes et du cartilage de la tête.

et le rhombencéphale (hindbrain). A un stade ultérieur, une segmentation supplémentaire va avoir lieu : le prosencéphale sera subdivisé en télencéphale et diencephale (Puelles and Rubenstein 2003), alors que le rhombencéphale sera divisé en métencéphale et myélencéphale. Le mésencéphale reste quant à lui un segment unique (Puelles 2007).

Chez les vertébrés, le SNC est dérivé des cellules neuroépithéliales (NE) du tube neural. Ces cellules sont relativement homogènes pendant les premières phases du développement et subissent de nombreuses divisions pour accroître leur population. Pendant la mise en place de la plaque et du tube neural, des morphogènes tels que l'AR, FGF, Shh, et BMP sont sécrétés. Ces morphogènes sont responsables de la mise en place de l'axe antéropostérieur et dorsoventral du tube neural, ce qui permet la spécification des premières régions du cerveau que sont le prosencéphale, le mesencéphale, le rhombencéphale et la moelle épinière. Le passage du stade prolifération au stade différenciation est dû à la transition d'un mode de division symétrique à un mode de division asymétrique. La voie de signalisation Notch est au centre de cette étape, c'est pourquoi je vais la détailler dans la partie suivante.

#### *2.1.2.2 La voie Notch : au cœur de la différenciation neuronale*

Alors que les voies de signalisation LIF-STAT ou MAPK-ERK sont importantes dans le maintien ou non de l'état pluripotent, une nouvelle voie de signalisation apparaît lors de la différenciation neuronale : la voie Notch. Cette voie est connue pour réguler le passage entre la phase de maintien des précurseurs neuraux et la phase de différenciation (pour revue (Louvi and Artavanis-Tsakonas 2006)). Le modèle classique de fonctionnement de cette voie est appelé « inhibition latérale ». De manière simplifiée, les cellules qui seront désignées pour devenir des neurones vont exprimer les ligands de la voie Notch. Ces ligands vont se fixer aux récepteurs Notch des cellules voisines. Le récepteur Notch une fois activé va entraîner l'inhibition de la différenciation de cette cellule.

La régulation Notch se fait ainsi par des interactions cellules-cellules. Les récepteurs Notch sont au nombre de quatre chez les mammifères. Présents dans une cellule, ils sont activés par leurs ligands exprimés dans les cellules voisines. Les ligands pour ces récepteurs sont Delta-like (Dll) 1, -3, -4 et Jagged (Jag) 1, -2) Après activation du récepteur par son ligand, son domaine intracellulaire (NICD) sera clivé, et transloqué dans le noyau. NICD pourra alors former un complexe avec les protéines RBP-J ou Maml (Mastermind-like) et activer la transcription de ses gènes cibles (pour revue (Kopan and Ilagan 2009))

Dans le système nerveux de l'embryon, les cibles de Notch les mieux caractérisées sont les protéines Hes (plus particulièrement Hes1 et Hes5) (Iso, Kedes et al. 2003; Kageyama, Ohtsuka et al. 2008). Les protéines Hes sont des protéines de types bHLH, qui vont réprimer l'expression de gènes tels que *Ascl1* ou *Neurog1,-2*. Les protéines *Ascl1* et *Neurog1,-2* sont importantes pour la différenciation neuronale (Nieto, Schuurmans et al. 2001). De ce fait, cette voie de différenciation sera réprimée dans les cellules contenant un récepteur Notch actif.

### 2.1.3 Les différents composants du système nerveux central

La structure du SNC est hautement complexe et reflète ainsi l'importance et la diversité des fonctions qu'il guide. Chez les vertébrés, le SNC est dérivé des cellules NE du tube neural. Au cours des premières phases de développement du cerveau, les cellules NE sont relativement homogènes et sont sujettes à une importante prolifération afin d'accroître leur population avant la phase de spécification. Ces cellules vont *in fine* produire l'ensemble des cinq types majeurs de cellules du SNC que sont les neurones, les astrocytes, les oligodendrocytes, les cellules épendymaires et les cellules souches neurales.

La différenciation en neurones représentant mon modèle d'étude, je reviendrai plus précisément sur cette différenciation dans la partie 2.1.4 en me basant sur le modèle du cortex cérébral.

Le premier type cellulaire formé après les neurones sont les astrocytes. Ces cellules jouent divers rôles dans le SNC. Outre les fonctions de support structural, les astrocytes sont impliqués dans la formation de la barrière sang/cerveau, ou encore dans les processus de réparation suite à un choc traumatique. D'un point de vue temporel, les astrocytes sont générés après la neurogenèse. Après l'apparition des astrocytes matures, ce sont des oligodendrocytes matures qui vont se former. Ces oligodendrocytes ont pour fonction principale d'apporter un support et une isolation des axones du SNC. Cette fonction est assurée via la création d'une gaine de myéline. Ainsi, un seul oligodendrocyte va permettre de « protéger » jusqu'à 50 axones.

Les cellules épendymaires et les cellules souches neurales apparaissent quant à elles dans le cortex post-natal. Les cellules épendymaires forment en réalité une barrière avec le fluide cébrospinal. Ces cellules permettent notamment d'assurer la fonction des niches de cellules souches adultes. En effet, il existe deux régions spécifiques au sein du cerveau adulte où un

processus de différenciation neuronale persiste. Ces deux niches contiennent des cellules souches neurales de type astrocytaire qui se divisent lentement. Ces cellules peuvent donner naissance à des progéniteurs intermédiaires qui vont pouvoir générer des neurones du bulbe olfactif ou de l'hippocampe (pour revue (Guerout, Li et al. 2014)).

#### *2.1.4 Biologie cellulaire de la neurogenèse*

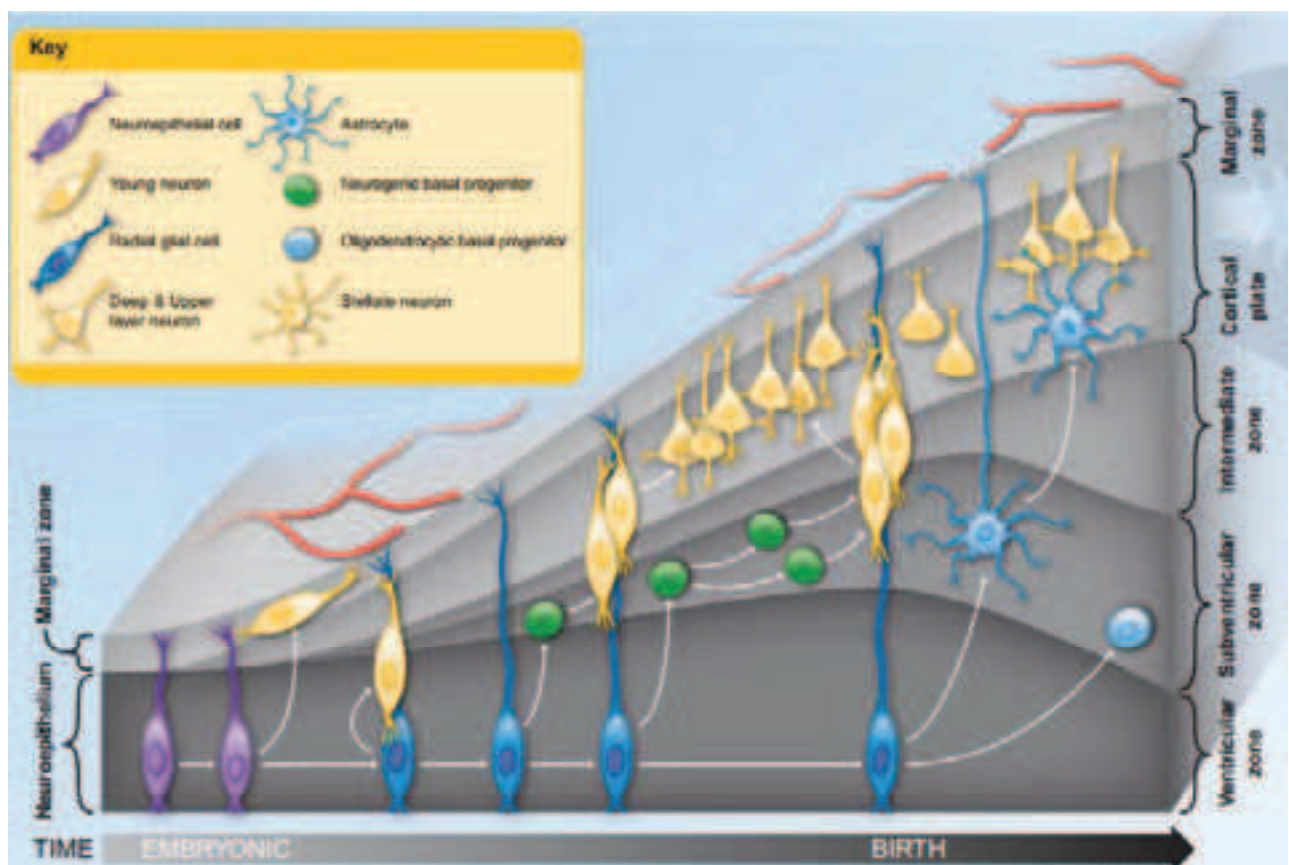
Au cours du développement, les cellules souches neurales sont à l'origine de l'ensemble des neurones du SNC ainsi que des cellules gliales, épendymaires et des cellules souches neurales adultes. L'une des propriétés des cellules souches neurales, comme pour toute cellule souche, est leur capacité d'autorenouvellement. Celui-ci peut avoir lieu soit suite à une division symétrique, qui va générer deux cellules filles identiques à la cellule mère, soit par division asymétrique, qui va donner naissance à une cellule fille similaire à la cellule d'origine et à une cellule d'un autre type (*Figure 10*). Ainsi, au cours du développement, les cellules NE, qui peuvent par leurs caractéristiques être considérées comme des cellules souches, vont dans un premier temps subir des divisions symétriques pour permettre leur prolifération. Dans un deuxième temps, elles vont se diviser asymétriquement de nombreuses fois pour assurer l'autorenouvellement et l'initiation de la différenciation. Ces divisions asymétriques vont produire soit directement un neurone, soit une cellule progénitrice qui va subir une dernière division symétrique pour générer deux neurones. A ce stade le terme de cellules post-mitotiques peut être employé. Ces cellules ont atteint leur phase ultime de différenciation.

##### *2.1.4.1 Les cellules neuroépithéliales*

Avant la neurogenèse, le tube neural est formé d'une couche unique de cellules NE, qui forment le neuroépithélium. Ce neuroépithélium apparaît comme pseudo-stratifié, du fait que les noyaux des cellules neuroépithéliales migrent le long de l'axe apico-basal au cours du cycle cellulaire. Ces cellules neuroépithéliales présentent des caractéristiques typiques de cellules épithéliales et sont fortement polarisées le long de leur axe apico-basal (Huttner and Brand 1997).

##### *2.1.4.2 Les cellules radiales gliales*

Lors de la génération des neurones, le neuroépithélium se transforme en un tissu avec de nombreuses couches cellulaires. Lors du passage de l'étape de prolifération à l'étape de différenciation, des marqueurs des astrocytes apparaissent et les cellules NE donnent



[www.abcam.com/cortical-neurogenesis](http://www.abcam.com/cortical-neurogenesis)

Figure 10 : Modèle de différenciation neuronale dans le cortex cérébral

naissance à un nouveau type de cellules : les cellules radiales gliales (RG). Ces cellules possèdent encore des propriétés résiduelles des cellules NE, mais présentent également de nouvelles caractéristiques rappelant celles des astrocytes (Huttner and Brand 1997; Campbell and Gotz 2002; Kriegstein and Gotz 2003). Les cellules RG vont au fur et à mesure des divisions totalement remplacer les cellules neuroépithéliales, et de ce fait, la grande majorité des neurones du cerveau seront dérivés de ces cellules (Malatesta, Hack et al. 2003). Parmi les caractéristiques neuroépithéliales que les cellules radiales conservent se trouvent la présence de marqueurs tels que la Nestine (Hartfuss, Galli et al. 2001), la polarité apico-basale, la migration du noyau le long de l'axe apico-basal. Contrairement aux cellules NE, les cellules RG présentent des caractéristiques astrogliales, notamment la présence de granules de glycogène et l'expression de marqueurs tels que Glas, S100b, GFAP et BLBP (Gadisseux and Evrard 1985; Kriegstein and Gotz 2003). La différenciation de la majorité de ces cellules radiales est restreinte à un seul type cellulaire : elles deviendront soit des astrocytes, soit des oligodendrocytes, mais dans la majorité des cas des neurones.

#### *2.1.4.3 Les progéniteurs basaux*

En dehors des cellules RG, un autre type de cellules apparaît au début de la neurogenèse : les progéniteurs basaux (Haubensak, Attardo et al. 2004; Miyata, Kawaguchi et al. 2004; Noctor, Martinez-Cerdeno et al. 2004). Ces derniers se distinguent des cellules NE et RG par leur localisation. Alors que les cellules NE et RG se situent au niveau du neuroépithélium qui va donner naissance à la zone ventriculaire, les progéniteurs basaux se situent dans la zone dite sub-ventriculaire. Ces progéniteurs peuvent apparaître suite à la mitose des cellules NE, mais plus souvent suite à la division symétrique ou asymétrique des cellules RG. Encore une fois, ce sont les gènes exprimés par ces progéniteurs qui les différencient des deux autres types de cellules décrits précédemment. Ainsi, TBR2, CUX1 et CUX2 sont des marqueurs spécifiques de ces cellules progénitrices (Nieto, Monuki et al. 2004; Zimmer, Tiveron et al. 2004; Englund, Fink et al. 2005). Les progéniteurs de cette zone sub-ventriculaire contribuent à la neurogenèse suite à une division symétrique qui va permettre de générer deux neurones (Haubensak, Attardo et al. 2004; Noctor, Martinez-Cerdeno et al. 2004). De ce fait, l'hypothèse a été émise que ces progéniteurs permettent, en apportant une division supplémentaire, d'augmenter le nombre de neurones produit à partir des cellules épithéliales de base. Cette idée est confirmée phylogénétiquement par la corrélation entre l'augmentation de taille de la zone sub-ventriculaire et l'augmentation du nombre de neurones dans le cortex

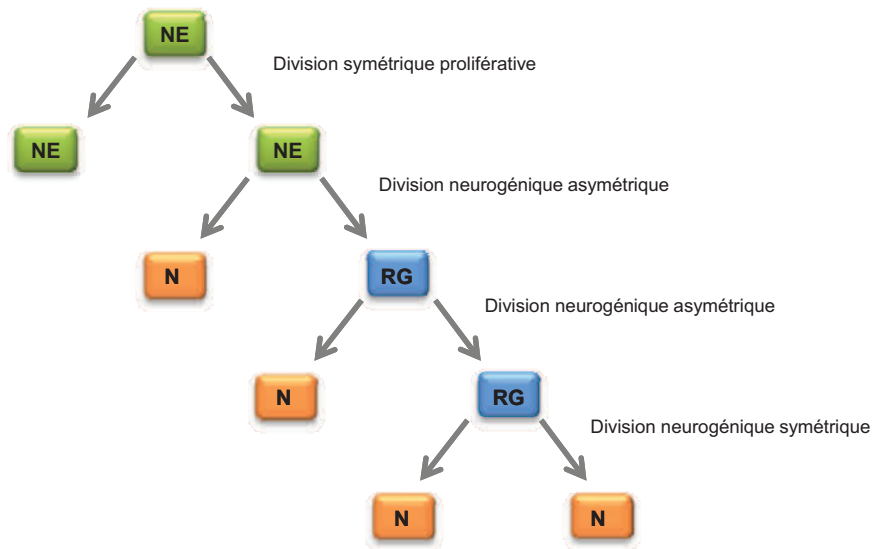
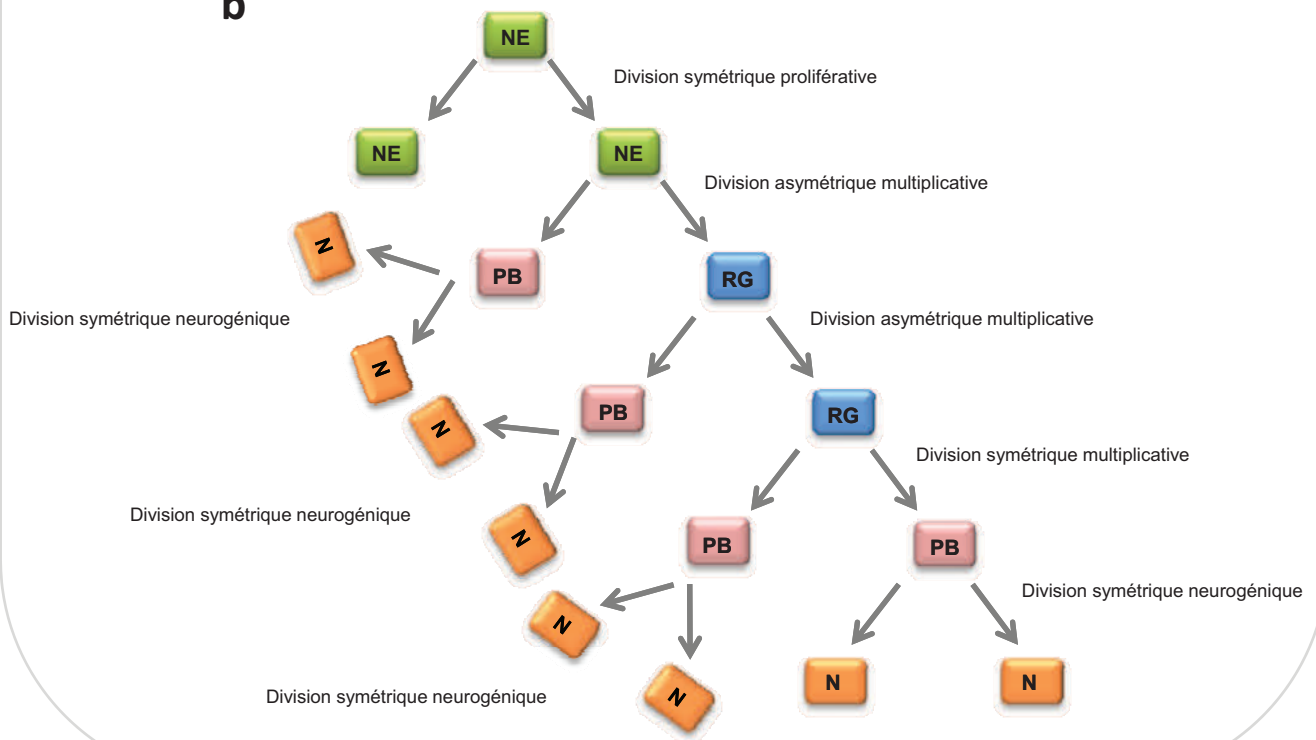
cérébral (*Figure 11*) (Smart, Dehay et al. 2002). Les neurones auxquels je m'intéresserai ultérieurement sont des neurones de type glutamatergiques, qui sont majoritairement retrouvés au niveau du cortex cérébral.

#### 2.1.4.4 *Cortex et neurones glutamatergiques*

Le néocortex des mammifères est une structure complexe, très bien organisée qui comprend six couches composées de centaines de types neuronaux différents. Ces couches diffèrent par la morphologie ou la densité des neurones qui la composent. Le cortex est la région du cerveau responsable des fonctions cognitives, de la perception sensorielle, de la conscience. Cette région a par conséquent subi une expansion prononcée au cours de l'évolution (Finlay and Darlington 1995).

Le mécanisme de développement cortical est similaire parmi les mammifères. Le cortex cérébral est développé à partir de la partie antérieure du tube neural, le prosencéphale, qui va se diviser en télencéphale et en diencephale (comme décrit ci-dessus). Lors du développement du néocortex, différentes zones du cerveau vont donner naissance à différents types de neurones. Ainsi, au stade embryonnaire 12.5, la zone ventrale du prosencéphale va donner naissance à des interneurones, qui vont faire des projections locales. Ces interneurones, qui ont pour neurotransmetteur le GABA, vont migrer au sein du cortex pour atteindre leur destination finale. La région dorsale du prosencéphale va quant à elle voir naître des neurones à projections (« projection neurons ») ou neurones excitateurs, possédant de grands axones. Ces neurones sont dits glutamatergiques et sont caractérisés par leur morphologie pyramidale. Ils vont permettre de transmettre l'information entre le néocortex et les autres régions du cerveau. Au cours du développement, les différents sous-types de neurones excitateurs sont produits au cours d'une séquence temporelle bien définie, puis migrent le long des cellules radiales pour atteindre les différentes couches du cortex. Les projections des neurones des différentes strates sont spécifiques, ainsi par exemple les neurones de la couche V vont avoir des projections en direction du mésencéphale, du rhombencéphale et de la moelle épinière.

Au sein du néocortex, certains mécanismes contrôlant la génération et la spécification des neurones ont été identifiés, mais la complexité de cette partie du cerveau fait que l'ensemble des mécanismes régissant cette zone sont loin d'être élucidés.

**a****b**

**Figure 11 : Vue simplifiée des différents intermédiaires de la différenciation neuronale.** Représentation du lien entre les cellules neuroépithéliales (NE), les cellules radiales gliales (RG), les progéniteurs basaux (PB) et les neurones (N). En l'absence de progéniteurs basaux (a), le nombre de neurones obtenus à partir d'une cellule neuroépithéliale est plus faible. L'utilisation de progéniteurs basaux comme intermédiaires (b) permet d'augmenter le nombre de neurones.

## 2.2 Modèle de différenciation neuronale des cellules ES *in vitro*

L'identification des mécanismes qui régissent le développement du cerveau fait partie des principaux défis de la neurobiologie en raison de la complexité du processus, que ce soit au niveau moléculaire, cellulaire ou anatomique. Sur la base de leurs propriétés d'autorenouvellement et de pluripotence, les cellules ES présentent un outil de choix pour essayer de fournir un modèle du développement chez les mammifères. De nombreuses études ont rapporté la différenciation dirigée des cellules ES dans des types spécifiques de neurones, mais il n'a pas été facile de relier cette différenciation aux mécanismes existant *in vivo*. Après avoir rappelé brièvement les différentes voies de signalisation qui régissent le maintien de la pluripotence et la différenciation des cellules ES, je m'intéresserai au protocole de différenciation neuronale glutamatergique suite à l'induction de la voie de l'acide rétinoïque.

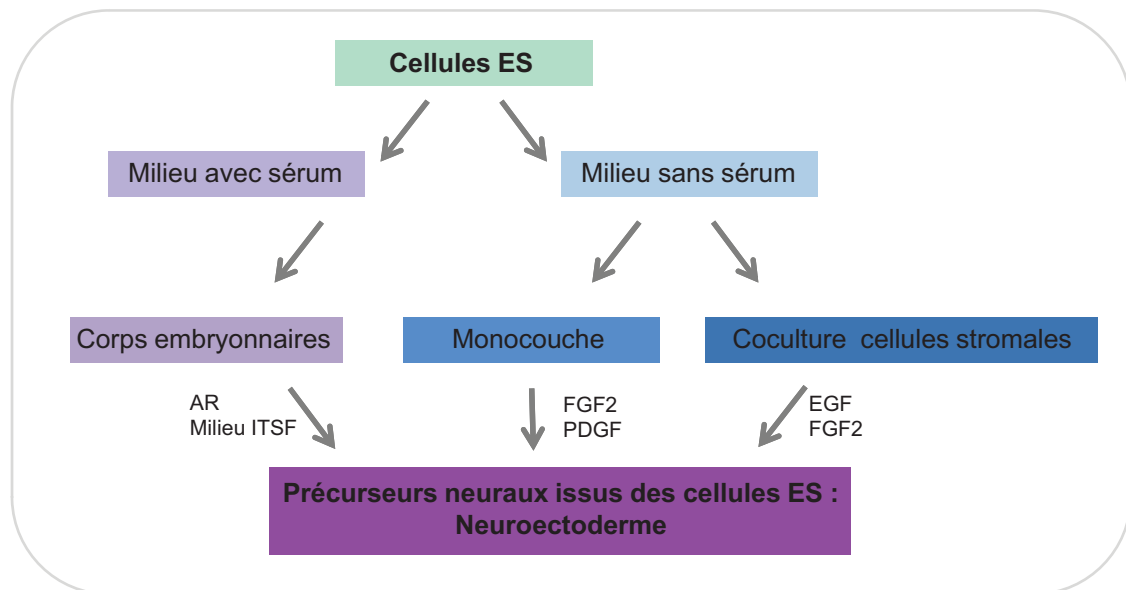
### 2.2.1 Importance de différentes voies de signalisation

Les cellules ES, cultivées dans un milieu minimum sans influence extérieure (LIF et feeders), expriment spontanément des marqueurs neuronaux tels que Sox1 et la Nestine (Smukler, Runciman et al. 2006), ce qui conforte l'idée que l'induction neurale se fait par défaut. Cependant, cette différenciation spontanée n'aboutit pas à une population homogène de neurones, mais donne plusieurs types cellulaires. L'activation des voies de signalisation FGF/IGF et celle de l'AR stimule cette différenciation neuronale (Ying, Nichols et al. 2003; Ying, Stavridis et al. 2003; Bibel, Richter et al. 2004).

### 2.2.2 Protocole de différenciation en neurones glutamatergiques *in vitro*

De nombreux protocoles existent pour la différenciation neuronale des cellules ES *in vitro*, tous basés sur la manipulation des voies de signalisation citées ci-dessus. La méthode des corps embryonnaires, la culture en monocouche ou encore la culture sur cellules stromales ont été utilisées afin de générer des neurones (*Figure 12*) (pour revue (Cazillis, Lelievre et al. 2005)). L'ensemble de ces protocoles montre que, comme ceci se passe au sein de l'embryon, *in vitro*, les cellules ES répondent à divers signaux responsables de la différenciation neuronale.

La description des différentes étapes de la différenciation neuronale *in vivo* faite précédemment suggère que la formation des cellules radiales gliales puis des neurones nécessite la génération d'un neuroépithélium polarisé. De ce fait, il n'est pas forcément évident que suite au traitement des cellules ES par l'acide rétinoïque, celles-ci donnent

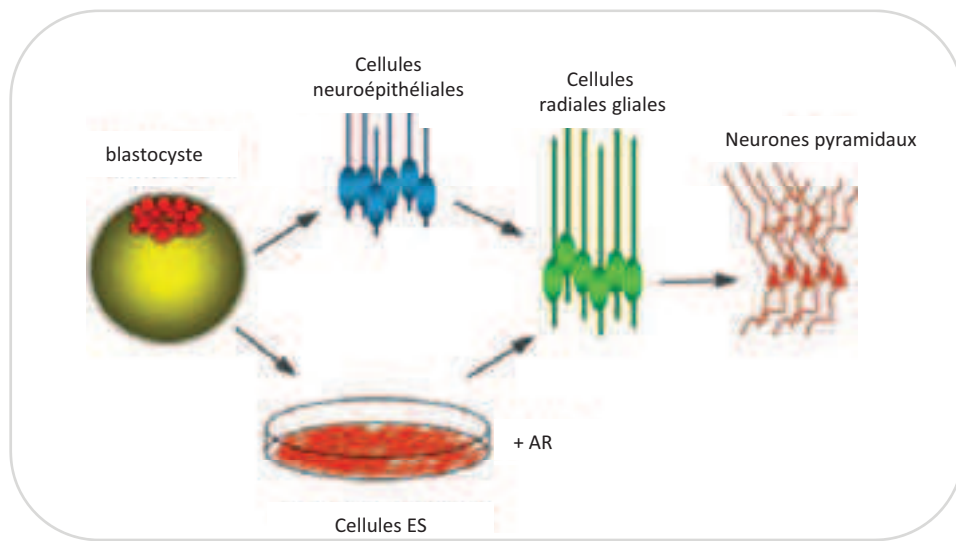


*D'après Cazillis, Lelievre et al 2005*

**Figure 12 : Méthodes de différenciation des cellules ES *in vitro*.** Différentes techniques ont été dérivées dans le but de différencier les cellules ES en précurseurs neuronaux. En utilisant un milieu contenant du sérum et la culture sous forme de corps embryonnaire, les précurseurs neuronaux peuvent être obtenus suite au traitement par l'AR ou encore par un milieu ITSF (Insuline, Transferrine, Sélénium, Fibronectine). En l'absence de sérum, l'utilisation de facteurs tel que EGF et FGF2 permet de favoriser la prolifération des précurseurs neuronaux.

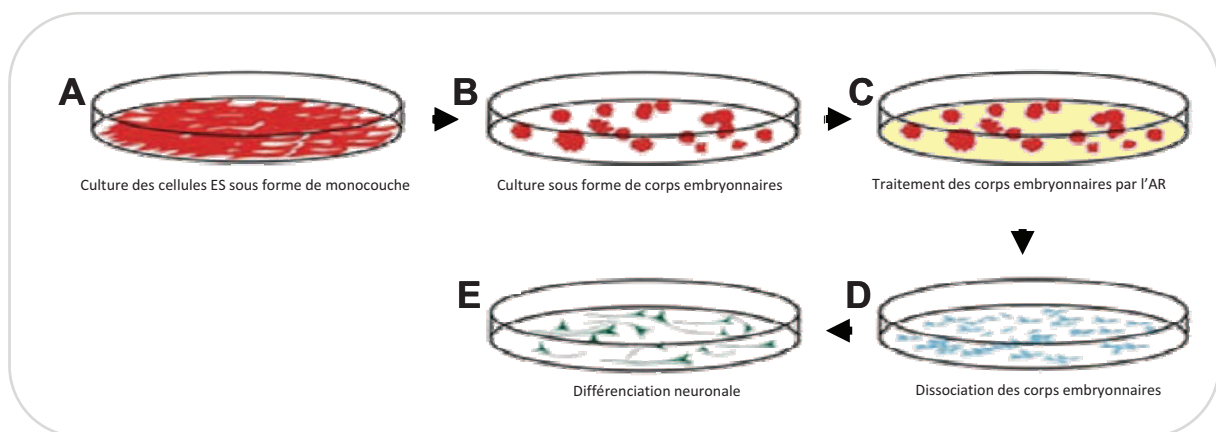
naissance à une population homogène de cellules RG. Pourtant cette population homogène de cellules radiales est bien observée lors de la différenciation neuronale induite par l'AR. Non seulement les cellules obtenues expriment les marqueurs connus des cellules RG, mais en plus, elles expriment Pax6, suggérant qu'elles représentent un sous-type de cellules radiales gliales similaires à celles développées notamment dans le cortex (Gotz and Barde 2005). Plus précisément, le protocole de différenciation basé sur l'induction neuronale par l'AR développé par Bibel et al permet l'obtention d'une population homogène de cellules radiales gliales qui vont se différencier en une population pure et homogène de neurones pyramidaux, similaires à ceux présents dans le cortex et l'hippocampe (*Figure 13*) (Bibel, Richter et al. 2004; Bibel, Richter et al. 2007). De plus, ce protocole nous permettra d'étudier l'action de l'AR au cours de la différenciation neuronale.

Les cinq étapes majeures dudit protocole sont illustrées dans la *Figure 14*. Comme décrit précédemment, le maintien des cellules ES dans un état indifférencié nécessite leur culture en présence de LIF ajouté dans le milieu de culture et de feeders. La première étape du protocole consiste donc en l'élimination des feeders de la culture. Cette déplétion peut être réalisée rapidement en trois passages successifs de trente minutes puisque les feeders sont plus denses et adhèrent plus rapidement que les cellules ES sur la boîte de culture. Le LIF étant maintenu dans le milieu de culture, les cellules ES restent dans un état indifférencié pendant plusieurs jours. Les cellules ES sont ensuite cultivées pendant huit jours sous forme d'EB et traitées avec l'AR les quatre derniers jours. Ces EB sont ensuite dissociés et les cellules réparties dans des boîtes de cultures tapissées de poly-D-ornithine/laminine en présence de N2. Ce milieu contient de nombreuses hormones, telles que l'insuline, la transferrine et la progestérone (Bibel, Richter et al. 2007), favorisant la neurogenèse et le maintien des neurones en culture. Le remplacement du milieu N2 par du milieu neurobasal permet de générer des neurones glutamatergiques et d'éliminer toutes les cellules non neuronales. Ainsi, les neurones obtenus par ce protocole de différenciation présentent les caractéristiques des neurones pyramidaux qui se trouvent notamment dans le cortex cérébral tel que décrit précédemment ou encore au niveau de l'hippocampe *in vivo*.



Gotz et Barde 2005

**Figure 13 : Différenciation neuronale *in vivo* et *in vitro*.** *In vivo*, les premières cellules constituant le tube neural, à l'origine du cerveau, sont les cellules neuroépithéliales. Ces cellules peuvent donner naissance à des cellules radiales gliales au cours du développement. De même, les cellules ES, dérivées de la masse cellulaire interne du blastocyste, peuvent être différenciées en cellules radiales gliales après induction par l'AR. Que ce soit dans la différenciation *in vivo* ou *in vitro*, les cellules radiales gliales peuvent donner naissance à des neurones pyramidaux, tels que les neurones glutaminergiques.



Bibel et al 2004

**Figure 14 : Protocole de différenciation neuronale des cellules ES.** Les cellules ES sont cultivées sous forme de monocouche, en conditions de prolifération idéales (**A**). A 80% de confluence, les cellules sont récupérées et ensemencées sur des boîtes non adhérentes, ce qui permet la formation des corps embryonnaires (EB : embryoid bodies). Les EB sont cultivés dans du milieu dépourvu de LIF (**B**) puis sont traités à l'AR pendant 4 jours (**C**). Après ces 4 jours de traitement, les cellules sont dissociées et remises en culture sur un support qui favorise leur adhésion ainsi que leur différenciation en neurones. A cette étape, le milieu de culture utilisé est du milieu N2, milieu dans lequel les cellules évoluent pour donner des cellules radiales gliales neurogéniques (**D**), qui sont dans un dernier temps différenciées en neurones glutaminergiques (pyramidaux) en remplaçant le milieu N2 par du milieu complet (**E**).

# ***Chapitre 3***

*La transcription  
et sa régulation*

## 3 Chapitre III : La transcription et sa régulation

### 3.1 Mécanismes généraux de l'initiation de la transcription

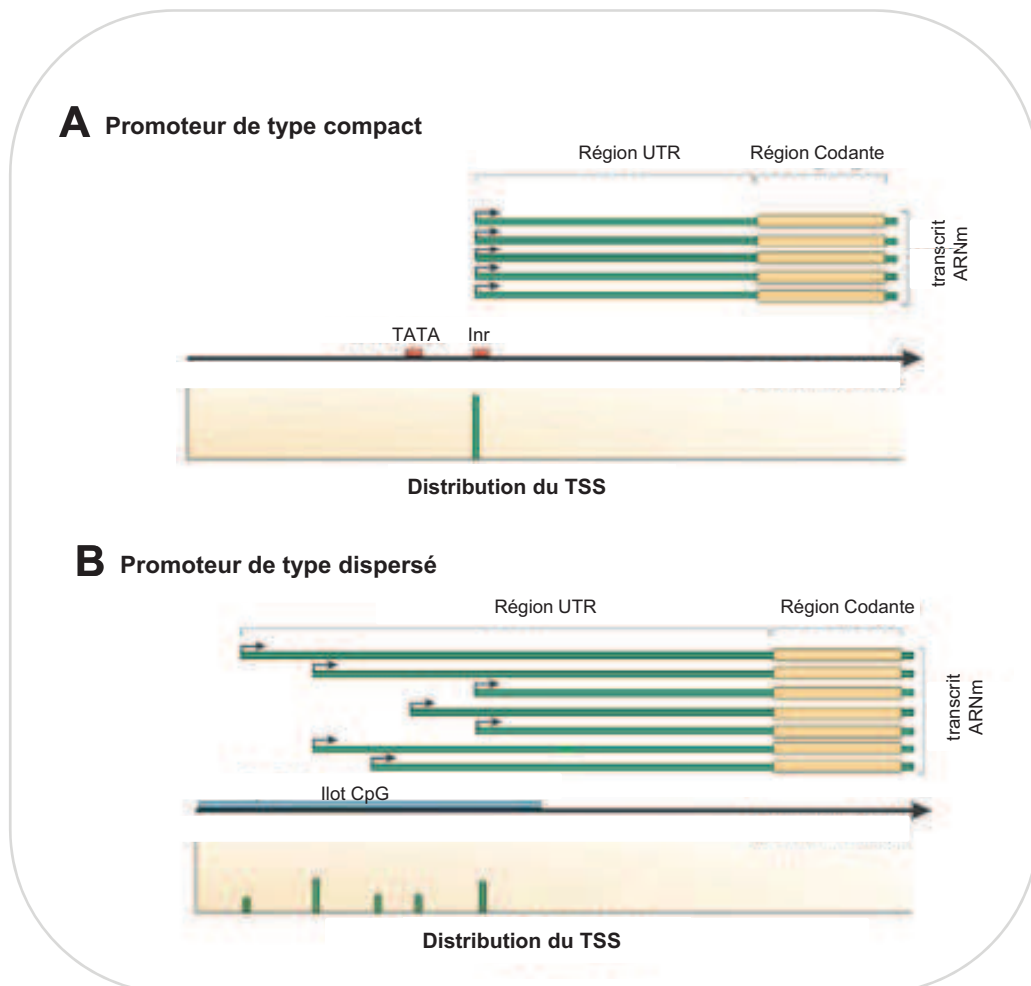
Chez les eucaryotes, l'ARN polymérase II (Pol II) est responsable de la transcription des gènes de classe II et assure la synthèse des ARNm et des petits ARN nucléaires. La réaction de transcription par la Pol II peut être divisée en trois étapes : l'initiation, l'élongation et la terminaison. L'initiation de la transcription nécessite la formation du complexe de pré-initiation (PIC). Le PIC est composé de la Pol II, du complexe Médiateur et des facteurs généraux de la transcription (GTF) TFIIA, -B, -D, -E, -F, -H, dont la plupart sont eux-mêmes des complexes multiprotéiques. L'assemblage du PIC s'effectue au niveau du promoteur minimal autour du site d'initiation de la transcription (TSS). Deux types de promoteurs minimaux existent : compact versus dispersé (*Figure 15*). Les promoteurs compacts contiennent un site d'initiation de la transcription unique, alors que les promoteurs dispersés contiennent différents sites d'initiation de la transcription sur un intervalle de 50 à 100 nucléotides. Le promoteur minimal compact peut comporter plusieurs éléments de séquence, comme l'élément TATA, un élément initiateur (Inr), un élément BRE (TFIIB Responsive Element), ou encore un élément en aval du site d'initiation de la transcription (DPE : Downstream Promoter Element) (*Figure 16*). En général, ces éléments sont caractéristiques des promoteurs qui sont hautement régulés pendant la différenciation ou en réponse à des voies de signalisation. La plupart des promoteurs cellulaires ne contiennent néanmoins aucun de ces éléments mais sont caractérisés par la présence de régions riches en GC appelées îlots CpG et sont des promoteurs de type dispersé (Sandelin, Carninci et al. 2007).

#### 3.1.1 Les différents composants du promoteur minimum

Les éléments promoteurs tels que la boîte TATA, l'élément BRE, Inr, MTE (Motif Ten Element) et DPE sont typiquement retrouvés dans les « core » promoteurs de type compact. Néanmoins, ces éléments du promoteur minimal ne sont pas universels. En réalité, chacun de ces éléments est présent uniquement dans un sous-ensemble de promoteurs.

##### 3.1.1.1 La boîte TATA

La boîte TATA est le premier élément promoteur de base à avoir été identifié. Cet élément tire son nom de sa séquence consensus, TATAWAAR, où le nucléotide T en amont est souvent situé à -30 ou -31 par rapport au site d'initiation de la transcription (Breathnach and



*D'après Sandelin, Carninci et al. 2007*

**Figure 15 : Comparaison des promoteurs de type compact et de type dispersé.** Alors que les promoteurs dits compacts, ou encore à TSS unique, comprennent un élément TATA et/ou Inr (A), les promoteurs de types dispersés (à TSS multiples) sont riches en îlots CpG.

Chambon 1981; Carninci, Sandelin et al. 2006; Ponjavic, Lenhard et al. 2006). La boîte TATA est reconnue et fixée par TBP (TATA Binding Protein), qui est une sous-unité du complexe TFIID. La boîte TATA est présente uniquement dans une partie des promoteurs compacts. De ce fait, cet élément est plutôt rare chez les vertébrés, puisque moins d'un tiers des promoteurs sont de type compact, et seule une fraction des promoteurs de ce type contient un élément TATA.

#### 3.1.1.2 BRE

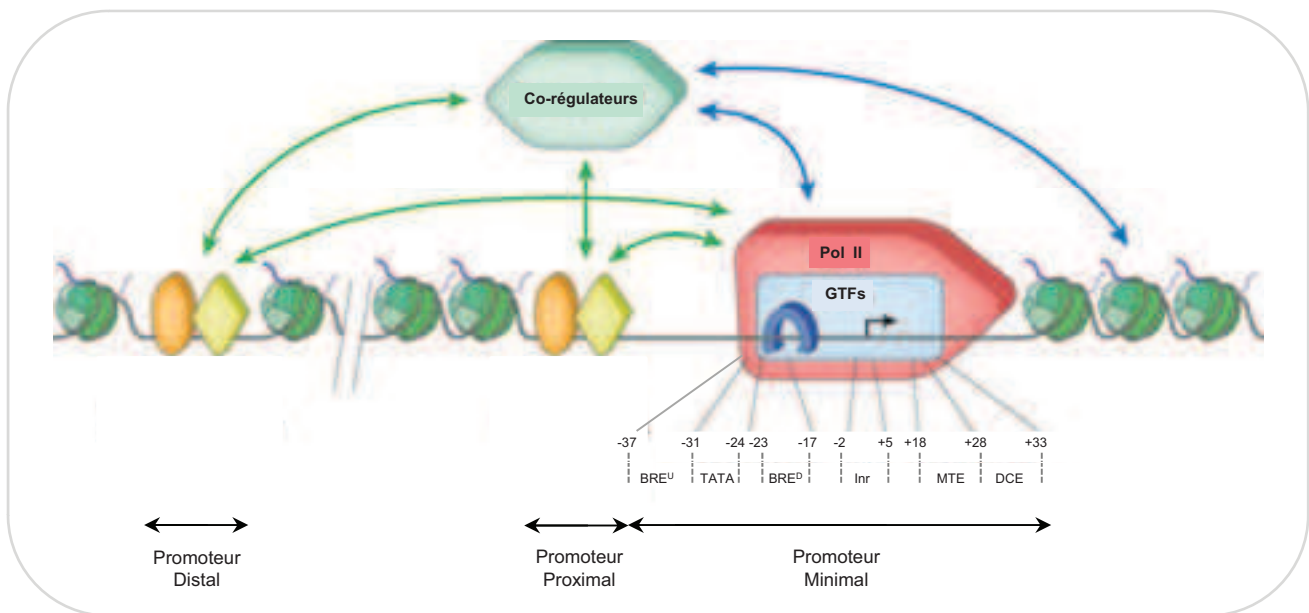
Le BRE a initialement été identifié comme étant un élément reconnu par TFIIB situé en amont de la boîte TATA (Lagrange, Kapanidis et al. 1998). Il a ensuite été découvert que TFIIB pouvait en réalité se fixer en amont (BRE<sup>U</sup> pour Upstream) ou en aval (BRE<sup>D</sup> pour downstream) de la boîte TATA (Deng and Roberts 2005; Deng and Roberts 2007). La séquence consensus de l'élément BRE<sup>U</sup> est SSRGCC (Lagrange, Kapanidis et al. 1998), celle de BRE<sup>D</sup> située juste en aval de la boîte TATA, est RTDKKKK (Deng and Roberts 2005). Selon le contexte du promoteur, les BRE<sup>U</sup> et BRE<sup>D</sup> peuvent agir de manière positive ou négative. Si ces séquences étaient dans un premier temps supposées importantes pour l'affinité de TFIIB avec le promoteur *in vitro* (Deng and Roberts 2005), des expériences de ChIP-chip *in vivo* semblent indiquer que la fixation de TFIIB ne nécessite pas forcément la présence des éléments BRE (Albert, Grote et al. 2010)

#### 3.1.1.3 Inr

Le motif initiateur (Inr) comprend le site d'initiation de la transcription (Smale and Baltimore 1989; Smale and Kadonaga 2003). Si le consensus de l'Inr varie selon les espèces (YYANWYY chez l'homme, TCAKTY chez la drosophile), le nucléotide A conservé dans la séquence représente souvent le site d'initiation de la transcription. L'Inr est capable de diriger la transcription seul ou en synergie avec la boîte TATA et/ou le DPE. L'Inr est un site de reconnaissance pour TFIID. Plus précisément, ses sous-unités TAF1 et TAF2 peuvent reconnaître ce motif (Thomas and Chiang 2006)

#### 3.1.1.4 DPE

Le DPE a été identifié comme un élément de reconnaissance de TFIID (Burke and Kadonaga 1996). Le DPE est conservé de la drosophile (séquence consensus RGWYVT (Kutach and Kadonaga 2000)) à l'homme et est situé à +28 / +33 en aval de site d'initiation de la transcription.



*D'après Fuda, Ardehali et al. 2009*

**Figure 16 : Les séquences de régulation des gènes de classe II.** Le promoteur se définit comme la région d'ADN qui dirige l'initiation de la transcription d'un gène. Dans les promoteurs de classe II, les éléments tels que la boîte TATA, l'élément Inr et l'élément DPE forment le promoteur minimal, ou « core » promoteur. Ces éléments permettent de lier les GTFs. Les régulateurs de la transcription (ovale orange et losange jaune), qui représentent soit des activateurs, soit des répresseurs, se lient à des régions du promoteur proximal ou du promoteur distal. Les activateurs, par l'intermédiaire d'un adaptateur appelé médiateur, permettent d'activer la transcription.

#### 3.1.1.5 MTE

Le MTE a été trouvé en associant des analyses bioinformatiques et biochimiques. Ce motif de 10bp a pour consensus la séquence CSARCSSAAC et se situe de +18 à +27 du site d'initiation chez la drosophile. Comme pour le DPE décrit précédemment, le MTE fonctionne coopérativement avec l'Inr, avec un espacement précis requis entre les deux éléments. Alors que le MTE fonctionne en synergie avec les éléments TATA et DPE, l'addition d'un MTE peut également compenser la perte de la transcription due aux mutations dans les éléments TATA et DPE.

Les études du MTE et du DPE étant réalisées chez la drosophile, il se peut que ces éléments aient un consensus plus large / moins restrictif chez les mammifères.

#### 3.1.2 Les îlots CpG

La méthylation de l'ADN est une modification covalente qui ne va en rien changer la séquence génomique. Chez les vertébrés, cette réaction se fait majoritairement dans un environnement génomique donné, c'est-à-dire lorsqu'un nucléotide C est suivi par un nucléotide G (C Phosphate G). Dans les cellules ES, les cytosines peuvent également être méthylées au niveau des dinucléotides CpA et CpT (Ramsahoye, Biniszkiewicz et al. 2000)

Un îlot CpG est défini comme une région de plus de 500bp dans laquelle le pourcentage de CG est supérieur à 55% (Gardiner-Garden and Frommer 1987; Takai and Jones 2002). Plus de 70% des promoteurs humains contiendraient un îlot CpG, ce qui fait de ces îlots le type de promoteur le plus fréquent (Saxonov, Berg et al. 2006). Inversement, près de 50% des îlots CpG seraient situés dans le TSS des gènes. Des études récentes ont mis en évidence des îlots CpG non associés avec des sites d'initiation de la transcription annotés, mais qui néanmoins présenteraient des fonctions promotrices (Illingworth, Gruenewald-Schneider et al. 2010; Maunakea, Nagarajan et al. 2010).

Les promoteurs contenant des îlots CpG sont souvent dépourvus des éléments du promoteur minimal cités ci-dessus, ce qui peut expliquer qu'ils possèdent souvent plusieurs TSS contrairement au promoteur contenant un élément TATA et/ou Inr qui eux possèdent un TSS fixe (Carninci, Sandelin et al. 2006) (*Figure 15*).

Lors du développement embryonnaire précoce, le taux de méthylation de l'ADN est considérablement réduit, puis, peu avant l'implantation, l'ADN est reméthylé. Cette

reméthylation est due à l'action d'enzymes appelés des DNA méthyltransférases (DNMT). Il faut noter que les îlots CpG sont largement exclus de cette reméthylation. L'idée est alors que la machinerie transcriptionnelle laisse une « empreinte » au cours de l'embryogenèse. En effet, cette machinerie transcriptionnelle empêcherait l'accès des DNMT et par conséquent la méthylation de l'ADN (Illingworth and Bird 2009).

En 2009, une autre modification des cytosines a été mise en évidence. L'hydroxyméthylation de la cytosine (5hmC : 5- hydroxymethylcytosine) est catalysée par les enzymes de la famille TET (Ten-Eleven Translocation) et est abondante dans certains types de cellules. Ces enzymes catalysent également l'oxydation de cette 5hmC en 5-formylcytosine (5fC) puis 5-carboxycytosine (5caC). La marque 5hmC sert ainsi d'intermédiaire dans la réaction de déméthylation médiée par l'enzyme TDG (thymine DNA glycosylase) (pour revue (Moore, Le et al. 2013)). La marque 5hmC pourrait également agir comme un signal pour des facteurs reconnaissant la chromatine. Cette modification serait relativement abondante notamment dans les cellules pluripotentes et dans les neurones (Pour revue (Wu and Zhang 2011)).

### 3.1.3 Régions régulatrices en dehors du promoteur

Bien que le promoteur minimal suffise pour assembler le PIC et diriger la transcription à partir d'ADN nu *in vitro*, la régulation de la transcription *in vivo* nécessite l'action des régulateurs de transcription. Ces derniers reconnaissent d'autres séquences régulatrices situées dans le promoteur proximal (proche du TSS) et surtout au niveau du promoteur distal (*Figure 16*). Ces éléments sont appelés enhancers ou silencers, suivant leur effet positif ou négatif sur la transcription et se situent parfois très loin en amont ou en aval du TSS. Ces séquences, reconnues par les facteurs de transcription, permettent une régulation fine de l'expression des gènes lors de la différenciation et de la spécification de l'identité cellulaire. Tous les facteurs de transcription n'interagissent pas directement avec l'ADN, mais peuvent agir via leurs partenaires ou en tant que sous-unité de complexes protéiques. Ces facteurs sont définis comme des coactivateurs ou corépresseurs et interagissent avec les enzymes de modification ou les complexes de remodelage de la chromatine, ainsi qu'avec les composants du PIC.

### 3.1.3.1 Régions proximales

Les séquences régulatrices proximales sont en général situées entre 40 et 110bp en amont du site d'initiation de la transcription. Ces séquences sont reconnues par des facteurs de transcription qui peuvent agir indirectement sur la machinerie transcriptionnelle en recrutant d'autres facteurs responsables de la modification de la chromatine ou directement en interagissant avec les éléments de la machinerie transcriptionnelle.

Ces séquences régulatrices peuvent avoir un effet activateur (UAS / Upstream Activating Sequence) ou répresseur (URS : Upstream Repressing Sequence) selon la protéine qui les reconnaît. Les séquences régulatrices proximales les plus connues sont les boîtes CCAAT et GGGCGG.

### 3.1.3.2 Régions distales

La régulation de l'expression d'un gène ne se fait pas uniquement par le biais de régions « adjacentes ». Cette régulation peut également être due à des séquences pouvant être distantes de plusieurs milliers de paires de bases du promoteur. Différents types d'éléments ont ainsi été décrits.

Les enhancers sont sûrement les plus étudiés parmi ces éléments. Ils peuvent être décrits comme l'élément prédominant pour la régulation de l'expression d'un gène. Lorsqu'ils sont placés en amont d'un promoteur minimal et d'un gène rapporteur, ils permettent de récapituler l'expression native du gène. Leur activité est la conséquence de la fixation de facteurs de transcription à des motifs ADN de 6 à 20 pb. *In vivo*, des boucles de la chromatine permettent de créer l'interaction entre les enhancers et les promoteurs proximaux de leurs gènes cibles (Akbari, Bousum et al. 2006). Les silenciers quant à eux répriment l'expression d'un gène via des interactions avec le promoteur proximal du gène ciblé (Ogbourne and Antalis 1998). Les éléments insulateurs permettent d'isoler les enhancers des gènes proches et de diriger leur activité vers les promoteurs minimaux ciblés (Bell, West et al. 2001).

Ces différents éléments peuvent être situés à plusieurs kilobases voir mégabases du gène régulé. De plus, il peut y avoir plusieurs enhancers pour un même gène. Le terme de « shadow » enhancer est alors utilisé, utilisant une métaphore pour indiquer que cet enhancer a été découvert après l'enhancer principal (qui est souvent le premier découvert, et qui se situe à proximité du gène). Ce shadow enhancer peut se trouver à une localisation inattendue, telle que dans l'intron d'un gène voisin. Des expériences chez la drosophile ont mis en évidence

l'importance fonctionnelle d'avoir plusieurs enhanceurs. Dans des conditions de développement normal, l'enhancer proximal ou distal est suffisant. En revanche si les embryons se développent à des températures élevées, la perte d'un des deux enhanceurs va entraîner des défauts de gastrulation (Perry, Boettiger et al. 2010).

L'identification de ces enhanceurs est possible par différentes méthodes. Des analyses biochimiques, telles que la mutation de l'enhancer, ou sont clonage en amont d'un gène rapporteur peuvent être utilisées pour caractériser ces éléments. De plus, des analyses bioinformatiques peuvent également être mise en œuvre pour identifier des enhanceurs potentiels. En effet, ces éléments contiennent souvent une forte densité de sites de fixation pour des facteurs de transcription, ce qui a permis de créer des algorithmes qui vont identifier ces régions (Berman, Nibu et al. 2002; Markstein, Markstein et al. 2002). Même si cette technique fonctionne, seul 10-30% des régions identifiées sont d'authentiques enhanceurs lorsqu'ils sont testés. Des expériences d'immunoprécipitation de la chromatine suivi de séquençage à haut débit (ChIP-seq) ciblant des facteurs de transcription ou des marques d'histones permettent également d'identifier ces enhanceurs. Je reviendrai dans une partie ultérieure sur les différentes marques d'histones associées aux différents éléments promoteurs / régulateurs.

### **3.2 Les acteurs protéiques de la transcription**

Chez les eucaryotes, l'ensemble des gènes codant pour des protéines sont transcrits par la Pol II. L'initiation de cette transcription et sa régulation nécessite six facteurs généraux de la transcription ainsi que des coactivateurs. Le complexe médiateur est le complexe coactivateur central puisqu'il permet de faire le lien entre les éléments de régulation en amont et la région promotrice. Je vais dans cette partie détailler quelques-uns des acteurs de la transcription, en commençant par l'enzyme, l'ARN polymérase II.

#### **3.2.1 L'ARN polymérase II et la machinerie de base**

##### **3.2.1.1 L'ARN polymérase II**

La Pol II est un complexe multiprotéique composé de 12 sous-unités. Ces sous-unités, dont certaines sont communes aux trois ARN polymérases et d'autres spécifiques de la Pol II, sont appelées RPB (RNA Polymerase B) (Young 1991). La partie catalytique de l'enzyme est formée par un noyau de 10 sous-unités à laquelle vient se lier la paire RBP4/7 (Armache, Kettenberger et al. 2003). La Pol II est capable de générer des ARN à une vitesse très rapide,

variant de 20 à 70 nucléotides par seconde. Néanmoins son activité n'est pas linéaire le long de la matrice d'ADN puisque la Pol II effectue un déplacement en avant préalable à l'intégration d'un nucléotide, puis effectue un mouvement de recul pour terminer la maturation de l'extrémité 3' de l'ARN (Wang, Bushnell et al. 2009)

La sous-unité RBP1 de la Pol II comprend un domaine CTD (C-Terminal Domain) sujet à des phosphorylations importantes pour la régulation de la transcription. Cette extrémité CTD est composée d'une répétition de la séquence heptamérique YSPTSPS, au sein de laquelle la sérine 2 et la sérine 5 représentent les sites de phosphorylation.

### *3.2.1.2 Les facteurs généraux de la transcription*

Afin de se fixer de manière stable sur les promoteurs, la Pol II nécessite la présence des GTF. Ces facteurs de transcription ont été découverts par des expériences de chromatographie. Le nom de ces facteurs est d'ailleurs associé avec la fraction chromatographique dans laquelle ils ont été retrouvés. Cette expérience a permis d'identifier TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIH, où TF signifie « facteur de transcription » et où le chiffre II fait référence à la Pol II (Sawadogo and Roeder 1985). La [table 1](#) donne un bref aperçu de la composition et des propriétés fonctionnelles de chaque GTF chez l'homme (pour revue (Thomas and Chiang 2006)).

## *3.2.2 Modèles d'assemblage du PIC*

*In vitro*, deux modèles d'assemblage du PIC ont été décrits : le modèle séquentiel et le modèle de l'holoenzyme.

### *3.2.2.1 Assemblage séquentiel*

La nature hiérarchique de l'assemblage des GTF au niveau du promoteur a été étudiée au début des années 1980 en utilisant des fractions chromatographiques d'extraits de cellules Héla. En 1989, l'équipe de Buratowski a déterminé l'ordre d'assemblage du PIC, en se basant sur des études d'électrophorèse en condition native et de footprinting. Plus précisément, ils ont déterminé que TFIID en premier reconnaissait la boîte TATA, suivi de manière séquentielle de l'entrée de TFIIA, TFIIB, la Pol II et TFIIE (Buratowski, Hahn et al. 1989) (à cette époque TFIIIF et TFIIH n'avaient pas encore été identifiés). Une fois l'ensemble des GTF identifié et purifié, le modèle d'assemblage séquentiel a été mis à jour. En premier lieu, TFIID se fixerait à la région promotrice. Interviendraient ensuite TFIIA puis TFIIB, qui permettraient de stabiliser TFIID au niveau du promoteur. La troisième étape consiste à

**Table 1 : Composition des différents GTF et leurs fonctions chez l'homme**

| <b>GTF</b> | <b>Composition protéique</b>                                                                 | <b>Fonction</b>                                                                                                                                                                                                                          |
|------------|----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| TFIIA      | P35 (α), p19 (β), p12 (γ)                                                                    | Antirépresseur, stabilisation du complexe TBP-TATA, coactivateur                                                                                                                                                                         |
| TFIIB      | p33                                                                                          | Sélection du site d'initiation, stabilisation du complexe TBP-TATA, recrutement de PolII/TFIIF                                                                                                                                           |
| TFIID      | TBP + TAFs (TAF1-TAF14)                                                                      | Coactivateur, activité kinase, histone acétyl transférase, liaison au promoteur                                                                                                                                                          |
| TFIIE      | p56 (α), p34 (β)                                                                             | Recrutement de TFIIH, formation d'un complexe ouvert, transition initiation/élongation                                                                                                                                                   |
| TFIIF      | RAP30 , RAP74                                                                                | Fixation de la Pol II, recrutement de TFIIE et TFIIH, sélection du site d'initiation, joue sur l'efficacité de l'élongation                                                                                                              |
| TFIIH      | p89 (XPB), p80 (XPD), p62, p52, p44, p40 (CDK7), p38 (Cycline H), p34, p32 (MAT1), p8 (TFB5) | Activité hélicase pour l'ouverture de l'ADN, activité de réparation de l'ADN, activité ATPase impliquée dans l'initiation de la transcription, activité kinase pour la phosphorylation du CTD de la PolII, activité E3 ubiquitine ligase |

*À partir de Thomas et Chiang 2006*

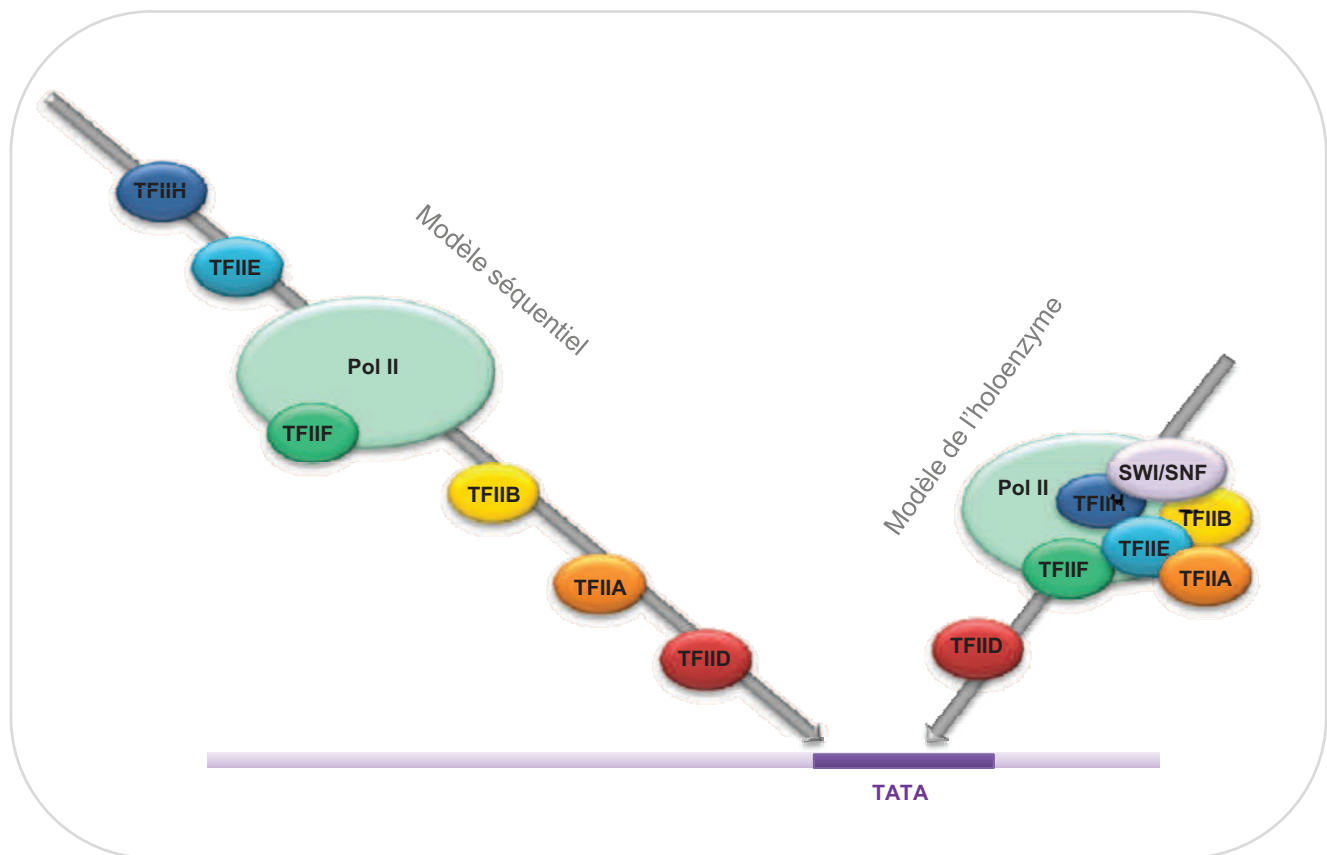
l'entrée de la Pol II et TFIIF. Après la formation d'un complexe stable TFIID-TFIIA-TFIIB-PolII/TFIIF, TFIIE puis TFIIH sont recrutés (pour revue (Thomas and Chiang 2006)). Cette suite d'étapes est définie sous le terme de modèle séquentiel (*Figure 17*).

#### 3.2.2.2 *Modèle de l'holoenzyme*

Le modèle alternatif de l'holoenzyme a été décrit pour la première fois chez la levure en 1994. En effet, la Pol II a pu être purifiée sous forme d'holoenzyme contenant la Pol II, certains GTF (TFIIB, TFIIF, TFIIH) et certaines protéines SRB (Suppressors of RNA polymerase B). Néanmoins, cette holoenzyme n'est pas capable d'initier la transcription *in vitro* sans l'ajout de TBP et de TFIIA (Koleske and Young 1994). En parallèle, plusieurs complexes, d'une composition légèrement différente de celle initialement décrite par l'équipe de Koleske ont été décrits. Ces complexes peuvent contenir l'histone acétyltransférase GCN5, le complexe de remodelage de la chromatine SWI/SNF ou d'autres éléments importants dans la réparation de l'ADN ou la maturation des mRNA. Cette variation entre les différents complexes identifiés peut s'expliquer entre autre par des méthodes de purification différentes. Néanmoins, une chose est commune entre l'ensemble des complexes identifiés : l'absence de TFIID. Ceci suggère que, de manière similaire au modèle procaryotique, un facteur, en l'occurrence TFIID ( $\sigma$  chez les procaryotes) permet de recruter l'holoenzyme Pol II ( $\alpha\beta\beta'$  chez les procaryotes) au niveau de la région promotrice (*Figure 17*).

#### 3.2.3 *Les cofacteurs de la transcription*

La formation du PIC commence généralement par la fixation de TFIID, suivi de l'arrivée des autres facteurs généraux de la transcription et de la Pol II, soit selon un modèle séquentiel, soit selon le modèle de l'holoenzyme (décrit ci-dessus). La formation de ce complexe au niveau de promoteur est suffisante pour observer un niveau de transcription basal. Néanmoins, pour obtenir une transcription « régulée », des cofacteurs sont souvent nécessaires. Ces derniers transmettent des signaux régulateurs entre les activateurs ou répresseurs spécifiques du gène et la machinerie générale de transcription. Différentes classes de cofacteurs existent, incluant le médiateur, les cofacteurs positifs (USA : Upstream stimulatory activity) et les cofacteurs négatifs (NC : Negative Cofactor). Ces éléments fonctionnent indépendamment ou en combinaison pour déterminer l'activité promotrice spécifiquement pour un gène ou un type cellulaire.



**Figure 17 : Deux modèles d'assemblage du PIC.** Deux modèles de formation du PIC existent. Dans un premier modèle, dit séquentiel, les facteurs de transcription et la Pol II sont recrutés dans un ordre bien défini au niveau du promoteur. Dans le modèle de l'holoenzyme, TFIID est recruté en premier, puis permet le recrutement d'un complexe formé des autres GTFs, de la PolII et d'autres facteurs tel que par exemple le complexe de remodelage de la chromatine SWI/SNF.

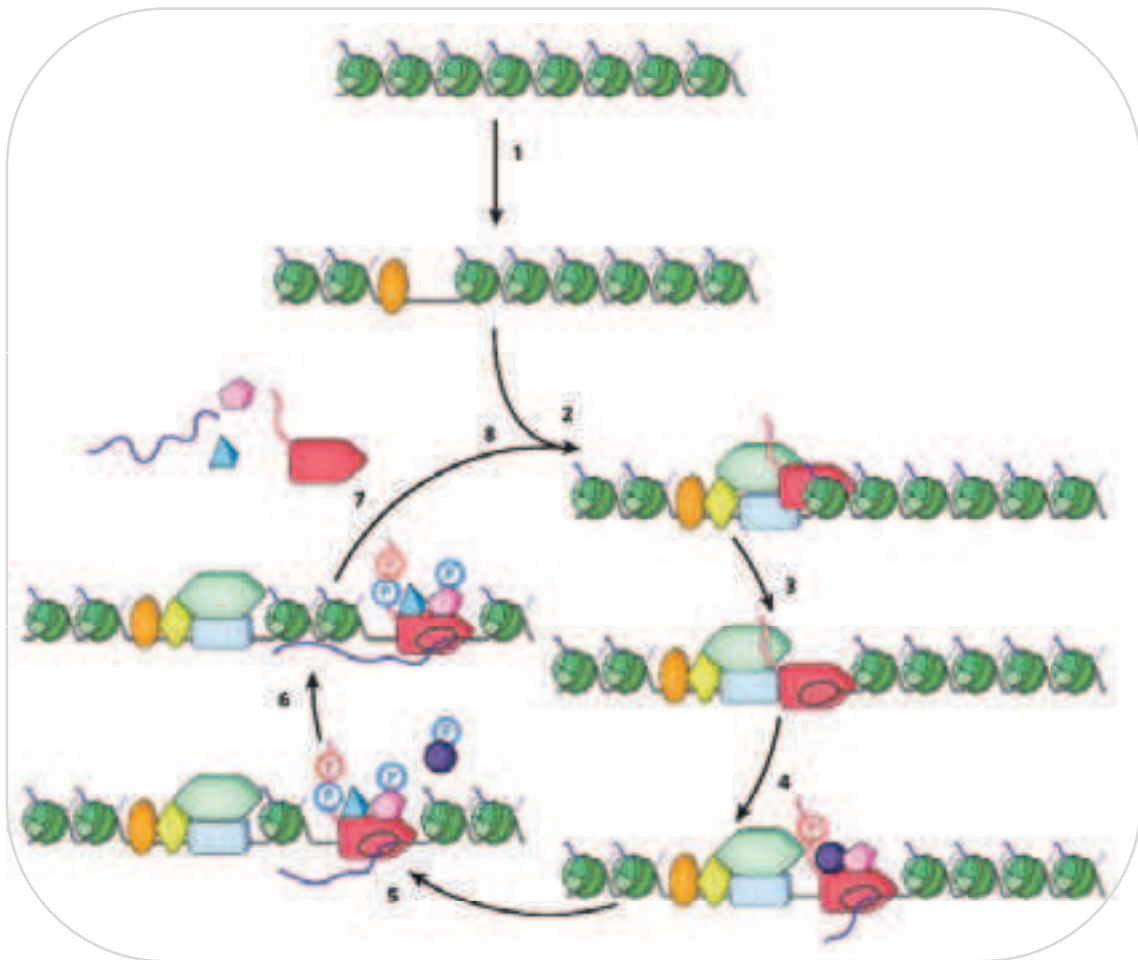
En guise d'exemple, le complexe médiateur sert d'intermédiaire entre les coactivateurs de la transcription et la machinerie de base. Ce complexe n'est pas nécessaire pour la transcription de base, mais est impliqué dans la transcription régulée. Il peut être recruté au cours de la transcription par certains récepteurs nucléaires, tels que notamment le récepteur de l'acide rétinoïque (RAR) que je décrirai ultérieurement (Pavri, Lewis et al. 2005).

### 3.3 Régulation de la transcription

La majorité des gènes de classe II sont soumis à une régulation transcriptionnelle qui peut être divisée en différentes phases, durant lesquelles la polymérase est recrutée au niveau du promoteur, initie la transcription, effectue l'élongation du transcrit, puis passe en phase de terminaison (pour revue (Core and Lis 2008)). Ces phases majeures peuvent en réalité être décomposées en huit étapes potentiellement soumises à une régulation.

Le cycle transcriptionnel commence par une modification de l'état de la chromatine rendant accessible les éléments du promoteur minimal. Cette étape est dirigée par les facteurs de transcription se fixant sur le promoteur distal et/ou proximal (étape 1). L'ouverture de la chromatine permet l'assemblage du PIC, suivi du recrutement de la Pol II (étape 2), ce qui entraîne l'ouverture du double brin d'ADN au niveau du TSS (étape 3). La Pol II entre ensuite en phase d'élongation et se détache du PIC. Suite à cette troisième étape, la Pol II peut s'arrêter dans un état de « pause » sur l'ADN (étape 4), dont la sortie peut aboutir à une terminaison précoce ou à l'élongation (étape 5). Dans le cas de l'élongation, la polymérase doit réaliser un transcrit de la totalité du gène, ce qui constitue la sixième étape pouvant subir une régulation. Les septièmes et huitièmes étapes qui peuvent moduler le taux de la transcription sont, respectivement, la terminaison et la ré-initiation d'un nouveau cycle de transcription (*Figure 18*) (Fuda, Ardehali et al. 2009).

L'idée que la Pol II puisse faire une pause suite à sa fixation sur l'ADN est récente, puisque préalablement l'hypothèse majeure proposait que le recrutement de la Pol II constitue la principale étape de contrôle. Les résultats récents montrent que, pour de nombreux gènes, la Pol II est recrutée mais maintenue dans un état de pause, ce qui empêche l'élongation et par conséquent l'expression du gène (Fuda, Ardehali et al. 2009). Les différents états de la Pol II sont déterminés par des modifications covalentes du CTD de la grande sous-unité de la Pol II. La transition entre l'initiation et l'état de pause est corrélée avec la phosphorylation des résidus Sérine 5 du CTD par l'activité kinase de la sous-unité CDK7 de TFIIH. L'étape



*D'après Fuda, Ardehali et al. 2009*

**Figure 18 : Régulation du cycle transcriptionnel.** De nombreuses étapes peuvent être soumises à régulation au cours de la transcription :

- 1) Fixation de facteurs de transcription au niveau des promoteurs proximal et distal
- 2) Ouverture de la chromatine et assemblage du PIC
- 3) Séparation du double brin d'ADN
- 4) Entrée de la PolII en phase précoce d'élongation
- 5) Sortie de l'état de pause avec soit une terminaison précoce, soit l'entrée en phase d'élongation
- 6) Transcription de la totalité du gène
- 7) Terminaison
- 8) Ré-initiation d'un nouveau cycle de transcription

d'élongation est quant à elle associée à la phosphorylation des résidus Sérine 2 du CTD par la sous-unité CDK9 de P-TEFb (Positive Transcription Elongation Factor b). Le contrôle de la transcription permet ainsi de passer très rapidement d'un état de répression à l'activation, sans étape d'assemblage du PIC (Fuda, Ardehali et al. 2009).

L'utilisation de nouvelles techniques telles que l'immunoprécipitation de la chromatine (ChIP) combinée à l'utilisation d'anticorps spécifiques des différents types de phosphorylation du CTD a permis de mettre en évidence le fait que la Pol II se trouvait en état de pause en aval du TSS, ainsi que les modifications qui sont associées à cet état. La mise en évidence des différentes étapes impliquées dans la régulation de chaque gène peut également être réalisée par ces techniques (Fuda, Ardehali et al. 2009).

### 3.4 Transcription et épigénétique

#### 3.4.1 *Vue d'ensemble des modifications d'histones*

Comme décrit ci-dessus, l'initiation de la transcription est dépendante de l'état de la chromatine, qui module l'accessibilité des facteurs de transcription au niveau du promoteur minimal. La structure chromatinienne est dynamique et régulée par différentes modifications covalentes au niveau des queues d'histone, souvent désignées comme modifications « épigénétiques ». Les méthylations, les acétylations, les ubiquitinations ou encore les phosphorylations font partie de ces modifications. L'ensemble des modifications possibles constitue le code des histones. La [table 2](#) donne une vue non exhaustive de différentes marques d'histones. De nombreuses autres marques non citées dans le tableau sont néanmoins connues ([Figure 19](#)) (Berger 2007; Li, Carey et al. 2007; Portela and Esteller 2010). Parmi ces modifications, celles des lysines et des arginines sont d'un intérêt particulier, puisque ces résidus sont sujets à différents types de modifications. De plus, en ce qui concerne la méthylation, les résidus peuvent être mono-, di- ou triméthylés. Les modifications d'histone jouent un rôle important dans la compaction de la chromatine, mais également dans la régulation de la transcription, la réparation ou la réplication. Le contexte chromatinien sera ainsi différent entre les différents types de cellules somatiques, ainsi qu'entre les cellules souches et les cellules différenciées.

Les modifications post-traductionnelles des histones peuvent remplir leurs fonctions du seul fait de leur présence sur les histones, mais peuvent également être reconnues par des intermédiaires protéiques. Ces intermédiaires possèdent des domaines capables de reconnaître

Table 2 : Exemples de modifications d'histones et de leur rôle dans la transcription.

| Type de modification | Position de la modification                             | Rôle transcriptionnel |
|----------------------|---------------------------------------------------------|-----------------------|
| Acétylation          | H3 K9, -14, -18, -56<br>H4 K5, -8, -13, -16<br>H2A, H2B | Activation            |
| Méthylation          | H3 R2, -17, -22<br>H4 R3                                | Activation            |
|                      | H3 K4, -36, -79                                         | Activation            |
|                      | H3 K9, -27<br>H4 K20                                    | Répression            |
| Phosphorylation      | H3 T3 / H3 S10, -28<br>H2A, H2B                         | Activation            |
| Ubiquitination       | H2B K120, -123                                          | Activation            |
|                      | H2A K119                                                | Répression            |
| Sumoylation          | H2B K6, -7<br>H2A K126                                  | Répression            |

D'après Berger 2007 ; Li et al 2007



D'après Portela et Esteller 2010

Figure 19 : Récapitulatif des différentes marques d'histones

précisément le type de modification et vont permettre le recrutement d'autres facteurs de régulation de la transcription ou de remodelage de la chromatine (Pour revue (Taverna, Li et al. 2007)).

### *3.4.2 Mise en place des principales marques d'histones*

#### *3.4.2.1 Acétylation*

L'acétylation des lysines est un processus hautement dynamique et régulé par deux groupes d'enzymes qui possèdent des actions opposées. L'acétylation est sous le contrôle des histones acétyltransférases (HAT) alors que la desacétylation est réalisée par des histones déacétylases (HDAC).

Les HATs utilisent comme cofacteur l'acétyl-coA et catalysent le transfert d'un groupement acétyl ((Roth, Denu et al. 2001). Ce transfert permet de neutraliser les charges positives des lysines, ce qui permet de réduire la force des interactions entre les histones et l'ADN (Wolffe and Pruss 1996). La famille des HATs peut être divisée en trois groupes en se basant sur l'homologie de la séquence en acides aminés ou sur la structure conformationnelle. Ces trois groupes sont GNAT (Gcn5-related N-acetyltransferases), MYST et CPB (CREB Binding Protein)/p300 (pour revue (Hodawadekar and Marmorstein 2007))

Des expériences de ChIP chez plusieurs espèces ont permis de mettre en évidence la présence de ces histones acétylés au niveau des régions promotrices et des séquences codantes des gènes (Schubeler, MacAlpine et al. 2004; Bernstein, Kamal et al. 2005). En guise d'exemple, l'acétylation de H4K16 joue un rôle dans la décondensation de la chromatine (Shogren-Knaak, Ishii et al. 2006) alors que celle de H4K8 est importante pour la réplication (Doyon, Cayrou et al. 2006).

#### *3.4.2.2 Phosphorylation*

Tout comme c'était le cas pour l'acétylation des lysines, la phosphorylation des histones est un phénomène dynamique. Elle a lieu au niveau des résidus sérine, thréonine et tyrosine. Le niveau de phosphorylation est sous le contrôle de kinases et de phosphatases, qui vont respectivement ajouter ou enlever cette modification (Oki, Aihara et al. 2007). Les kinases vont transférer un groupe phosphate à partir de l'ATP, ce qui va ajouter des charges négatives aux histones. Les histones ainsi phosphorylés sont associés à une transcription activée. Ces phosphorylations sont également importantes pour la réparation ou l'apoptose. En guise d'exemple, la phosphorylation de la sérine 14 de H2B est importante pour l'induction de

l'apoptose (Cheung, Ajiro et al. 2003). La phosphorylation de la thréonine 11 de l'histone H3 est quant à elle importante dans la réparation, puisque cette thréonine est déphosphorylée au niveau des lésions de l'ADN (Shimada, Niida et al. 2008).

#### 3.4.2.3 Méthylation

La méthylation des histones se fait au niveau des résidus lysines et arginines. Contrairement à l'acétylation et à la phosphorylation, cette marque n'affecte en rien la charge de la protéine histone. Le degré de complexité de cette modification est néanmoins plus élevé puisque les lysines peuvent être mono-, di- ou triméthylés, et les arginines peuvent être mono- ou diméthylés. Les enzymes agissant sur les lysines sont des histones méthyltransférases (HMT) à domaine SET (Su(var)3-9, Enhancer of zest, Trithorax). Les arginines quant à elles sont méthylées par des arginines méthyltransférases (PRMT : Protein Arginine Methyltransferase) (pour revue (Zhang and Reinberg 2001)). Tout comme pour les autres marques décrites ci-dessus, la méthylation des histones est réversible, la déméthylation étant catalysée par des histones déméthylases (HDM).

La méthylation des histones joue un rôle complexe dans la régulation de la transcription. En effet, en fonction du résidu méthylé mais également du nombre de méthylations, la fonction peut être différente. Ainsi, H3K9me3 ou H3K27me3 sont des marques de répression, alors que H3K4me3, H3K36me3 et H3K79me3 sont des marques d'activation (pour revue (Martin and Zhang 2005)). De plus, alors que H3K4me3 est situé au niveau des promoteurs, H3K4me2 sera retrouvé dans la région codante, alors que H3K4me est enrichi à la fin de la phase codante ou dans les enhancers (Zhang, Bernatavichute et al. 2009).

Si la majorité des résidus ne contiennent qu'un seul type de modification, certains peuvent en contenir deux ; ainsi, comme cité précédemment, H3K27me3 est une marque de répression. L'acétylation de ce même résidu au niveau des enhancers est en revanche synonyme de transcription active (Wang, Zang et al. 2008).

# ***Chapitre 4***

*La régulation de la  
transcription par les  
récepteurs de l'acide  
rétinoïque*

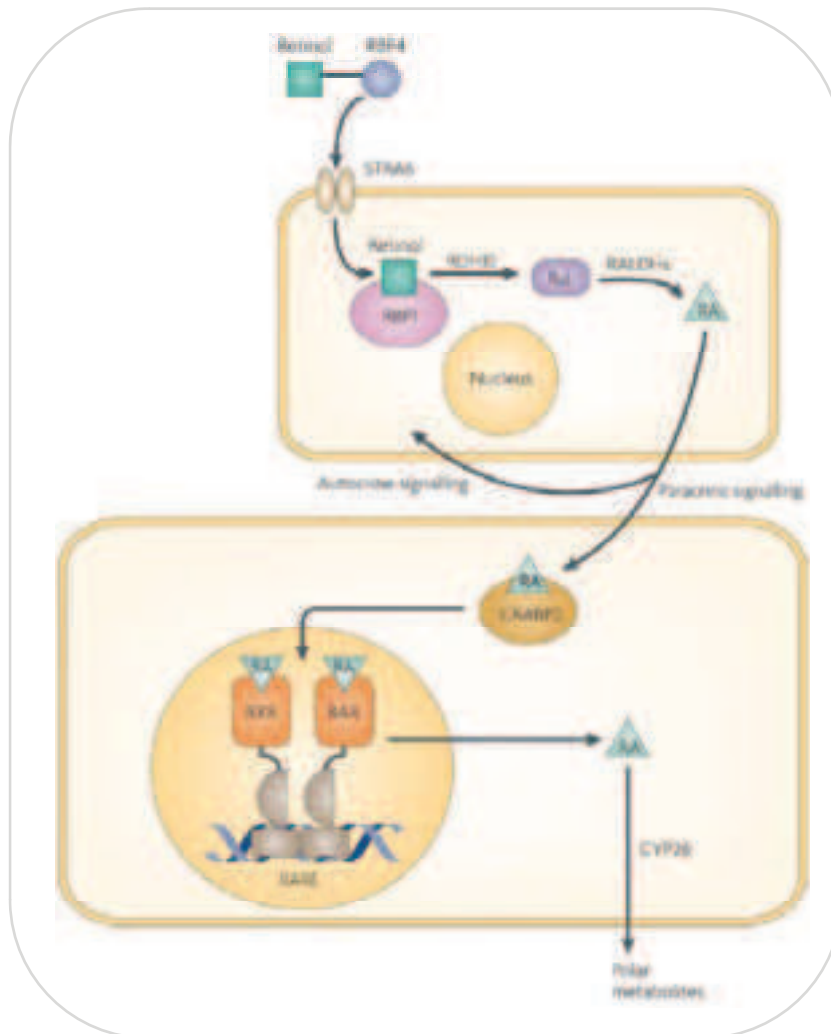
## 4 Chapitre IV : La régulation de la transcription par les récepteurs de l'AR

### 4.1 La voie de l'acide rétinoïque

L'AR tout-trans, le dérivé actif de la vitamine A (ou rétinol), joue des rôles multiples lors du développement embryonnaire et dans l'homéostasie de plusieurs tissus adultes (Bastien and Rochette-Egly 2004; Mark, Ghyselinck et al. 2006; Niederreither and Dollé 2008). La définition des axes embryonnaires, la régionalisation du SNC, le contrôle de la neurogenèse font partie des fonctions de l'AR. A la fin des années 1980, il a été démontré que l'AR était un ligand de plusieurs récepteurs nucléaires, et que de ce fait cette molécule pouvait agir directement sur le contrôle de l'activité transcriptionnelle de ses gènes cibles (pour revue (Niederreither and Dollé 2008)).

La voie de synthèse de l'acide rétinoïque ainsi que les enzymes impliqués ont été décrits durant la dernière décennie (Niederreither and Dollé 2008). Le rétinol, plus couramment connu sous le nom de vitamine A, est pris en charge par la protéine RBP4 (Retinol-Binding Protein 4), puis est importé dans la cellule grâce à STRA6 (Stimulated by Retinoic Acid gene 6), un récepteur membranaire reconnaissant RBP4 (*Figure 20*). Une fois dans le cytoplasme, deux étapes majeures ont lieu : dans un premier temps, le rétinol est transformé en rétinaldéhyde (Ral) par RDH10 (Rétinol DésHydrogénase 10) ou par l'Alcool DésHydrogénase 7 (ADH7). Ensuite, le rétinaldéhyde est transformé en AR par une RétinAlDéhyde DésHydrogénase (RALDH<sub>1-3</sub>). L'import de l'AR dans le noyau se fait par l'intermédiaire de la protéine CRABP2 (Cellular Retinoic-Acid Binding Protein). La fixation de l'AR au dimère RAR-RXR (Retinoid X Receptor, récepteur du rétinoïde X) permet de moduler l'expression des gènes cibles. Chacun de ces deux récepteurs existe sous la forme de trois isotypes différents ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) (Mangelsdorf and Evans 1995). L'action de l'acide rétinoïque via ces récepteurs peut se faire de manière autocrine ou paracrine (Maden 2007).

La vitamine A, et par conséquent son dérivé actif l'AR, sont indispensables au développement pré et post-natal ainsi qu'à la vie adulte, en assurant des fonctions dans la croissance, la survie ou encore la vision (Mark, Ghyselinck et al. 2006). Au cours de l'embryogenèse, l'AR régule la mise en place de l'axe antéropostérieur, agit sur la croissance et la mise en place du cerveau postérieur, est impliqué dans le développement des membres,



Maden, 2007

**Figure 20 : la voie de signalisation de l'acide rétinoïque.** Le rétinol est reconnu par RBP4, puis incorporé dans la cellule grâce à STRA6. Une cascade enzymatique permet la synthèse d'AR, qui agira de manière autocrine ou paracrine. Suite à l'import de l'AR dans le noyau par CRABP2, sa reconnaissance du dimère RAR-RXR permet de moduler l'expression génique. Dans le cytoplasme, l'AR est catabolisé par l'enzyme CYP26.

mais peut également réguler la morphogenèse et la différenciation d'organes tels que le cœur, les reins ou le pancréas (Niederreither and Dollé 2008). L'importance de l'AR au cours du développement a été mise en évidence à travers différentes expériences de génétique chez la souris. Comme décrit ci-dessus, l'AR agit par le biais des récepteurs RAR-RXR. La délétion des gènes codant pour ces récepteurs par recombinaison homologue ou encore la délétion de certains enzymes de la voie de synthèse de l'AR affecte les structures tissulaires ainsi que les organes de l'embryon.

Outre son rôle au cours de l'embryogenèse, l'AR peut également jouer un rôle dans les tissus adultes. Des expériences de mutation somatique spatio-temporelle des récepteurs RAR et RXR ont été réalisées dans divers tissus. Cette approche a permis de montrer par exemple que RAR $\gamma$  est nécessaire à la formation des corps lamellaires (vésicules présentes dans le cytoplasme des pneumocytes) ou encore que RXR $\alpha$  est nécessaire au cycle des follicules pileux et à l'homéostasie de l'épiderme (Ghyselinck, Chapellier et al. 2002). L'AR est également impliqué dans le maintien de l'état différencié des neurones adultes, ainsi que dans la régénération nerveuse (Mark, Ghyselinck et al. 2006; Maden 2007; Niederreither and Dollé 2008). L'inactivation de la voie de signalisation de l'AR a pour conséquence la dégénérescence des neurones moteurs, ainsi que le développement de maladies neurodégénératives, telles qu'Alzheimer ou Parkinson (Maden 2007). L'élaboration de traitements médicaux pour de telles dégénérescences nécessite dans un premier temps la compréhension des mécanismes moléculaires régis par l'AR.

## 4.2 Les récepteurs de l'acide rétinoïque

Les récepteurs de l'acide rétinoïque font partie de la grande famille des récepteurs nucléaires. Les petites molécules lipophiles (hormones stéroïdiennes, rétinoïdes, acides gras libres) jouent un rôle important dans la reproduction, le développement et le métabolisme chez les eucaryotes (Nagy and Schwabe 2004; Bain, Aaron F. Heneghan et al. 2007). Ces molécules agissent sur des récepteurs nucléaires (NRs), qui sont des facteurs de transcription activés par leur ligand, et permettent ainsi la régulation de l'expression de réseaux géniques complexes (Bain, Aaron F. Heneghan et al. 2007). Ces récepteurs forment une famille très diversifiée par rapport à leur rôle physiologique (Nagy and Schwabe 2004). La superfamille des NRs peut être divisée en trois classes : les récepteurs stéroïdiens (Classe I), les récepteurs thyroïdiens et rétinoïdiens (Classe II) et les récepteurs orphelins (Classe III) (Bain, Aaron F. Heneghan et al. 2007). Malgré la division en 3 classes, les NRs présentent des domaines

conservés, dont le plus conservé est le domaine central de liaison à l'ADN (DBD) (Nagy and Schwabe 2004).

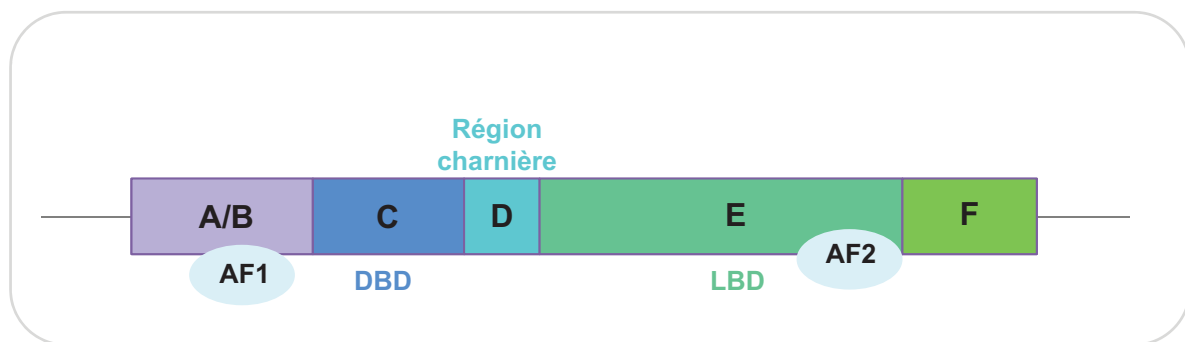
#### 4.2.1 *Structure des récepteurs à l'acide rétinoïque*

Les récepteurs de l'acide rétinoïque (RAR) ont une organisation structurale similaire à celle retrouvée pour l'ensemble des récepteurs nucléaires. Cette structure modulaire est composée de six régions notées de A à F (*Figure 21*) (Gronemeyer and Laudet 1995; Chambon 1996). Ce sont les régions C et E qui contiennent respectivement les domaines de liaison à l'ADN et au ligand (LBD, Ligand Binding Domain). Ces deux domaines étant particulièrement conservés et par conséquent importants dans la fonctionnalité du récepteur, je les détaillerai ultérieurement.

La région A/B est située du côté N-terminal de la protéine. Cette région peut contenir ou non un domaine de transactivation AF-1 (Activation function 1), et est également connue sous le nom de NTD. Cette région est importante dans le contrôle de la transcription des gènes cibles de l'AR (Nagpal, Saunders et al. 1992; Nagpal, Friant et al. 1993). Alors que la région A est très variable entre les sous-types et les isoformes de RAR, la région B est quant à elle plutôt bien conservée.

La région D très faiblement conservée est considérée comme une région charnière entre le LBD et le DBD. Cette région permet une rotation du DBD, ce qui permettrait à l'ensemble DBD-LBD d'adopter différentes conformations tout en évitant les problèmes d'encombrement stérique. Cette région contient également des signaux de localisation nucléaire.

La région F, qui représente l'extrémité C-terminale du récepteur, est présente chez RAR mais absente chez RXR. Cette région est hautement variable en longueur et en séquence entre les différents RAR. Les fonctions de cette région sont à ce jour peu comprises. Il semblerait néanmoins qu'en absence de ligand, cette région permette de stabiliser l'hélice 12 de RAR $\alpha$  et ainsi d'inhiber le recrutement de coactivateurs et de favoriser celui de corépresseurs (Farboud and Privalsky 2004). Cette région comporte également différents sites de phosphorylation qui pourraient être importants pour les propriétés des RAR (Bastien, Adam-Stitah et al. 2000).



**Figure 21 : Représentation schématique de la structure d'un récepteur nucléaire.** Le récepteur contient deux domaines structuraux très bien caractérisés et conservés qui sont le site de fixation à l'ADN (DBD) et le site de fixation du ligand (LBD). La partie Nter du récepteur contenant la fonction AF-1 est quant à elle moins conservée entre les différentes classes de récepteurs. Une deuxième fonction AF est retrouvée dans le domaine E.

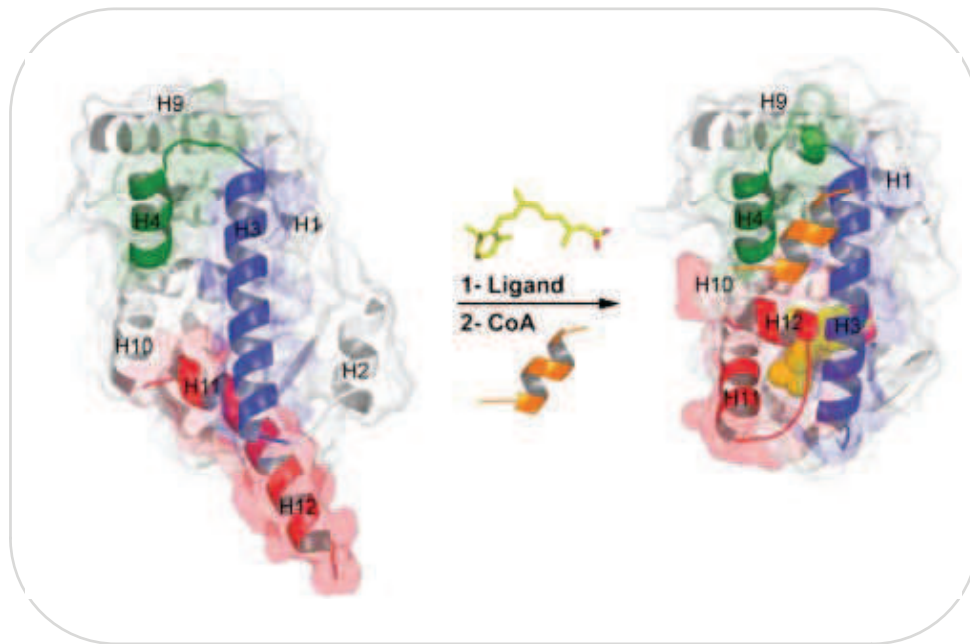
#### 4.2.2 Le LBD

Depuis 1995, date de caractérisation de la première structure cristallographique du LBD de RXR, de nombreuses autres structures de LBD ont été déterminées par cristallographie ou par RMN. Sans exception, toutes les structures étudiées présentent une organisation similaire. Ainsi, le LBD est un domaine protéique globulaire comportant douze hélices  $\alpha$  (Figure 22) (Bourguet, Vivat et al. 2000; Perez, Bourguet et al. 2012). Du point de vue de l'organisation tridimensionnelle, les hélices forment trois feuillets antiparallèles. La partie inférieure du domaine présente une cavité hydrophobe dans laquelle les ligands peuvent se fixer. Un feuillet  $\beta$  à deux brins et l'hélice carboxy-terminale du récepteur (hélice 12) se situent de part et d'autre de cette cavité (Bourguet, Vivat et al. 2000; Nagy and Schwabe 2004).

L'étude globale de la superfamille des NRs montre que la taille de la poche de fixation du ligand varie et que cette variation de taille reflète la fonction du récepteur. Pour les récepteurs stéroïdiens qui reconnaissent uniquement une famille précise de petites molécules, la poche est plus petite et tapissée d'acides aminés polaires permettant de faire des liaisons hydrogènes spécifiques avec leurs ligands naturels. En revanche, les récepteurs de classe III interagissent avec des ligands très variés et ont donc tendance à avoir une poche plus importante qui peut accueillir des ligands de structures différentes (Bain, Heneghan et al. 2007).

#### 4.2.3 Le DBD

La première structure d'un DBD de la famille des NRs lié à l'ADN a été déterminée pour le récepteur des glucocorticoïdes (GR). Elle sert, aujourd'hui encore, de modèle pour l'ensemble de la famille des NRs. Le domaine de liaison à l'ADN se replie en une structure globulaire faisant intervenir deux doigts de zinc, chaque atome de zinc étant coordonné par quatre résidus cystéine. La mutation de ces cystéines a montré que les atomes de zinc sont nécessaires au maintien du repliement de la protéine et qu'ils sont par conséquent importants pour l'interaction avec l'ADN. Outre ces deux doigts de zinc, la structure contient deux hélices  $\alpha$  : l'hélice en N-terminal interagit avec le sillon majeur de l'ADN, elle va donc conférer à la protéine sa spécificité de reconnaissance, alors que l'hélice en C-terminal, qui se situe spatialement au-dessus et perpendiculaire à la première hélice, contribue à la stabilité de la structure (Bain, Heneghan et al. 2007).



*Perez et al 2012*

**Figure 22 : Représentation du LBD du récepteur RXR.** Suite à la fixation de leur ligand, les récepteurs nucléaires changent la conformation de leur LBD. Le LBD est composé de trois feuillets d'hélices  $\alpha$  . La fixation du ligand au niveau d'une cavité non polaire dans la partie inférieure du récepteur induit un remaniement du LBD. L'hélice H12 (en rouge) va venir fermer la cavité contenant le ligand ce qui entraîne un changement conformationnel du LBD et lui permet d'interagir avec des coactivateurs.

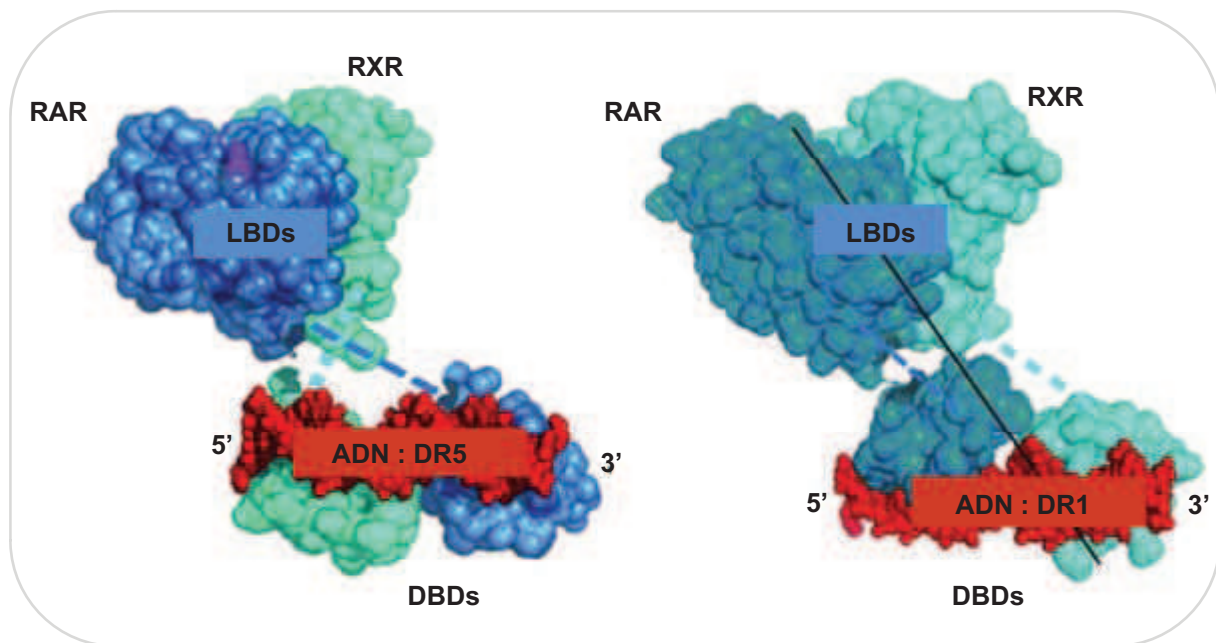
### 4.3 Reconnaissance de l'ADN

Les récepteurs nucléaires reconnaissent une séquence d'ADN hexanucléotidique de séquence consensus est 5'-RGKTCA-3'. Or, la spécificité de reconnaissance ne dépend pas uniquement de la séquence primaire, mais également de la géométrie que celle-ci adopte (Khorasanizadeh and Rastinejad 2001). En effet, la plupart des NRs se fixent à l'ADN sous forme d'homodimères (classe I) ou d'hétérodimères avec RXR (classe II) via un site de reconnaissance constitué de deux séquences hexanucléotidiques (Bain, Heneghan et al. 2007). Les NRs de classe I reconnaissent une organisation palindromique des sites, par exemple le récepteur des œstrogènes se lie aux ERE (Estrogen Response Elements) de structure 5'-AGGTCA-N<sub>3</sub>-TGACCT-3', tandis que les récepteurs de classe II reconnaissent des répétitions directes (DRs) espacées de 1 à 5 nucléotides. Néanmoins, il existe des exemples de sites formés par des DR avec d'autres espacements ou par des répétitions inversées (IR), par exemple des IR0 (5'-GGGTCATGACCC-3') pour l'hétérodimère RAR/RXR (Khorasanizadeh and Rastinejad 2001). Au fil des études, il a été confirmé que l'hétérodimère RAR/RXR pouvait reconnaître une grande variété d'éléments de réponse (Delacroix, Moutier et al. 2010). De plus, cet espacement entre les éléments de répétition est directement lié à la conformation que va adopter le récepteur au niveau de l'ADN. Ainsi, au niveau d'un DR5, RXR occupe l'hémisite situé en 5', alors qu'au niveau d'un DR1, c'est le DBD de RAR qui occupe cet hémisite (*Figure 23*) (Rochel, Ciesielski et al. 2011). L'espacement entre les éléments de réponse peut ainsi être directement relié au rôle du récepteur, puisque la conformation du LBD sera différente.

L'ensemble des informations et descriptions relatives à la fixation des récepteurs nucléaires, plus spécifiquement du dimère RAR/RXR au niveau de l'ADN, est décrit dans le chapitre de livre figurant en annexe (How the RAR-RXR heterodimer recognizes the genome – *Urban Sylvia, Tao Ye and Irwin Davidson*)

### 4.4 Modèle d'activation des gènes cibles de l'AR

Le modèle de fonctionnement des récepteurs nucléaires appartenant à la classe II indique qu'en absence de ligand, les NRs sont fixés sur l'ADN et recrutent des corépresseurs de la transcription. L'arrivée du ligand entraînera un changement conformationnel du LBD, qui a pour conséquence la séparation des corépresseurs et le recrutement des coactivateurs. Ce modèle a néanmoins été remis en question par des études récentes qui montrent que les RARs



D'après Rochel and al., 2011

**Figure 23 : Organisation différentielle du dimère RAR-RXR en fonction de l'espacement des hémisites.** En fonction qu'il reconnaisse un DR5 ou un DR1, l'organisation structurale du dimère RAR-RXR va changer. Au niveau d'un DR1, le LBD et le DBD de RAR seront en 5'. Au niveau d'un DR5, la flexibilité des régions charnières (en pointillés) permet l'inversement des DBD sur l'ADN, tout en conservant le LBD de RAR du côté 5' de l'élément de réponse.

peuvent être présents au niveau de la membrane cytoplasmique (Rochette-Egly and Germain 2009) ou tout simplement dans le noyau mais dissocié de l'ADN (Bruck, Vitoux et al. 2009).

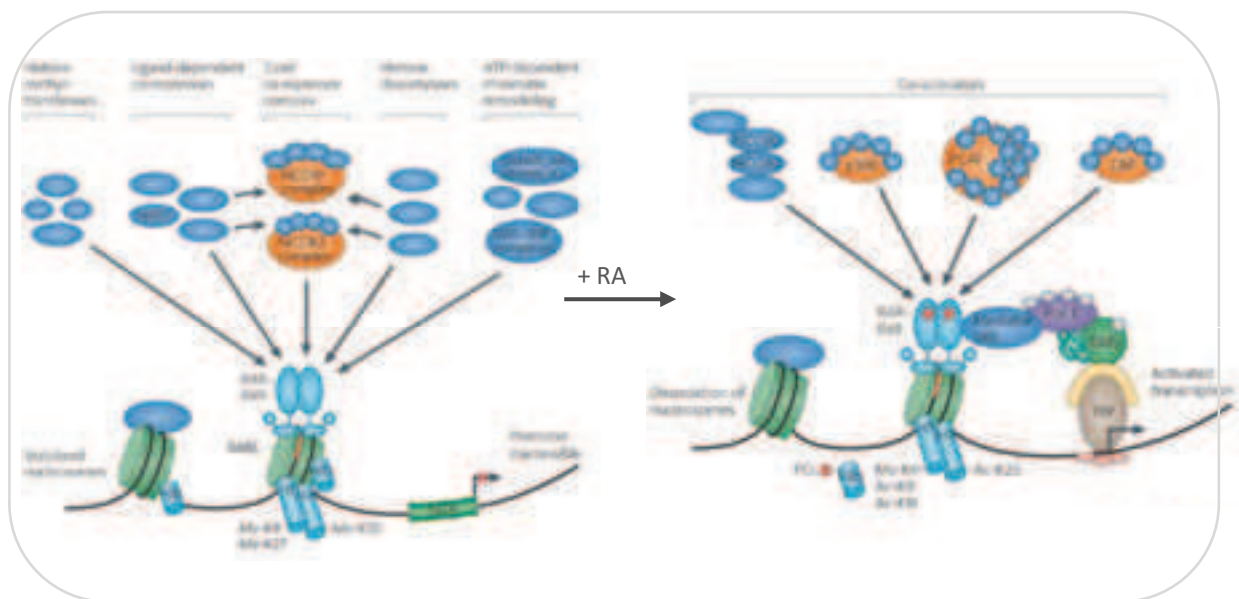
L'action des récepteurs nucléaires se fait grâce au recrutement de protéines régulatrices, regroupées sous les termes de coactivateur et corépresseur selon leurs fonctions (Perissi and Rosenfeld 2005). De manière générale, les corépresseurs se fixent au niveau de l'hélice 3 du récepteur en absence du ligand, grâce à leur séquence LxxxIxxx(I/L) (*Figure 24*). Deux corépresseurs principaux ont été caractérisés : NCoR1 (Nuclear-Receptor Corepressor 1) et NCoR2, aussi connu sous le nom de SMRT (Silencing Mediator for Retinoid and Thyroid hormone receptors). Tous deux font partie de complexes multiprotéiques présentant des activités enzymatiques responsables de la répression transcriptionnelle (*Figure 24*). Ces interactions permettent d'expliquer la répression des gènes cibles des NRs de classe II en absence du ligand.

La fixation du ligand entraîne un changement conformationnel de l'hélice 12 qui est à l'origine de la libération des corépresseurs et du recrutement des coactivateurs (Nagy and Schwabe 2004). Plus de 300 coactivateurs des NRs ont été identifiés ([www.nursa.org](http://www.nursa.org)). Ces molécules contiennent souvent le motif LxxLL qui interagit avec le LBD du récepteur (Nagy and Schwabe 2004). Comme les corépresseurs, les coactivateurs sont des complexes multiprotéiques contenant les activités enzymatiques nécessaires à l'activation transcriptionnelle, notamment les modifications covalentes des queues des histones par les HATs et les HMTs. Par ailleurs, il existe des corépresseurs qui sont recrutés par les NRs en présence du ligand d'une manière semblable aux coactivateurs. RIP140, LCOR, PRAME et TIF1a font partie de ces corépresseurs (Rochette-Egly and Germain 2009).

La figure 24 montre que le passage de la répression à l'activation de la transcription résulte uniquement de la fixation du ligand. En réalité, ceci est une vue très simplifiée, puisque le processus d'activation résulte d'une série d'interactions dynamiques et séquentielles avec une multiplicité de coactivateurs. Ces interactions peuvent varier selon le type de promoteur et le type cellulaire (Perissi and Rosenfeld 2005).

#### **4.5 L'acide rétinoïque dans la différenciation neuronale**

Lors de la différenciation neuronale des cellules ES (voir chapitre 2), le programme d'expression des gènes va varier suite au traitement par l'AR. Des gènes cibles directes des RARs vont être très rapidement induits suite au traitement par l'AR. Parmi ces gènes se



Niederreither et Dollé, 2008

**Figure 24 : Mécanisme de régulation de la transcription par les récepteurs nucléaires de classe II.**

Le mécanisme de régulation de la transcription est relativement similaire pour l'ensemble de la famille de NRs. En absence de ligand, des co-répresseurs (NCoR1 et NCoR2) formant des complexes avec des HDACs ou des HMTs sont recrutés, ce qui stabilise la structure compacte de la chromatine. La fixation du ligand, dans ce cas l'acide rétinoïque, entraîne le départ des co-répresseurs et le recrutement de complexes multiprotéiques co-activateurs. Dans le cas du dimère RAR-RXR, la phosphorylation des récepteurs est nécessaire à leur activité.

trouvent notamment *Rarb2*, des gènes impliqués dans le métabolisme de l'AR, tel que les gènes *Cyp26a1*, *Cyp26b1*, *Cyp26c1*. Des gènes impliqués dans la mise en place de l'axe antéro-postérieur tel que les gènes *Hox* (*Hoxa1*, *Hoxa3*, *Hoxa5*, *Hoxb1*, *Hoxb4*) font également parti des gènes rapidement induits, tout comme d'autres gènes important dans l'initiation de la neurogenèse que sont *Meis2*, *Gbx2* ou *Dlx3* (Moutier, Ye et al. 2012).

En plus de ces gènes qui représentent la réponse précoce à l'AR, un deuxième groupe de gènes est induit plus tardivement. Ces gènes codent notamment pour des facteurs de transcription tel que les facteurs *Irx* (*Irx1*, -3, -5), les facteurs de la famille *Hes* (*Hes3*, *Hes5*), de la famille *Zic* (*Zic1* et *Zic4*) ou encore *Brn2*. Ces gènes ne représentent pas des cibles directes des RAR, et sont donc induits de manière secondaire et indirecte suite au traitement pas l'AR. L'induction de ces gènes lors de la différenciation neuronale indique leur potentiel rôle dans ce processus.

# ***Chapitre 5***

*Zic1 et Brn2,*

*deux facteurs clés de la  
différenciation neuronale*

## 5 Chapitre V : Zic1 et Brn2, deux facteurs clés de la différenciation neuronale ?

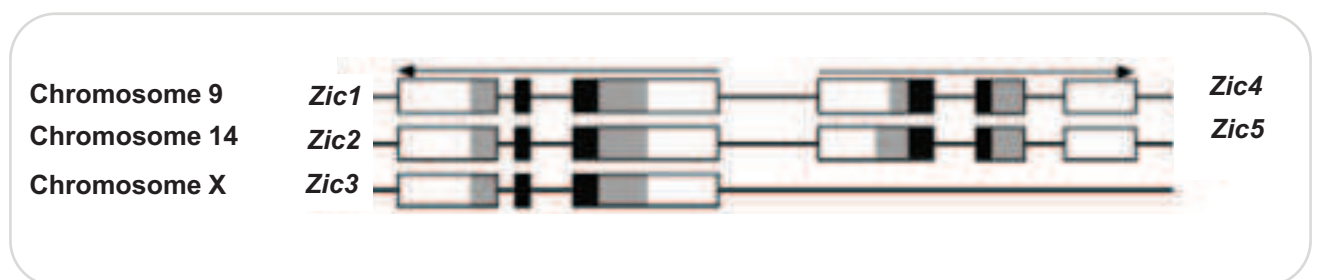
### 5.1 Zic1

La famille des facteurs de transcription Zic (Zinc finger of cerebellum) contient cinq membres (Zic1-5) chez l'homme, la souris ou le zénope (sept membres chez le poisson zèbre), chacun comportant cinq domaines doigt de Zinc (C2H2). Le premier doigt de Zinc est spécifique de chaque membre de la famille. Les doigts de Zinc 2-5 sont très bien conservés entre les différents membres de la famille, avec plus de 85% d'homologie. Les cinq membres de la famille sont hautement conservés de la souris à l'homme et des phénotypes associés à la mutation de l'ensemble des membres ont été caractérisés dans au moins une des espèces (Grinberg and Millen 2005).

Durant le développement embryonnaire, les gènes *Zic* sont exprimés dans le système nerveux dorsal ainsi que dans les somites (pour revue (Merzdorf 2007)). L'expression des différents membres de la famille se superpose, ce qui laisse présager, vu la forte homologie, que la mutation d'un membre de cette famille puisse être compensée par l'expression des autres membres. Alors qu'un problème au niveau du tube neural est visible quelque soit le membre de la famille Zic muté, il existe aussi des phénotypes spécifiques à chaque membre. Ainsi, la perte de fonction de *Zic2* produit une holoprosencéphalie (malformation du prosencéphale), la mutation de *Zic3* entraîne des malformations cardiaques congénitales, et celle de *Zic5* des problèmes au niveau de la crête neurale (Grinberg and Millen 2005).

#### 5.1.1 Structure des gènes *Zic*

Les cinq gènes membres de la famille *Zic*, chez l'humain ou chez la souris, sont regroupés au niveau de trois loci génomiques. *Zic1* et *Zic4*, ainsi que *Zic2* et *Zic5* sont présents sous forme de tandem, respectivement au niveau des chromosomes 9 et 14 chez la souris (chromosomes 3 et 13 chez l'homme). En revanche, *Zic3* est présent sur le chromosome X dans les deux espèces (*Figure 25*). Les divergences entre les membres de la famille ZIC sont localisées au niveau du premier doigt de Zinc. Ce premier doigt de Zinc est relativement bien conservé entre les membres *Zic1-3*, mais est plus divergent pour *Zic4* et *Zic5*. Cette divergence permet de diviser les membres de la famille Zic en deux classes, division qui est confirmée par la présence d'un domaine « ZOC » ( *Zic1-3 odd-paired conserved*) uniquement



*D'après Merzdorf 2007*

**Figure 25 : Organisation chromosomale des cinq membres de la famille *Zic* chez la souris.**

pour les trois premiers membres de la famille Zic. L'ensemble des membres de la famille contient également un domaine conservé en amont des 5 doigts de zinc (ZF-NC : Zinc finger N-flanking region). Seul ZIC2 présente une région poly-alanine. L'arrangement génomique des membres de la famille ZIC ainsi que la conservation de la séquence et de l'organisation protéique suggèrent qu'un seul gène ancestral a subi une série de duplications pour obtenir les cinq gènes existants chez les mammifères (pour revue (Aruga, Kamiya et al. 2006)).

### 5.1.2 Expression et régulation des gènes Zic

La présence de gènes sous forme de tandem dans le génome est souvent associée à une co-expression des deux gènes du tandem. *Zic2* et *Zic5*, tout comme *Zic3* sont exprimés de manière concomitante au début de l'embryogénèse chez la souris : tous trois sont exprimés dans l'ectoderme et le mésoderme au stade gastrula (Furushima, Murata et al. 2000; Elms, Scurry et al. 2004). Au cours de l'organogenèse, ces gènes sont exprimés dans des régions spécifiques du neur ectoderme et du SNC en développement, ainsi que des somites. De même, l'expression de *Zic1* et *Zic4* apparaît également au début de l'organogenèse, avec une transcription dans des régions spécifiques telles que le prosencéphale et les somites ((Elms, Scurry et al. 2004; Gaston-Massuet, Henderson et al. 2005)). Peu de données sont connues quant à la régulation de Zic, notamment *in vivo*. Aucune protéine régulant les gènes *Zic* n'a été mise en évidence *in vivo*, même si des données dans les cellules ES laissent présager que l'expression de *Zic3* est sous le contrôle de NANOG (Lim, Loh et al. 2007) et inversement, que *Zic3* permet d'activer le promoteur de NANOG (Lim, Hong et al. 2010). Des expériences dans d'autres organismes indiquent que différentes voies de signalisation et de multiples facteurs de transcription permettent de réguler l'expression des gènes *Zic* (pour revue (Merzdorf 2007)).

### 5.1.3 Fonctions biologiques

Comme déjà mentionné, des mutations de perte de fonction des gènes *ZIC* chez l'homme sont à l'origine d'une grande variété de défauts congénitaux (pour revue (Merzdorf 2007)). Des expériences chez la souris montrent que très souvent les phénotypes retrouvés chez cet organisme modèle suite à une mutation des gènes *Zic* sont similaires aux problèmes observés chez l'humain. Des mutations de *Zic1* et *Zic4*, isolées ou doubles, ont pour conséquence l'apparition de défauts au niveau du cerve l l um, qui sont à rapprocher des malformations Dandy-Walker observées chez l'homme (Grinberg, Northrup et al. 2004). Chez l'homme, une holoprosencéphalie est observée lorsqu'une mutation est présente au niveau de *Zic2*. De

manière similaire chez la souris, une perte de fonction de *Zic2* entraîne des problèmes au stade gastrula qui vont résulter en un défaut de formation du prosencéphale (Warr, Powles-Glover et al. 2008).

L'observation des défauts congénitaux ou plus généralement des phénotypes associés à la perte de fonction des facteurs *Zic* permet de déterminer les processus biologiques qui nécessitent l'expression de *Zic*, mais ne révèle pas le mécanisme moléculaire d'action des facteurs *Zic*. Sont-ils de simples cofacteurs ou sont-ils des facteurs de transcription se fixant directement à l'ADN ? Il semblerait que les deux cas de figure existent. Des expériences de ChIP-chip dans les cellules ES ont permis d'identifier un motif consensus de fixation pour *Zic3* (5'-CC(C)/(T) GCTGGG-3') et ont mis en évidence que *Nanog* était la cible directe ce facteur (Lim, Hong et al. 2010). Ceci qui démontre que *Zic3* peut agir comme un facteur de transcription en se liant à l'ADN. D'autres expériences montrent que les protéines *Zic* peuvent agir avec les facteurs de transcription TCF pour réprimer la voie de signalisation Wnt, sans interaction directe des protéines *Zic* avec l'ADN (Pourebahim, Houtmeyers et al. 2011; Fujimi, Hatayama et al. 2012). Dans ce cas les protéines *Zic* adoptent un rôle de corépresseur.

#### 5.1.4 Les particularités de *ZIC1*

La majorité des études de *Zic1* ont été réalisées chez le xénope. *Zic1* code pour un facteur de transcription exprimé dans la plaque neurale présomptive. Son domaine d'expression se restreindra plus tard à la partie dorsale du tube neural et à la crête neurale (Kuo, Patel et al. 1998; Nakata, Nagai et al. 1998). *Zic1* joue un rôle important dans la structuration de la plaque neurale et est important pour la formation de la crête neurale, des somites et du cervelet (Cornish, Hassan et al. 2009). Chez le xénope, *Zic1* est l'un des premiers marqueurs moléculaires de la détermination neurale à être exprimé (Tropepe, Li et al. 2006). L'activation de l'expression de *Zic1* induit l'expression de marqueurs de la crête neurale et de la partie dorsale du tube neural à la fois dans la calotte animale mais également pour les embryons entiers. Au niveau de l'ectoderme plus ventral, *Zic1* induit la formation d'agrégats cellulaires similaires aux précurseurs de la crête neurale et supprime le destin ventral de ces cellules (Kuo, Patel et al. 1998).

Au niveau du neurectoderme, Msx1 induit l'expression de Zic1 et de Pax3, puis ces deux facteurs de transcription vont activer Slug de manière dépendante de Wnt et Foxd3 (Sasai, Mizuseki et al. 2001; Monsoro-Burq, Wang et al. 2005; Sato, Sasai et al. 2005).

De nombreuses études portent donc sur le rôle de Zic1 au niveau de la crête neurale mais l'expression de Zic1 dans la partie dorsale du tube neural et la conséquence de cette expression sur la différenciation reste plus floue. Chez la souris, les mutants homozygotes pour Zic1 présentent une hypoplasie cérébelleuse et l'absence d'un lobule au niveau du lobe antérieur (Aruga, Minowa et al. 1998). Ceci montre l'importance de Zic1 au niveau du développement du cerveau.

## 5.2 Brn2

BRN2 fait partie de la famille des facteurs de transcription à domaine POU. Les membres de cette famille peuvent être divisés en six classes et sont les régulateurs du destin cellulaire de nombreuses lignées. En fonction du membre de la famille, l'expression peut être ubiquitaire, ou réduite à un type cellulaire spécifique. (*Table 3*) (pour revue (Cook and Sturm 2008)).

Les membres de la famille POU sont impliqués dans des fonctions biologiques diverses et variées. En guise d'exemple, OCT4 intervient dans la programmation des cellules souches embryonnaires, alors que OCT1 et OCT2 sont impliqués dans la réponse immunitaire (Scholer 1991; Ryan and Rosenfeld 1997). Malgré l'importance des fonctions biologiques de ces facteurs, leur nombre n'est pas très élevé chez les différentes espèces : quinze, cinq et quatre facteurs POU chez l'homme, la drosophile et le nématode respectivement (Venter, Adams et al. 2001). La diversité du nombre de fonctions assurées par ces facteurs peut alors s'expliquer par des contrôles à différents niveaux, tel que la fixation au niveau de l'ADN, les modifications post-traductionnelles ou encore l'interaction avec d'autres facteurs de transcription. Je m'intéresserai dans la partie suivante à la fonction de mon facteur de transcription d'intérêt, Brn2, puis je détaillerai la structure des membres de la famille POU. Cette structure peut en effet être directement reliée à la fonction de ces facteurs de transcription.

**Table 3 : Les différents membres de la famille POU  
et leur domaine d'expression**

| Classe | Gène                 | Protéine    | Expression                           |
|--------|----------------------|-------------|--------------------------------------|
| POU1F  | <i>Pou1f1</i>        | Pit1        | Glande pituitaire<br>(hypophyse)     |
| POU2F  | <i>Pou2f1</i>        | Oct1        | Ubiquitaire                          |
|        | <i>Pou2f2</i>        | Oct2        | Lymphoïde                            |
|        | <i>Pou2f3</i>        | Epoc1       | Peau                                 |
| POU3F  | <i>Pou3f3</i>        | Brn1        | Système nerveux<br>Reins             |
|        | <b><i>Pou3f2</i></b> | <b>Brn2</b> | <b>Système nerveux</b>               |
|        | <i>Pou3f4</i>        | Brn4        | Tube neural                          |
|        | <i>Pou3f6</i>        | Oct6        | Blastocyste,<br>Cellules ES, Cerveau |
| POU4F  | <i>Pou4f1</i>        | Brn3a       | Système visuel                       |
|        | <i>Pou4f2</i>        | Brn3b       | Système visuel                       |
|        | <i>Pou4f3</i>        | Brn3c       | Système auditif                      |
| POU5F  | <i>Pou5f1</i>        | Oct4        | Cellules ES, Ovocytes                |
|        | <i>Pou5f2</i>        | Sprm1       | Sperme                               |
| POU6F  | <i>Pou6f1</i>        | Brn5        | Système nerveux                      |
|        | <i>Pou6f2</i>        | Rpf1        | Divers                               |

*D'après Cook et Sturm 2008*

### 5.2.1 Fonction du facteur Brn2

#### 5.2.1.1 Brn2 dans la différenciation neuronale *in vivo*

L'importance des facteurs de la famille POU dans la différenciation neuronale est mise en évidence phénotypiquement dans les souris présentant un KO pour ces facteurs. Brn1 et Brn2 sont exprimés dans le cerveau de l'embryon, notamment dans le néocortex (He, Treacy et al. 1989). Néanmoins, le KO individuel de chaque facteur présente un phénotype uniquement dans des régions très limitées du cerveau. Dans les mutants de Brn2, une perte de neurones est observée au niveau de l'hypothalamus, région dans laquelle Brn1 n'est pas exprimé (Nakai, Kawano et al. 1995; Schonemann, Ryan et al. 1995). A l'inverse, dans les mutants de Brn1, c'est la morphologie au niveau de l'hippocampe qui est altérée, or Brn2 est à peine détectable dans cette région. Au niveau du néocortex, où les deux membres de la famille sont présents, aucun phénotype n'est observé chez les souris où un seul gène est muté. Ceci suggère une redondance de fonction entre ces deux facteurs confirmée par le double mutant. Contrairement à ce qui est observé pour les simples mutants, le phénotype visible dans ces doubles mutants est drastique. Les doubles mutants homozygotes pour Brn1 et Brn2 meurent dans l'heure après la naissance. Les malformations cérébrales chez le double mutant sont sévères : le bulbe olfactif présente une hypoplasie, le cervelet est moins plissé et les cellules de Purkinje y sont plus denses. Le néocortex quant à lui, est beaucoup moins épais et son plissement paraît totalement désorganisé (Sugitani, Nakai et al. 2002).

L'un des défauts majeurs observés dans les souris mutantes pour Brn1 et Brn2 se situe au niveau de la mise en place des différentes couches du cortex. Il existe deux voies de signalisation importantes dans la migration radiale des neurones corticaux. La première voie fait intervenir la protéine Reelin, exprimé par les cellules Cajal-Retzius situé au niveau de la zone marginale (pour revue (Rice and Curran 2001)). La deuxième voie est médiée par Cdk5 et ses sous unités régulatrices Cdk5r1 et Cdk5r2 (mieux connues sous le nom de p35 et p39 respectivement) (pour revue (Smith, Greer et al. 2001)). Des expériences réalisées en 2002 mettent en évidence les problèmes de structuration du cortex dans les doubles mutants Brn1/Brn2 et suggèrent que les deux voies citées ci-dessus pourraient être à l'origine de ce phénotype (McEvelly, de Diaz et al. 2002; Sugitani, Nakai et al. 2002). De plus, des études ont montré une interaction entre la voie Reelin et la voie Notch précédemment décrite (Hashimoto-Torii, Torii et al. 2008). L'importance de cette voie Notch dans la mise en place

du cortex a été démontrée par Dominguez et al. qui ont mis en évidence que les membres de la famille Pou3f pouvaient agir directement sur Hes5. (Dominguez, Ayoub et al. 2013).

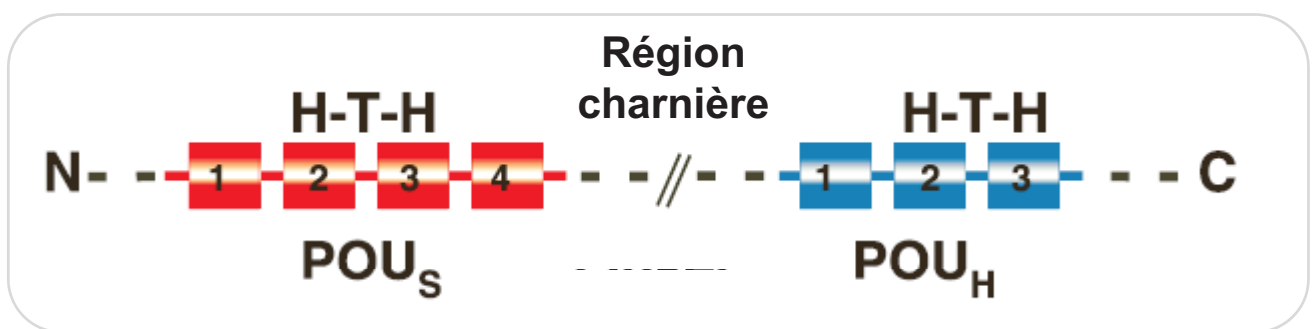
#### 5.2.1.2 *Brn2 dans la différenciation neuronale in vitro*

Les premières expériences démontrant l'importance de Brn2 dans la différenciation neuronale *in vitro* ont été réalisées dans des cellules de carcinome embryonnaire murin P19. Ces cellules peuvent, tout comme les ES, se différencier en neurones suite au traitement par l'AR. En utilisant ce système de différenciation, les auteurs ont montré que l'inactivation de Brn2 empêchait les cellules de se différencier en neurones (Fujii and Hamada 1993).

L'importance de Brn2 dans la neurogenèse *in vitro* a depuis été confirmée. En 2013 Lodato & al suggèrent qu'une surexpression de Brn2 permettrait de promouvoir la différenciation neuronale (Lodato, Ng et al. 2013). En 2010, les expériences de transdifférenciation identifiaient également ce facteur pour la conversion directe d'un fibroblaste en neurone (Vierbuchen, Ostermeier et al. 2010; Wapinski, Vierbuchen et al. 2013).

#### 5.2.2 *Structure de Brn2*

Les facteurs de transcription de la famille POU ont été regroupés dans une même famille grâce à leur domaine POU, hautement conservé entre les différents membres et qui leurs permet de reconnaître l'ADN (Herr, Sturm et al. 1988; Sturm and Herr 1988). Ce domaine est en réalité divisé en deux unités de fixation de l'ADN, POU<sub>S</sub> et POU<sub>H</sub>, qui sont reliées entre elles par un bras flexible variable (*Figure 26*) (Andersen and Rosenfeld 2001). Le domaine POU<sub>S</sub>, pour « POU spécifique » est situé en N-terminal et se compose d'environ 75 acides aminés. Le domaine en C-terminal, appelé POU homéo (POU<sub>H</sub>), est composé de 60 acides aminés (pour revue (Ryan and Rosenfeld 1997)). Chacun de ces deux domaines forme une structure en hélice-tour-hélice (HTH). Chez les mammifères, il existe six classes de protéines POU : POU1F à POU6F. Ces classes ont été établies en fonction de la séquence en acide aminés du domaine POU, mais également en fonction de la taille de la région charnière (Ryan and Rosenfeld 1997). La structure modulaire formée par les deux sous-domaines POU et la région charnière permet aux protéines de la famille POU de reconnaître une grande variété de motifs ADN ainsi que différents transactivateurs, tout cela dépendant notamment du positionnement des sites fixés par POU<sub>S</sub> et POU<sub>H</sub> (Scully, Jacobson et al. 2000; Remenyi, Tomilin et al. 2001).



*D'après Cook et Sturm 2008*

**Figure 26 : Organisation structurale des protéines à domaine POU.**

### 5.2.3 Reconnaissance de l'ADN

Une première étude a mis en évidence la liaison spécifique du domaine POU avec une région octamérique de l'ADN de type 5'-ATGCAAAT-3'. La région charnière hautement flexible entre les deux sous-domaines POU permet cependant une réorientation de ces derniers, ce qui leur autorise la reconnaissance de régions octamériques divergentes (Herr and Cleary 1995). Des études cristallographiques de la protéine OCT1 liée à la séquence octamérique ont montré que les deux domaines POU<sub>S</sub> et POU<sub>H</sub>, se fixaient chacun au niveau du sillon majeur de l'ADN, mais sur les faces opposées de l'hélice d'ADN. De ce fait, l'octamère peut être divisé, avec POU<sub>S</sub> qui se fixe à la séquence 5'-ATGC et POU<sub>H</sub> qui se fixe à 3'-AAAT (Klemm, Rould et al. 1994). Cette étude cristallographique n'a pas permis de visualiser la région charnière, confirmant que cette région n'est pas rigide, dans ce cas. Il semblerait que dans le cas des protéines de la classe POU3F, dont fait partie BRN2, cette région peut se replier en une petite hélice  $\alpha$  (Blaud, Vossen et al. 2004).

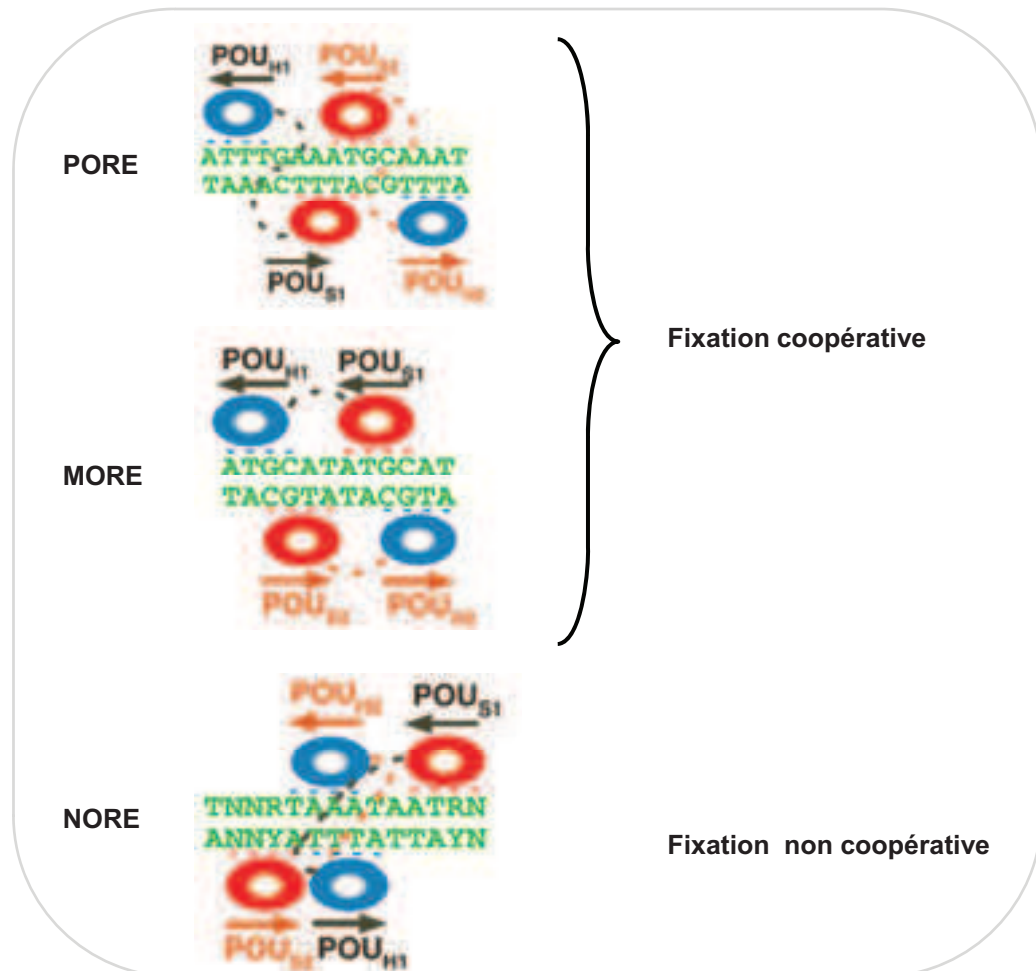
La région charnière qui relie les deux sous-domaines POU va donc permettre une grande flexibilité/variabilité au niveau de la reconnaissance des séquences d'ADN. Cette séquence d'ADN reconnue va également entraîner une activité différentielle des protéines POU, comme je le décrirai dans la partie suivante.

### 5.2.4 Mécanismes de régulation de la liaison et de l'activité des protéines POU

#### 5.2.4.1 Séquence du site de fixation

Les protéines de la famille POU ont initialement été décrites comme se fixant sous forme de monomère à une séquence de type octamère au niveau de l'ADN. Plus récemment, leur capacité d'homo- et d'hétérodimérisation a été mise en évidence. Des expériences *in vitro* ont permis de montrer que les protéines POU étaient capables de dimériser de manière coopérative au niveau de l'ADN (Rhee, Gruber et al. 1998).

En réalité, plusieurs motifs de dimérisation ont été décrits pour Brn2 (*Figure 27*). Le motif PORE (Palindromic Oct factor Recognition Element), ATTTGAAATGCAAA, a été identifié dans l'enhancer du gène de l'ostéopontine (OPN) murine (Botquin, Hess et al. 1998). Ce motif a été décrit comme un élément de réponse ADN au facteur OCT4. OCT4 peut d'ailleurs fixer le motif PORE sous forme de monomère ou de dimère. La dimérisation peut être promue ou diminuée par le remplacement d'un seul nucléotide. *In vitro*, le motif PORE se comporte comme un élément « général » de reconnaissance des facteurs OCT (Botquin, Hess et al.



*D'après Cook et Sturm 2008*

**Figure 27 : Organisation structurale des domaines POU en fonction de la séquence ADN reconnue**

1998). Afin d'étudier plus en détails la capacité de dimérisation des facteurs OCT, les séquences liées à l'élément de réponse Pit1 du gène de la prolactine ont été caractérisées. L'élément de réponse MORE (More Palindromic Oct factor Recognition Element) ainsi caractérisé a finalement été retrouvé dans de nombreux promoteurs (Tomilin, Remenyi et al. 2000). Tomilin & al ont également mis en évidence que le dimère OCT1 fixé au niveau du MORE n'est pas capable d'interagir avec le cofacteur OBF1, alors que la fixation au niveau du PORE permet l'interaction avec ce cofacteur. Ceci signifierait que, suivant la conformation adoptée par le dimère OCT1 (c'est-à-dire du motif fixé), l'activité transcriptionnelle pourrait être différente. Ce sont les structures cristallographiques des domaines POU d'OCT1 avec les séquences PORE et MORE qui ont permis de comprendre comment une même chaîne polypeptidique pouvait former un dimère et présenter une interface protéine-protéine différente (Remenyi, Tomilin et al. 2001). Un troisième motif, le NORE (N-Oct3 Responsive Element), TNNRTAAATAATRN, a été identifié suite à la modélisation moléculaire du dimère de BRN2. Ce motif est notamment retrouvé dans les promoteurs des gènes *Aadc* et *Crh* (Alazard, Bland et al. 2005). Le motif NORE diffère des motifs PORE et MORE puisque dans ce cas les domaines POU<sub>H</sub> fixent des régions chevauchantes. De plus, la liaison des dimères aux motifs MORE et PORE est coopérative, alors que dans le cas du NORE, elle est non coopérative.

#### 5.2.4.2 Modifications post-traductionnelles

Des modifications post-traductionnelles telles que la phosphorylation, l'ubiquitination, la sumoylation ou l'oxydation ont été décrites pour les protéines de la famille POU. En guise d'exemple concernant Brn2, l'oxydation d'une cystéine de Brn2 diminue sa liaison aux octamères de séquence consensus, mais également aux octamères de séquence divergente. Ces résultats ont été obtenus par l'addition d'ions cuivre, fer, de peroxyde d'hydrogène ou de diamide dans les expériences de liaison à l'ADN (DNA binding assays). Cet effet peut être reversé par l'ajout de l'agent réducteur dithiothréitol (Smith, Brightwell et al. 1998).

La phosphorylation de nombreux membres de la famille POU a également été décrite comme ayant un impact sur la régulation de la liaison à l'ADN. Cette phosphorylation a été étudiée pour PIT1, mais également pour OCT1, OCT2 et BRN5 (Kapiloff, Farkash et al. 1991; Grenfell, Latchman et al. 1996; Kasibhatla, Tailor et al. 1999; Pevzner, Kraft et al. 2000). Tout comme pour l'oxydation, la phosphorylation de BRN2 a également été étudiée. Dès 1996, il a été proposé que BRN2 puisse être phosphorylée par la protéine kinase A (PKA)

en réponse à l'AMPC. Cette phosphorylation entraîne l'exclusion de BRN2 du noyau, ce qui empêchera son interaction avec ses séquences cibles au niveau de l'ADN (Eisen 1996). Plus récemment, la phosphorylation de la protéine BRN2 humaine au niveau d'une sérine du domaine POU<sub>H</sub> a été mise en évidence (Nieto, Joseph et al. 2007). Cette phosphorylation de la sérine 362 inhiberait la fixation aux motifs PORE et NORE de l'ADN. En revanche, la phosphorylation n'a aucun impact sur la liaison aux motifs MORE, ce qui indique que cette modification post-traductionnelle peut permettre une sélection des promoteurs fixés. Ce phénomène a également été décrit au niveau du gène *Nestin* de la souris (Sunabori, Tokunaga et al. 2008). Plus précisément, cette équipe a mis en évidence par des expériences de « time-lapse » que l'expression du gène *Nestin* était augmentée durant la phase G1-S du cycle cellulaire. En revanche, son expression est beaucoup plus faible quand les cellules sont en phase G2-M. La phosphorylation d'un régulateur en amont, dépendante du cycle cellulaire, en l'occurrence BRN2, est à l'origine de cette expression différentielle du gène *Nestin* au cours du cycle cellulaire. La phosphorylation de BRN2 réduit sa capacité à se lier à l'élément enhancer du gène *Nestin*. Il en résulte une diminution de la transcription de ce gène en phase G2-M.

L'importance de cette phosphorylation de la sérine 362 de BRN2 a finalement été démontrée dans le mélanome. En effet, une expérience de surexpression de la protéine sauvage ou de la protéine mutée au niveau du site de phosphorylation a permis de montrer des effets propres à chacune. En effet, la surexpression de la version sauvage induit la prolifération et réprime la migration des mélanocytes *in vitro*. En revanche, la surexpression de la version mutée réprime à la fois la prolifération et la migration (Berlin, Denat et al. 2012). Ceci démontre une nouvelle fois que les modifications post traductionnelles de BRN2 sont en soit un mécanisme de régulation.

#### 5.2.4.3 Localisation

La navette nucléocytoplasmique a été décrite pour deux membres de la famille POU3F. OCT6 et BRN4 contiennent dans leur domaine POU une région hydrophobique riche en leucine qui permet leur passage entre le noyau et le cytoplasme. L'export du noyau est dépendant de CRM1 (Baranek, Sock et al. 2005). Le blocage de cet export grâce à un inhibiteur de CRM1, la leptomycine B, entraîne une réduction de l'activité des gènes régulés par OCT6. Ceci indique que ce transfert nucléocytoplasmique est nécessaire au maintien de compétence de transactivation de la protéine OCT6. Le domaine responsable de ce transfert

nucléocytoplasmique est conservé dans de nombreuses protéines POU, notamment de la famille POU3F, à laquelle appartient BRN2. La localisation cytoplasmique de Brn2 a été visualisée au cours du développement de la moelle épinière chez la souris (Tanaka, Kamachi et al. 2004). Néanmoins, il n'a pas été prouvé que le mécanisme permettant cette localisation soit similaire à celui décrit pour Oct6.

#### *5.2.5 Partenaires protéiques des protéines à domaine POU*

Chez les vertébrés, des études structure-fonction suggèrent que les propriétés des protéines proneurales divergent du fait de leur interaction avec différents cofacteurs. Souvent, les familles de facteurs de transcription coordonnent leurs activités par interactions avec des coactivateurs, corépresseurs, ou en formant des complexes avec d'autres facteurs de transcription (Remenyi, Scholer et al. 2004). Comme je l'ai décrit précédemment, le recrutement des corégulateurs des protéines à domaine POU peut être déterminé par la séquence d'ADN reconnue ou par l'accessibilité de la surface d'interaction pour d'autres protéines. Les facteurs de transcription à domaine POU peuvent interagir ou du moins coopérer avec différentes autres protéines, dont je vais citer quelques exemples dans les paragraphes suivants.

##### *5.2.5.1 Facteurs généraux de la transcription*

Les facteurs de transcription de la famille POU sont capables d'interagir avec d'autres protéines, interaction nécessaire à la formation d'un site de nucléation et à la génération d'une unité de transcription. En guise d'exemple, OCT1 et OCT2 peuvent interagir avec les facteurs généraux de la transcription incluant la TBP. Cette interaction se fait de manière directe, et ne dépend donc pas de la présence d'ADN ou d'autres cofacteurs (Zwilling, Annweiler et al. 1994). Concernant BRN2, une expérience de « double hybride » a permis de montrer son interaction avec TBP. De plus, des expériences de « GST pulldown » ont révélé que des contacts protéiques directs avaient lieu entre BRN2 et TBP, mais également entre BRN2 et TFIIB (Smit, Smith et al. 2000).

##### *5.2.5.2 POU-SOX*

En plus du recrutement de GTF, les protéines à domaine POU peuvent également interagir avec d'autres familles de facteurs de transcription spécifiques d'un lignage. L'une de ces familles est composée de facteurs de transcription SOX qui sont comme les protéines POU, des protéines extrêmement importantes pour la détermination et la spécification cellulaire

(Wegner 2005). L'action combinée des protéines POU et SOX constitue un code de régulation transcriptionnelle pour de nombreux gènes exprimés dans le SNC (Bailey, Klos et al. 2006). Ce code POU-SOX stipule qu'une protéine à domaine POU ne va interagir qu'avec un partenaire SOX spécifique afin de créer une réponse transcriptionnelle, et a été observé pour la première fois dans les cellules gliales (Kuhlbrodt, Herbarth et al. 1998; Kuhlbrodt, Herbarth et al. 1998).

L'hétérodimérisation entre OCT4, facteur POU spécifique des cellules embryonnaires, et de SOX2 est sûrement l'exemple de collaboration le mieux caractérisé. L'interaction d'OCT4-SOX2 est considérée comme étant un code combinatoire dans le développement précoce de l'embryon. En effet, le niveau d'expression de ces deux facteurs permet de diriger l'établissement des trois feuillets de l'embryon de mammifère (Avilion, Nicolis et al. 2003). OCT4 et SOX2 peuvent notamment interagir au niveau de deux enhancers distincts que sont FGF4 et UTF1 et à différent degrés de coopérativité (Remenyi, Lins et al. 2003). Cette différence peut expliquer une transcription différentielle des gènes cibles, tels que FGF4 et UTF1, basée sur le niveau relatif de protéines OCT4 et SOX2 présent dans les cellules. La différence peut également s'expliquer par la séquence des éléments de reconnaissance. En effet, dans l'enhancer de FGF4, il y a trois paires de base entre les sites de liaison POU et SOX, alors ces sites sont directement adjacents dans l'enhancer d'*Utf1*. Outre *Fgf4* et *Utf1*, les gènes *Oct4* et *Sox2* contiennent également des motifs POU et SOX, indiquant qu'une boucle d'autorégulation module l'expression de ces gènes.

Plus récemment, l'idée d'un échange de facteurs POU pour un même facteur SOX a été émise. Ainsi, SOX2 pourrait former un dimère avec OCT4 dans les cellules ES, puis un dimère avec BRN2 au niveau des mêmes promoteurs pour spécifier la différenciation neuronale (Lodato, Ng et al. 2013).

#### 5.2.5.3 Famille des facteur PAX

Les membres des facteurs de transcription POU et SOX sont capables d'interagir avec la famille de facteurs de transcription PAX. La structure cristallographique OCT1/SOX2/enhancer de *Fgf4* a révélé qu'une région Cter du domaine HMG de SOX2 pouvait interagir avec une région similaire et hautement conservée du domaine POU<sub>S</sub>. Il a été démontré que cette surface d'interaction au niveau de SOX2 pouvait être reconnue par PAX6 (Remenyi, Lins et al. 2003). Dans un type cellulaire où l'ensemble de ces protéines seraient

exprimées, les protéines PAX pourraient donc servir de corégulateurs pour l'interaction entre les protéines POU et SOX.

Un autre membre de la famille PAX, PAX3, est exprimé dans les cellules de mélanocytes. Des interactions entre BRN2 et PAX3, BRN2 et SOX10 ainsi SOX10 et PAX3 ont pu être mises en évidence par des expériences de double hybride (Smit, Smith et al. 2000).

#### 5.2.5.4 *ASCL1*

Les protéines proneurales jouent un rôle central dans la régulation du processus de neurogenèse (Bertrand, Castro et al. 2002). ASCL1 et les protéines de la famille des neurogénines, qui sont les principales protéines proneurales chez les mammifères, sont impliquées dans l'engagement vers la voie neuronale, la migration cellulaire, la sortie du cycle cellulaire ou encore l'inhibition de la différenciation de cellules voisines par l'activation de la voie Notch (Bertrand, Castro et al. 2002; Nakada, Hunsaker et al. 2004; Ge, He et al. 2006). Chez la souris, il a été montré que BRN1 et BRN2 peuvent interagir avec le facteur de transcription proneural ASCL1. L'activation de la transcription des gènes cibles passe par la fixation à un motif octamère de type « E-box »(CANNTG) identifié au départ dans l'enhancer du gène *Delta1*. Ce gène code pour l'un des ligands de la famille des récepteurs Notch. Les analyses bioinformatiques ont permis d'identifier 21 autres gènes dont les promoteurs contiennent également ce motif reconnu par ASCL1/BRN. Parmi ces gènes se trouvent de nombreux membres de la voie de signalisation Notch, mais également deux membres de la famille ZIC, à savoir ZIC1 et ZIC4 (Castro, Skowronska-Krawczyk et al. 2006).

Récemment, des expériences d'immunoprécipitation de la chromatine couplée au séquençage à haut débit (ChIP-seq) ont confirmé ce lien entre BRN2 et ASCL1. En effet, en surexprimant les facteurs BRN2, ASCL1 et MYT1L (myelin transcription factor 1-like) dans des fibroblastes, ces derniers adoptent des caractéristiques neuronales. Un ChIP dirigé contre BRN2 dans ces conditions a montré que de nombreux sites contenaient un motif E-box, suggérant la coopération entre ASCL1 et BRN2 pour engager ces fibroblastes dans une voie neuronale (Wapinski, Vierbuchen et al. 2013).

# ***Résultats***

# ***Projet de thèse***

## Projet de thèse

Lors de mon arrivée dans le laboratoire du Docteur Irwin Davidson, l'un des axes de recherches consistait en l'étude du rôle de l'AR ainsi que de ses récepteurs dans différents processus physiologiques. La cartographie des sites de fixation de RAR avait préalablement été réalisée dans les fibroblastes embryonnaires murins (MEF) ainsi que dans les cellules ES (Delacroix et al). Cette expérience a mis en évidence la variabilité et la spécificité de liaison des RAR entre différentes lignées cellulaires et démontre également que l'action de l'AR est différente selon le processus physiologique impliqué.

Afin d'aller plus loin dans ces travaux, nous avons choisi d'utiliser un processus de différenciation dépendant de l'AR. Comme je l'ai décrit précédemment, l'AR est important pour le développement de l'embryon. Outre son rôle *in vivo*, l'AR est également connu pour son importance dans la différenciation neuronale *in vitro*.

Mon travail de thèse a dans un premier temps consisté à reproduire et optimiser le protocole de différenciation neuronale publié précédemment (Bibel, Richter et al. 2004; Bibel, Richter et al. 2007). L'ensemble des paramètres pouvant impacter la réussite de cette différenciation a du être soigneusement établi afin que ce protocole puisse être réalisé de manière reproductible, quasiment en routine.

Une fois ce protocole mis en place, j'ai collaboré au projet initié par Delacroix & al, qui étudiait les cibles des RAR dans un processus de différenciation *in vitro* similaire à ce qui est observé *in vivo*. Pour ce projet, j'ai été amenée à réaliser des expériences de ChIP-seq et de RNA-seq dans les cellules ES en cours de différenciation, afin de fournir au laboratoire une liste de séquences cibles directes des RAR. Ces résultats de ChIP-seq ont permis d'identifier de nouveaux types de motifs et d'espacement des hémisites (DR) reconnus par le dimère RAR/RXR. Des analyses complémentaires d'EMSA et de « CAT ELISA » ont été réalisées au laboratoire et ont permis de valider les DR trouvés et de vérifier leur fonctionnalité (Annexe manuscrit n°2).

Les données de RNA/ChIP-seq que j'ai obtenues ont également servi pour la publication des résultats de l'équipe de Cécile Rochette-Egly, c'est pourquoi ce manuscrit se situe également en annexe (Annexe Manuscrit n°3)

Une fois ce protocole de différenciation à l'AR entièrement maîtrisé, nous nous sommes intéressés à plusieurs facteurs de transcription dont l'expression était induite dans ce processus de différenciation. De nombreux candidats ont été choisis dans un premier temps, puis une sélection plus stricte a été faite suite aux premiers résultats d'expériences. Afin de déterminer le rôle de nos facteurs candidats, nous avons choisi d'utiliser une approche de shARN afin de réduire individuellement l'expression de chaque facteur. Parmi les facteurs sélectionnés initialement se trouvaient Hes3 et Hes5, Gbx2, Dbx1, Msx1, les membres de la famille Irx (Irx1, -3, -5), Ptf1a ou encore Brn2. L'induction de l'expression est visible pour l'ensemble de ces facteurs au niveau mRNA lors de la différenciation neuronale. Néanmoins, bien que connus pour leurs rôles dans la neurogenèse *in vivo* et/ou *in vitro*, les réseaux transcriptionnels pour ces facteurs sont très peu connus. L'idée du projet consistait à utiliser en parallèle des expériences de RNA-seq et de ChIP-seq afin de déterminer le réseau transcriptionnel activé par le facteur d'intérêt choisi.

Alors que pour certains facteurs je n'ai obtenu aucune réduction d'expression par utilisation de shARN, d'autres facteurs dont l'expression était significativement impactée ne montraient aucun phénotype associé. Ce type de résultat souligne que ces facteurs n'étaient pas capables d'impacter la différenciation neuronale individuellement. Le facteur de transcription qui nous a donné le résultat le plus marquant a été Brn2, puisque nous avons confirmé que ce facteur jouait un rôle clé dans la différenciation neuronale. L'utilisation des techniques de ChIP-seq et de RNA-seq nous a permis de définir un ensemble de facteurs cibles directes ou indirectes de Brn2. Parmi les facteurs identifiés dans ces analyses se trouvent d'autres facteurs de transcription avec un profil d'expression similaire à celui de Brn2, qui pourraient par conséquent également avoir un rôle dans la différenciation neuronale.

Alors qu'il n'existait aucune donnée de ChIP-seq lorsque j'ai commencé mon travail sur Brn2, pendant mes études, plusieurs expériences dans d'autres modèles ont été publiées (Lodato, Ng et al. 2013; Wapinski, Vierbuchen et al. 2013). L'intégration des résultats publiés avec les données que j'ai obtenues nous a permis d'identifier un nombre restreint de gènes cibles de Brn2, fixés dans les différents types de différenciation neuronale. Parmi ces gènes, qui peuvent en quelque sorte être considérés comme le réseau transcriptionnel minimum activé par Brn2 se trouve notamment Zic1.

De la même manière que pour Brn2, Zic1 est un facteur important dans la différenciation neuronale *in vivo* mais pour lequel il n'existe pas de données à haut débit permettant d'identifier son réseau transcriptionnel. L'ensemble des résultats concernant l'étude de ces deux facteurs est décrit dans le manuscrit présent dans cette partie.

Dans les résultats complémentaires faisant suite à ce manuscrit je vous présenterais les résultats obtenus initialement pour l'étude du facteur Hes3 ainsi que les optimisations de protocole réalisées afin d'obtenir une bonne réduction d'expression de notre gène cible par l'approche lentivirale.

***Manuscrit n°1***

## **A Brn2-Zic1 axis specifies the neuronal fate of retinoic acid treated embryonic stem cells.**

Sylvia Urban, Dominique Kobi, Stéphanie LeGras, Tao Ye, and Irwin Davidson#.

Department of Functional Genomics and Cancer

Institut de Génétique et de Biologie Moléculaire et Cellulaire.

CNRS/INSERM/UDS.

1 Rue Laurent Fries,

67404 Illkirch Cédex. France.

FAX: 33 3 88 65 32 01. TEL: 33 3 88 65 34 40 (45)

# To whom correspondence should be addressed

E-mail: [irwin@igbmc.fr](mailto:irwin@igbmc.fr)

Running Title: A Brn2-Zic1 axis.

Key words: Ascl1, Fgf2, Sox2, neurogenesis.

## **Abstract.**

Murine embryonic stem (ES) cells grown as embryoid bodies (EBs) and treated with all-trans retinoic acid (RA) differentiate into a homogenous population of glutamatergic neurons. This process is initiated by activation of a set of target genes by the RA-receptors that in turn activate a complex set of transcriptional regulatory networks leading to neurogenesis. Here we show that the POU-Homeodomain transcription factor Pou3f2 (Brn2) is essential for the neuronal differentiation programme. ShRNA mediated Brn2 silencing abrogates neuronal fate in favour of a poorly defined cell population expressing various mesoderm and endoderm markers. By integrating RNA-seq following Brn2 silencing with Brn2 ChIP-seq, we show that Brn2 binds to the promoters of the *Ascl1* and *Neurog1* genes and activating their expression and subsequently regulating many genes of the neurogenic programme. We integrated Brn2 ChIP-seq data from RA-treated EBs and P19 cells with public datasets from ES cells differentiated into neuronal precursors by Fgf2 treatment and fibroblasts trans-differentiated into neurons by ectopic Brn2 expression. This analysis showed that Brn2 binds a distinct but overlapping set of genomic loci in each system, but defines a set of common binding sites and target genes amongst which is transcription factor Zic1. We show that shRNA mediated silencing of Zic1 in RA-treated EBs abrogates neural fate and we define a hierarchical Brn2-Zic1 axis essential to specify the neural fate of RA-treated ES cells.

## **Introduction.**

Retinoic acid (RA) plays an essential role in embryonic development and in the homeostasis of many adult tissues (Niederreither and Dolle, 2008) (Mark et al., 2006). The action of RA is mediated by the three nuclear receptor isotypes  $RAR\alpha$ ,  $RAR\beta$  and  $RAR\gamma$ , each of which encodes several isoforms generated by alternative splicing (Niederreither and Dolle, 2008). The RARs heterodimerise with RXRs and bind to a set of response elements in the regulatory elements of their target genes comprising two half sites with the consensus sequence 5'A/GGG/TTCA-3' (Khorasanizadeh and Rastinejad, 2001); (Rochel, 2011). RA-response elements with different spacing and topologies of the half sites have been described, the most characterised of which are direct repeats with 2 or 5 base pair spacing, DR2 and DR5, although DR1 and DR8 have also been described along with inverted repeat (IR0) (Balmer and Blomhoff, 2005; Moutier et al., 2012). In absence of ligand, RAR-RXRs repress transcription through association with a set of co-repressors, whereas in the presence of ligand a conformational change in the ligand-binding domain releases the co-repressors and allows association with a set of co-activator proteins to promote activation of target genes (Rosenfeld et al., 2006). RAR activity is also modulated by phosphorylation and several phosphorylation sites have been characterised (Al Tanoury et al., 2014; Chebaro et al., 2013).

Mouse embryonic stem cells have been extensively studied as a model for a variety of differentiation pathways. A major consideration in such studies is to obtain a highly homogenous and pure population of differentiated cells. ES cells can be differentiated into a highly homogeneous population of glutamatergic neurons when grown as embryoid bodies and treatment with RA (Bibel et al., 2007; Bibel et al., 2004; Gotz and Barde, 2005). This

pathway mimics generation of cortical glutamatergic neurons from Pax6-expressing progeny in the developing mouse cortex (Gotz and Barde, 2005). In this pathway, the initial trigger for differentiation is the activation of a set of target genes by the RARs. While RAR $\alpha$  and RAR $\beta$  are dispensable for differentiation, the RAR $\gamma$  isotype is essential, in particular the RAR $\gamma$ 2 isoform that comprises a phosphorylation site in its N-terminal region (Al Tanoury et al., 2014). ES cells can also be differentiated into neuronal precursor cells by other techniques, such as by treatment with Fgf2 (Okabe et al., 1996; Wichterle et al., 2002)

While the initial events in RA-induced differentiation have been well characterised (Mahony et al., 2011) (Moutier et al., 2012) (Al Tanoury et al., 2014), the downstream transcription factors required for assuring neural fate are unknown. RA-treated embryoid bodies first differentiate into Pax6-expressing radial-glial-type progenitor cells. Pax6 plays an essential role in this process since differentiation of ES cells lacking Pax6 generates GABAergic neurons expressing high levels of the neurotrophin receptor p75NTR causing their rapid death (Nikolopoulou et al., 2007). Pax6 is therefore critical for the type neurons generated, but does not affect the neuronal fate of the ES cells.

Pou3f2 also known as N-Oct3 or Brn2 (hereafter Brn2) is a member of the Pou-Homeodomain family of transcription factors with an important role in neurogenesis (Schonemann et al., 1998) (Ryan and Rosenfeld, 1997). Brn2 is expressed in sub-ventricular zone progenitor cells throughout the neuraxis and is down-regulated upon differentiation, but remains expressed in cortical layers II/III and in V pyramidal neurons (Hagino-Yamagishi et al., 1997). *Brn2* knockout neonates showed neuronal loss only in the hypothalamic supraoptic and paraventricular nuclei where the closely related *Brn1* is not expressed (Nakai et al. 1995; Schonemann et al. 1995). In contrast, *Brn1* mutants display changes in brain morphology only in the hippocampus that weakly expresses Brn2 (Sugitani et al., 2002). Both Brn1 and Brn2 are expressed in the neocortex, and no developmental

defects were seen whether either of the two were inactivated. Knockout of both Brn1 and Brn2 resulted in severe proliferation defects of cortical progenitor cells and migration defects of upper-layer neurons, ultimately leading to a disorganized and thinned cortex (McEvelly et al., 2002) (Sugitani et al., 2002). In addition, Brn2 also plays an important role in Schwann cell development (Jaegle et al., 2003). Moreover, the essential role of Brn2 in neurogenesis is highlighted by the fact that it can be used in combination with Ascl1 and Myt1l to reprogramme fibroblasts into a neuronal fate *in vitro* (Vierbuchen et al., 2010) (Wapinski et al., 2013). Despite the critical role of Brn2 in neurogenesis, its target genes and mode of action remain poorly characterised.

Brn2, like the other members of the POU domain family of transcription factors, share the POU DNA-binding domain (DBD). The POU domain consists of two DNA-binding units (POUs for POU specific and POUh for POU homeodomain) connected by a flexible linker (Phillips and Luisi, 2000; Ryan and Rosenfeld, 1997). The POU and POUh domains each comprise a helix-turn-helix structure of which the third helix recognises the DNA and provides sequence specificity (Cook and Sturm, 2008; Klemm et al., 1994). This molecular structure allows POU proteins to recognize several types of binding element depending on the spacing and the positioning adopted by the two subdomains (Klemm and Pabo, 1996). Brn2 binds as a dimer to at least three distinct types of sequence designated as PORE (palindromic Oct recognition element), MORE (more palindromic Oct recognition element) and NORE (N Oct-3 recognition element) as well as the simpler OCT motif (Alazard et al., 2005; Botquin et al., 1998; Millevoi et al., 2001; Nieto et al., 2007; Tomilin et al., 2000).

In this study, we show that Brn2 is essential for RA-induced neuronal differentiation of ES cells. Integration of RNA-seq and ChIP-seq data identifies a set of Brn2 regulated target genes involved in neurogenesis in these cells. We performed an integrative analysis of current Brn2 ChIP-seq data from several neurogenesis models to identify a common set of Brn2

occupied sites and potential target genes that form the core of the Brn2-regulated neurogenesis network. From this analysis, we identified transcription factor Zic1 as a critical downstream target regulated by Brn2 in RA-treated ES cells. Zic1 is essential for neuronal differentiation and Brn2 and Zic1 cooperate to regulate a set of neurogenic genes. We thus identify a novel Brn2-Zic1 axis as essential for specifying the neuronal fate of RA-treated ES cells.

## **Results.**

### **POU3F2 (BRN2) is essential for neuronal identity of RA-differentiated embryoid bodies.**

Treatment of ES cells grown as embryoid bodies (EBs) by retinoic acid induces a large set of genes and triggers a developmental programme leading to differentiation of a large proportion of the cells as glutamatergic neurons (Fig. 1A)(Bibel et al., 2007). We have previously defined the profile of RAR-RXR binding sites over the EB genome and identified genes that are direct targets of retinoic acid signalling, many of which are themselves transcription factors responsible for regulating further downstream target genes (Moutier et al., 2012). We therefore investigated which of the early or late induced target genes may be critical for RA-induced differentiation. Examination of the RNA-seq data and quantitative reverse transcription PCR (RT-qPCR) analysis identified several transcription factors whose expression is induced at various times after RA treatment (Supplemental Fig. 1). RA treatment repressed expression of *Pou5f1* (*Oct4*) and *Nanog*, but had less effect on that of *Sox2* that continues to be expressed at later stages of differentiation. Expression of *Pax6*, *Pax3*, *Nestin*, and *Mapt* is clearly detectable between 24-48 hours after RA addition (Supplemental Fig. 1), whereas that of others such as *Ascl1*, *Msx3*, *Dcx1* and *Pou3f2* (Fig. 1B) is evident only at later stages between 48-and 96 hours. Several of these factors are involved in diverse aspects of neurogenesis *in vivo* [see for example, (Castro and Guillemot, 2011)]

[Ramos and Robert, 2005]]. In particular Brn2 mRNA and protein are strongly expressed only between 72-96 hours after RA treatment as neuronal precursors are in the process of being specified (Fig. 1B and C).

To test the requirement of Brn2 in RA-induced neuronal differentiation of EBs, we used an shRNA approach in which ES cells were infected with lentivirus vectors expressing an shRNA directed against Brn2 (ShBrn2) or a control shRNA (ShC). The ES cells were selected with puromycin and then differentiated as RA-treated EBs. Knockdown of Brn2 expression after 96 hours of RA treatment could be seen both by immunoblot and RT-qPCR (Fig. 1D and E). Loss of Brn2 led to a marked reduction of the number of surviving neuronal cells after dissociation of the EBs and plating on poly-ornithine-laminin coated plates (Fig. 1F). Moreover the surviving cells did not show the typical morphology of early neurons with few long axonal projections and a diminished number of cells with rounded and refringent cell bodies. These residual cells died after further 2-3 days of culture (data not shown).

Expression of several neuronal marker genes normally induced in WT ES cells, were repressed in the shBrn2 knockdown cells (Fig. 1G). To investigate more globally the consequences of Brn2 silencing on gene expression, we performed RNA-seq on the 96 hour RA-treated EBs from Brn2 and control shRNA infections. The expression of around 280 genes was repressed, while that of 275 genes was up-regulated (Supplemental Dataset 1). Ontology analysis of these genes showed that Brn2 was required to assure the expression of a large set of genes involved in neurogenesis and transcription regulation (Fig. 1H and Supplemental Fig. 2A). Amongst the genes whose expression was diminished are critical regulators of neurogenesis such as *Ascl1*, *Neurog1*, *Dcx*, *Dcc*, *Zic1/2/4/5*, *Slit1* and *Uncx*. Brn2 silencing affects Notch signalling as expression of *Dll1*, *Dll3*, *Jag1* and the downstream targets *Hes3* and *Hes5* are all diminished (Supplemental Fig. 2A). In contrast, ontology analysis of genes that are up-regulated indicated their association with terms such as

interferon-mediated immunity, and mesoderm development (Fig. 1I, and Supplemental Fig. 2B). These data show that Brn2 is an upstream regulator of the network that specifies neuronal identity of differentiating ES cells. In absence of Brn2, surviving cells rather express a diverse set of genes including those associated with mesoderm and endoderm differentiation and show up-regulated expression of *Foxd1* that is a marker for reprogramming of cell fate (Koga et al., 2014). These cells also show a more general defect characterised by the activation of the interferon response.

### **Brn2 directly regulates genes critical for neuronal differentiation.**

To identify genes that are directly regulated by Brn2, we performed ChIP-seq profiling of its genome occupancy in RA-induced EBs. Under these conditions, >2100 occupied sites were identified. Brn2 bound sites were dispersed throughout the genome in both inter- and intragenic regions, but showed a strong enrichment in promoter regions (-1kb/+100bp) compared to the representation of this region in the genome (Fig. 2A and Supplemental Dataset 2). In agreement with this, Brn2-occupied sites were enriched around the TSS (Fig. 2B). For example, Brn2-occupied sites were found in the promoter regions of *Ascl1*, *Neurog1*, and upstream of *Pou6f1* and *Epha10* (Fig. 2C).

Brn2 has been shown to bind several different types of recognition motifs such as the NORE, MORE and PORE elements as well as the simpler OCT motif (ATGCAAAT). Analysis of the sequences under the Brn2 occupied peaks indicated the presence of few fully consensus OCT sites, but a much larger number of sites with a single mismatch (Fig. 2D). Similarly, few consensus NORE sites were identified, but allowing a single mismatch up to 266 motifs were found. It should be noted that this motif is rather degenerate and allowing an additional mismatch leads to a very loose consensus such that the significance of these sites is questionable. More strikingly, no PORE motifs were found. Instead, the most predominant

motif was the MORE with 143 consensus elements and up to 459 taking into account a single mismatch. MORE+1 motifs with a single base pair insertion between the palindromic pseudo-repeats were also strongly represented in the Brn2 bound sequences. The consensus sequences for these elements were also derived from MEME *de novo* motif discovery analysis of 600 occupied sites that identified MORE and MORE+1 sequences amongst the most significantly enriched motifs. Both of these motifs were centered under the peak consistent with them acting as the Brn2 binding site (Fig. 2E). This approach confirms that MORE and MORE+1 are the most represented motifs bound by Brn2 in RA-treated EBs. Note also that an OCT motif with a single mismatch is comprised within the MORE elements and thus many of the detected OCT elements overlap with the MOREs.

MEME analysis also revealed enrichment of a ZFX-like binding motif in the Brn2 occupied regions. This motif was more abundant than either the MORE or MORE+1 elements, but was not located at the center of the binding site (Fig. 2E), being rather located flanking the MORE motifs. Interestingly, Brn2 sites with ZFX-like motifs were strongly enriched at the proximal promoter and TSS, but were more rare at distant sites (Supplemental Fig. 2C). We therefore examined the association between the Brn2-bound motifs and those of other transcription factors that may be enriched at Brn2-occupied sites. Co-occurrence motif analysis (Fig. 2F) revealed enriched association between the MORE and MORE+1 motifs and motifs for Sp1, Meis transcription factors that are important in neurogenesis (Agoston et al., 2014; Yamada et al., 2013), the DR1 motif for the Rar/Pparg-Rxr heterodimers, Runx1 and Foxd3. There is however selectivity as MORE is associated with motifs for other nuclear receptors, notably the DR5 for Rar-Rxr, and sites for Nr5a2 and Nr2f1 (COUP-TF1), as well as for Trp53/Trp63, whereas MORE+1 is associated with Tead1 and Rest both of which have known functions in neurogenesis (Ballas and Mandel, 2005; Cao et al., 2008). Together

these data show that Brn2 may cooperate with nuclear receptors, Rest, Tead1 and factors binding to the ZFX-like motif to regulate expression of genes essential for neuronal function.

We next integrated the RNA-seq and ChIP-seq analysis to identify genes associated with Brn2-occupied sites and changes in gene expression following Brn2 knockdown. Using a window of +/- 40 kb with respect to the TSS, 2616 genes were annotated as associated with Brn2 bound sites (Supplemental Fig. 2D). These genes are strongly enriched in several functions including nucleic acid metabolism, mRNA transcription and splicing as well as cell cycle and neurogenesis. (Supplemental Fig. 2E and F). Nevertheless, of these only 36 showed diminished expression upon Brn2 silencing, and 13 showed up-regulation (Supplemental Fig. 2D and G). Amongst the direct down-regulated targets are genes coding for transcription factors *Ascl1*, *Neurog1*, another neurogenic POU-factor *Pou6f1*, *Zic1* and *Zic4* and *Ebf3*. These 36 genes are associated with 48 Brn2-occupied sites showing enrichment around the TSS (Supplemental Fig. 2H) and a majority are associated with MORE or MORE+1 motifs of which 8 of 36 are associated with a combination of MORE/MORE+1 and/or ZFX-like motifs. In contrast, only 13 up-regulated genes are associated with Brn2-bound sites that show no specific preference for their location relative the TSS. Brn2 therefore directly activates expression of neurogenic transcription factors that can then regulate other genes to promote neuronal differentiation.

We also performed Brn2 gain of function experiments to identify additional potential target genes. We infected ES cells with a lentivirus vector where the Brn2 cDNA can be expressed under the control of 5XDR5 RA-response elements. The cells were selected with puromycin and differentiated as embryoid bodies. In these cells ectopic Brn2 expression is under the control of RA and was induced prematurely 24 hours after RA treatment and persisted over the 96 hours (Fig. 3A and B). Brn2 overexpression had little effect on the number of neurons and their morphology (data not shown), but enhanced expression of

several of its target genes was seen. Interestingly, although Brn2 was prematurely expressed by 24 hours, an overexpression of its target genes was only observed at 96 hours (Fig. 3C).

We performed RNA-seq analysis in control RA-induced EBs and from EBs prematurely expressing Brn2 showing that 393 genes were up-regulated and 327 were down-regulated (Supplemental Dataset 1). Ontology analysis of these genes showed that premature Brn2 expression enhanced expression of a set of genes involved in neurogenesis (Fig. 3D). Amongst these are important neurogenesis regulators such as *Dcx*, *Dcc*, *Mapt*, *Slit1*, *Zic1/2/4/5*, as well as several other Pou domain factors (*Pou2f2*, *Pou4f1*, *Pou6f1*). Genes down-regulated upon Brn2 premature expression are notably enriched in mesoderm factors (Fig. 3E) like *Nkx2.1*, *Nkx6.1*, *Nkx6.2* and *Gata3*. Thus, gain of Brn2 function promotes expression of genes involved in neurogenesis and the neural fate while repressing other fates.

Interestingly, comparison of deregulated genes in Brn2 loss and gain of function experiments showed significant overlap. While only 3 genes were up-regulated in both experiments (i.e. showing the same change in both gain and loss of function, data not shown), none were down-regulated. On the other hand, 134 (47%) genes down-regulated by loss of Brn2 were up-regulated after Brn2 overexpression (Fig 3F). Similarly, 47 (17%) genes up-regulated after loss of Brn2 were down-regulated by overexpression (Fig. 3G). Integration of RNA-seq data from the gain of function experiment with ChIP-seq data, showed that of the 393 genes whose expression was induced by Brn2 overexpression, 44 are associated with a Brn2 binding site (Fig. 3H and I). The same is true for 17 of the genes whose expression was repressed by Brn2 overexpression. Of the 134 genes whose expression is induced by gain of function and repressed by loss of function 15 were associated with Brn2 occupied sites, but only 2 of the 47 genes that appeared as repressed by Brn2 (Fig. 3J)

The results of the gain of function experiments are complementary to those of Brn2 knockdown as they define a larger set of genes whose expression is controlled both directly and indirectly by Brn2 during differentiation.

### **Comparison of Brn2 genomic localisation in ES cells induced to differentiate into neurons by different pathways.**

The above data define a set of genes that are potentially directly regulated by Brn2 to induce neuronal differentiation after RA-treatment. However ES cells can be induced to differentiate into neurons by other protocols. Lodato et al., analysed Brn2 genomic occupancy in ES cells induced to differentiate into NPCs after withdrawal of Fgf2 (Lodato et al., 2013). In this analysis, they determined that Sox2 and Brn2 bind to active H3K27ac-marked enhancer elements and cooperate to induce the expression of a series of genes required for neuronal differentiation.

In order to facilitate the comparison with the data presented here in RA-induced EBs, we re-analysed their data set from Fgf2-induced NPCs with the same criteria used to analyse our data presented above and kept only peaks that were commonly called in their replicate samples to obtain a stringent set of peaks. Lodato reported 6574 Brn2 occupied sites in Fgf2-induced NPCs (Lodato et al., 2013). Reanalysis, identified only 3005 peaks in the more stringent set (Fig. 4A), of which only 323 (i.e. around 10%) are common to both the Lodato and our data sets (Fig. 4B). Similarly, we analysed a second data set in which Lodato overexpressed Brn2 in an inducible manner using the Tet0 system in ES cells (Lodato et al., 2013). Reanalysis of this data, using the Brn2 ChIP in the control line not expressing ectopic Brn2, identified 2031 peaks, compared to the 8400 reported by Lodato, of which 303 are also occupied in RA-induced EBs (Figs. 4A and B). We also reanalysed the Brn2 ChIP-seq data using the input data set from the ectopic expressing line and in this case identified only 1541

peaks of which 286 are occupied in in RA-induced EBs. Moreover, there is also limited overlap between the Brn2 sites in Fgf2-induced NPCs and following ectopic expression as only 202 (200 using the input as control for the Tet0 data) sites are occupied under all conditions.

Analysis of the sequences bound by Brn2 in the Lodato et al, data sets revealed that the Fgf2-induced NPCs data set is enriched in a degenerate MORE motif (Fig. 4C) similar to what is observed for the TetO dataset. An additional OCT motif is however found in the TetO dataset (Fig. 4D). The 202 sites that are occupied under all conditions are strongly enriched in MORE-type motifs, similar to those seen in the RA-induced EBs (Fig. 4E).

Lodato et al, reported an association between Brn2-occupied sites and those bound by Sox2. Re-analysis of the Sox2 ChIP-seq data set identified 10207 sites compared to the 16685 previously reported and we clearly detected the Sox2 consensus motif at these sites (Figs. 4A and F). Nevertheless, while there was a significant overlap between the Brn2 and Sox2 sites in the Fgf2-induced NPC data, there was little overlap with ectopically expressed Brn2 and only 43 sites shared with Sox2 were seen when combining both data sets (Fig. 4G).

As Lodato reported that Brn2 and Sox2 associated with sites in active H3K27ac-marked enhancers, we examined whether Brn2 binding sites in the RA-induced EBs and the commonly occupied sites are enriched at active enhancer elements using the H3K27ac data set published previously for Fgf2-induced NPCs by Creighton et al (Creighton et al., 2010). We found that only 50 Brn2 binding sites (2,4%) from RA-induced EBs are located in enhancers marked as active in Fgf2-induced NPCs (Fig. 4H). Similarly, using data for the H3K4me1 mark, we identified 105 Brn2 sites located in poised enhancers (4,9%). Of the 202 shared sites, 19 (9.4%) are located in H3K27ac active enhancers and 43 (21%) are located in H3K4me1-marked poised enhancers. Furthermore, of the 202 Brn2 bound shared sites only 20 (10%) also showed Sox2 occupancy and 4 are associated with active enhancers.

Together these data show that Brn2 occupies a distinct set of target sites in Fgf2- and RA-induced EBs. While there is association of Brn2 sites with Sox2 in Fgf2-induced NPCs, this is not seen at Brn2 sites in RA-induced EBs nor at the sites shared between the two model systems. It is nevertheless possible to identify a shared set of sites highly enriched in MORE-type motifs and associated with genes essential for neuronal differentiation.

### **Brn2 is essential for P19 cell neuronal differentiation.**

As we showed above that Brn2 occupies differing sets of sites when ES cells are stimulated to differentiate into neurons by distinct signals, we extended our study to P19 embryonal carcinoma cells as an additional model for neuronal differentiation. P19 cells differentiate into a diverse set of neurons when grow as RA-treated aggregates for 4 days before plating in neuronal specific medium (Fig. 5A). In these cells, Brn2 expression is strongly induced 24 hours after RA-treatment and its expression persists in the aggregates throughout the four days (Fig. 5B-C). To test the role of Brn2 in these cells, we used lentiviral infection to express shRNA against *Brn2* and differentiated the infected cell population. Brn2 expression was strongly repressed after 4 days of RA-treatment (Fig. 5D-E) and the resulting cell population failed to differentiate into typical neuronal like cells (Fig. 5F). Brn2 is therefore essential for RA-induced neuronal differentiation of both ES and P19 cells [see also (Fujii and Hamada, 1993)]. We also made cells constitutively expressing ectopic Brn2 to determine whether its expression could bypass the requirement for RA treatment. However, no neuronal differentiation could be seen in absence of RA showing that Brn2 is necessary, but not sufficient for differentiation (data not shown).

Expression of many neuronal markers was repressed in the Brn2 silenced P19 cells (Fig. 5G). Further RNA-seq analysis, after plating as neurons identified 1653 down-regulated genes (Supplemental Dataset 1). Ontology analysis showed that these genes are strongly

enriched in neuronal markers (Fig. 5H and Supplemental Fig. 3A). More than 1500 up-regulated genes were also identified comprising numerous genes associated with mesoderm development including several BMPs amongst which *Bmp4* plays an important role in embryonic mesoderm formation (Hogan, 1996) (Fig. 5I and Supplemental Fig. 3B).

We next compared the genes down-regulated after *Brn2* silencing in 4 day RA-induced EBs and RA-treated P19 aggregates. Of the 280 genes down-regulated by *Brn2* silencing in RA-induced EBs, 195 are also down-regulated in *Brn2* silenced P19 cells (Supplemental Fig. 3C). Thus, despite the differences between these two model systems, *Brn2* regulates many genes in common in these two cell lines. The larger number of regulated genes in the P19 cells probably reflects the fact that these were analysed 4 days after induction of *Brn2* expression in re-plated neurons, while in RA-induced EBs, we analysed gene expression only 24-48 hours after *Brn2* expression appears. The shared genes therefore likely correspond to the initial stages of the *Brn2*-induced gene expression programme. Of the 275 genes up-regulated by sh*Brn2* silencing, only 67 were up-regulated in P19 cells and are enriched in terms associated with interferon-mediated immunity (Supplemental Fig. 3D).

To determine how *Brn2* regulates gene expression in P19 cells, we attempted to perform ChIP-seq in RA-treated P19 aggregates similar to what we had performed in RA-induced EBs. However, expression of *Brn2* after 4 days of RA induction is significantly lower in P19 cells (Agg4) compared to RA-treated EBs (Supplemental Fig. 3E), and we were unsuccessful in obtaining a meaningful ChIP-seq profile. As inducible *Brn2* expression in Fgf2-treated ES cells was shown to stimulate neuronal differentiation, we engineered P19 cells to express ectopic *Brn2* under the control of 5 DR5 elements such that its expression was induced by RA at the same time as the endogenous protein. These P19 cells differentiated into neurons indistinguishable from native P19 cells (data not shown). We therefore performed ChIP-seq with these cells 24 hours after RA treatment to identify the first set of

genes bound and potentially regulated by Brn2. This approach identified 1743 occupied sites (Fig. 6A and Supplemental Dataset 2). Unlike RA-induced EBs, Brn2 occupied sites in P19 cells showed no enrichment at the promoter or close to the TSS, but were located at distal inter and intragenic regions (Fig. 6B-C). Analysis of the sequences associated with these sites revealed an absence of PORE motifs, few NORE motifs, but a preponderance of MORE and MORE+1 and OCT motifs (Fig. 6D). MEME analysis also identified a MORE-type motif centered beneath the peak as the most enriched sequence (Fig. 6E). Thus, like RA-induced EBs, Brn2 preferentially occupies MORE like motifs in P19 cells.

We also compared occupancy of ectopic Brn2 with genes that are down-regulated in shBrn2 silenced cells. Using a window of +/-40kb with respect to the peak, 673 genes are associated with the 1743 Brn2-occupied sites Supplemental Fig. 3F). Of these, 91 are down-regulated upon Brn2 silencing, while 46 are amongst the up-regulated genes (Supplemental Fig. 3F). Even if this analysis is of limited value as we compared Brn2 occupancy at 24 hours with RNA-seq in plated neurons 5 days later, we could still identify 91 genes amongst which are *Neurog1*, *Zic1* and *Zic4* previously identified in RA-induced EBs (Supplemental Fig. 3G and 3H). Moreover, as we performed the ChIP-seq analysis after 24 hours, it is likely that these genes represent the first to be activated by Brn2 following its induction by RA-treatment.

We next compared genome occupancy in P19 cells and RA-induced EBs. Of the 2127 sites occupied in RA-induced EBs, 287 were occupied in P19 cells (Fig. 7A and B). Analysis of the sequences at these sites by MEME identified a large number with one or several MORE and MORE+1 motifs (Fig. 7A and C). These common sites are associated with 202 potential target genes enriched in neurogenesis functions (Fig. 7D) of which 10 are commonly down-regulated upon Brn2 silencing in RA-induced EBs and P19 cells (Fig.7). Amongst these are *Neurog1*, *Pou6f1*, *Zic1* and *Zic4*. Brn2 therefore occupies a distinct set of sites in the two cell

types, with only a limited number of common sites. These sites comprise MORE-type motifs and are associated with a small number of neurogenesis genes that define an early induced gene expression programme common to both cell types.

To extend this analysis, we integrated the data from RA-induced EBs, with that of Fgf2-induced NPCs and P19 cells to compare Brn2 function in three distinct models of neurogenesis. We first compared the P19 data with that of Lodato et al. 500 Brn2 sites are commonly occupied in Fgf2-induced NPCs and P19 cells and 453 are common between the P19 cells and the ES cells in which ectopic Brn2 is expressed under the control of Tet0 (Fig. 7F). Of these sites, 298 are occupied under all 3 conditions. Comparing the 202 sites occupied in ES cells under all conditions (see Fig. 4B) with 298 identified above identified 175 that are shared under all conditions in all model systems. This integrative analysis allows us to define a set of Brn2 occupied sites and potentially regulated genes that form the core of the neurogenesis programme common to each model system.

It has recently been shown that Brn2, Ascl1 and Myt1l cooperate to induce differentiation of fibroblasts into neurons (Vierbuchen et al., 2010). To investigate the mechanisms involved in this process, Wapinski et al, performed Brn2 ChIP-seq in fibroblasts in which these 3 factors had been ectopically expressed to induce differentiation (Wapinski et al., 2013). In these experiments 2344 peaks were identified in fibroblasts and 5278 peaks in the resulting reprogrammed NPCs. An integrative comparison of this data with our RA-induced EBs and P19 cell data identified 86 sites commonly occupied in these data sets (data not shown). Finally, we compared these 86 sites with the 175 common to the RA and Fgf2-induced NPCs and P19 data sets to identify a set of 72 Brn2 occupied sites that are common to all data sets from 4 different model systems (Fig. 8A). These sites are preferentially located in gene promoters or introns (Fig. 8A), are enriched around the TSS (Fig. 8B), and are highly

enriched in MORE and MORE+1 motifs (Fig. 8C). These 72 Brn2 binding sites are associated with and potentially regulate 91 target genes (Supplemental Dataset 3).

### **ZIC1 is essential for neuronal differentiation of RA-treated EBs.**

Amongst the sites that are common to all current Brn2 ChIP-seq data sets, one prominent site lies 5' to the *Zic1* and *Zic4* genes (Fig 8D). The Zic proteins have previously described roles in neural development (Ali et al., 2012; Merzdorf, 2007) and *Zic1* expression is strongly down-regulated in shBrn2 knockdown RA-induced EBs. These observations suggest that *Zic1* may be an important downstream effector of Brn2 function. To test this idea, we used lentivirus vectors to express shRNA directed against *Zic1* to silence its expression. Loss of *Zic1* strongly impaired neuronal differentiation, with cell death and the remaining cells showing a loss of the typical neuron-like morphology (Fig. 9A and B). This was confirmed by RNA-seq identifying 101 down-regulated genes strongly enriched in neurogenesis functions (Fig. 9C and Supplemental Dataset 1). The 133 up-regulated genes on the other hand show diverse functions (Fig.9D).

We next compared the genes down-regulated by Brn2 and *Zic1* silencing showing that 49 of the 101 genes down-regulated by sh*Zic1* were also down-regulated by shBrn2 (Fig. 9E). These genes are strongly enriched in neurogenesis functions, but do not include Brn2 itself. Genes involved in neurogenesis such as *Epha8*, *Cdh8*, *Olig3* or *Zic3* are also found amongst the 52 genes regulated uniquely by *Zic1*, showing that *Zic1* activates a complimentary set of neurogenic genes. In contrast, 33 of the 133 genes induced by sh*Zic1* knockdown are also induced by shBrn2 knockdown, but are not associated with neurogenesis (Fig. 9F). Thus, loss of *Zic1* does not repress *Brn2* expression, thereby situating Brn2 upstream of *Zic1* consistent with the idea that it binds to a site adjacent to *Zic1* to activate its expression. Consequently, *Zic1* and Brn2 may cooperate to regulate a set of downstream targets such as *Pax3*, *Pax7*, *Msx3*, *Zic4* or *Ncan* that are associated with Brn2-occupied sites and are down-regulated by

silencing of *Brn2* and *Zic1*. At present we know of no ChIP-grade antibodies for *Zic1* and thus we are unable to determine which genes are associated with *Zic1*-occupied sites. Nevertheless, as described above, we detected a ZFX-like motif enriched at many *Brn2* bound sites that may be recognised by *Zic1*, in agreement with possible cooperation of these factors.

## **Discussion.**

### **Brn2 is upstream of a neurogenic transcriptional network in RA-treated EBs.**

Here we show that expression of *Brn2* is essential for the neurogenic fate of RA-induced EBs. ShRNA-mediated *Brn2* silencing abrogated the neuronal fate of RA-treated EBs. RA-treated EBs differentiate first into Pax6-expressing radial-glial type NPCs. *In vivo*, radial glial cells are essential for the development of most cortical neurons (Gotz and Barde, 2005; Malatesta et al., 2003). Given the important role of *Brn2* (and *Zic1*) in the developing cortex it is not surprising that *Brn2* plays such a critical role in RA-treated EBs *in vitro*. However, two features distinguish the *in vitro* and *in vivo* scenarios. Firstly, *in vivo* *Brn1* and *Brn2* play redundant roles in the cortex and a clear phenotype is observed only upon double knockout (Sugitani et al., 2002). In contrast, *Brn2* plays a predominant role *in vitro* in EBs cells despite expression of the closely related *Brn1* (*Pou3f3*). *Brn1* mRNA is strongly up-regulated in differentiating EBs and its expression is not significantly affected by *Brn2* silencing. These data indicate that *Brn1* expression cannot compensate for loss of *Brn2* in ES cells *in vitro* a result that contrasts with the partially redundant roles of these factors *in vivo*. Secondly, while *Brn2* is expressed in RA-treated EBs and their derived radial glial cells *in vitro*, it is not expressed in the ventricular zone in the developing cortex, but rather at later stages of differentiation from sub-ventricular zone (Hagino-Yamagishi et al., 1997). *Brn2* does not therefore appear to be expressed in radial glial cells of the ventricular zone *in vivo*.

Despite these differences, the *in vitro* experiments identified a large set of neurogenic genes whose expression requires Brn2 expression. To identify the direct targets, we integrated the RNA-seq data with our ChIP-seq data and found that Brn2 binds to and directly regulates several critical neurogenic factors. Brn2 binds to the proximal promoters of *Ascl1* and *Neurog1* and expression of both of these genes is strongly down regulated by Brn2 silencing. *Ascl1* and *Neurog1* are critical transcription factors in neurogenesis and the loss of their expression strongly contributes to the loss of neural fate. Interestingly however, *Ascl1* expression initiates somewhat before that of *Brn2* in differentiating EBs suggesting that Brn2 is not essential for its initial activation, but is required to maintain its expression at later stages.

Notch signalling is strongly affected by Brn2 silencing in EBs as expression of *Dll1*, *Dll3*, *Jag1*, *Hes3* and *Hes5* were all strongly down-regulated. It has previously been shown that *Ascl1* and Brn2 cooperate to activate *Dll1* *in vivo* in the chick neural tube and in P19 cells (Castro et al., 2006). This cooperativity was dependent on a promoter element comprising an OCT motif flanked by 2 E-boxes as binding sites for *Ascl1*. This type of element is also found in the promoter of *Dll3* and other effectors of Notch signalling as well as other neurogenic genes several of which are down-regulated by Brn2 silencing in RA-induced EBs and/or P19 cells. Nevertheless, in RA-induced EBs and P19 cells, we do not observe Brn2 occupancy of the OCT motif in the *Dll1* promoter by ChIP-seq, nor of the OCT-E-box elements in the *Dll3* promoter or in the other identified genes (although it may be argued that this may be seen at a later stage in P19 cells). In RA-induced EBs, activation of these genes is dependent on Brn2, but this is an indirect effect probably due to their activation by *Ascl1* and *Neurog1* that are direct targets of Brn2. Hence in contrast to the cooperative model suggested by Castro for activation of these target genes *in vivo* (Castro et al., 2006), in RA-induced EBs activation of Notch signalling involves in hierarchical process in which Brn2 activates *Ascl1* and *Neurog1*

that in turn activate the downstream genes of the Notch pathway. Moreover, we note that our data indicate that the OCT motif is not a preferred site for Brn2 that rather binds MORE-type elements. The elements described above do not have the full MORE consensus sequence. It is therefore possible that another POU factor occupies these sites and cooperates with Ascl1 in RA-induced EBs. Furthermore, we did not detect E-box binding factors amongst those that are enriched at MORE and MORE+1 sites. These observations do not support the model of Brn-Ascl1 cooperativity in the differentiation of RA-induced EBs, although this appears to be the case *in vivo* and in fibroblast reprogramming. Finally, we note that other genes of the Notch pathway, including *Notch1* and *Notch2*, *Jag2*, *Hes6* and *Hey2* are associated with Brn2 binding sites and although they are not overtly regulated in differentiating EBs regulation of this pathway appears as a major function of Brn2.

While loss of function identified a set of genes directly regulated following Brn2 silencing, we also performed gain of function experiments by ectopic Brn2 expression as the shRNA-mediated silencing is incomplete and it is possible that the remaining low Brn2 levels could be sufficient to activate certain genes resulting in an under estimate of the number of its targets. Brn2 overexpression identified a further 44 genes whose expression was induced and had associated Brn2-occupied sites. Aside from the genes that are directly regulated by Brn2 in RA-induced NPCs, examination of potential target genes based on their proximity to Brn2-occupied sites and that may be regulated in other tissues and cells revealed many sites in the Hox clusters suggesting that Brn2 may play a role in modulating Hox gene expression.

**Definition of a potential core neurogenic Brn2-regulated regulatory network for neurogenesis identifies a Brn2-Zic1 axis.**

As with most transcription factors, only a small number of Brn2-occupied sites are likely to be functional. In an attempt to define which sites and target genes may be the most

critical for Brn2 function in neurogenesis, we compared our ChIP-seq data with other public data sets. We were surprised to find that Brn2 genome occupancy was very different in Fgf2- and RA induced NPCs. Lodato defined a series of Brn2-Sox2 occupied enhancers involved in NPC differentiation (Lodato et al., 2013), while we found little evidence for association with Sox2 at Brn2 sites in RA-induced EBs and few of these were located in enhancers marked with H3K27ac in Fgf2-induced NPCs. This reflects the fact that only a subset of Brn2 sites is commonly occupied in the two populations and probably also the existence of a distinct set of enhancers in the two populations. Moreover, integration of data from P19 cells and reprogrammed fibroblasts indicated that only a small set of Brn2 sites are common between any of the data sets and we identified only 72 sites common to all.

It has been suggested that Brn2 acts together with Sox2 at a set of distal enhancers. Nevertheless, analysis of the Brn2 sites associated with genes that are regulated by Brn2 in RA-induced EBs showed that these sites are strongly enriched around the TSS suggesting that they are located in the proximal promoter rather than distal enhancers. As mentioned above this is the case for the sites located at the critical *Ascl1* and *Neurog1* genes. In addition, these sites are not associated with Sox2 sites. Moreover, the 72 sites that are common to all systems are enriched around the TSS. Thus, while Brn2 may function at distal enhancers it also plays an important functional role at the proximal promoter.

Some of these differences may be accounted for dynamic genome occupancy of Brn2 that may be constantly changing its repertoire of target genes as differentiation proceeds and/or they may reflect inherent differences in the pathways that induce neurogenesis in the different models. Irrespective of the underlying causes, the common set of sites revealed by the above comparison may reveal a core set of sites and target genes playing a critical role in the neurogenesis programme controlled by Brn2. Amongst the set of 91 Brn2 target genes that are potential targets in all of the analysed data sets is the atypical cadherin Fat4. Fat4

expression is strongly induced in RA-treated EBs and is directly regulated by Brn2, while the related Fat1 gene is expressed in RA-treated EBs and associated with a Brn2-occupied site, but is not regulated by Brn2 knockdown. Fat4 is involved in planar cell polarity signalling and mutations in human FAT4 and its receptor DCHS1 are associated with Van Maldergem syndrome characterised by abnormal positioning of neurons in the cerebral cortex of patients (Cappello et al., 2013). Knockdown of these proteins in the mouse show they are important in neuro-progenitor cell proliferation, maintenance of the progenitor state and positioning of cells within the proliferative zones of the developing brain. Similarly, Fat4 is also important for facial migration of branchiomotor neurons (Zakaria et al., 2014) and for cranial neural tube formation (Saburi et al., 2012). All of these data point to an important role of Fat4 in the differentiation and migration of cortical neurons, hallmarks of Brn2 function during brain development in mouse. These observations suggest that Fat4 may be a critical downstream target gene of Brn2 in EBs and *in vivo* in cortex development.

Amongst the other common Brn2 targets is Pou6f1 (Brn5) that is highly expressed in the brain. Pou6f1 is also expressed in myelinating Schwann cells whose development is regulated by Brn2 (Jaegle et al., 2003). Pou6f1 may therefore be a direct downstream target of Brn2 in these cells.

Our comparison further identified a common and prominent Brn2 site located in the locus encoding Zic1 and Zic4 and expression of these genes was down-regulated following Brn2 knockdown. As Zic1 and Zic4 have been associated with Dandy-Walker syndrome and development of the cerebellum (Merzdorf, 2007); (Inoue et al., 2007); (Grinberg et al., 2004), we asked whether they played a role in the neurogenic fate of RA-treated EBs. Indeed, we found that Zic1 silencing abrogated neuronal differentiation, but that Brn2 expression was not significantly down-regulated. These results show that Brn2 is essential, but not sufficient to specify neuronal fate and that it must cooperate with its downstream target Zic1. Amongst

the target genes that are regulated by Brn2 and Zic1 are *Wnt1* and *Wnt3* consistent with the known role of Zic1 in regulating Wnt expression in neuroectoderm patterning and development of the neural crest (Merzdorf, 2007) (Merzdorf and Sive, 2006). These results thus reveal the critical role of a Brn2-Zic1 axis is specifying the neuronal fate of RA treated ES cells.

Interestingly, we noted that a GC-rich ZFX-like motif was often found in association with the MORE motif. This motif shows only limited homology to previously defined Zic-binding motifs identified by *in vitro* Selex and EMSA experiments and for the moment no ChIP-seq data are available for Zic1 that would define its real *in vivo* consensus binding sequence. A full analysis of the role of Zic1, analogous to that made for Brn2, is hampered by the lack of a ChIP-grade antibody. Nevertheless, we are currently testing several of these sites for their ability to bind Zic1 by EMSA to determine if the co-occurrence of the MORE and Zfx-like motifs may constitute a sequence element that would explain the Brn2-Zic1 cooperativity in neurogenesis.

## **Materials and Methods**

### **Cell Culture and generation of stable cell lines**

Lentiviral cell lines were generated using shRNA vectors obtained from Sigma (Mission sh-RNA series) in the PLKO vector. Viral supernatants were obtained by transfection of HEK293T cell lines with the PLKO-shBrn2 (shBrn2 sequence : 5'-AGGCGGATCAAACCTCGGATTT-3'), PLKO-shZic1 (shZic1 sequence : 5'-CCAGCGCTGCTTTCAATTCTA-3')). Co-transfection with PLP1, PLP2 and PLP-VSVG vectors (Sigma) was used to produce viral particles. Supernatants of co-transfected HEK293T cells were collected 48h after transfection and used to infect ES or P19 cells. Stable ES/P19 cells were selected 2 days after infection with puromycin (1µg/ml). Western blots were performed with anti-Brn2 antibody (Santa Cruz SC-6029)

For generation of ES or P19 lines overexpressing Brn2, cells were co-transfected with a vector containing 3HA-Brn2 under the control of an RA-inducible promoter (5x DR5) and a second vector encoding puromycin resistance. Transfected cells were selected with 1µg of puromycin and overexpression of tagged protein was verified by western blot using the 12CA5 HA antibody (Roche, Basel, Switzerland).

### **Neuronal differentiation of ES and P19 cells**

ES J1 cell differentiation was performed as previously described (Bibel et al., 2007; Bibel et al., 2004). Briefly, ES cells were grown and amplified on feeder cells in DMEM (4,5g/l glucose) w/GLUTAMAX-I containing 15%FCS and LIF. After feeder depletion,  $4 \cdot 10^6$  ES cells were plated on non adherent bacterial dishes, in medium containing 10% serum to form EBs that were treated with  $5 \cdot 10^{-6}$ M of all-trans retinoic acid for 4 days. Samples were collected at several time points to do further RNA or protein extraction. Neuronal differentiation was completed by dissociation of EBs using Trypsin-EDTA 0,05% (Invitrogen) and  $2,5 \cdot 10^6$  cells were plated on a 25mm<sup>2</sup> plate coated with poly-DL-ornithine and Laminine. After 1 day in DMEM/HamF12 medium complemented with BSA and N2-supplement, early-differentiated neurons were obtained.

P19 cells were differentiated into neurons using a protocol adapted from previously described experiments (Gao et al., 2001). Briefly, P19 cells are amplified in DMEM (1g/l glucose) medium with 10% of FCS. To perform neuronal differentiation,  $3 \cdot 10^6$ M cells were plated in P10 bacterial dish (Greiner), in presence of  $5 \cdot 10^{-6}$ M RA. After 4 days, aggregates were dissociated in trypsin (2,5%) and  $2,5 \cdot 10^6$  cells plated on a 60mm<sup>2</sup> plate coated with poly-L-lysine.

### **mRNA preparation, quantitative PCR and RNA-seq.**

RNA isolation was performed according to standard procedure using the kit from Sigma. qRT-PCR was carried out with SYBR® Green I (Qiagen) and Multiscribe Reverse Transcriptase (Invitrogen) and monitored using a LightCycler® 480 (Roche). Primer sequences for cDNA were designed using Primer3Plus software. *Gapdh* expression was used as to normalise the results.

For RNA-sequencing, mRNA was isolated from two independent experiments. Messenger-RNA-seq was performed essentially as described (Herquel et al., 2013) with libraries of template molecules suitable for high throughput DNA sequencing created using “TruSeq™ RNA Sample Preparation v2 Kit” (Illumina). Reads were mapped onto the mm9 assembly of the *Mus-musculus* genome by using Tophat version 1.4.1 (Trapnell et al., 2009) (Trapnell and Salzberg, 2009) and Bowtie version 0.12.7 (Langmead et al., 2009). Quantification of gene expression was performed using HTSeq-0.5.3p5. Comparisons were performed using the statistical method proposed by Anders & Hubner (Anders and Huber, 2010). Resulting p-values were adjusted for multiple testing by using the Benjamini & Hochberg method (Benjamini and Hechtlinger, 2014; Benjamini and Yekutieli, 2005). Significantly deregulated genes were selected using a log2 fold change >1 or <-1 and an Adjusted Pvalue cutoff of 0,1. Gene ontology analysis were performed using the functional annotation clustering function of DAVID (<http://david.abcc.ncifcrf.gov/home.jsp>)

### **ChIP seq analysis**

ChIP-seq experiments were performed on chromatin from EBs or P19 aggregates dissociated by douncing directly in crosslinking solution (PFA 0,4%). Sonication and ChIP were performed as previously described (Herquel et al., 2013). The ChIP-grade Brn2 antibody (Sc 6029-X) was used for immunoprecipitation.

ChIP-seq libraries were prepared and sequenced as previously described (Choukrallah et al., 2012) (Herquel et al., 2013). Peak detection was performed using the MACS software

(<http://liulab.dfci.harvard.edu/MACS/>) (Zhang et al., 2008) using the corresponding input as control with default settings. *De novo* motif search was done using MEME-ChIP software (<http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi>). ChIP-seq peak annotation was performed using Homer, individual motif occurrences were detected by FIMO, both softwares are available on IGBMC Galaxy platform ([www.galaxeast.fr](http://www.galaxeast.fr)). Co-occurrence analysis was performed as follows : The analysis was initiated by detecting known motifs using FIMO within regions of interest given as input. From FIMO results, two perl scripts computed the frequency of occurrence of each motif and of co-occurring motifs. Two motifs were considered as co-occurring if they are present in the same sequence. This analysis was done once on the input BED file and n times on randomly selected regions. Region size distribution of the randomly selected regions was the same as for the input BED file. Randomly selected regions were used to compute an expected distribution of motif occurrence or motif co-occurrence. The significance of the motif occurrence and co-occurrence was estimated through the computation of a Z-score computed as :  $z=(x-\mu)/\sigma$ , where x is the observed value (number of motif occurrence or co-occurrence),  $\mu$  is the mean of the number of occurrences or co-occurrences of motifs (computed over controls) and  $\sigma$  is the standard deviation of the number of occurrences or co-occurrences of motifs (computed over controls). Z-scores were subsequently turned into p-values and a multi-testing correction was applied (Benjamini and Hechtlinger, 2014; Benjamini and Yekutieli, 2005). Common peaks in the different data sets were called using the 'Intersect' function of Galaxeast that detects peaks with at least one overlapping nucleotide within a window of 200 nucleotides centered on the peak summit.

## **Acknowledgements.**

We thank, C. Keime and A. Velt for RNA-seq analysis, B. Jost, and all the staff of the IGBMC high throughput sequencing facility, a member of “France Génomique” consortium (ANR10-INBS-09-08) and M. Hestin and G. Rossi for assistance with ES cell culture. This work was supported by grants from the CNRS, the INSERM, the Ligue Nationale et Départementale Région Alsace contre le Cancer and the Institut National du Cancer (INCa). ID is an ‘équipe labellisée’ of the Ligue Nationale contre le Cancer ,the ANR-10-LABX-0030-INRT French state fund through the Agence Nationale de la Recherche under the frame programme Investissements d’Avenir labelled ANR-10-IDEX-0002-02..

## **Author contributions.**

S.U. performed and analysed all the experiments on RA-treated EBs and analysed the data from the P19 cells and public data sets together with D.K and S.LG. D.K. performed the experiments on the P19 cells. S.LG. and T.Y assisted in bioinformatics analysis of the motif co-occurrence and peak detection respectively. I.D. S.U. and D.K. conceived experiments, analysed data, I.D. and S.U. wrote the paper.

## **References.**

- Agoston, Z., Heine, P., Brill, M.S., Grebbin, B.M., Hau, A.C., Kallenborn-Gerhardt, W., Schramm, J., Gotz, M., and Schulte, D. (2014). Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development* *141*, 28-38.
- Al Tanoury, Z., Gaouar, S., Piskunov, A., Ye, T., Urban, S., Jost, B., Keime, C., Davidson, I., Dierich, A., and Rochette-Egly, C. (2014). Phosphorylation of the retinoic acid receptor RARgamma2 is crucial for the neuronal differentiation of mouse embryonic stem cells. *J Cell Sci* *127*, 2095-2105.

- Alazard, R., Blaud, M., Elbaz, S., Vossen, C., Icre, G., Joseph, G., Nieto, L., and Erard, M. (2005). Identification of the 'NORE' (N-Oct-3 responsive element), a novel structural motif and composite element. *Nucleic Acids Res* 33, 1513-1523.
- Ali, R.G., Bellchambers, H.M., and Arkell, R.M. (2012). Zinc fingers of the cerebellum (Zic): transcription factors and co-factors. *Int J Biochem Cell Biol* 44, 2065-2068.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Ballas, N., and Mandel, G. (2005). The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol* 15, 500-506.
- Balmer, J.E., and Blomhoff, R. (2005). A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J Steroid Biochem Mol Biol* 96, 347-354.
- Benjamini, Y., and Hechtlinger, Y. (2014). Discussion: An estimate of the science-wise false discovery rate and applications to top medical journals by Jager and Leek. *Biostatistics* 15, 13-16; discussion 39-45.
- Benjamini, Y., and Yekutieli, D. (2005). Quantitative trait Loci analysis using the false discovery rate. *Genetics* 171, 783-790.
- Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* 2, 1034-1043.
- Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci* 7, 1003-1009.
- Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M.K., Vriend, G., and Scholer, H.R. (1998). New POU dimer configuration mediates antagonistic control of an

- osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev* 12, 2073-2090.
- Cao, X., Pfaff, S.L., and Gage, F.H. (2008). YAP regulates neural progenitor cell number via the TEA domain transcription factor. *Genes Dev* 22, 3320-3334.
- Cappello, S., Gray, M.J., Badouel, C., Lange, S., Einsiedler, M., Srour, M., Chitayat, D., Hamdan, F.F., Jenkins, Z.A., Morgan, T., *et al.* (2013). Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development. *Nat Genet* 45, 1300-1308.
- Castro, D.S., and Guillemot, F. (2011). Old and new functions of proneural factors revealed by the genome-wide characterization of their transcriptional targets. *Cell Cycle* 10, 4026-4031.
- Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J.A., Nguyen, L., Gossler, A., Gottgens, B., *et al.* (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev Cell* 11, 831-844.
- Chebaro, Y., Amal, I., Rochel, N., Rochette-Egly, C., Stote, R.H., and Dejaegere, A. (2013). Phosphorylation of the retinoic acid receptor alpha induces a mechanical allosteric regulation and changes in internal dynamics. *PLoS Comput Biol* 9, e1003012.
- Choukrallah, M.A., Kobi, D., Martianov, I., Pijnappel, W.W., Mischerikow, N., Ye, T., Heck, A.J., Timmers, H.T., and Davidson, I. (2012). Interconversion between active and inactive TATA-binding protein transcription complexes in the mouse genome. *Nucleic Acids Res* 40, 1446-1459.
- Cook, A.L., and Sturm, R.A. (2008). POU domain transcription factors: BRN2 as a regulator of melanocytic growth and tumourigenesis. *Pigment Cell Melanoma Res* 21, 611-626.

- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* *107*, 21931-21936.
- Fujii, H., and Hamada, H. (1993). A CNS-specific POU transcription factor, Brn-2, is required for establishing mammalian neural cell lineages. *Neuron* *11*, 1197-1206.
- Gao, X., Bian, W., Yang, J., Tang, K., Kitani, H., Atsumi, T., and Jing, N. (2001). A role of N-cadherin in neuronal differentiation of embryonic carcinoma P19 cells. *Biochem Biophys Res Commun* *284*, 1098-1103.
- Gotz, M., and Barde, Y.A. (2005). Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons. *Neuron* *46*, 369-372.
- Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W.B., and Millen, K.J. (2004). Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat Genet* *36*, 1053-1055.
- Hagino-Yamagishi, K., Saijoh, Y., Ikeda, M., Ichikawa, M., Minamikawa-Tachino, R., and Hamada, H. (1997). Predominant expression of Brn-2 in the postmitotic neurons of the developing mouse neocortex. *Brain Res* *752*, 261-268.
- Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R., *et al.* (2013). Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. *Nature structural & molecular biology* *20*, 339-346.
- Hogan, B.L. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev* *10*, 1580-1594.

- Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K., and Aruga, J. (2007). Zic1 and Zic3 regulate medial forebrain development through expansion of neuronal progenitors. *J Neurosci* 27, 5461-5473.
- Jaegle, M., Ghazvini, M., Mandemakers, W., Piirsoo, M., Driegen, S., Levavasseur, F., Raghoenath, S., Grosveld, F., and Meijer, D. (2003). The POU proteins Brn-2 and Oct-6 share important functions in Schwann cell development. *Genes Dev* 17, 1380-1391.
- Khorasanizadeh, S., and Rastinejad, F. (2001). Nuclear-receptor interactions on DNA-response elements. *Trends Biochem Sci* 26, 384-390.
- Klemm, J.D., and Pabo, C.O. (1996). Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev* 10, 27-36.
- Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., and Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* 77, 21-32.
- Koga, M., Matsuda, M., Kawamura, T., Sogo, T., Shigeno, A., Nishida, E., and Ebisuya, M. (2014). Foxd1 is a mediator and indicator of the cell reprogramming process. *Nat Commun* 5, 3197.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* 9, e1003288.

- Mahony, S., Mazzoni, E.O., McCuine, S., Young, R.A., Wichterle, H., and Gifford, D.K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome biology* 12, R2.
- Malatesta, P., Hack, M.A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., and Gotz, M. (2003). Neuronal or glial progeny: regional differences in radial glia fate. *Neuron* 37, 751-764.
- Mark, M., Ghyselinck, N.B., and Chambon, P. (2006). Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu Rev Pharmacol Toxicol* 46, 451-480.
- McEvelly, R.J., de Diaz, M.O., Schonemann, M.D., Hooshmand, F., and Rosenfeld, M.G. (2002). Transcriptional regulation of cortical neuron migration by POU domain factors. *Science* 295, 1528-1532.
- Merzdorf, C.S. (2007). Emerging roles for zic genes in early development. *Dev Dyn* 236, 922-940.
- Merzdorf, C.S., and Sive, H.L. (2006). The zic1 gene is an activator of Wnt signaling. *Int J Dev Biol* 50, 611-617.
- Millevoi, S., Thion, L., Joseph, G., Vossen, C., Ghisolfi-Nieto, L., and Erard, M. (2001). Atypical binding of the neuronal POU protein N-Oct3 to noncanonical DNA targets. Implications for heterodimerization with HNF-3 beta. *Eur J Biochem* 268, 781-791.
- Moutier, E., Ye, T., Choukrallah, M.A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D., *et al.* (2012). Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *J Biol Chem* 287, 26328-26341.

- Niederreither, K., and Dolle, P. (2008). Retinoic acid in development: towards an integrated view. *Nat Rev Genet* 9, 541-553.
- Nieto, L., Joseph, G., Stella, A., Henri, P., Burlet-Schiltz, O., Monsarrat, B., Clottes, E., and Erard, M. (2007). Differential effects of phosphorylation on DNA binding properties of N Oct-3 are dictated by protein/DNA complex structures. *J Mol Biol* 370, 687-700.
- Nikoletopoulou, V., Plachta, N., Allen, N.D., Pinto, L., Gotz, M., and Barde, Y.A. (2007). Neurotrophin receptor-mediated death of misspecified neurons generated from embryonic stem cells lacking Pax6. *Cell Stem Cell* 1, 529-540.
- Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech Dev* 59, 89-102.
- Phillips, K., and Luisi, B. (2000). The virtuoso of versatility: POU proteins that flex to fit. *J Mol Biol* 302, 1023-1039.
- Ramos, C., and Robert, B. (2005). msh/Msx gene family in neural development. *Trends Genet* 21, 624-632.
- Rochel, N.C., F. Godet, J. Roessle, M. Peluso-Iltis, C. Svergun, DI. Moras, D. (2011). Common architecture of nuclear receptor heterodimers on DNA direct repeats. *Nature Structural and Molecular Biology* *in press*.
- Rosenfeld, M.G., Lunyak, V.V., and Glass, C.K. (2006). Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev* 20, 1405-1428.
- Ryan, A.K., and Rosenfeld, M.G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev* 11, 1207-1225.

- Saburi, S., Hester, I., Goodrich, L., and McNeill, H. (2012). Functional interactions between Fat family cadherins in tissue morphogenesis and planar polarity. *Development* *139*, 1806-1820.
- Schonemann, M.D., Ryan, A.K., Erkman, L., McEvilly, R.J., Bermingham, J., and Rosenfeld, M.G. (1998). POU domain factors in neural development. *Adv Exp Med Biol* *449*, 39-53.
- Sugitani, Y., Nakai, S., Minowa, O., Nishi, M., Jishage, K., Kawano, H., Mori, K., Ogawa, M., and Noda, T. (2002). Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. *Genes Dev* *16*, 1760-1765.
- Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M., and Scholer, H.R. (2000). Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell* *103*, 853-864.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105-1111.
- Trapnell, C., and Salzberg, S.L. (2009). How to map billions of short reads onto genomes. *Nat Biotechnol* *27*, 455-457.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* *463*, 1035-1041.
- Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., *et al.* (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* *155*, 621-635.
- Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell* *110*, 385-397.

Yamada, T., Urano-Tashiro, Y., Tanaka, S., Akiyama, H., and Tashiro, F. (2013). Involvement of crosstalk between Oct4 and Meis1a in neural cell fate decision. *PLoS One* 8, e56997.

Zakaria, S., Mao, Y., Kuta, A., Ferreira de Sousa, C., Gaufo, G.O., McNeill, H., Hindges, R., Guthrie, S., Irvine, K.D., and Francis-West, P.H. (2014). Regulation of neuronal migration by dchs1-fat4 planar cell polarity. *Curr Biol* 24, 1620-1627.

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

### **Legends to Figures.**

**Figure 1.** Brn2 is required for neuronal differentiation of RA-treated EBs. **A.** The protocol for neuronal differentiation of mouse ES cells is schematised. **B-C.** Brn2 expression in RA-treated EBs was measured by RT-qPCR (B) or by immunoblot (C). **D-E.** The down-regulated expression of Brn2 was measured by RT-qPCR (D) and by immunoblot (E). **F.** Phase contrast microscopy images of cells expressing the indicated shRNAs 24 hours after dissociation of the EBs. Magnification X20. **G** Relative gene expression in shBrn2 and shC cells was measured by RT-qPCR at the indicated times. **H-I** Ontology analysis of genes whose expression was up-or down-regulated following Brn2 knockdown.

**Figure 2.** Genome occupancy of Brn2 in RA-treated EBs. **A.** Pie-chart showing genomic distribution of Brn2 sites with enrichment in the promoter region. **B.** Brn2 sites are enriched around the TSS. Graph shows a plot of the frequency of Brn2 occupied sites versus the distance to TSS. **C.** UCSC screenshots of the indicated loci illustrating representative Brn2 binding sites. **D.** Table indicating the frequency of occurrence of the indicated sequence

motifs at the Brn2-occupied sites. **E.** Analysis of the sequences at 600 Brn2 binding sites using MEME-ChIP identifies the MORE and MORE+1 motifs as the two most enriched motifs centered over the peak. A more degenerate Zfx-like motif was also detected. **F.** The co-occurrence of transcription factor binding motifs with MORE and MORE+1 at Brn2-occupied sites is represented as a heatmap matrix. Arrows indicate motifs for several transcription factors that are relevant for neurogenesis.

**Figure 3.** Gain of Brn2 function identifies additional target genes. **A-B.** RT-qPCR and anti-HA western blot analysis show Brn2 overexpression in RA-treated EBs at the indicated times. **C.** RT-qPCR shows increased expression of Brn2 target genes upon gain of function and stronger repression of POU5F1 expression. **D-E.** Ontology analysis of genes showing enhanced activation or repression upon Brn2 gain of function. **F-G.** Comparison of genes showing complementary expression profiles upon loss and gain of Brn2. **H.** Comparison of RNA-seq and ChIP-seq data identifies additional genes up- or down-regulated by Brn2 gain of function associated with Brn2 occupied sites. **I.** List of genes that are potential direct targets identified in panel H. **J.** Identification of 15 genes that show complimentary expression patterns in Brn2 loss and gain of function and association with Brn2-occupied sites.

**Figure 4.** Distinct profiles of Brn2 genomic occupancy in RA-treated EBs and Fgf2-induced NPCs. **A.** Re-analysis of public data sets using stringent MACS criteria reduces the number of detected Brn2 peaks. Table shows the previously reported numbers of detected sites and the results of the more stringent analysis. **B.** Venn diagrammes indicating overlap between the Brn2 sites in the indicated datasets. **C.** MEME-ChIP analysis of Brn2-occupied sites in NPCs identifies an enriched MORE-like motif. **D.** MEME-ChIP analysis of Brn2-occupied sites following ectopic Brn2 expression in ES cells reveals enrichment in MORE and MORE+1 and OCT motifs. **E.** MORE-like motifs are strongly enriched in the 202 Brn2 sites shared

between RA-treated EBs and NPCs and ES cells expressing ectopic Brn2. **F.** Sox2 binding sequence derived from MEME-ChIP analysis of Sox2 bound sites in NPCs. **G.** Venn diagrams indicating overlap between the Brn2- and Sox2-occupied sites in the indicated data sets. **H.** Table showing overlap between Brn2 binding sites and H3K4me1 and H3K27ac-marked enhancers.

**Figure 5.** Brn2 is required for RA-induced neuronal differentiation of P19 cells. **A.** The differentiation protocol for P19 cells is schematised. **B-C.** RT-qPCR and western blot analysis of Brn2 expression during P19 cells differentiation. **D-E.** RT-qPCR and western blot analysis of showing reduced Brn2 expression in P19 cells expressing shRNA. **F.** Phase contrast microscopy images of P19 cells expressing the indicated shRNAs 24 hours after dissociation of aggregates. Magnification X20. **G.** Comparative gene expression in differentiating P19 cells and EBs following shBrn2 knockdown. **H-I.** Ontology analysis of genes whose expression was up-or down-regulated following Brn2 knockdown.

**Figure 6.** Genome occupancy of ectopically expressed Brn2 in RA-treated P19 cells. **A.** Pie-chart showing distribution of Brn2 sites in the genome **B.** Brn2 sites are not enriched around the TSS. Graph shows a plot of the frequency of Brn2 occupied sites versus the distance to TSS. **C.** UCSC screenshots of the indicated loci illustrating representative Brn2 binding sites. **D.** Table indicating the frequency of occurrence of the indicated sequence motifs at the Brn2-occupied sites. **E.** Analysis of the sequences at 600 Brn2 binding sites using MEME-ChIP reveals enrichment in MORE-like motifs centered over the peak.

**Figure 7.** Comparison of Brn2-occupied sites in RA-treated EBs and P19 cells. **A.** Venn diagrams indicating the overlap between the Brn2 sites in the indicated data sets. A MEME-ChIP analysis of the commonly occupied sites shows enrichment in MORE-like motifs. **B.** UCSC screenshot showing representative examples of common and specific peaks. **C.** Table indicating the frequency of the indicated motifs at common and specific sites. **D.**

Ontology of the genes associated with the Brn2 sites commonly occupied in EBs and P19 cells. **E.** Venn diagrammes indicating the overlap between genes associated with Brn2-occupied sites and down-regulation following shBrn2 silencing in EBs and P19 cells. **F.** Venn diagrammes indicating the overlap between genes associated with Brn2-occupied sites in P19 cells and NPCs with or without ectopic Brn2 expression.

**Figure 8.** Identification of a common core set of Brn2 occupied sites and target genes. **A.** Pie-chart showing distribution of the core set of 72 Brn2 sites in the genome **B.** A subset of core sites are enriched around the TSS. Graph shows a plot of the frequency of Brn2 occupied sites versus the distance to TSS. **C.** Core sites are highly enriched in MORE-like motifs. **D.** Representative UCSC screenshots of the indicated loci illustrating a conserved Brn2 binding site in the *Zic1-Zic4* locus.

**Figure 9.** *Zic1* is essential for neuronal differentiation of EBs. **A.** Phase contrast microscopy images of cells expressing the indicated shRNAs 24 hours after dissociation of the EBs. Magnification X20. **B.** Down-regulation of *Zic1* expression measured by RT-qPCR. **C-D.** Ontology analysis of genes whose expression was up or down-regulated following *Zic1* knockdown. **E-F.** Venn diagrammes illustrating overlap between genes up and down-regulated by shBrn2 and sh*Zic1* silencing along with ontology analysis.

### **Legends to Supplemental Figures and Datasets.**

**Supplemental Figure 1.** Gene expression in differentiating EBs. Each graph shows the relative gene expression levels of the indicated genes and the indicated times.

**Supplemental Figure 2.** Gene expression changes upon shBRN2 silencing in EBs. **A-B.** List of genes down and up-regulated by shBrn2 silencing in the indicated ontology clusters. **C.** Pie charts showing genomic distribution of Brn2-occupied sites with the indicated motifs. **D.** Venn diagrammes illustrating overlap between up and down-regulated genes and genes

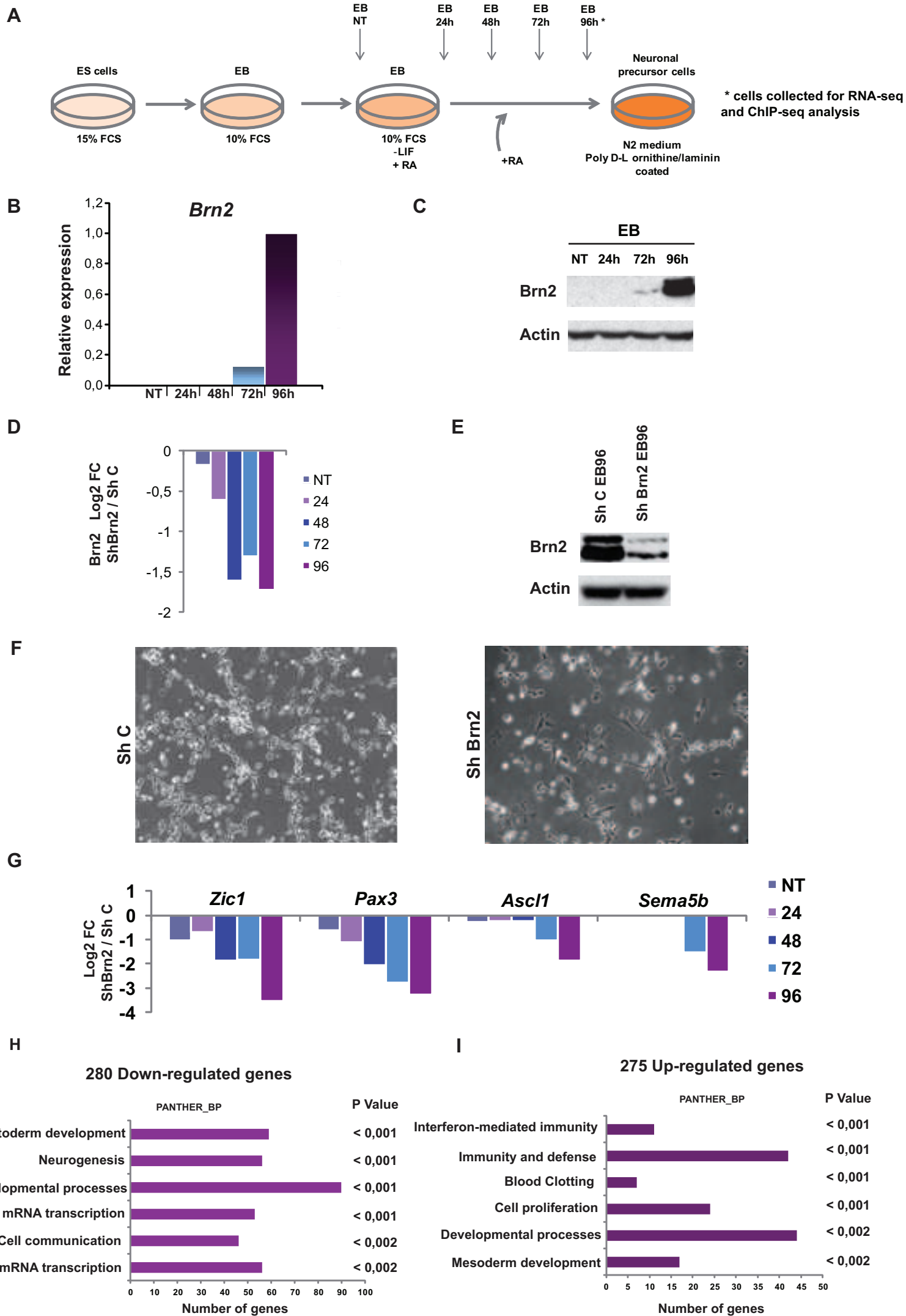
associated with Brn2 occupied sites. **E.** Ontology analysis of genes associated with Brn2 occupied sites. **F.** List of genes with associated Brn2-occupied sites assigned to the neurogenesis ontology cluster. **G.** List of potential direct Brn2 target genes. **H.** Brn2-occupied sites at potential direct targets are enriched around the TSS.

**Supplemental Figure 3.** Gene expression changes upon shBRN2 silencing in P19 cells. **A-B.** List of genes down and up-regulated by shBrn2 silencing in the indicated ontology clusters. **C-D.** Venn diagrammes illustrating overlap between up and down-regulated genes in EBs and P19 cells, along with the ontology analysis of the shared genes. **E.** Western blot analysis of Brn2 expression levels in EBs and P19 aggregates. **F.** Venn diagrammes illustrating overlap between up and down-regulated genes and genes with associated Brn2-occupied sites in P19 cells. **G.** Ontology analysis of the 100 potential direct target genes. **H.** List of genes from the neurogenesis cluster in panel G.

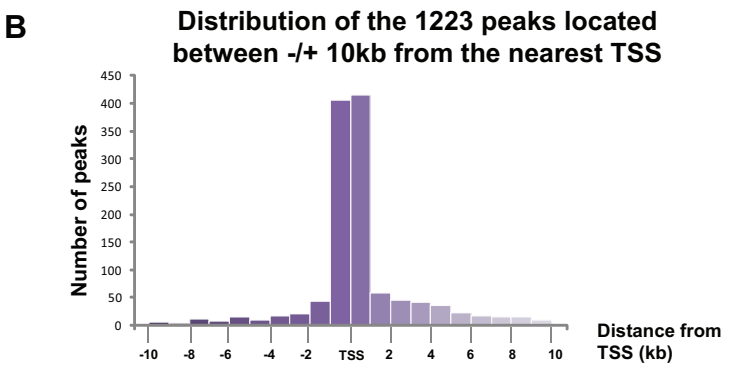
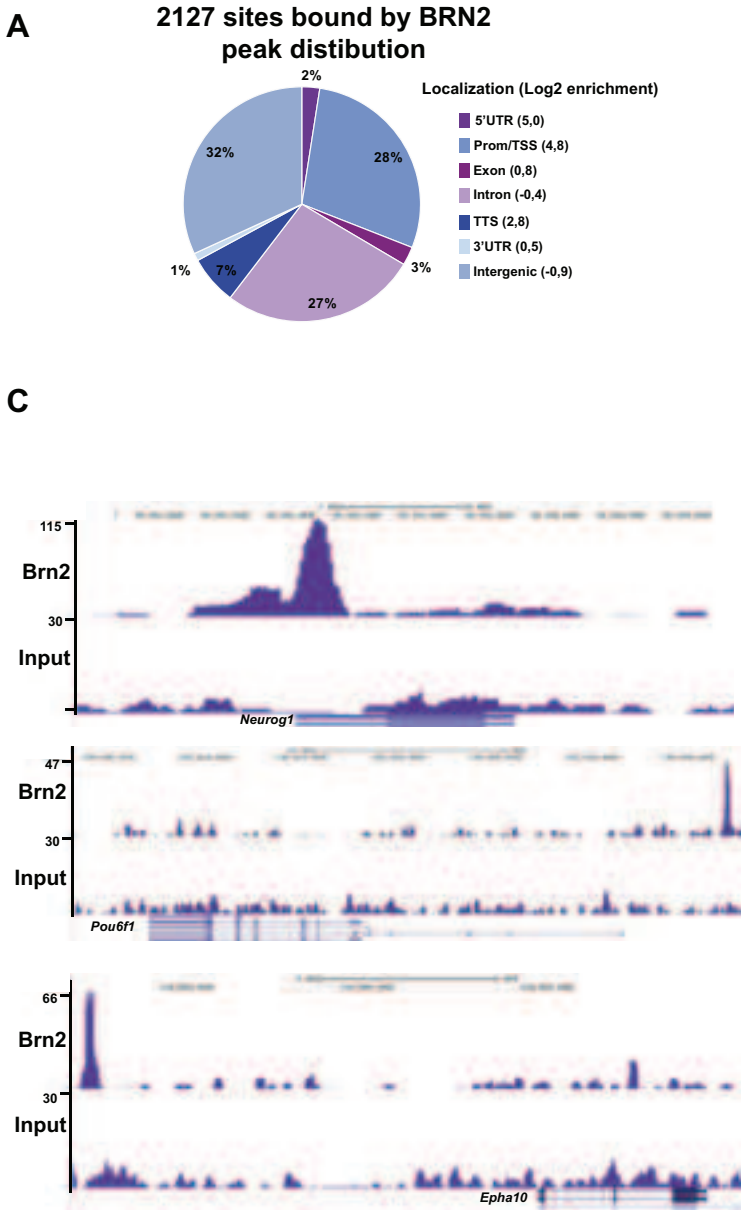
**Supplemental Dataset 1.** Results of RNA-seq following shBrn2 knockdown in EBs and P19 cells and shZic1 knockdown in EBs. Odd numbered pages show the results of RNA-seq in the indicated cell types. Up- and down-regulated genes are shown on separate pages. In each case the Ensembl ID is indicated along with the normalised RPKM values for the duplicate shBrn2 (Zic1) and ShC (control) cells, gene name, gene description, Log2 change in expression and the adjusted p-value. The even numbered pages show the ontology analysis of the up- and down-regulated genes from each condition. The results represent the Panther-BP-All ontology class from David.

**Supplemental Dataset 2.** Results of Brn2 ChIP-seq in EBs and P19 cells. Pages 1 and 2 show the results of the MACS-GREAT analysis for EBs and P19 cells respectively. In each case are shown: the peak coordinates, length summit, number of reads in peak, fold enrichment and false discovery rate from MACS and : the GREAT annotation of each peak with associated genes and distance from TSS.

**Supplemental Dataset 3.** Common Brn2-occupied sites. Page 1 shows the MACS-GREAT analysis of the 72 Brn2-occupied sites common to all data sets. Page 2 shown the list of 91 genes associated with these 72 sites.



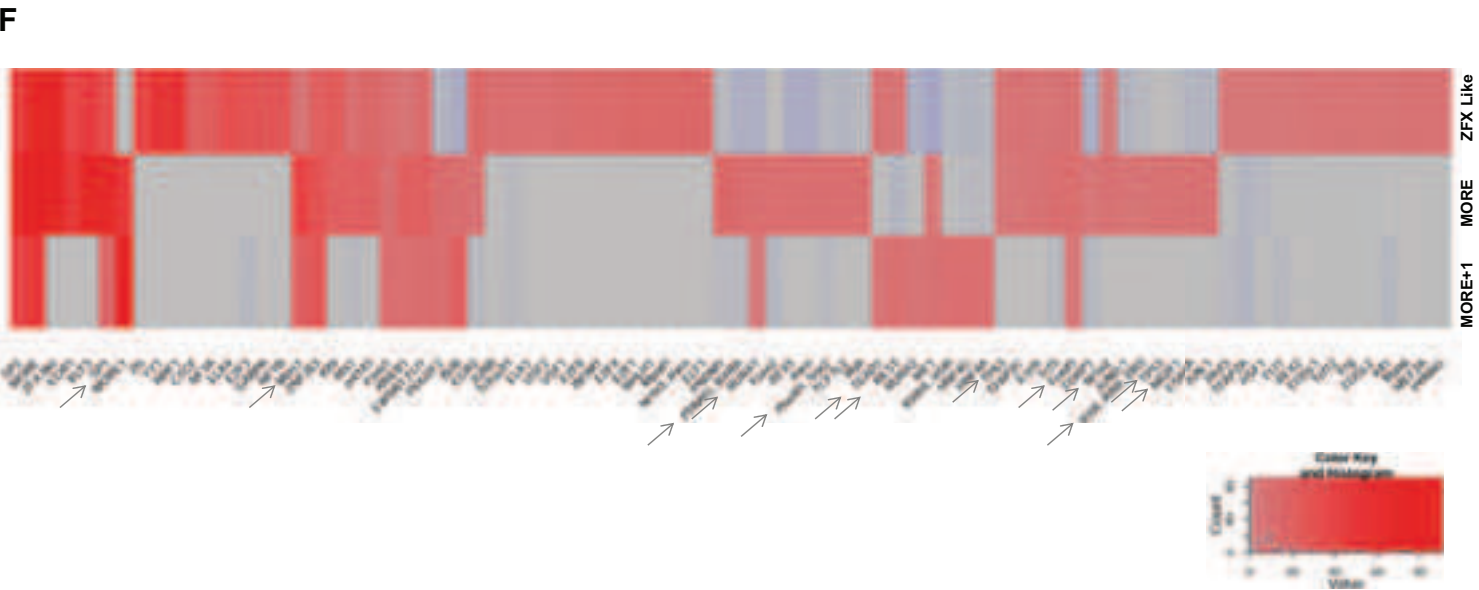
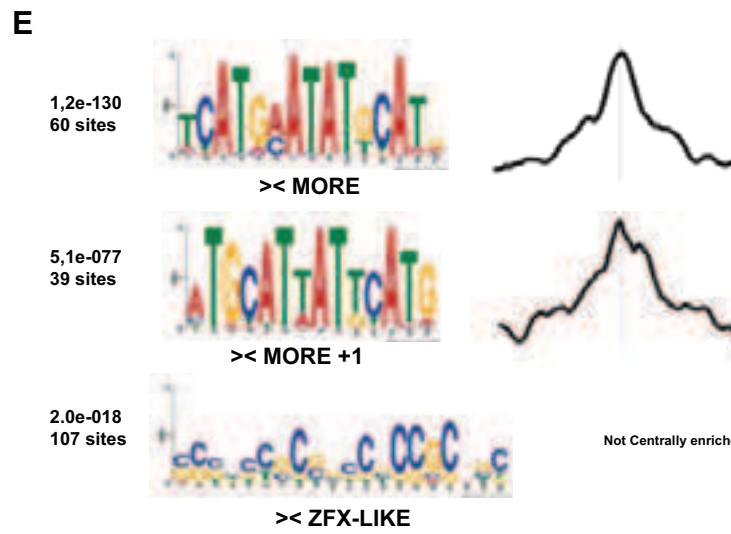
Urban et al., Fig. 1



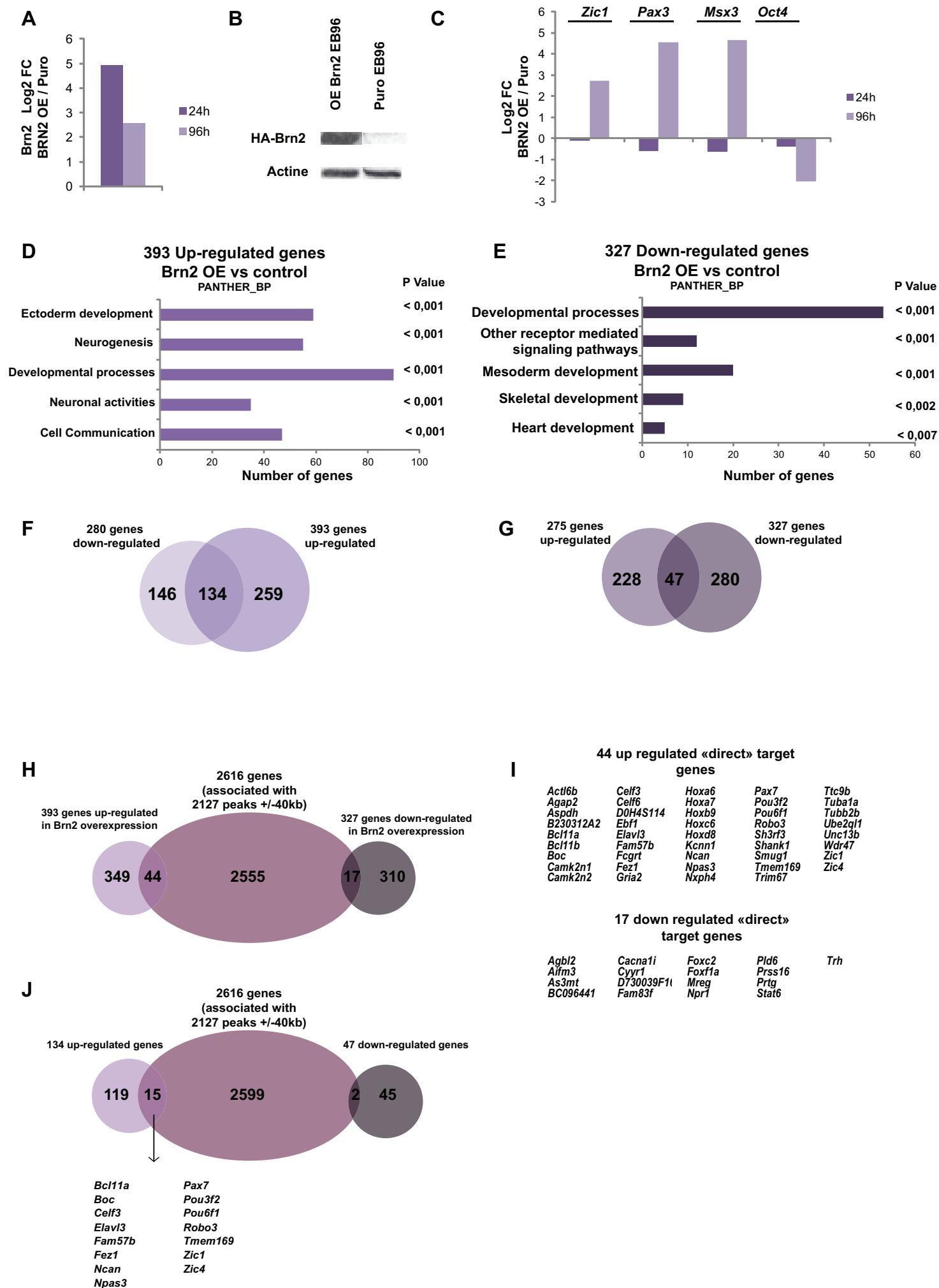
**D**

2127 peaks (200bp wide = summit  $\pm$  100bp)

|                           | 0 mismatch |       | 1 mismatch |       | 2 mismatch |       |
|---------------------------|------------|-------|------------|-------|------------|-------|
|                           | peak       | Motif | Peak       | Motif | Peak       | Motif |
| ATGCAAAT (OCT)            | 49         | 62    | 496        | 684   | 1552       | 4409  |
| ATRNATATRC AWR (MORE)     | 118        | 143   | 259        | 459   | 517        | 1179  |
| ATRNATWATRC AWR (MORE +1) | 68         | 84    | 159        | 249   | 316        | 543   |
| ATTGAAATGCAAAT (PORE)     | 0          | 0     | 1          | 1     | 3          | 3     |
| TNNNTAAATA AWR (NORE)     | 20         | 21    | 193        | 226   | 949        | 1912  |

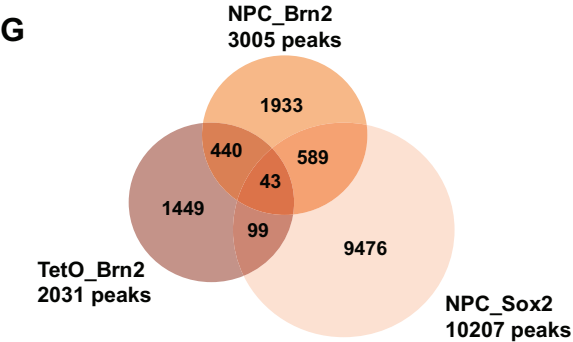
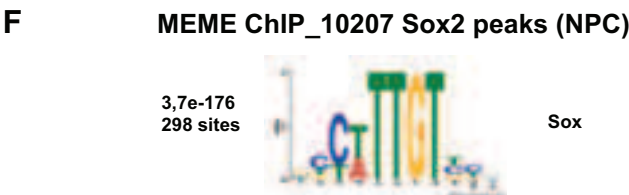
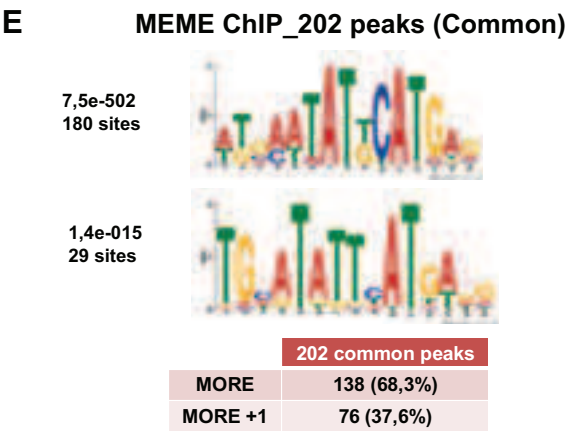
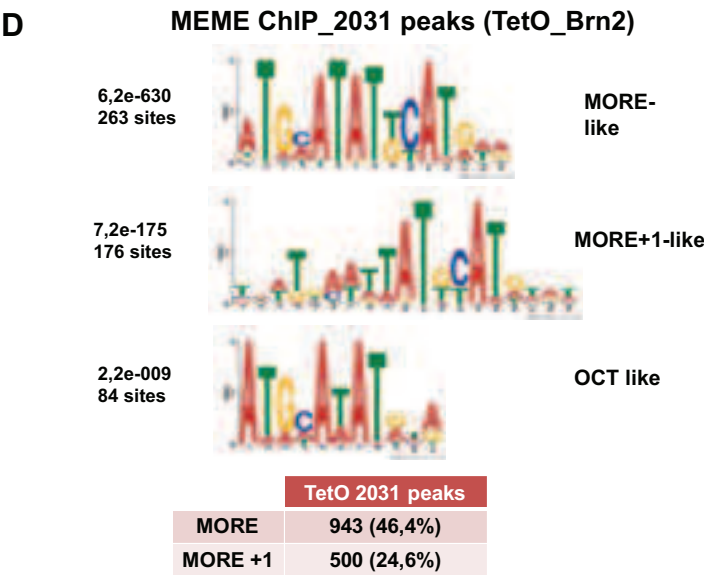
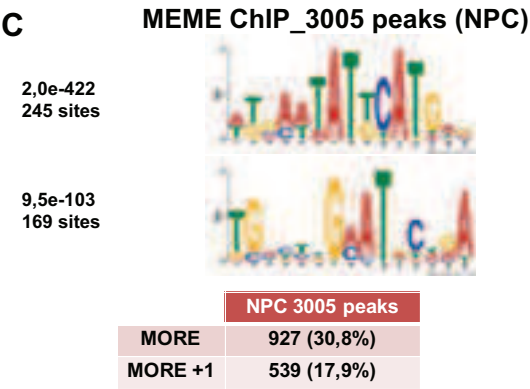
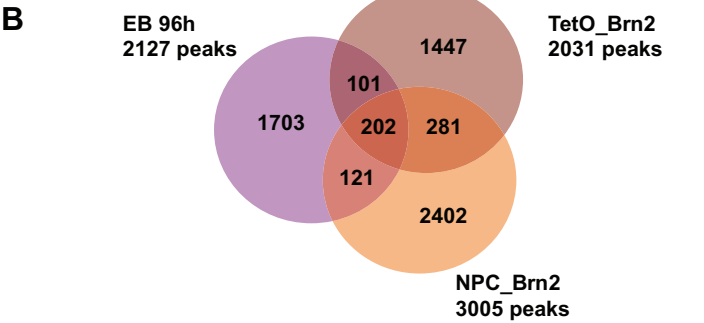


Urban et al., Fig. 2



**A**

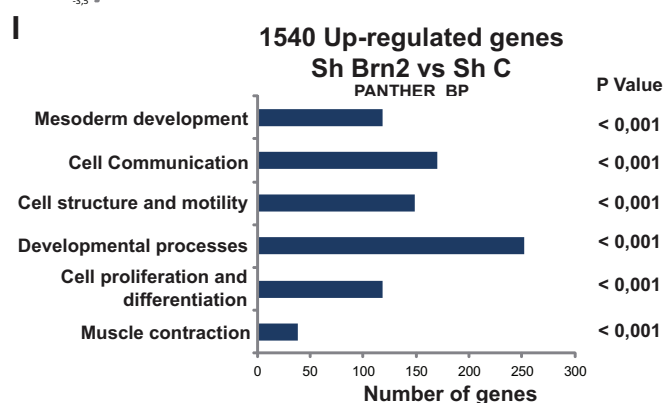
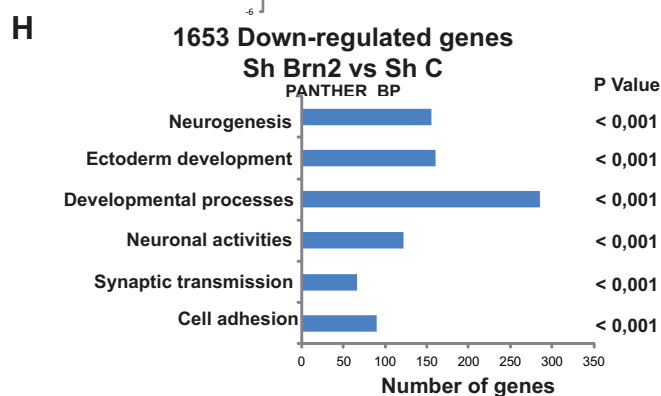
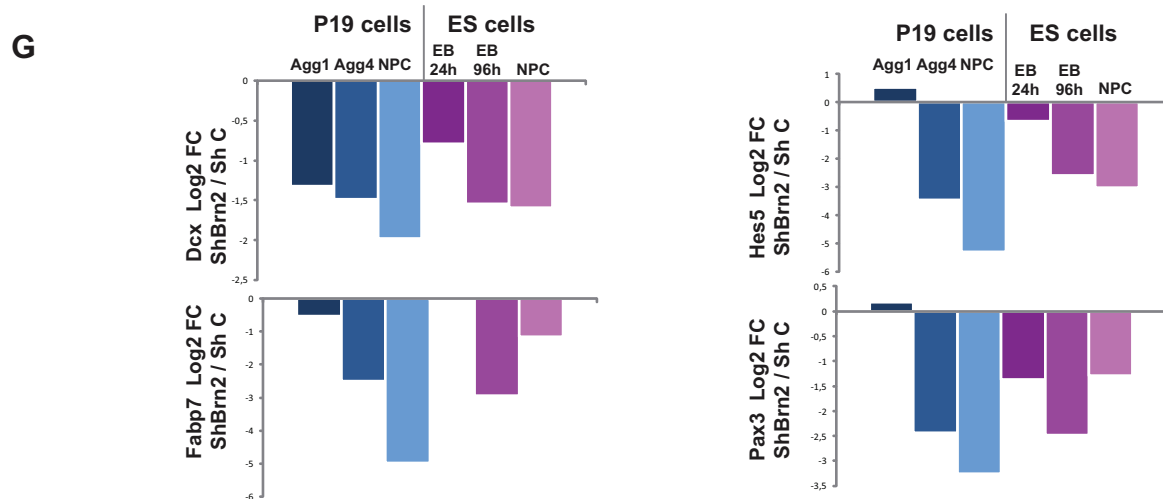
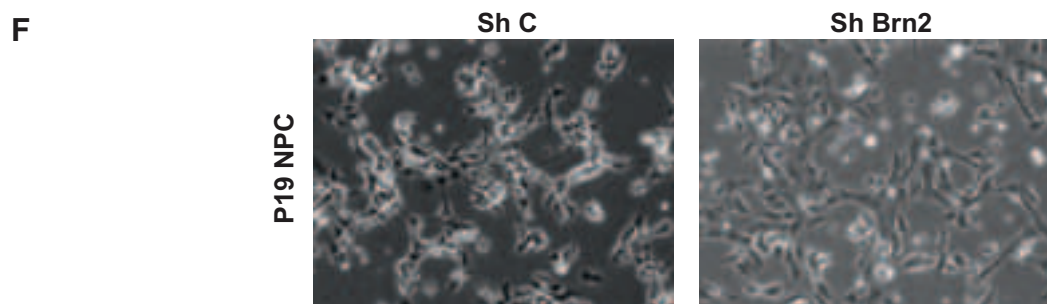
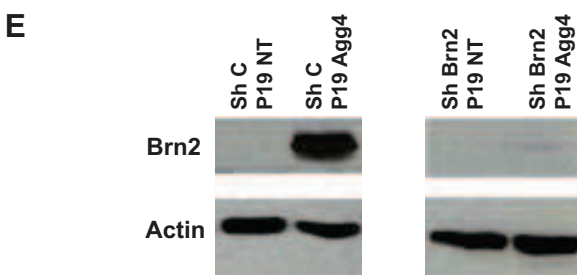
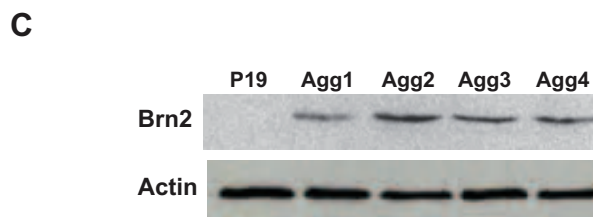
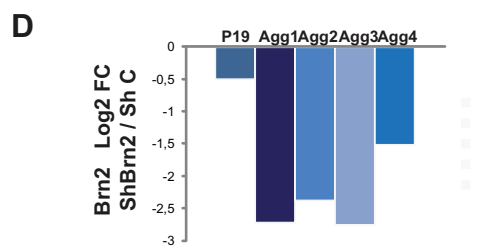
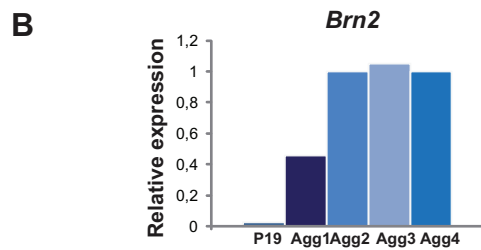
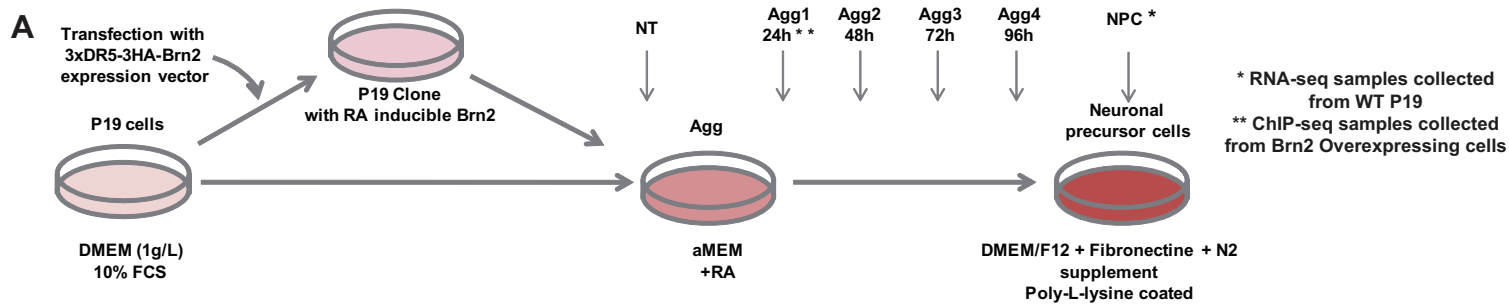
|                  | Lodato analysis | Our analysis |
|------------------|-----------------|--------------|
| Lodato_TetO_Brn2 | 8401 peaks      | 2031 peaks   |
| Lodato_NPC_Brn2  | 6574 peaks      | 3005 peaks   |
| Lodato_NPC_Sox2  | 16685 peaks     | 10207 peaks  |



**H**

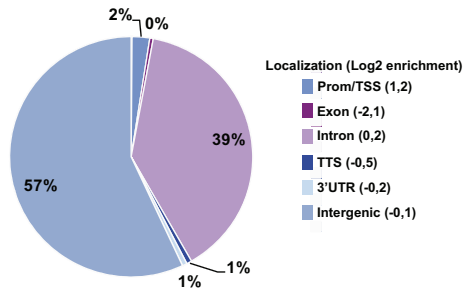
|                          | 2127<br>RA-treated EB | 202 Common<br>Brn2 | 20 Common<br>Brn2/Sox2 |
|--------------------------|-----------------------|--------------------|------------------------|
| H3K4me1 +<br>(H3K27Ac -) | 105<br>(4,9%)         | 43<br>(21,3%)      | 11<br>(55%)            |
| H3K4me1 +/-<br>H3K27Ac + | 50<br>(2,4%)          | 19<br>(9,4%)       | 4<br>(20%)             |

Urban et al., Fig. 4

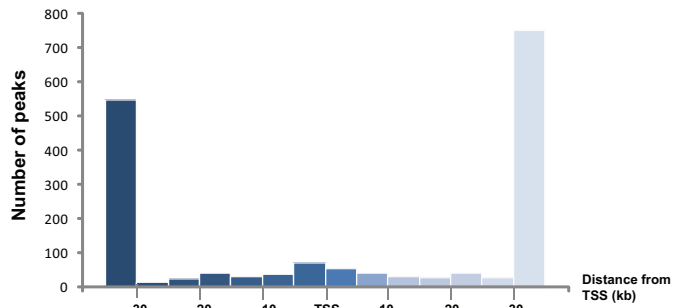


**A**

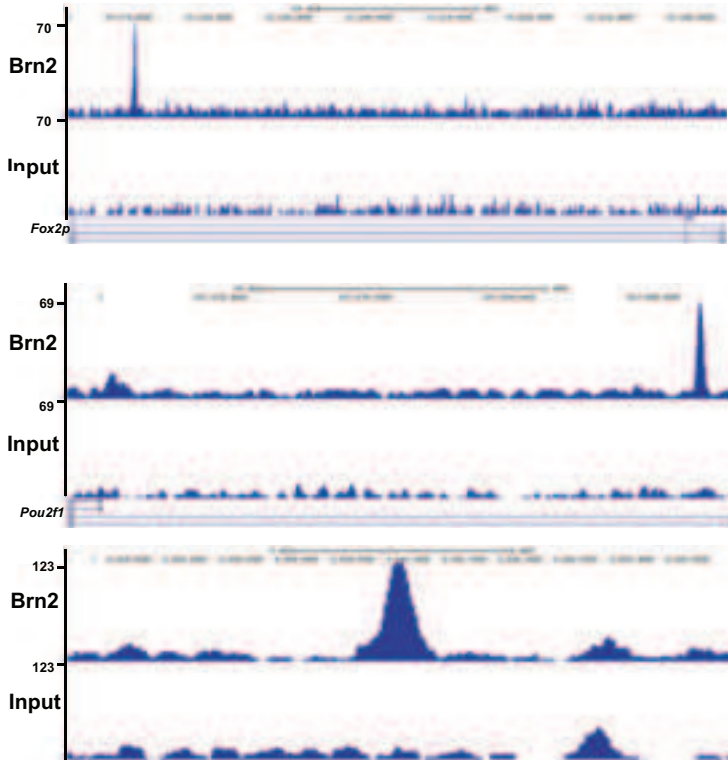
### 1743 sites bound by BRN2 peak distribution



# B



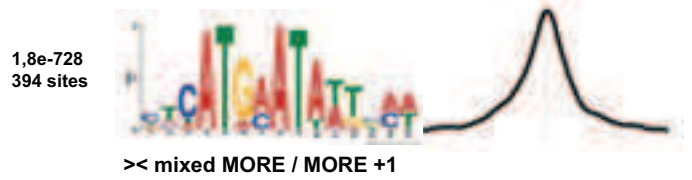
**C**

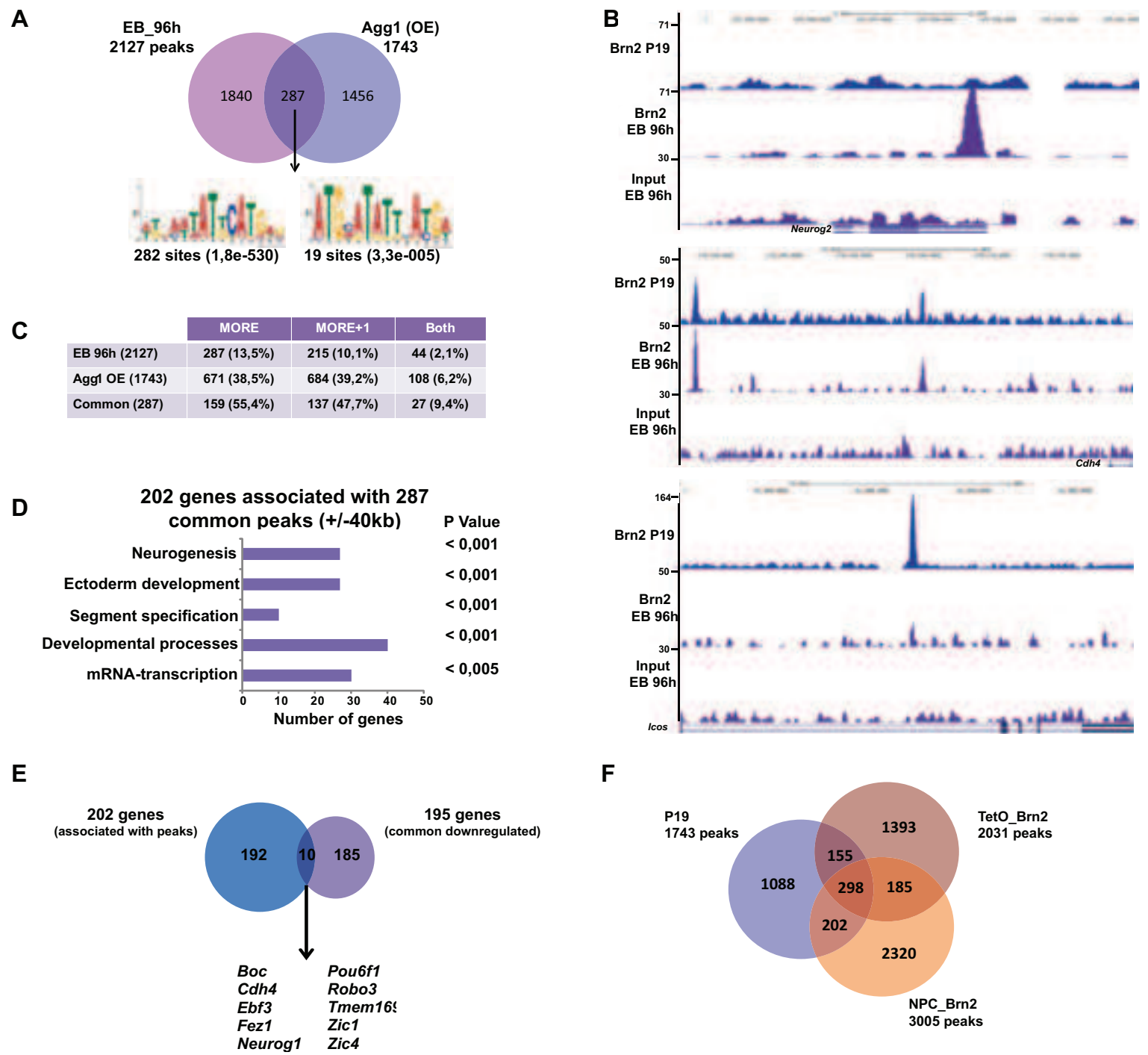


D

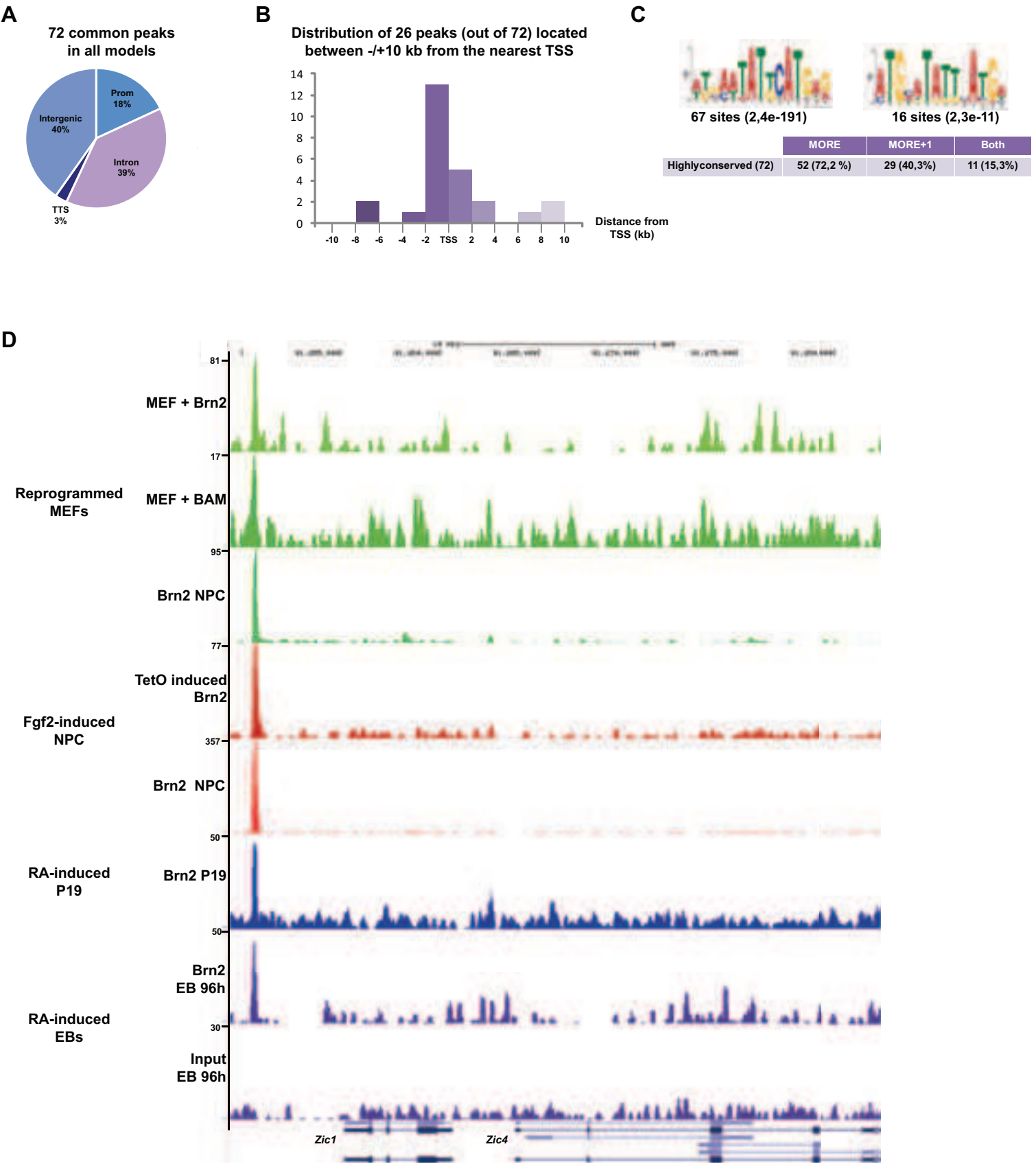
|                          | 1743 peaks (200bp wide = summit +/- 100pb) |       |            |       |            |       |
|--------------------------|--------------------------------------------|-------|------------|-------|------------|-------|
|                          | 0 mismatch                                 |       | 1 mismatch |       | 2 mismatch |       |
|                          | peak                                       | Motif | Peak       | Motif | Peak       | Motif |
| ATGCAAT                  | 126                                        | 132   | 917        | 1281  | 1699       | 7025  |
| ATRNATATRCAWR (MORE)     | 165                                        | 198   | 512        | 845   | 954        | 2116  |
| ATRNATWATRCAWR (MORE +1) | 142                                        | 163   | 495        | 692   | 827        | 1537  |
| ATTTGAATGCAAT (PORE)     | 0                                          | 0     | 4          | 4     | 29         | 29    |
| TNNNTAAATAWR (NORE)      | 14                                         | 15    | 262        | 292   | 1335       | 2945  |

## E

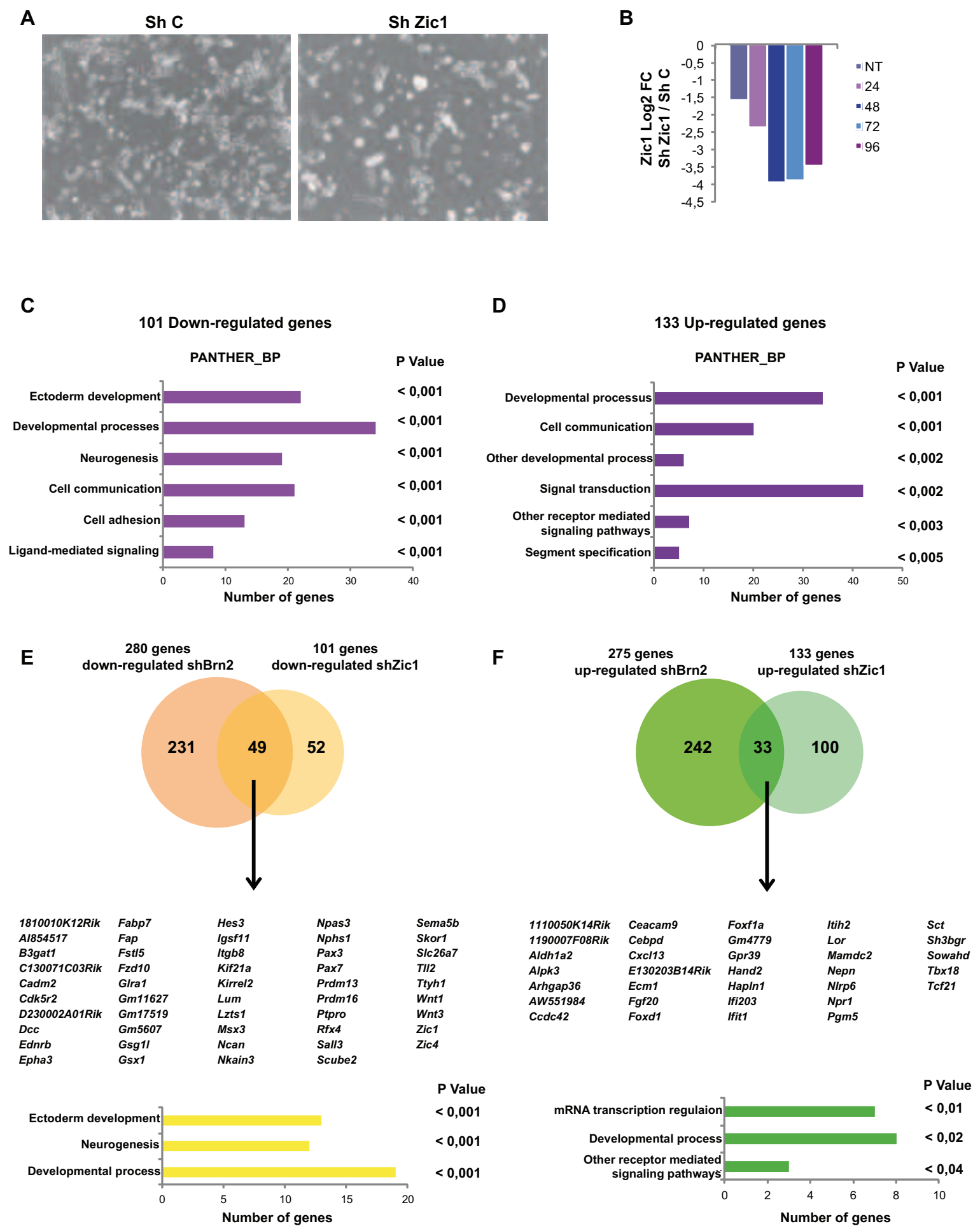




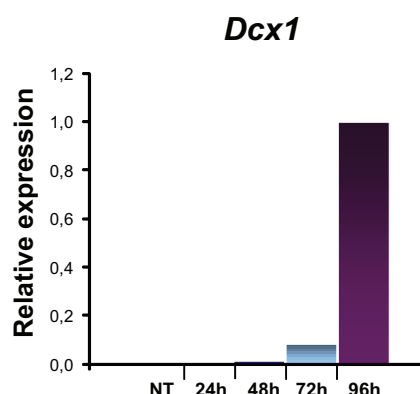
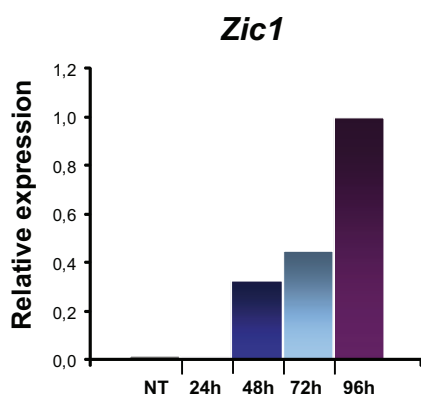
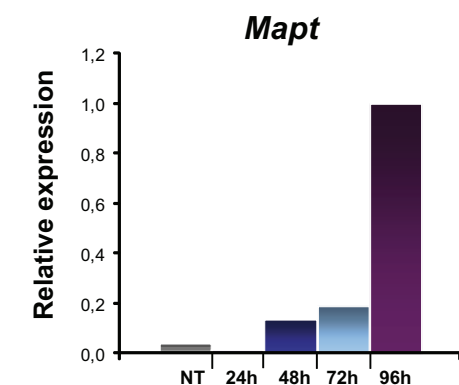
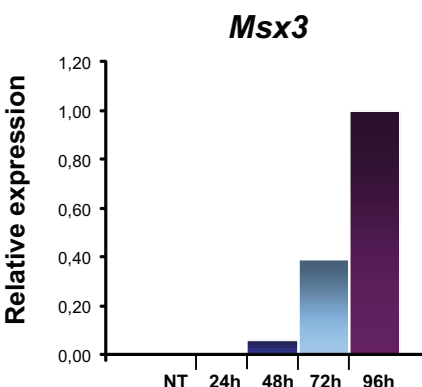
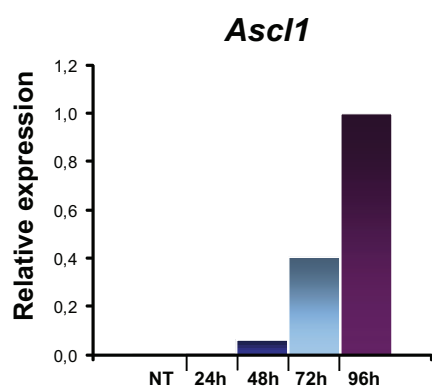
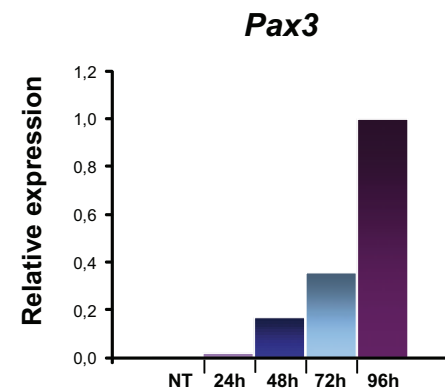
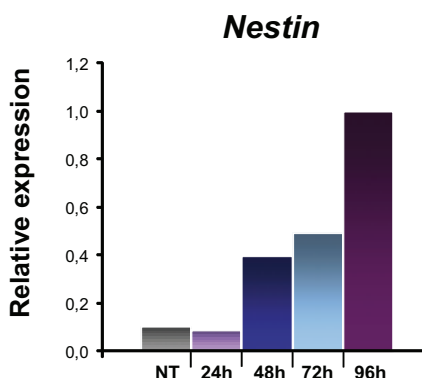
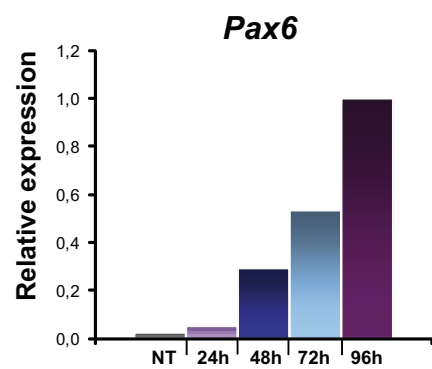
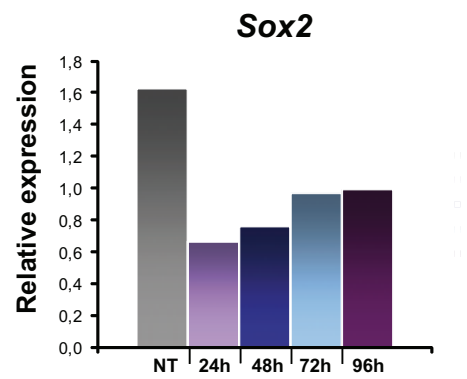
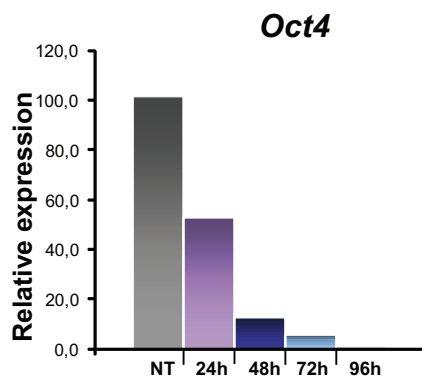
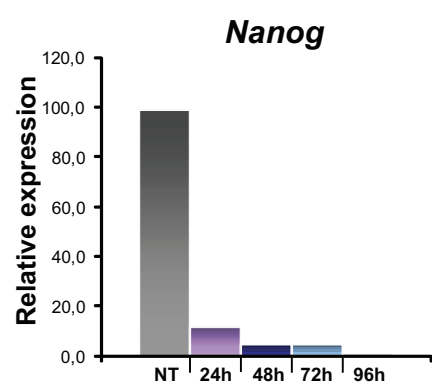
Urban et al., Fig. 7



Urban et al., Fig. 8



Urban et al., Fig. 9



A

## GO Neurogenesis cluster

|                |               |                |               |
|----------------|---------------|----------------|---------------|
| <i>Ascl1</i>   | <i>Dner</i>   | <i>Irx5</i>    | <i>Pax7</i>   |
| <i>Bhlhe22</i> | <i>Dtx4</i>   | <i>Irx6</i>    | <i>Pax8</i>   |
| <i>Bhlhe23</i> | <i>Ednrb</i>  | <i>Jag1</i>    | <i>Ppfia2</i> |
| <i>Celsr2</i>  | <i>Elavl3</i> | <i>L1cam</i>   | <i>Prox1</i>  |
| <i>Celsr3</i>  | <i>En1</i>    | <i>Lhx1</i>    | <i>Reln</i>   |
| <i>Chl1</i>    | <i>Epha10</i> | <i>Lhx5</i>    | <i>Rtn1</i>   |
| <i>Cntn1</i>   | <i>Epha3</i>  | <i>Megf10</i>  | <i>Sema5b</i> |
| <i>Crb1</i>    | <i>Foxd3</i>  | <i>Ncan</i>    | <i>Slit1</i>  |
| <i>Dbx1</i>    | <i>Glra1</i>  | <i>Nell2</i>   | <i>Stmn2</i>  |
| <i>Dbx2</i>    | <i>Gsx1</i>   | <i>Neurog1</i> | <i>Uncx</i>   |
| <i>Dcc</i>     | <i>Hes3</i>   | <i>Nlgn1</i>   | <i>Zic1</i>   |
| <i>Dcx</i>     | <i>Hes5</i>   | <i>Pax2</i>    | <i>Zic2</i>   |
| <i>Dll1</i>    | <i>Irx1</i>   | <i>Pax3</i>    | <i>Zic4</i>   |
| <i>Dll3</i>    | <i>Irx3</i>   | <i>Pax5</i>    | <i>Zic5</i>   |

B

## GO Immunity and defense cluster

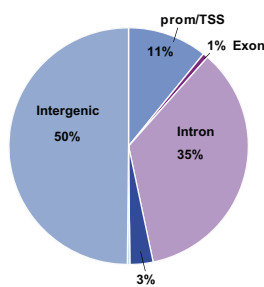
|                |                    |                 |               |
|----------------|--------------------|-----------------|---------------|
| <i>Angptl4</i> | <i>Gbp1</i>        | <i>Irf35</i>    | <i>Plscr2</i> |
| <i>Anxa8</i>   | <i>Gbp2</i>        | <i>Ifit1</i>    | <i>Ptgs2</i>  |
| <i>C1qtnf2</i> | <i>Gbp3</i>        | <i>Ifitm1</i>   | <i>Ptk2b</i>  |
| <i>C1ra</i>    | <i>Gbp4</i>        | <i>Ifitm3</i>   | <i>Ptx3</i>   |
| <i>C1s</i>     | <i>Gbp6</i>        | <i>Ifitm5</i>   | <i>Saa1</i>   |
| <i>Ccr7</i>    | <i>Gsta3</i>       | <i>Itgb7</i>    | <i>Saa2</i>   |
| <i>Cd74</i>    | <i>H2Eb1</i>       | <i>Lgals3bp</i> | <i>Sh2d4a</i> |
| <i>Cysltr2</i> | <i>H2Oa</i>        | <i>Lif</i>      | <i>Sla2</i>   |
| <i>Dapp1</i>   | <i>I830012O16R</i> | <i>Oas1g</i>    | <i>Spib</i>   |
| <i>Fgb</i>     | <i>Ifi203</i>      | <i>Oas12</i>    |               |
| <i>Fgg</i>     | <i>Ifi2711</i>     | <i>Plaur</i>    |               |

## GO Mesoderm development cluster

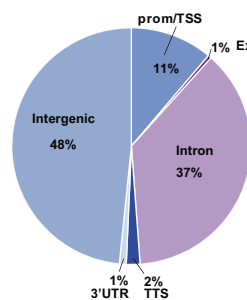
|               |               |               |              |
|---------------|---------------|---------------|--------------|
| <i>Afp</i>    | <i>Foxf1a</i> | <i>Nkx2</i>   | <i>Tnni1</i> |
| <i>Angpt4</i> | <i>Foxf2</i>  | <i>Nkx8</i>   | <i>Tnni2</i> |
| <i>Btn1a1</i> | <i>Hand2</i>  | <i>Postn</i>  |              |
| <i>Foxa3</i>  | <i>Hapln1</i> | <i>Prl2a1</i> |              |
| <i>Foxd1</i>  | <i>Nfe2</i>   | <i>Prl7b1</i> |              |

C

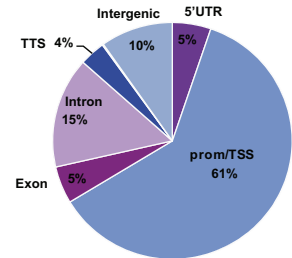
287 peaks containing MORE motif



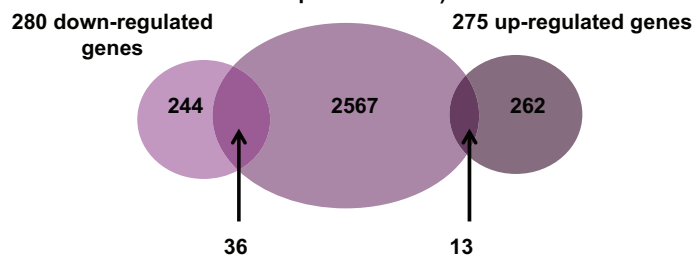
215 peaks containing MORE+1 motif



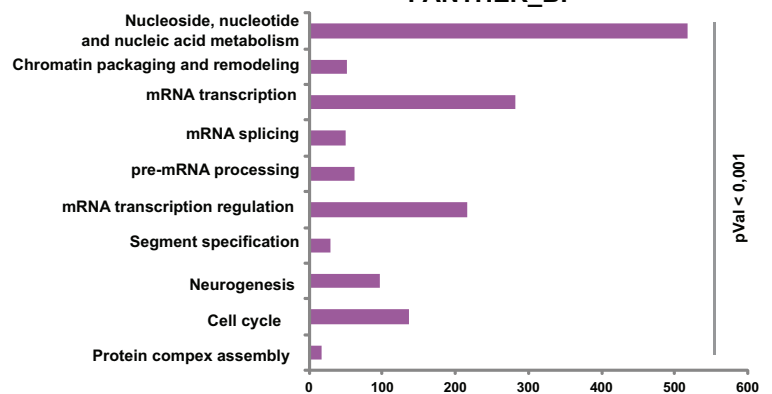
631 peaks containing ZFX-like motif



D

2616 genes  
(associated with  
2127 peaks +/-40kb)

E

2616 genes  
PANTHER\_BP

F

|                     |               |               |              |               |                |                 |                    |
|---------------------|---------------|---------------|--------------|---------------|----------------|-----------------|--------------------|
| <i>6230409E13Ri</i> | <i>Elavl3</i> | <i>Hif1a</i>  | <i>Hoxb4</i> | <i>Hoxd10</i> | <i>Mafb</i>    | <i>Ntng1</i>    | <i>Sema3a</i>      |
| <i>Acer2</i>        | <i>Epha10</i> | <i>Hoxa1</i>  | <i>Hoxb5</i> | <i>Hoxd11</i> | <i>Msi1</i>    | <i>Olfm2</i>    | <i>Sh3bp4</i>      |
| <i>Ascl1</i>        | <i>Ephb4</i>  | <i>Hoxa10</i> | <i>Hoxb6</i> | <i>Hoxd12</i> | <i>Ncan</i>    | <i>Otx1</i>     | <i>Tnfrsf14</i>    |
| <i>Basp1</i>        | <i>Evx2</i>   | <i>Hoxa2</i>  | <i>Hoxb7</i> | <i>Hoxd4</i>  | <i>Neo1</i>    | <i>Pafah1b1</i> | <i>Tulp1</i>       |
| <i>Bcan</i>         | <i>Fat1</i>   | <i>Hoxa4</i>  | <i>Hoxb8</i> | <i>Hoxd8</i>  | <i>Neurog1</i> | <i>Pax3</i>     | <i>Ulk2</i>        |
| <i>Bnc2</i>         | <i>Foxb1</i>  | <i>Hoxa5</i>  | <i>Hoxb9</i> | <i>Hoxd9</i>  | <i>Neurog2</i> | <i>Pax6</i>     | <i>Vax1</i>        |
| <i>Cdk5rap2</i>     | <i>Foxc2</i>  | <i>Hoxa6</i>  | <i>Hoxc4</i> | <i>Irx3</i>   | <i>Nnat</i>    | <i>Pax7</i>     | <i>Yes1</i>        |
| <i>Cntfr</i>        | <i>Foxf1a</i> | <i>Hoxa7</i>  | <i>Hoxc5</i> | <i>Irx5</i>   | <i>Notch1</i>  | <i>Pcdh18</i>   | <i>Zfmx3</i>       |
| <i>Crb2</i>         | <i>Foxl1</i>  | <i>Hoxa9</i>  | <i>Hoxc6</i> | <i>Jag2</i>   | <i>Notch4</i>  | <i>Pcdh7</i>    | <i>Zfp609</i>      |
| <i>Dtx4</i>         | <i>Hes5</i>   | <i>Hoxb1</i>  | <i>Hoxc8</i> | <i>Jarid2</i> | <i>Noto</i>    | <i>Pstpip2</i>  | <i>Zfp91, Cntf</i> |
| <i>Dvl2</i>         | <i>Hes6</i>   | <i>Hoxb2</i>  | <i>Hoxc9</i> | <i>Lhx6</i>   | <i>Nrcam</i>   | <i>Rax</i>      | <i>Zic1</i>        |
| <i>Efna2</i>        | <i>Hey2</i>   | <i>Hoxb3</i>  | <i>Hoxd1</i> | <i>Madd</i>   | <i>Ntf5</i>    | <i>Robo2</i>    | <i>Zic4</i>        |

G

36 downregulated «direct target»

|               |               |                |                |
|---------------|---------------|----------------|----------------|
| <i>Ascl1</i>  | <i>Eml1</i>   | <i>Lin7a</i>   | <i>Pou3f2</i>  |
| <i>Bcl11a</i> | <i>Epha10</i> | <i>Msx3</i>    | <i>Pou6f1</i>  |
| <i>Boc</i>    | <i>Fam57b</i> | <i>Myb</i>     | <i>Rab30</i>   |
| <i>Cdh4</i>   | <i>Fat4</i>   | <i>Ncan</i>    | <i>Rhbd13</i>  |
| <i>Celf3</i>  | <i>Fez1</i>   | <i>Neurog1</i> | <i>Robo3</i>   |
| <i>Chrna3</i> | <i>Hes5</i>   | <i>Nova1</i>   | <i>Tmem169</i> |
| <i>Dtx4</i>   | <i>Irx3</i>   | <i>Npas3</i>   | <i>Zfp536</i>  |
| <i>Ebf3</i>   | <i>Irx5</i>   | <i>Pax3</i>    | <i>Zic1</i>    |
| <i>Elavl3</i> | <i>Lfng</i>   | <i>Pax7</i>    | <i>Zic4</i>    |

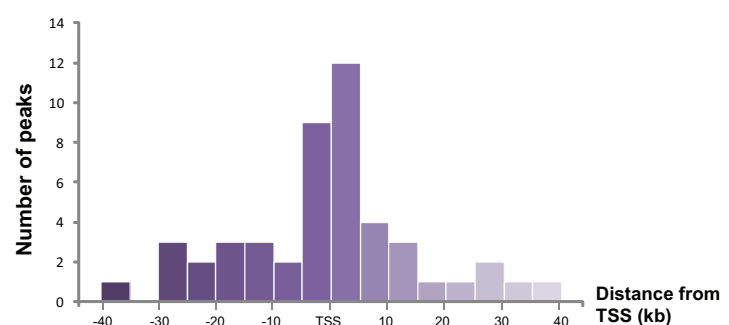
13 upregulated «direct target»

|                      |              |               |
|----------------------|--------------|---------------|
| <i>1190007F08Rik</i> | <i>H2-Oa</i> | <i>Spib</i>   |
| <i>Btn1a1</i>        | <i>Ifi35</i> | <i>Tm4sf5</i> |
| <i>Fbxo15</i>        | <i>Napsa</i> | <i>Vgll2</i>  |
| <i>Fgf20</i>         | <i>Nfe2</i>  |               |
| <i>Foxf1a</i>        | <i>Npr1</i>  |               |

Aaa : Associated with a peak containing MORE motif  
 Aaa : Associated with a peak containing MORE+1 motif  
 Aaa : Associated with a peak containing ZFX like motif

H

distribution of 48 peaks associated with 36 down regulated genes



A

## GO Neurogenesis cluster

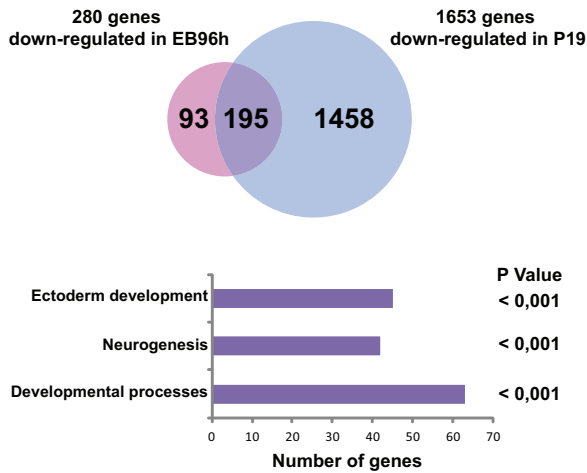
|               |        |         |          |          |       |
|---------------|--------|---------|----------|----------|-------|
| 6230409E13Rik | Efnb3  | Hoxb7   | Neurog2  | Pcdhbb2  | Stmn3 |
| Ablim3        | Elavl2 | Hoxb8   | Nfasc    | Pcdhbb3  | Tlx1  |
| Acer2         | Elavl3 | Hoxb9   | Ngfr     | Pcdhbb4  | Tlx3  |
| Arx           | Elavl4 | Irx3    | Nlgn1    | Pcdhbb5  | Tnn   |
| Ascl1         | En2    | Irx4    | Nlgn3    | Pcdhbb6  | Tnr   |
| Bcan          | Epha10 | Irx5    | Notch1   | Pcdhbb7  | Trim2 |
| Cdh23         | Epha5  | Isl1    | Npdc1    | Pcdhbb8  | Ulk2  |
| Cdh8          | Epha8  | Ism1    | Nrcam    | Pcdhbb9  | Unc5d |
| Cdhr3         | Ephb1  | Jag2    | Nrg3     | Pcdhbb10 | Uncx  |
| Celsr2        | Ephb2  | Jarid2  | Ntn      | Pcdhbb11 | Vax2  |
| Celsr3        | Ephb6  | L1cam   | Ntn4     | Pcdhbb12 | Zfhx4 |
| Chl1          | Fcho1  | Lhx1    | Ntrk2    | Pcdhbb13 | Zic1  |
| Cntn1         | Fgf15  | Lhx5    | Ntrk3    | Pcdhbb14 | Zic2  |
| Cntn2         | Fgfr3  | Lsmp    | Odz2     | Pcdhbb15 | Zic3  |
| Cntn3         | Fgfr4  | Madd    | Olfm1    | Pcdhbb16 | Zic4  |
| Crb1          | Foxb1  | Mam13   | Olfm2    |          | Zic5  |
| Cxcr4         | Foxj1  | Megf11  | Olfm3    |          |       |
| Dach1         | Gbx1   | Mpped1  | Opcml    |          |       |
| Dcc           | Gbx2   | Msi1    | Pax2     |          |       |
| Dcx           | Gira2  | Msi2    | Pax3     |          |       |
| Dll1          | Gpm6b  | Ncan    | Pax7     |          |       |
| Dll3          | Gsx1   | Negr1   | Pax8     |          |       |
| Dll4          | Hes5   | Nell1   | Pcdhbb10 |          |       |
| Dner          | Hes6   | Nell2   | Pcdhbb11 |          |       |
| Dscam         | Hey1   | Neurod1 | Pcdhbb12 |          |       |
| Dtx1          | Hoxa6  | Neurod4 | Pcdhbb13 |          |       |
| Dtx4          | Hoxb5  | Neurod6 | Pcdhbb14 |          |       |
| Ednrb         | Hoxb6  | Neurog1 | Pcdhbb15 |          |       |
|               |        |         | Pcdhbb16 |          |       |

B

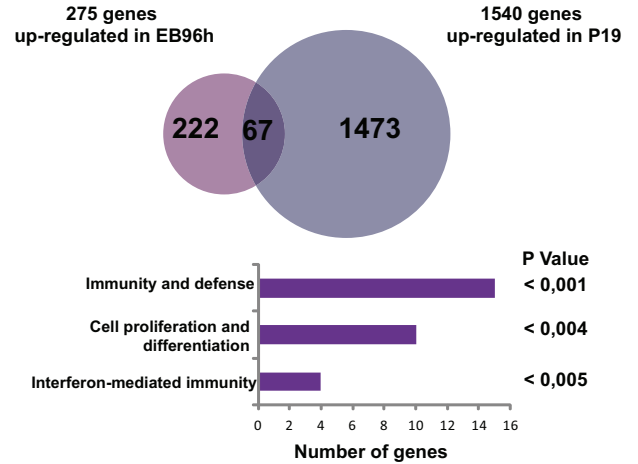
## GO Mesoderm development

|         |         |        |              |          |
|---------|---------|--------|--------------|----------|
| Barx1   | Bmp5    | Foxf1a | Myl7         | Tnni1    |
| Epha1   | Figf    | Foxf2  | Myl9         | Tnfrsf15 |
| Epha2   | Cmklr1  | Foxh1  | Nog          | Utrn     |
| Epha7   | Col5a1  | Foxo1  | Nfatc2       | Vdr      |
| Fos     | Col5a3  | Foxp2  | Odz3         | Zfp1     |
| Farp1   | Col12a1 | Foxp4  | Ogn          | Zfp1     |
| Flt1    | Col18a1 | Fhl1   | Popdc2       |          |
| Flt4    | Cs1     | Fhl2   | Pcsk5        |          |
| Gata3   | Cyr61   | Fhl3   | Runx1        |          |
| Gli1    | Ddah1   | Frk    | Sema3a       |          |
| Hlx     | Dlx5    | Hand2  | Sema3c       |          |
| Klf4    | Dusp2   | Hhex   | Sema3d       |          |
| Lpp     | Dmd     | Il6st  | Stat6        |          |
| Nkx2    | Egr3    | Kdr    | Anxa2        |          |
| Pdlim5  | Eng     | Ltpb1  | Fbn2         |          |
| Pdlim7  | Efemp1  | Ltpb2  | Mx2          |          |
| Igsf5   | Efemp2  | Ltpb4  | Tuft1        |          |
| Foxg1   | Epyc    | Lmod1  | Ski          |          |
| Tbx2    | Fbn1    | Mfge8  | Spry1        |          |
| Lyn     | Fgf1    | Mbn1   | Spry2        |          |
| Adam19  | Fbln2   | Mbn12  | Tnmd         |          |
| Adamts1 | Fbln5   | Myocd  | Tenc1        |          |
| Alpl    | Foxa1   | Mef2c  | Tll1         |          |
| Amotl2  | Foxa2   | Myom1  | Tgfb         |          |
| Apln    | Foxc1   | Myh10  | Tpm1         |          |
| Bgn     | Foxc2   | Myh11  | Tpm2         |          |
| Bmp2    | Foxd1   | Myh1   | Gm7809, Tpm4 |          |
| Bmp4    | Foxd2   | Myh9   | Tnni2        |          |

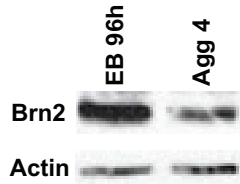
C



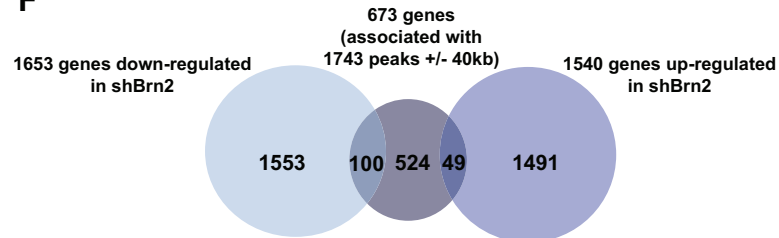
D



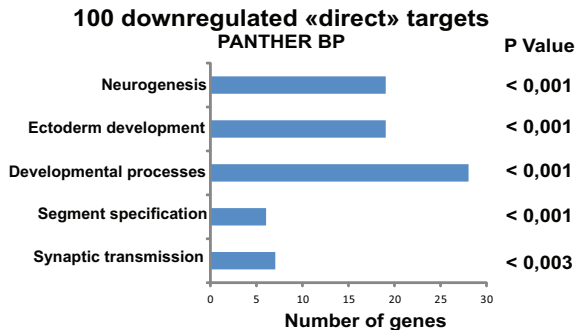
E



F



G



H

|       |         |       |
|-------|---------|-------|
| Hoxa6 | Jarid2  | Prox1 |
| Hoxb5 | Mam13   | Robo2 |
| Hoxb6 | Neurog1 | Tlx1  |
| Hoxb7 | Nfasc   | Zic1  |
| Hoxb8 | Notch1  | Zic4  |
| Hoxb9 | Nrcam   |       |
| Ism1  | Olfm2   |       |

# *Résultats complémentaires*

## Résultats Complémentaires

Avant de focaliser notre travail uniquement sur le rôle de Brn2 puis ultérieurement de Zic1 dans la différenciation neuronale, j'ai eu l'occasion de tester le rôle de nombreux autres facteurs de transcription. Sur l'ensemble des facteurs testés, Brn2 est celui qui donnait d'après nous le résultat le plus intéressant et exploitable, c'est pourquoi nous avons défini ce facteur comme prioritaire dans nos analyses. De plus, le protocole d'infection lentivirale et de différenciation neuronale nécessite plus d'un mois entre le début de l'expérience et le résultat. Il était donc difficile de travailler sur l'ensemble des facteurs sélectionnés initialement.

Cette partie comprendra dans un premier temps les résultats des expériences préliminaires effectuée pour les facteurs de transcription Hes3, puisque comme vous avez pu le voir, ce facteur semble être affecté suite à la perte d'expression de Brn2. Il m'a donc paru intéressant de discuter des résultats préliminaires que j'avais obtenus pour ce facteur. Je vais ensuite vous présenter une expérience réalisée sur Brn2, dans laquelle nous avons optimisé notre protocole de différenciation (voir matériel et méthodes ci-dessous). Afin d'améliorer notre efficacité de silencing, j'ai utilisé un protocole de différenciation légèrement raccourci. Comme vous pourrez le voir dans le cas du facteur Brn2, ce protocole permet d'avoir des résultats plus significatifs et c'est donc celui-ci qui a été utilisé pour les expériences présentées dans le manuscrit ci-dessus.

### 1. Matériel et méthodes

L'ensemble des techniques utilisées pour ces résultats complémentaires sont similaires à celle décrites dans le manuscrit ci-dessus. Une seule modification a été appliquée au protocole de différenciation, les données présentées dans le manuscrit en question ayant été réalisées avec le protocole optimisé. Le protocole normal de différenciation tel qu'il est décrit consiste à cultiver les EB pendant 4 jours dans du milieu sans LIF. Pendant cette période, les cellules continuent de se multiplier et forment les trois feuillets de l'embryon. Les EB sont ensuite traités pendant quatre jours supplémentaires à l'AR, période durant laquelle les cellules vont être spécifiées à devenir des neurones. Après ces quatre jours, les cellules sont dissociées et mises en culture dans un milieu neuronal (DMEM-HamF12, N2supplement, BSA) sur des boîtes tapissées de poly-DL-ornithine/laminine. Dans ces conditions elles adoptent rapidement une morphologie de cellules radiales gliales. Dans les résultats suivants ce

protocole sera appelé 4/4 (pour 4 jours de différenciation sans AR, suivi de 4 jours de traitement). Un protocole de différenciation optimisé, appelé 2/4 (pour 2 jours de culture des EB sans AR puis 4 jours de traitement) permet également d'obtenir une population homogène de cellules radiales gliales après dissociation des corps embryonnaires. Ce protocole 2/4 est en tout point similaire au protocole 4/4, à la seule exception de cette première phase de culture des EB qui est raccourcie.

## 2. Hes3

### 2.1. Description du facteur de transcription Hes3

Parmi les gènes dont l'expression est fortement induite après le traitement par l'AR se trouvent ceux qui codent pour les facteurs de transcription Hes3 et Hes5. Les gènes *Hes* des mammifères sont les orthologues des gènes Hairy et Enhancer of Split de la drosophile, qui codent pour des répresseurs de la transcription. Cette famille comprend sept membres, parmi lesquels Hes1, Hes3 et Hes5 qui sont exprimés dans les cellules souches neurales (Kageyama, Hatakeyama et al. 2006; Kageyama, Ohtsuka et al. 2008). Les facteurs de transcription de la famille Hes possèdent à leur extrémité amino-terminale un domaine conservé formé par une région basique, suivie d'un motif hélice-boucle-hélice (bHLH). Ce domaine permet la formation d'un dimère et la reconnaissance de l'ADN. De plus, le domaine basique en position amino-terminale contient un résidu proline, impliqué dans la reconnaissance d'une séquence d'ADN particulière appelée N-box (5'-CACNAG-3'). Outre ce domaine amino-terminal impliqué dans la reconnaissance de l'ADN, les facteurs de transcription Hes possèdent un domaine carboxy-terminal contenant la séquence Trp-Arg-Pro-Trp (WRPW domain) nécessaire au recrutement des co-répresseurs et à leur fonction transcriptionnelle. Ainsi, le co-répresseur TLE/Grg, homologue de Groucho de la drosophile, interagit avec ce domaine WRPW et permet de recruter des histones déacétylases et par conséquent de réprimer la transcription (Kageyama, Ohtsuka et al. 2005).

La fonction des facteurs Hes dans la prolifération, la différenciation et la spécification des cellules souches neurales a été étudiée *in vivo* (Kageyama, Ohtsuka et al. 2008). De plus, comme je l'ai décrit précédemment, les facteurs Hes sont les effecteurs de la voie Notch, connu pour être importante dans l'initiation du processus de différenciation. Les facteurs de transcription Hes1, Hes3, et Hes5 sont impliqués dans ce processus de neurogenèse. Des expériences de double invalidation *in vivo* ont permis d'associer les facteurs Hes aux différentes étapes de la neurogenèse décrites ci-dessus (*Figure 1*). Ainsi, Hes1 et Hes3 sont nécessaires pour le maintien des cellules neuroépithéliales, alors que Hes1 et Hes5 régulent le maintien des cellules radiales gliales. Après la neurogenèse, les cellules radiales gliales peuvent se différencier en astrocytes et ce processus est également sous le contrôle des facteurs Hes1 et Hes5. L'étude de souris transgéniques montre que la surexpression des gènes *Hes* lors des stades précoces du développement neuronal maintient les cellules souches neurales et empêche leur différenciation en agissant comme des antagonistes des facteurs

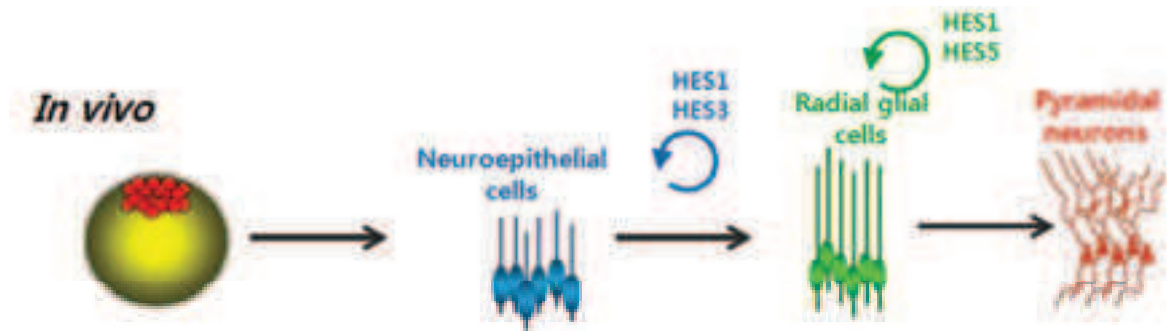


Figure 1 : Implication des facteurs de transcription Hes à différents stades de la différenciation neuronale.

bHLH pro-neuraux que sont *Ascl1*, *Atoh1* (Atonal Homolog 1) et les *Neurog*. En revanche, à des stades ultérieurs, cette surexpression promeut la formation des astrocytes, ceci indiquant qu'en fonction du stade de développement, les facteurs *Hes* peuvent avoir des activités opposées (Kageyama, Ohtsuka et al. 2008).

## 2.2.Résultats

### 2.2.1. Réduction de l'expression du gène *Hes3* au niveau ARN

Une approche similaire à celle utilisée pour les facteurs de transcription *Brn2* et *Zic1* décrits ci-dessus a été mise en œuvre pour tenter de réduire l'expression du gène *Hes3*.

Au cours de la différenciation neuronale des cellules ES, l'expression du gène *Hes3* est fortement induite après 24h de traitement des EB par l'AR, puis diminue aux temps ultérieurs dans les EB. Au stade neurones, l'expression de *Hes3* est quasiment nulle (*Figure 2A*). L'utilisation d'un shARN ciblant *Hes3* permet de réduire significativement l'expression de sa cible, et ceci à l'ensemble des temps de différenciation (*Figure 2A*).

### 2.2.2. La perte d'expression de *HES3* n'abolit pas la différenciation neuronale

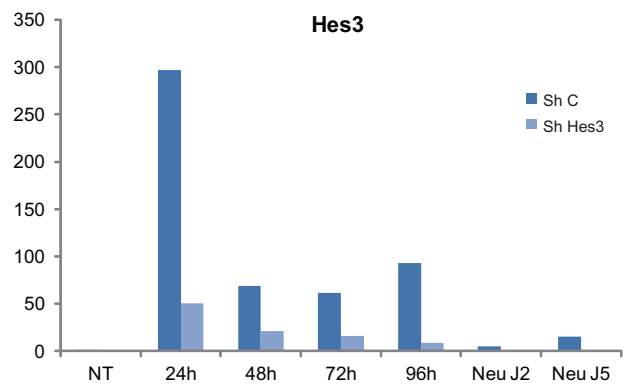
Afin de vérifier si la baisse d'expression de *Hes3* détectée par RT-qPCR avait un impact sur la différenciation neuronale, nous avons réalisé un immunomarquage après 13 jours de différenciation. Les corps neuronaux ainsi que les axones sont marqués grâce à un anticorps dirigé contre TUJ1 (Neuron-specific class III beta-tubulin). Une baisse du nombre total de neurones différenciés est observée pour la lignée sh*Hes3*, mais ces neurones ne présentent aucune différence morphologique en comparaison des neurones de la lignée contrôle (*Figure 2B*).

### 2.2.3. L'expression de facteurs neuronaux est altérée par la perte de *Hes3*

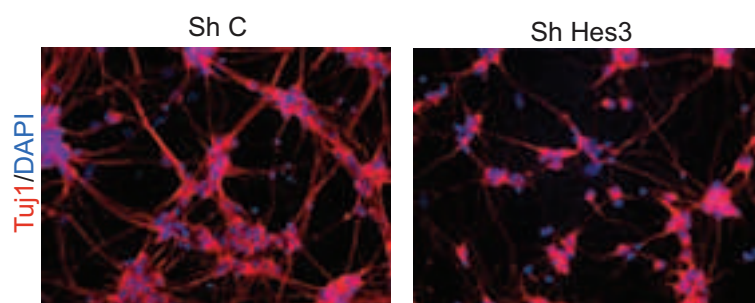
Nous avons décidé de mesurer la conséquence de la perte d'expression de *Hes3* sur l'activation de certains marqueurs neuronaux. Avant tout, les gènes *Hes* étant importants dans les étapes initiales de la différenciation et dans le choix du maintien des cellules à un stade

*Figure 2 : Rôle de la baisse d'expression de Hes3 sur la différenciation neuronale induite par l'AR. Après le traitement des EB par l'acide rétinoïque, l'expression au niveau ARN de Hes3 est fortement induite dans la lignée contrôle après 24h de traitement. L'utilisation d'un shARN dirigé contre Hes3 permet de réduire significativement son expression, et ce à l'ensemble des points testés (A). Ces deux lignées différencient en neurones, même s'il semble y avoir un nombre légèrement réduit de neurones dans la lignée shHes3 (B). L'expression des marqueurs de neurones (Pax6, Dcx, Vglut) semble être réduite suite à la réduction d'expression de Hes3. En revanche, Olig2 et CNP, marqueurs de cellules gliales, semblent légèrement induits (C).*

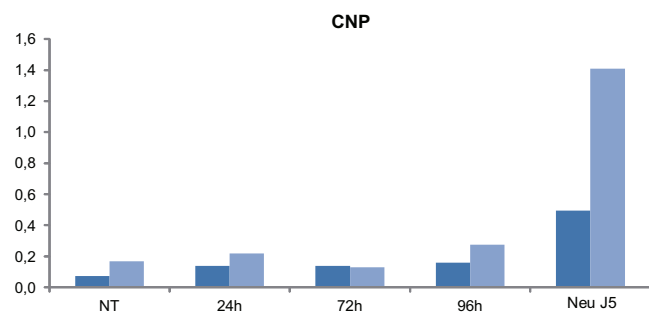
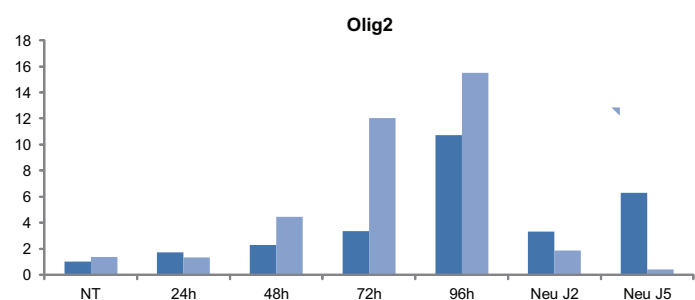
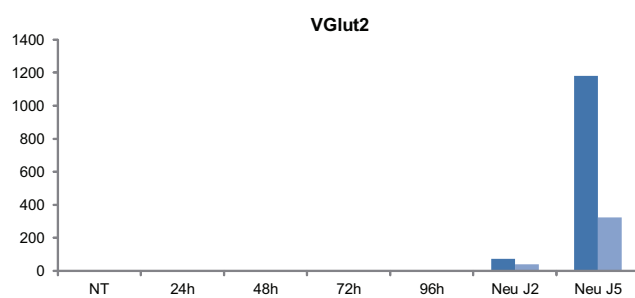
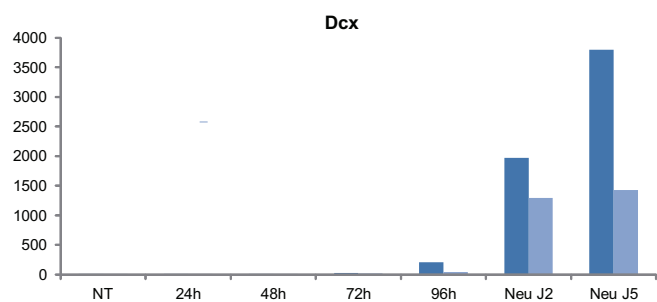
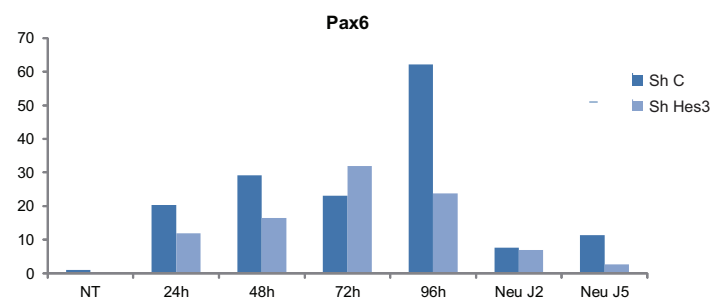
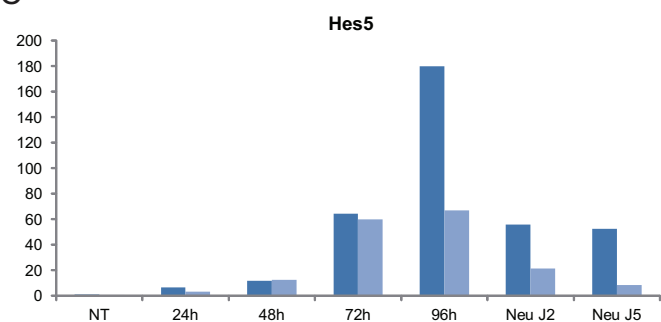
A



B



C



progéniteur ou non, nous avons choisi de nous intéresser à *Hes5*, qui intervient en aval de *Hes3* dans cette voie de différenciation. Comme le montre les résultats de la [figure 2C](#), l'expression de *Hes5* n'est pas affectée par l'absence de *Hes3* dans les premières heures de la différenciation neuronale. L'impact de la perte d'expression de *Hes3* sur *Hes5* n'apparaît qu'au temps 96h et est maintenu dans les neurones.

Nous nous sommes ensuite concentrés sur le marqueur des cellules radiales gliales qu'est *Pax6*. L'expression de ce facteur apparaît très tôt dans la différenciation neuronale, et continue d'augmenter jusqu'à 96h. La perte d'expression de *Hes3* semble affecter à un niveau relativement moindre celle de *Pax6*. *Dcx*, un marqueur légèrement plus tardif de la différenciation neuronale (marqueur des neurones immatures), apparaît quant à lui uniquement après 96h de traitement par l'AR des EB, et est fortement exprimé dans les neurones. Comme pour *Pax6*, il semblerait que la perte d'expression de *Hes3* affecte faiblement l'expression de *Dcx*. Nous nous sommes également intéressés à l'un des marqueurs « final » de la différenciation en neurones glutamatergiques, à savoir le récepteur *VGlut2*. Comme pour les marqueurs plus précoces, l'expression de *VGlut2* semble diminuer suite à la perte d'expression de *Hes3*. Pour finir, sachant que les facteurs Hes sont importants pour la différenciation en cellules gliales, nous nous sommes intéressés à l'expression d'*Olig2* et de *CNP* (2', 3'-cyclic-nucleotide 3'-phosphodiesterase). L'expression de ces deux facteurs est augmentée dans les cellules de la lignée sh*Hes3*. Il semblerait donc que *Hes3* soit important pour spécifier un devenir de neurones aux cellules, et empêcher la différenciation en cellules gliales.

### 2.3.Discussion et perspectives

*In vivo*, les facteurs Hes sont connus pour réguler différentes étapes du processus de différenciation neuronale. Alors que *Hes3* est important pour le maintien des cellules neuroépithéliales, *Hes5* est important à un stade ultérieur.

La connaissance du rôle de *Hes3* *in vivo* laisse présager d'une différenciation plus rapide en l'absence de ce facteur, puisque ce dernier est important pour le maintien et la prolifération des progéniteurs neuraux. D'après nos résultats, la différenciation neuronale n'est pas significativement affectée, et l'expression des marqueurs tels que *Pax6* ou *Dcx* ne semble pas décalée dans le temps, ce qui laisserait présager d'une différenciation précoce. En revanche, la diminution de l'expression de *VGlut2* et l'augmentation de celle de *CNP* pourrait suggérer

une différenciation plutôt gliale que neuronale. Néanmoins, malgré un nombre réduit de neurones, leur différenciation semble belle et bien complète.

Afin de comprendre réellement le rôle de Hes3 dans la différenciation neuronale, des études plus poussées seraient nécessaires. Dans un premier temps, il serait intéressant d'utiliser un protocole de différenciation optimisé afin d'avoir un meilleur silencing (Voir partie suivante sur l'optimisation du protocole). Ensuite, l'utilisation d'un anticorps dirigé contre Hes3 permettrait d'avoir une meilleure indication du niveau protéique dans la lignée shHes3. L'étude d'autres marqueurs au cours de la différenciation neuronale ou une étude à haut débit permettrait alors peut être de comprendre un peu plus en détail le rôle de Hes3 dans cette voie de différenciation. De plus, un immunomarquage avec des marqueurs précoces et tardifs de cellules gliales devrait être fait pour confirmer ou infirmer l'augmentation de ce type de cellules en absence de Hes3.

L'analyse de Hes3 serait dans tous les cas à associer avec une analyse similaire pour le facteur Hes5, ces deux membres de la famille Hes étant associés fonctionnellement *in vivo*.

### 3. BRN2 : optimisation du protocole de différenciation

#### 3.1.Objectif de l'optimisation

Au cours des expériences de surexpression de BRN2, nous nous sommes aperçu que cette surexpression était souvent perdue au niveau des EB. Nous avons donc tenté de raccourcir le protocole de différenciation, afin de limiter ces pertes. Le protocole raccourci de deux jours permet également d'obtenir des précurseurs neuronaux. Afin d'homogénéiser l'ensemble des expériences, nous avons donc reconduit les expériences avec le shBrn2 dans des conditions raccourcies, similaires aux conditions de la surexpression. Néanmoins, avant de procéder au changement de protocole, nous avons déjà réalisé des expériences de shARN dans les conditions normales de différenciation. L'observation des différences sur les lignées shBrn2 nous a convaincu que le protocole raccourci était plus approprié à notre démarche expérimentale.

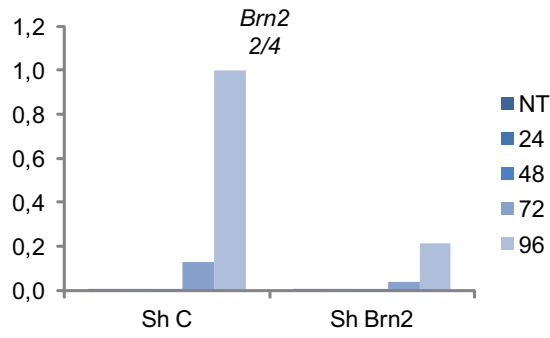
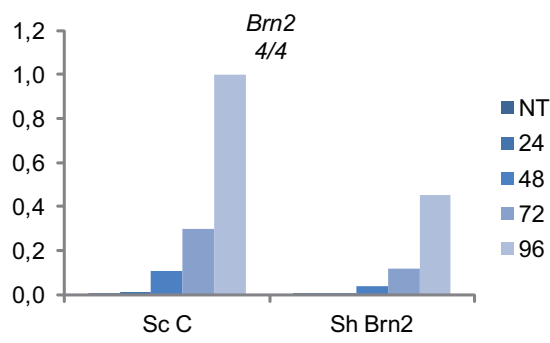
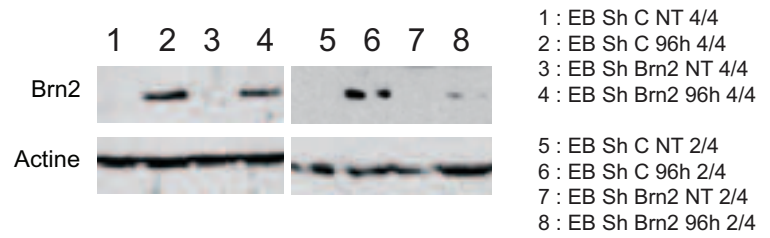
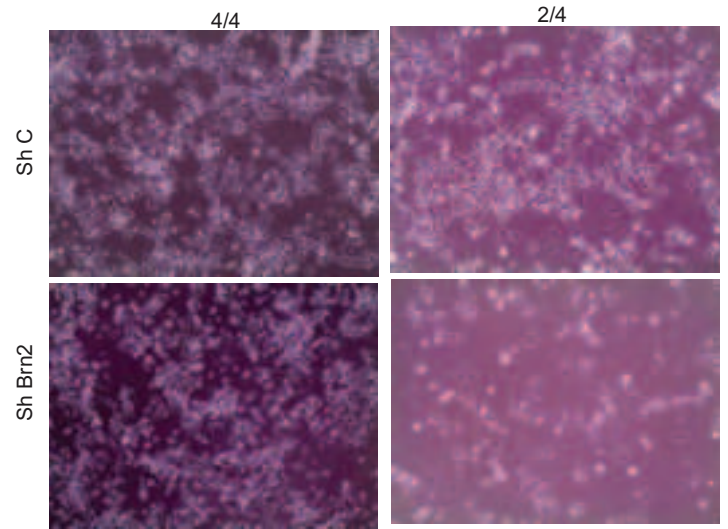
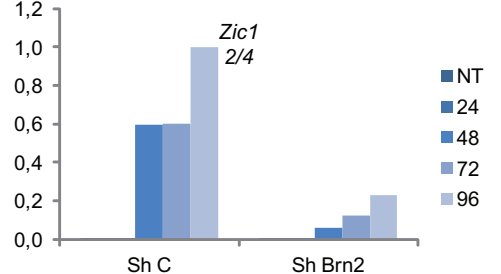
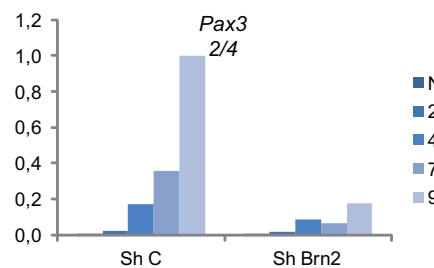
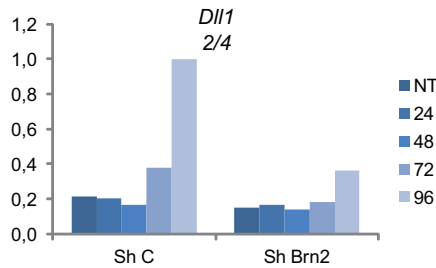
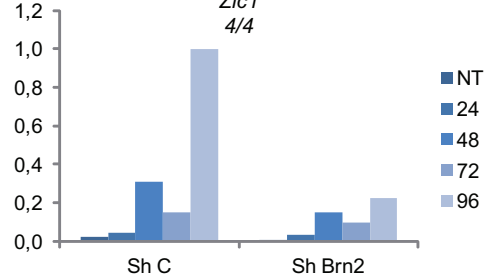
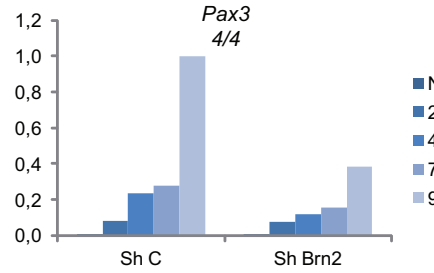
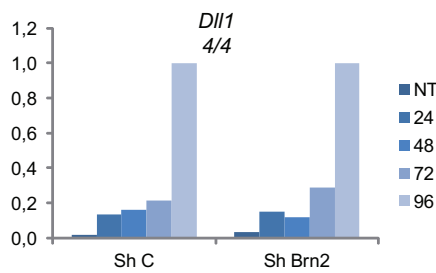
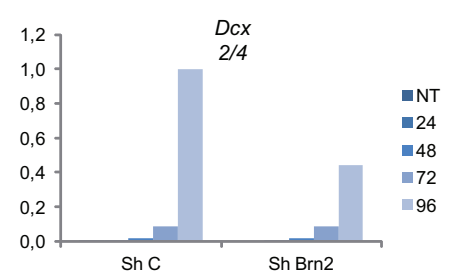
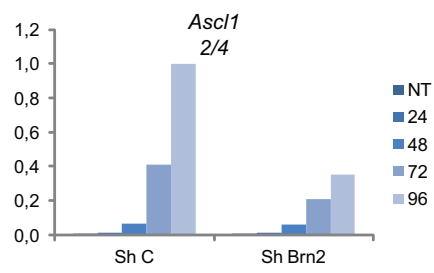
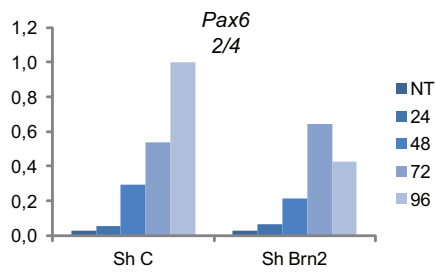
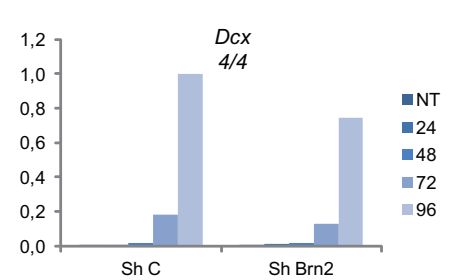
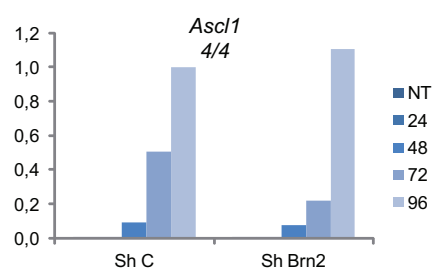
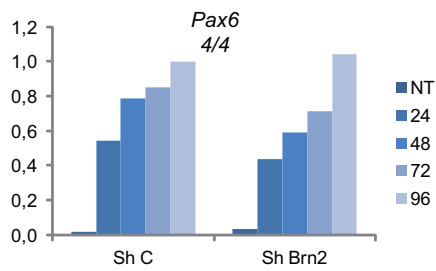
#### 3.2.Résultats comparatifs des deux protocoles de différenciation

##### 3.2.1. Réduction du niveau d'expression

Le niveau d'expression de *Brn2* est dans un premier temps mesuré par RT-qPCR lors de la différenciation neuronale. Le profil d'expression de *Brn2* est similaire dans les deux lignées contrôles, confirmant que le raccourcissement du protocole de différenciation n'altère pas l'expression des gènes durant cette différenciation (*Figure 3A*). L'étude de l'expression de *Brn2* dans la lignée shBrn2 montre qu'après 96h de traitement par l'acide rétinoïque, la baisse d'expression est de 2x pour le protocole 4/4. En revanche, le protocole 2/4 permet d'améliorer cette réduction d'expression et d'atteindre un facteur 5 de différence (*Figure 3A*).

Tout comme ce qui est observé au niveau de l'ARN, l'expression protéique de Brn2 est altérée significativement pour le protocole 2/4, alors que seule une faible baisse d'expression est visible pour le protocole 4/4 (*Figure 3B*).

*Figure 3 : Impact de la durée du protocole de différenciation sur l'efficacité de l'approche lentivirale. Après le traitement des EB par l'acide rétinoïque, l'expression de Brn2 est induite progressivement entre les EB non traités (NT) et traités pendant 96h à l'AR. L'expression mesurée par RT-qPCR montre une baisse significative dans le protocole 2/4, et un effet moindre dans le protocole 4/4 (A). Cette baisse d'expression est confirmée au niveau protéique par Western blot (B) ainsi qu'au niveau du phénotype (C). L'expression de tout un ensemble de marqueurs neuronaux est affectée par la baisse d'expression de Brn2 dans la différenciation 2/4. La différence d'expression de ces marqueurs entre la lignée sh C et sh Brn2 lors de la différenciation 4/4 est moindre, voire absente (D).*

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

### 3.2.2. Observation du phénotype

Lors de la différenciation dite 4/4, seule une très faible réduction de la différenciation est observée pour la lignée shBrn2. Néanmoins, de neurones et projections axonales sont toujours visibles. En revanche, l'utilisation du protocole 2/4, pour lequel la baisse d'expression de Brn2 est vraiment significative, permet d'observer une abolition de la différenciation neuronale. Les cellules résiduelles n'ont pas les caractéristiques des neurones (cellules réfringentes, corps cellulaires assez petits et présence d'axones) (*Figure 3C*). Ceci indique que le niveau d'expression résiduel de Brn2 est directement relié à la capacité des cellules à se différencier en neurones.

### 3.2.3. Impact sur certains gènes cibles

Pour l'ensemble des gènes cibles testés par RT-qPCR, le profil d'expression du shC 4/4 est globalement similaire au profil d'expression 2/4. Pour *Pax6*, marqueur de cellules radiales gliales, *Ascl1*, ou *Dcx1* (marqueur de neurones immatures), l'expression n'est pas altérée dans la différenciation 4/4. En revanche, dès que le silencing de Brn2 est meilleur, un impact est visible sur l'expression de ces facteurs. D'autres facteurs tels que *Zic1* et *Pax3* sont quant à eux affectés au niveau de l'ARNm et ce malgré une perte minime de l'expression de Brn2 (*Figure 3D*). Ces gènes sont donc sensibles à une faible variation du niveau d'expression de Brn2.

## 3.3. Conclusion

L'utilisation du protocole de différenciation tel qu'il est décrit (Bibel, Richter et al. 2004; Bibel, Richter et al. 2007) ne nous permet pas d'avoir de résultats très encourageants dans notre approche shARN. En revanche, l'optimisation du protocole qui consiste en un raccourcissement de deux jours permet d'obtenir des résultats significatifs, tant d'un point de vue perte d'expression que phénotype. Ce protocole n'affecte en rien le profil d'expression de l'ensemble des gènes testé, la différenciation suite au traitement pas l'AR se déroulant donc normalement.

# ***Discussion***

## Discussion

### 1. Brn2 spécifie l'identité neuronale des cellules ES.

Le rôle de Brn2 dans le processus de différenciation neuronale a été mis en évidence à la fois *in vivo* avec les souris KO mais également *in vitro* dans des cellules P19. Néanmoins, ces différentes études n'ont pas permis d'identifier l'ensemble des gènes régulés par Brn2 ni ses gènes cibles directes.

Par une approche combinant des analyses de RNA-seq et de ChIP-seq, nous avons mis en évidence les gènes dérégulés suite aux variations d'expression de Brn2, ainsi que les gènes fixés directement par ce facteur. Comme le laissaient présager les données précédemment publiées, Brn2 est important pour l'expression d'un ensemble de gènes impliqués dans la neurogenèse. Les gènes de la voie de signalisation Notch semblent particulièrement affectés (*Dll1*, *Dll3*, *Jag1* qui sont les ligands de Notch, ainsi que *Hes3* et *Hes5* qui sont deux effecteurs de la voie), ce qui impliquerait un défaut dans l'une des premières étapes de la neurogenèse. De plus, la migration des neurones semble également affectée suite à la perte ou à la surexpression de Brn2. Lors de la mise en place du cortex, deux voies de signalisation, la voie Cdk5 et la voie Reelin, semblent importantes pour la migration des neurones. Ces deux voies sont affectées dans nos expériences puisque l'expression de *Reelin*, *Cdk5r1* et *Cdk5r2* est affectée dans au moins une de nos conditions (Perte d'expression et/ou surexpression de Brn2). L'importance de ces deux voies avait déjà été décrite pour Brn2. Les gènes régulés par Brn2 déjà identifiés *in vivo* sont donc confirmés dans nos expériences, mais de nombreux autres gènes important dans la neurogenèse s'ajoutent à cette liste (*Dbx1*, -2, *Gad1*, -2, *Irx1*, -3, -5 ...)

### 2. Un faible nombre de site de fixation hautement conservés

Alors que nos expériences nous ont permis d'identifier de nombreux sites de fixation dans les cellules ES, la comparaison avec d'autres modèles de différenciation ainsi qu'avec nos données dans la lignée P19 a permis de mettre en évidence un faible nombre de sites conservés entre les différents lignées / processus de différenciation. Alors que des sites potentiels de fixation de Brn1 et Brn2 ont été définis précédemment notamment par recherche bioinformatique des séquences de fixation au niveau des gènes, peu de ses sites sont réellement fixés dans nos données. En guise d'exemple, Castro et al. ont mis en évidence que

les régions à proximité des gènes *Cdk5r1* et *Cdk5r2* contenaient des sites de fixation potentiels pour Brn1 et Brn2. La dérégulation de l'expression de ses gènes étant visible dans nos expériences, l'hypothèse qu'ils constituent des cibles directes de Brn2 aurait pu être émise. Néanmoins, aucun des ChIP analysés ne met en évidence la fixation de Brn2 au niveau de *Cdk5r1*. En revanche, *Cdk5r2* est fixé, mais uniquement dans les cellules P19. Ce site ne fait donc pas partie des cibles hautement conservées quelque soit la voie de différenciation neuronale mis en œuvre.

Les gènes associés aux sites de fixation hautement conservés dans les différents processus de différenciation représentent sans doute les cibles minimales reconnues par Brn2 et importantes dans la différenciation. Parmi ces gènes, qui constituent en quelque sorte le programme minimal nécessaire à la différenciation neuronale se trouvent notamment *Dlx1*, -2, *Pou6f1*, *Zic1*, -4. De manière très intéressante, parmi les sites qui sont hautement conservés, un très grand nombre contient le motif consensus défini pour Brn2, à savoir le MORE.

De nombreuses modifications post traductionnelles de Brn2 ont été mises en évidence, et pourraient être une des raisons expliquant les cibles différentielles entre l'ensemble des lignées. Une autre raison pourrait venir des autres facteurs protéiques qui peuvent permettre le recrutement de Brn2. Alors que sur les sites hautement conservés, les motifs MORE ou MORE+1 sont souvent retrouvé seuls, la présence de nombreux sites pour d'autres facteurs de transcription est visible à proximité des sites fixés par Brn2 dans les cellules ES. Il semblerait donc que Brn2 soit important pour initier le programme de différenciation neuronale. En revanche, les différents processus de différenciation ne mènent pas au même type de neurones finaux, ce qui peut donc expliquer un programme spécifique dans chaque type de différenciation. La présence de sites de fixation pour d'autres facteurs de transcription pourrait alors permettre d'envisager une collaboration avec ces facteurs spécifiquement dans la différenciation en neurones glutamatergiques.

De plus, Lodato et al. ont précédemment décrit que les sites de fixations de Brn2 étaient particulièrement enrichi au niveau des éléments enhancers (Lodato, Ng et al. 2013). La comparaison de nos données de ChIP dans les EB 96h avec leurs données pour les enhancers dans les NPC ne donne néanmoins qu'une faible corrélation. Nous sommes actuellement en train de réaliser, dans les EB96h, un ChIP pour les marques d'histones H3K27Ac et H3K4me, ce qui nous permettra d'identifier les enhancers spécifiques à notre processus de

différenciation. Encore une fois, en fonction du programme de différenciation mis en œuvre il se peut que les enhancers fixés soient différents, et la comparaison des résultats dans les NPC avec ceux dans les EB96h n'est alors que subjective. Dans tous les cas, il semblerait que Brn2 ne soit pas exclusivement localisé au niveau des enhancers, puisque dans les EB traités 96h de nombreux sites sont retrouvés au niveau de promoteurs proximaux. Il en est de même lorsque l'analyse est faite en utilisant les 72 sites hautement conservés entre les lignées, ce qui indique l'importance de la fixation de Brn2 au niveau des promoteurs proximaux.

### **3. Absence de redondance fonctionnelle entre les protéines Brn1 et Brn2**

Les études faites *in vivo* ont permis de mettre en évidence une redondance fonctionnelle de Brn1 et de Brn2. En effet, le double KO de Brn1 et Brn2 est nécessaire afin d'observer un phénotype drastique dans le cortex. En suivant cette logique, la simple réduction de l'expression de Brn2 aurait pu ne donner aucun phénotype dans notre processus de différenciation. Or le phénotype observé suite à la baisse d'expression de Brn2 seul est significatif. De plus, l'idée de la redondance fonctionnelle laisserait présager une augmentation de l'expression de Brn1 lorsque celle de Brn2 est réduite (et inversement). Or, dans nos expériences de RNA-seq, *Brn1* n'est pas significativement affecté. Néanmoins, l'expression de *Brn1* semble parallèle à celle de Brn2. En effet, dans la lignée shBrn2 l'expression de Brn1 est également diminuée, alors que dans la lignée qui surexprime Brn2, l'expression de Brn1 est faiblement augmentée. Afin de confirmer cela, un Western blot utilisant un anticorps spécifique contre Brn1 devrait être réalisé.

Qui plus est, Brn1 a un profil d'expression similaire à celui de Brn2 lors de la différenciation neuronale. Quel est alors le rôle de Brn1 dans ce processus. De manière similaire à ce qui a été réalisé pour Brn2, l'identification du réseau transcriptionnel de Brn1 pourrait être réalisée. La perte de Brn1 aurait-elle un phénotype aussi drastique que celle de Brn2 ? Est-ce que Brn2 serait capable de pallier l'absence de Brn1 ? Brn1 se fixerait-il aux cibles potentielles identifiées par analyses bioinformatique (telle que Cdk5r1 cité précédemment) ? Cette expérience permettrait de répondre à ces questions, et peut être de comprendre comment ces deux protéines sont régulées. Brn1 n'est visiblement pas une cible directe de Brn2, l'inverse serait-il alors vrai ?

#### 4. Quel rôle pour Brn5 ?

Outre Brn1 et Brn2, une autre protéine de la même famille semble particulièrement intéressante dans nos études : Brn5 (Pou6f1). L'expression de Pou6f1 est drastiquement affectée que ce soit lors de la perte d'expression de Brn2 dans les ES, lors de la surexpression de Brn2 dans les ES, ou encore lors de la perte d'expression de Brn2 dans les cellules P19. De plus, le site de fixation de Brn2 au niveau de Pou6f1 est hautement conservé dans l'ensemble des ChIP analysés. Le site de fixation de Brn2 au niveau du locus Pou6f1 contient un élément MORE, et un élément MORE +1. L'ensemble de ces résultats d'analyses laisse présager que Brn5 joue un rôle important dans notre processus de différenciation. Très peu de données concernant Brn5 sont connues à ce jour. Tout comme Brn2, Brn5 est exprimé lors du développement du cerveau ainsi qu'au niveau du cortex adulte (Cui and Bulleit 1998). Très peu d'études existent néanmoins pour ce facteur. Tout comme ce qui a été réalisé pour Brn2, il serait alors intéressant d'allier RNA-seq et ChIP-seq afin de déterminer le rôle de Brn5 dans la différenciation neuronale. L'analyse combinée des données pour Brn1, Brn2 et Brn5 permettrait de déterminer plus précisément la fonction de chacun de ces facteurs dans le processus de différenciation neuronale.

#### 5. La phosphorylation de Brn2 est-elle importante dans le processus de différenciation neuronale ?

L'ensemble des protéines POU possèdent un site de phosphorylation hautement conservé entre les différents membres, ce qui laisse supposer l'importance de cette phosphorylation dans la fonction de ces protéines (Andersen and Rosenfeld 2001). Des expériences récentes ont mis en évidence le rôle de la phosphorylation de Brn2 dans la migration et la prolifération des mélanocytes (Berlin, Denat et al. 2012). Il serait intéressant de déterminer si cette phosphorylation a également un impact sur la différenciation neuronale. Avant tout, la question qui se pose est de savoir si la version de Brn2 endogène exprimée lors de la différenciation neuronale est phosphorylée ou non. Cette phosphorylation de Brn2 au niveau de la sérine 362 se fait par l'enzyme protéine kinase A (PKA). Une première étape serait bien évidemment de vérifier en utilisant un anticorps contre cette enzyme, sa présence lors de la différenciation neuronale. Les niveaux d'expression des gènes *Prkaca*, *Prkacb*, *Prkx* codant pour les sous unités catalytiques de la PKA sont très élevés dans les EB96h (RNA-seq). Ceci laisse présager la présence de la PKA et donc la potentielle phosphorylation de BRN2. De plus, les études précédentes ont mis en évidence que la version de Brn2 phosphorylée est

incapable de se fixer aux motifs NORE et PORE. L'absence de sites de fixation au niveau de ce type de motifs dans notre étude laisse présager que Brn2 soit phosphorylée. L'utilisation d'un anticorps anti-phospho-sérine pourrait permettre de déterminer si la protéine endogène est phosphorylée, mais aucune information n'est à ce jour connu quand à d'autres sites potentiels de phosphorylation, et il n'est donc pas exclu que Brn2 soit phosphorylée à un autre site.

Afin d'étudier en détail l'importance de cette phosphorylation lors de la différenciation neuronale, il faudrait envisager de générer une lignée de cellules ES KO pour Brn2 et de tester la possibilité de compensation par expression de la version sauvage ou mutée de Brn2.

## **6. Zic1, une cible directe de Brn2 ?**

Zic1 semble d'après nos expériences être un facteur clé de la différenciation neuronale. Si Zic1 joue un rôle important dans le développement neuronal chez la souris, à ce jour, l'ensemble de ses cibles directes ne sont pas connues. Il n'existe pour l'instant aucun anticorps spécifique pour ce membre de la famille Zic, ce qui empêche d'identifier le réseau transcriptionnel et les gènes cibles directes de ce facteur. La génération d'une lignée de cellules sur-exprimant une version étiquetée de Zic1 serait une solution afin de réaliser à la fois des expériences de ChIP-seq, mais également des expériences de co-immunoprécipitation, afin d'identifier les partenaires protéiques de Zic1 dans ce processus de différenciation.

De plus, Zic1 est souvent associé à Zic4 dans la différenciation. Premièrement, leur profil d'expression est similaire lors de la différenciation neuronale des cellules ES. Deuxièmement, ces deux membres de la famille Zic sont associés avec le syndrome de Dandy-Walker chez l'homme. Nous nous sommes intéressés dans nos expériences spécifiquement à Zic1, car son niveau d'expression était supérieur à celui de Zic4. Néanmoins, les deux gènes ont un profil d'expression similaire lors de la différenciation neuronale. Quel serait le phénotype associé à la perte d'expression de Zic4 ? Il n'y a visiblement pas de redondance fonctionnelle entre ces deux gènes, puisque la perte d'expression de Zic1 entraîne une diminution de l'expression de Zic4. Le silencing de Zic4 pourrait nous permettre de savoir si l'inverse est également vérifié, et de comprendre un peu plus spécifiquement la fonction de ces deux membres de la famille Zic. Les modèles animaux de mutant de Zic1 et Zic4 se multiplient (xénope, teleost, zebrafish, médaka) mais l'identification des séquences reconnues par ces deux facteurs Zic n'est à ce jour pas réalisée.

## 7. Une alternative à l'approche lentivirale : le système Crispr-Cas9

Depuis les dernières années, de nouvelles techniques ont permis une avancée importante dans la modification et l'édition du génome de plusieurs organismes. Différents systèmes ont été utilisés afin de cibler ces modifications dans le génome. Les nucléases à doigt de Zinc (ZFN : Zinc Finger Nuclease), les nucléases TALEN (Transcription Activator-Like Effector Nuclease), et plus récemment les nucléases de type CRISPR-Cas9 (CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats) ont été utilisées (pour revue (Gaj, Gersbach et al. 2013)).

L'approche lentivirale que nous avons mise en œuvre permet d'obtenir une lignée stable présentant une réduction de l'expression du gène ciblé. Nous avons vu au cours de nos expériences que cette stabilité était toutefois relative, puisque cette perte d'expression peut être moindre quand le temps de l'expérience se rallonge. De plus, qui dit perte d'expression, dit également expression résiduelle. Ainsi, il restera toujours une faible expression de la protéine ciblée, ce qui peut expliquer des effets phénotypiques moindre.

Dans un premier temps, l'approche par CRISPR-Cas9 pourrait être utilisée afin de réaliser des lignées de cellules ES KO pour les facteurs Brn2 et Zic1. L'absence totale de l'expression de ces deux protéines permettrait sans doute d'identifier des gènes cibles supplémentaires, qui n'était pas significativement dérégulés dans l'approche lentivirale. De plus, comme je l'ai décrit précédemment, générer une lignée KO permettrait également d'étudier le rôle de la phosphorylation de Brn2. Réexprimer dans une lignée KO soit la version sauvage, soit la version mutée au niveau des sites de phosphorylation nous aiderai à comprendre le rôle de cette phosphorylation. De plus, ce système ne permet pas uniquement de générer des lignées KO, mais également d'éditer les gènes et de créer une mutation spécifique. Il serait alors possible de modifier directement la séquence génique de Brn2 afin d'exprimer une protéine « endogène » ne pouvant pas être phosphorylée.

Pour finir, il est assez difficile avec l'approche lentivirale d'imaginer un double voir un triple silencing. Le système CRISPR-cas9 a récemment permis d'agir sur cinq gènes en parallèle dans une lignée de cellules ES (Wang, Yang et al. 2013). Cette approche permettrait alors de réaliser une double inactivation de Brn1 et Brn2, voir même une triple inactivation Brn1, -2, -5. Il en va de même pour les facteurs de la famille Zic, Zic1 et Zic4. De plus, comme je l'ai évoqué précédemment, nous avons essayé par approche lentivirale de cibler de

nombreux autres facteurs de transcription tel que Hes3 et Hes5, Gbx2, Dbx1, Msx1, les membres de la famille Irx (Irx1, -3, -5) mais sans résultat concluant. Cette absence de résultats peut s'expliquer par une mauvaise réduction de l'expression par approche lentivirale ou par une redondance fonctionnelle entre les différents membres d'une famille. L'utilisation du système CRISPR-cas9 pourrait pallier à ces problèmes et permettre de comprendre le rôle de nombreux facteurs de transcription dans la différenciation neuronale des cellules ES induite par l'AR.

# ***Annexes***

***Publication annexe***

***n°1***

## **How the RAR-RXR heterodimer recognizes the genome**

Sylvia Urban, Tao Ye and Irwin Davidson#

Department of Functional Genomics and Cancer

Institut de Génétique et de Biologie Moléculaire et

Cellulaire.

CNRS/INSERM/UDS.

1 Rue Laurent Fries,

67404 Illkirch Cédex.

France.

FAX: 33 3 88 65 32 01. TEL: 33 3 88 65 34 40 (45)

# To whom correspondence should be addressed

E mail: [irwin@igbmc.fr](mailto:irwin@igbmc.fr)

Running Title: RAR genomic occupancy

Key words: Vitamin D, thyroid hormone, estrogen, chromatin

immunoprecipitation.

## 1. Introduction

All-*trans*-retinoic acid (atRA) is the naturally active vitamin A metabolite known to exert a wide range of effects on vertebrate development and to play critical roles in the homeostasis and physiopathology of many adult tissues (Mark, et al., 2006, Mark, et al., 2009) (see the chapters of sections XX and XX). This simple molecule exerts most of its pleiotropic effects through interaction with three members of the nuclear receptor superfamily, the all-*trans*-retinoic acid receptors (RAR)  $\alpha, \beta, \gamma$  that heterodimerize with the retinoid X receptors (RXR)  $\alpha, \beta, \gamma$  to act as RA-dependent transcriptional regulators that either activate or repress target genes (Rochette-Egly and Germain, 2009). Nuclear receptors are key molecules that can be targeted by synthetic drug molecules in human diseases such as cancer, psoriasis, osteoporosis, or type II diabetes.

Nuclear receptors (NRs) share a common modular structural organization comprising an N-terminal domain variable in length and sequence harboring a ligand-independent activation function (AF-1), the conserved DNA-binding domain (DBD) and the C-terminal ligand-binding domain (LBD) that comprises the ligand-dependent activation function AF-2. The LBD is a key regulatory domain containing the ligand-binding pocket and multiple interaction surfaces for homo- or heterodimerization and interactions with corepressors, coactivators and other cofactors that participate in signal transduction to the basal

transcriptional machinery (Perissi and Rosenfeld, 2005) (Bulyanko and O'Malley, 2010) (see chapters XX and XX - Rochette-Egly, Moras). The mechanisms of NR function in general and RAR function in particular have come under intense study at both the structural and functional level. From this large body of work a general model proposes that the ability of NRs to modulate target gene transcription results from complex and dynamic interactions between the receptors and multiple coactivator/corepressor complexes. In the absence of ligand, many NRs associate with corepressor complexes with histone deacetylase activity (HDACs) that modify chromatin to establish and maintain a repressed transcriptional state (Glass, C. K. and Rosenfeld, 2000, Nagy, et al., 1999). The presence of ligand induces a structural transition in the LBD leading to release of the corepressors and the formation of a novel interaction surface for coactivators including histone acetyl- and methyl-transferases, as well as chromatin remodelling complexes or components of the basal transcription machinery (Aoyagi and Archer, 2008). This general model holds true for many of the class I (acting independently of RXR) and class II (heterodimerizing with RXR) NRs, whereas orphan NRs with no identified ligand are rather thought to function as constitutive activators.

In this brief chapter, we will not cover the extensive literature on the mechanisms of gene activation and repression, but rather discuss how high throughput genomics

has helped to understand the first essential step in the whole gene regulatory process, namely how NRs in general and the RAR-RXR heterodimer in particular interact with the genome.

## **2. Hormone response elements**

Nuclear receptors bind specific sequences on DNA called hormone response elements, either as homodimers such as the steroid hormone receptors [e.g. estrogen receptor (ER)], as heterodimers with RXR [e.g. RARs, vitamin D3 receptor (VDR) thyroid hormone receptor (TR) or peroxisome proliferator activated receptor (PPAR)], or more rarely as monomers. Amongst the NRs, the RXRs are unique as they can act both as homodimers or heterodimers, although there is evidence that RARs can form homodimers *in vitro* and perhaps function in an RXR-independent manner in testis Sertoli cells *in vivo* (Osz, et al., 2012, Vernet, et al., 2006). The NR DBDs comprise two coordinated zinc ions and two alpha helices arranged in a perpendicular fashion one of which, the recognition helix, makes extensive contacts within the major groove to mediate sequence specific DNA recognition (Khorasanizadeh and Rastinejad, 2001). An interesting feature of NRs is their ability to recognize DNA elements with diverse spacing and topologies. Class II NRs that bind as heterodimers with RXR recognize direct repeats (DR) of the consensus half site sequence 5'-RGKTCA-3' separated by 1-5 nucleotides (DR1-DR5). This was originally formulated as the 1-5 rule where RAR-RXR

dimers bind DR1, DR2 and DR5, VDR-RXR binds DR3, TR-RXR binds DR4, while RXR homodimers also bind DR1 (Mangelsdorf, et al., 1995, Umesono, et al., 1991). Indeed the first identified RA-response elements (RAREs) correspond to DR2 and DR5 elements {de The, 1990 #8021}(Balmer and Blomhoff, 2005). In contrast, steroid receptor homodimers such as ER recognize inverted repeat elements spaced by 3 nucleotides (IR3) {Carroll, 2006 #4222; Welboren, 2009 #4588}.

### **3. VDR and TR occupy DR3 and DR4 elements in cultured cells**

While the first functional DR2 and DR5 type RAREs were identified many years ago, the number of functionally characterized RAREs has remained limited until recently. A similar situation exists for vitamin D-response elements (VDREs) and thyroid hormone-response elements (TREs), where a small number of such elements have been characterized in the promoters of target genes, but their global nature and genome distribution were poorly investigated. The recent advent of genome-wide techniques for investigating transcription factor distribution such as chromatin immunoprecipitation coupled to array hybridization (ChIP-chip) or to deep sequencing (ChIP-seq) allowed verification of the validity of the 1-5 rule for NRs on a genome wide scale.

Genome-wide analysis of VDR-RXR binding by ChIP-seq in human lymphoblastoid cell lines identified 623 VDR binding sites in absence of ligand and more than 2700 in the presence of

calcitriol, showing that the presence of ligand stimulated VDR genomic occupancy (Ramagopalan, et al., 2010). Calcitriol also stimulated the expression of 226 genes in these lines, of which 23% had a VDR binding site within 5 kilobase pairs (kb) of the transcription start site (TSS). Analysis of the VDR bound sequences showed a strong enrichment in DR3 elements with only a much smaller number of DR4 elements, but little or no representation of other spacing or topologies. Moreover, strongly occupied sites had a higher proportion of fully consensus elements compared to weaker occupied sites. The DR3 is therefore a major element recognized by the VDR-RXR heterodimer in the genome (Fig. 8.1).

A second study on genome-wide VDR localization was performed in THP1 monocytic leukemia cells (Heikkinen, et al., 2011). This independent analysis also revealed that VDR genome occupancy was strongly stimulated by ligand with a large increase in the number of binding sites in which DR3 was strongly enriched, and this despite the fact that exposure to ligand was short (40 minutes) in the case of Heikkinen et al., compared to 36 hours in the study of Ramagopalan et al. Nevertheless, comparisons of the two studies revealed that the repertoire of vitamin D3-regulated genes is rather distinct in the two cell models used (Carlberg, et al., 2012). Moreover, although the overall number of VDR-occupied sites was comparable in the two studies, after ligand stimulation only around 18% of the VDR occupied sites were common to the two

cell types used. While some of this difference may be due to the different times of exposure to ligand, there must also be a significant degree of cell-specific occupancy by the VDR of its cognate elements, a phenomenon also seen for the RAR-RXR heterodimer (see below).

A more surprising outcome from the comparison of the data sets is that while the DR3 is the only strongly enriched half site spacing, only around 30% of VDR-occupied sites comprise DR3s (Carlberg, et al., 2012). This fraction drops to only 14% of the sites that are occupied in the absence of ligand. *De novo* motif searches at the sites lacking DR3 elements failed to reveal other highly represented sequence motifs that would act as VDR binding elements. Thus, how VDR binds to sites lacking DR3 elements that constitute a large proportion of the overall repertoire remains to be fully determined.

The genome wide distribution of the TR $\alpha$ 1 and TR $\beta$ 1 receptors has also been investigated in C17.2 neural cells using a variation of ChIP-seq (ChAP-seq) that uses affinity purification of the tagged receptors to identify their binding sites (Chatonnet, et al., 2013). This analysis identified around 1000 sites selectively bound by TR $\alpha$ 1, but only 16 bound specifically by TR $\beta$ 1 and 395 bound by both receptors. Analysis of the bound sequences identified DR4 as the only enriched spacing and topology, although previous studies had suggested that IR0 and ER (everted repeat) 6 may also act as TREs

(Desvergne, 1994, Glass, C.K., et al., 1988). Interestingly, while the occupied DR4 elements conform well to the consensus sequence, the T in position 4 of the second half site bound by the TR is poorly conserved, when compared to the DR3 bound by VDR, suggesting that this nucleotide is not an important determinant of recognition by the TR as it is for other NRs (Fig. 8.1). Integration of the genome occupancy data with gene expression profiling in the presence and absence of thyroid hormone revealed that a vast majority of the induced genes displayed DR4-containing bound sites within 30 kb of the TSS. In contrast, no correlation of TR genome occupancy with down-regulated genes was observed, suggesting that in these cells TRs act uniquely as transcriptional activators. Moreover, this analysis showed that receptor selective transcription regulation does not evidently correlate with receptor selective genome occupancy. These results confirm that DR4 is the most enriched element recognized by the TR-RXR heterodimer in the genome.

#### **4. PPAR and LXR prefer non-consensus elements**

While the VDR- and TR-RXR heterodimers preferentially bind the consensus DR3 and DR4 motifs, respectively, other NRs bind much more degenerate motifs. Genome wide profiling of PPAR $\gamma$  in 3T3L1 cells induced to differentiate into adipocytes shows that occupied sites are enriched in a highly degenerate DR1 motif (Nielsen, et al., 2008) (Fig. 8.1). A similar conclusion

was drawn from genome-wide profiling of PPAR $\alpha$  in mouse liver that again revealed a degenerate DR1 as the preferred site for PPAR-RXR heterodimers (Boergesen, et al., 2012). Analysis of LXR-RXR heterodimer binding in mouse liver revealed enrichment of DR4 previously suggested as a response element for LXR-RXR heterodimers (Apfel, et al., 1994). Comparison of the DR4 elements bound by TR-RXR and by LXR-RXR indicates that the LXR-occupied sites show a more degenerate consensus sequence compared to TR (Fig. 8.1). In addition however, LXR-RXR also bound DR1, and to a lesser extent IR1 and ER2 elements. Many sites bound by LXR-RXR are also bound by RXR in livers of LXR knockout mice, thus in absence of the LXR receptor. These sites are enriched in degenerate DR1 elements and are also occupied independently by the PPAR-RXR heterodimer. The binding of the LXR-RXR and PPAR-RXR heterodimer to a number of shared sites results in cross-talk between the two signaling pathways in terms of gene expression, where agonists of PPAR $\alpha$  can diminish the effect of LXR agonists on genes with shared binding sites. Similarly, many of the shared PPAR-LXR-RXR-occupied sites are also binding sites for the NR HNF4 $\alpha$  in liver. The consensus binding sequence for PPAR $\alpha$  derived from these experiments is in fact very similar to the consensus derived for HNF4 $\alpha$  (Schmidt, et al., 2010) (Fig. 8.1). HNF4 $\alpha$  is one of the most promiscuous binders amongst the NRs and amongst transcription factors in general, with up to 60000

occupied sites in the genome. Interestingly, a comparison of HNF4 $\alpha$  genomic occupancy in the livers from many species (a study known as "zoo-ChIP") revealed the existence of sites that are occupied in many species, and those that are specific for a given species (Schmidt, et al., 2010). This type of study provides insights into the co-evolution of transcription factor binding and gene regulation.

Hence, while genome-wide studies confirm that consensus DR3 and DR4 elements are the preferred binding elements for the VDR- and TR-RXR heterodimers, they show that the PPAR- and LXR-RXR heterodimers rather bind to more degenerate non-consensus binding elements in cells and tissues. They also reveal that cross-talk between the PPAR- and LXR-mediated signaling pathways takes place at the level of DNA binding through the competitive occupancy of these receptors of common degenerate DR1 elements.

## **5. Genomic binding of the RAR-RXR heterodimer**

As mentioned above, RAREs were first defined as DR2 and DR5 elements. Nevertheless, initial *in vitro* studies suggested that RAR-RXR can bind other DR-spacing (Mader, et al., 1993) and, thus, the diversity of spacing and topology in the repertoire of genomic elements bound by the RAR-RXR heterodimer may be much more complex than seen for other NRs. Genome-wide occupancy by the RAR-RXR heterodimers has been

investigated in several model systems both by ChIP-chip and by ChIP-seq.

**5.1. Antagonism between RA and estrogen signaling in breast cancer cells.** In a first genome-wide study, the binding of RARs in MCF7 breast cancer cells was investigated. In breast cancer cells, RA negatively regulates cell proliferation and can induce the expression of pro-apoptotic genes. In contrast, activation of the ER by estrogen promotes proliferation and survival of breast cancer cells. To understand these antagonistic effects, a large scale ChIP-chip study was performed using eGFP tagged RAR $\alpha$  and RAR $\gamma$  expressed from integrated BAC transgenes in MCF7 cells (Hua, et al., 2009). This approach allowed identification of a large set of sites co-occupied by both receptors with only a smaller number of sites occupied in an isotype-specific manner. Analysis of the DNA sequences at the occupied sites revealed that the most commonly represented motif was DR5. Nevertheless, enrichment in DR0, DR1, DR10, IR0, ER8 and ER2 were also observed, suggesting that the RAR-RXR heterodimer can bind to a diverse set of spacing and topologies.

Comparison of these data sets with ChIP-chip datasets for ER in the same cells indicated a large overlap between the ER and RAR occupied sites. At genes that are antagonistically regulated by RA and estrogen, 71% of the ER and RAR binding sites overlap such that one of the half sites is shared,

leading to competitive binding between the two cognate receptors, thus accounting for the antagonistic regulatory effects (Hua, et al., 2009), a situation similar to what is seen for PPAR $\alpha$  and LXR as described above. At other genes however, the RAR and ER binding sites do not directly overlap, suggesting that the antagonistic effects results from long-range interactions between the two receptors. These observations provide molecular mechanisms to explain the antagonistic effects of RA and estrogen on the physiological properties of breast cancer cells. This approach has then been extended to 24 NRs and a large number of other transcription factors, providing a more integrated model of the gene regulatory networks that govern the properties of breast cancer cells (Kittler, et al., 2013).

## **5.2. Cell-specific genomic occupancy by RAR-RXR.**

RA has pleiotropic effects in cells cultured *in vitro* and in tissues *in vivo*. For example, RA acts to induce cell cycle arrest in MCF7 cells or cell lines and blasts from patients with acute promyelocytic leukemia harboring a PML-RAR $\alpha$  translocation (details in chapter XX de Thé). ChIP-seq studies of the PML-RAR fusion protein in such cells indicate binding to a wide variety of DR, IR and ER elements, suggesting that the fusion protein has a more relaxed DNA binding specificity than native RAR $\alpha$  (Martens, et al., 2010). In contrast, RA signaling promotes hepatocyte proliferation and hepato-

carcinogenesis in livers of mice lacking the RAR-corepressor TRIM24 (TIF1 $\alpha$ ) (Khetchoumian, et al., 2007). In such animals, hepatocellular carcinoma (HCC) can be abrogated by inactivation of a single allele of the *Rara* gene, showing that de-regulated RA signaling has a pro-oncogenic effect in hepatocytes. In these cells, TRIM24 represses the activity of the RAR bound to DR2 and DR5 elements within the long terminal repeat of the VL30 class of endogenous retroviruses (Herquel, et al., 2013). Upon inactivation of TRIM24, the VL30 elements are expressed, followed by the reverse transcription of the VL30 derived RNA that then accumulates as single and double stranded cDNA within the cytoplasm of the hepatocytes, activating the interferon response genes to set up a chronic inflammatory state. In this way, RA signaling is converted into a chronic inflammatory signal that promotes HCC development.

Another example where cross-talk between RA and other signaling pathways promotes cell growth is found in mouse embryonic fibroblasts (MEFs). In MEFs, RA activates the transforming growth factor beta (TGF $\beta$ ) signaling pathway to promote cell proliferation (Fadloun, et al., 2007). To better understand this cross-talk, ChIP-chip was performed in MEFs stably expressing hemagglutinin (HA)-flag tagged RAR $\alpha$  and RAR $\gamma$  (Delacroix, et al., 2010). These experiments identified a set of sites occupied by both RARs in the regulatory regions

of critical effectors of the TGF $\beta$  signaling pathway such as the TGF $\beta$ 3 ligand and the connective tissue growth factor (CTGF), a mitogen that cooperates with TGF $\beta$  signaling to promote fibroblast proliferation. Thus, RARs directly activate critical components of the TGF $\beta$  signaling pathway to promote cross-talk between the two pathways.

Analysis of the DNA bound by the RARs in these experiments surprisingly revealed very few DR1, DR2 and DR5 elements with fully consensus half sites. Instead, most sites comprised one or more mismatches with respect to the consensus (Delacroix, et al., 2010). Electrophoretic mobility shift assays (EMSA) showed that many of these non-consensus elements could directly bind RAR-RXR heterodimers and revealed the importance of the sequences of the half-sites, the spacers as well as the flanking regions. For example, the region upstream of the *Inmt* gene that is regulated by RA in MEFs comprises a non-consensus DR5 element 5'-GGGTCA-5-CAGTCA in which the highly conserved G in position 2 of the second half site (underlined) is mutated. EMSA with various mutant forms of this element shows that high affinity binding requires the presence of the GGG in the first half site as replacement with AGG or AGT leads to strongly decreased binding. In non-consensus elements, therefore, the presence of mutations in one half site imposes strong constraints in the sequence of the other half site. On the other hand, in the case of DR1 elements, the nature of the

spacer nucleotide is critical as fully consensus elements with a G or A as spacer bind with much higher affinity than with a C spacer that has a strongly reduced affinity.

One explanation for the pleiotropic effects of RA is that the RARs bind different sets of genomic sites and therefore regulate distinct target genes in different cells and tissues. This hypothesis was first shown experimentally by the comparison of the RAR bound sites in MEFs and in mouse embryonic stem (ES) cells. ChIP-chip in undifferentiated ES cells identified 462 RAR-bound sites and comparison with the 354 sites bound in MEFs showed only 58 that were commonly occupied in both cells types (Delacroix, et al., 2010). Thus, cell-specific occupancy of RAREs results in the regulation of different sets of target genes in these two cell lines. Although differences in chromatin organization may play an important role in regulating accessibility of sites to the RARs, the full molecular mechanisms that determine cell-specific genome occupancy remain to be determined.

### **5.3 RAR-RXR binding elements show diverse spacing and topology**

As mentioned above, the RAR-occupied sites identified by ChIP-chip experiments comprised surprisingly few canonical DR1, 2 or 5 elements. This may be accounted for by the presence of DR1, DR2 or DR5 elements with non-canonical half site sequences or elements comprising half sites with non-canonical spacing and topology. Indeed, pre-genomic *in vitro*

studies had suggested that RAR-RXR may bind to RAREs with other DR-spacing (Mader, et al., 1993). To address this issue on a much larger data set, ChIP-seq was used to profile RAR occupancy in ES cells grown as embryoid body precursors engaged in neuronal differentiation, and in F9 embryonal carcinoma cells (Moutier, et al., 2012). ES cells grown as embryoid bodies (EBs) and treated for 4 days with all-*trans*-RA (hereafter abbreviated RA) can be differentiated into a highly homogeneous population of glutaminergic neurons (Bibel, et al., 2007), whereas RA-treated F9 cells differentiate into primitive endoderm-like cells (Linder, et al., 1981).

RAR ChIP-seq in EBs treated for 2 hours with RA identified 13385 RAR occupied loci (Moutier, et al., 2012). Analysis of the bound sequences showed that DR0 is the most represented spacing, followed by DR2, DR5, and DR8 together with IR0 and IR9. DR0, not the canonical DR5, is thus the most highly represented DR element at RAR-occupied loci in EBs and this holds true irrespective of whether the sites are strongly or more weakly occupied. In addition, the more strongly occupied sites showed a higher proportion of elements, irrespective of spacing, with fully consensus half sites than weaker occupied sites. Thus unlike, PPAR-RXR and LXR-RXR described above, RAR-RXR preferentially bind elements with consensus half sites. In F9 cells treated for 2 hours with RA, more than 32000 RAR-occupied loci could be identified. Analysis of the top 13385 peaks showed that around 70% were also occupied in EBs and

consequently that DR0 is the most frequently occupied motif followed by DR2, DR5, DR8, DR1, IR9 and IR0.

EMSA and isothermal titration calorimetry experiments with purified recombinant RAR-RXR confirmed that RAR-RXR binds directly to DR0, IR0 and DR8 with affinities comparable to that of the canonical DR5. More surprisingly, high affinity binding was also observed using a set of "pseudo DR0" elements identified from the ChIP-seq data where the second half site strongly diverges from the consensus sequence (Fig. 8.1).

The above study revealed that DR8 was a novel element frequently occupied by RAR in both EBs and F9 cells. This was unexpected as only little evidence existed for the presence of such elements at RA-regulated genes (Chang, et al., 2006) (Bouterfa, et al., 1995). These elements can be classified as simple DR8s where the two half sites are separated by 8 nucleotides, or composite DR8s that are formed from the juxtaposition of a DR2 and DR0 with the two outermost half sites forming the DR8 (Fig. 8.1). *In vitro*, RAR-RXR bind the DR2 and DR8 spacing of the composite elements with high affinity, while the DR0 shows a lower binding (Moutier, et al., 2012). In contrast, in EBs the DR2 and DR0 are preferentially occupied compared to the DR8 (Moutier, et al., 2012). The significance of this specific half-site organization is at present unknown, but one hypothesis is that an RAR-RXR heterodimer may bind to the DR2 or DR0, allowing an orphan receptor monomer to simultaneously bind the additional

5' or 3' half site. A similar situation exists for the IR9, most of which appear to be composite elements formed from a DR0 and an IR3 in which the second half site of the DR0 is the first half site of the IR3 and the two outermost sites form the IR9 (Fig. 8.1).

#### **5.4 Allosteric regulation of RAR-RXR function by the DNA binding element.**

The results of these various studies of RAR-RXR genomic occupancy indicate that the RAR-RXR heterodimer can bind *in vivo* to a wider variety of half site spacing than other NRs. As discussed above, VDR-RXR exhibits a strong specificity for DR3, TR-RXR for DR4 (Ramagopalan, et al., 2010), and PPAR-RXR for degenerate DR1 (Nielsen, et al., 2008). The increased flexibility of the RAR-RXR dimer may be explained by the lack of defined secondary structure of the hinge connecting the DNA-binding (DBD) and ligand-binding (LBD) domains of the RAR and RXR (Rochel, 2011). In contrast, the  $\alpha$ -helical structure of the corresponding hinge region in VDR and TR likely impose a constraint on half-site spacing (Orlov, et al., 2012, Rochel, 2011).

What are the biological consequences of this increased flexibility in DNA recognition? Functional analysis of these different elements indicates that DR1, DR2, DR5, DR8, IR0 and IR9 all act as independent RAREs capable of activating expression from a minimal promoter in the presence of RA (Moutier, et al., 2012). In contrast, DR0 elements do not

activate transcription in this assay. These observations strongly suggest that the nature of the DNA binding element exerts an allosteric control on the transcriptional activity of the heterodimer such that it is transcriptionally active when bound to elements such as DR5 or DR8, but not when bound to DR0. The use of such artificial reporters does not exclude the possibility that RAR-RXR bound to DR0 elements in their native context may function as transcription activators due to synergistic interactions with other transcription factors. They nevertheless highlight the different transcriptional properties of the RAR-RXR when bound to different response elements. These observations reinforce previous examples of allosteric control of RAR-RXR function (Kurokawa, et al., 1994) and are reminiscent of those showing that the DNA binding element can act as an allosteric regulator of glucocorticoid receptor (Meijsing, et al., 2009).

Crystal structures of several NR DBDs and LBDs in presence and absence of ligand have been obtained more than 15 years ago, however a major challenge has been to understand the overall conformation of the heterodimers when bound to DNA elements (see chapter XX Moras). This technical challenge has only been resolved in recent years both by small angle X-ray scattering (SAXS) and cryo-electron microscopy. The use of these techniques revealed the organization of the RAR-RXR bound to DR1, DR2 and DR5 and of VDR-RXR bound to DR3 (Orlov, et al., 2012, Rochel, 2011). In the case of RAR-RXR these

experiments confirm previous results showing that when bound to DR5 the RAR occupies the 3' half-site, but when bound to DR1, RAR occupies the 5' half-site. Despite this switch in DNA recognition, the heterodimer adopts a very similar extended asymmetric conformation in the two situations due to the flexibility of the hinge region. This raises the question of the conformation of the RAR-RXR heterodimer on DR0, IR0 and DR8 elements. Namely, which receptor binds the 5' and 3' half sites in each configuration, and how does the nature of the binding element affect overall receptor conformation to modulate co-regulator positioning and its transcriptional properties? Future biophysical and structural experiments will precisely define the conformation of the RAR-RXR heterodimer bound to these different elements, highlighting the conformational changes that occur and how they modulate transcriptional properties.

Another important and outstanding question concerns the physiological significance of this diversity of binding elements. Does it matter whether a given promoter or enhancer contains a DR5 or an IR0, are these and the other RAR-occupied elements interchangeable to some extent? In artificial reporter assays there are no obvious differences, with the exception of the transcriptionally silent DR0, but it is possible that in the native context the ability of the RAR-RXR heterodimer to interact correctly with other transcription factors is modulated by the nature of the binding element to

fine tune the RA response. Testing this idea is challenging as ultimately it would require replacement of a given element with another by homologous recombination within an animal's genome, to test the importance of such a change on the RA response *in vivo*.

## **6. Summary and conclusions**

Studies of NR binding to their response elements have evolved from the original "one by one" gene/promoter approach in the pre-genomic era to global genome-wide studies. Early studies on a limited number of target genes showed that many hormone response elements comprise combinations of half-site sequences with various spacing and topologies. These examples were used to derive rules for DNA recognition that allow discrimination of regulatory elements by different receptors.

In this short chapter we have discussed how these rules have stood up to the test of high throughput genomics. Many of the pre-genomic rules and findings have stood the test of going genome-wide. This is globally the case for the ER, VDR-RXR and TR-RXR, that preferentially occupy the predefined IR3, DR3 and DR4 elements respectively, although a subclass of VDR occupied sites do not contain DR3 elements and how these sites are recognized remains to be understood. In contrast, for other NRs like PPAR and LXR genome-wide studies revealed a preference for non-canonical sites, whereas RAR-RXR prefer elements with canonical half sites. Lastly, for the RAR-RXR

heterodimer a surprising diversity of elements is occupied that reveal its flexibility.

One outstanding question is the biological significance of the binding site diversity for RAR-RXR. Simple reporter assays revealed allosteric control of RAR activity by the DNA binding element raising the possibility that the different types of elements may not be fully interchangeable and that the nature of the elements(s) at a given RAR-RXR target gene may play an essential role in the fine-tuning of RA-signaling. Future studies should reveal whether this is indeed the case

## **Acknowledgments**

This work was supported by grants from the CNRS, the INSERM, the Université de Strasbourg, the Association pour la Recherche contre le Cancer, the Ligue Nationale contre le Cancer, the ANR (project RARE-scales) and the INCa. ID is an "équipe labellisée" of the Ligue Nationale contre le Cancer.

## **References**

S. Aoyagi and T. K. Archer. 2008. Dynamics of coactivator recruitment and chromatin modifications during nuclear receptor mediated transcription. *Molecular and Cellular Endocrinology* 280(1-2):1-5

R. Apfel, D. Benbrook, E. Lernhardt, M. A. Ortiz, G. Salbert and M. Pfahl. 1994. A novel orphan receptor specific for a subset of thyroid hormone-responsive elements and its interaction with the retinoid/thyroid hormone receptor subfamily. *Molecular and Cellular Biology* 14(10):7025-7035

J. E. Balmer and R. Blomhoff. 2005. A robust characterization of retinoic acid response elements based on a

comparison of sites in three species. *Journal of Steroid Biochemistry and Molecular Biology* 96(5):347-354

M. Bibel, J. Richter, E. Lacroix and Y. A. Barde. 2007. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols* 2(5):1034-1043

M. Boergesen, T. A. Pedersen, B. Gross, S. J. van Heeringen, D. Hagenbeek, C. Bindsboll, S. Caron, F. Lalloyer, K. R. Steffensen, H. I. Nebb, J. A. Gustafsson, H. G. Stunnenberg, B. Staels and S. Mandrup. 2012. Genome-wide profiling of liver X receptor, retinoid X receptor, and peroxisome proliferator-activated receptor alpha in mouse liver reveals extensive sharing of binding sites. *Molecular and Cellular Biology* 32(4):852-867

H. L. Bouterfa, F. J. Piedrafita, D. Doenecke and M. Pfahl. 1995. Regulation of H1(0) gene expression by nuclear receptors through an unusual response element: implications for regulation of cell proliferation. *DNA Cell Biol* 14(11):909-919

Y. A. Bulynko and B. W. O'Malley. 2010. Nuclear Receptor Coactivators: Structural and Functional Biochemistry. *Biochemistry*

C. Carlberg, S. Seuter and S. Heikkinen. 2012. The first genome-wide view of vitamin D receptor locations and their mechanistic implications. *Anticancer Research* 32(1):271-282

Y. S. Chang, J. Y. Cho, H. A. Cho, H. J. Kim, J. Chang, C. M. Ahn, S. K. Kim and S. K. Kim. 2006. 9-cis retinoic acid induces insulin-like growth factor binding protein-3 through DR-8 retinoic acid responsive elements. *Cancer Biol Ther* 5(6):586-592

F. Chatonnet, R. Guyot, G. Benoit and F. Flamant. 2013. Genome-wide analysis of thyroid hormone receptors shared and specific functions in neural cells. *Proceedings of the National Academy of Sciences of the United States of America* 110(8):E766-775

L. Delacroix, E. Moutier, G. Altobelli, S. Legras, O. Poch, M. A. Choukrallah, I. Bertin, B. Jost and I. Davidson. 2010. Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Molecular and Cellular Biology* 30(1):231-244

B. Desvergne. 1994. How do thyroid hormone receptors bind to structurally diverse response elements? *Molecular and Cellular Endocrinology* 100(1-2):125-131

A. Fadloun, D. Kobi, L. Delacroix, D. Dembele, I. Michel, A. Lardenois, J. Tisserand, R. Losson, G. Mengus and I. Davidson. 2007. Retinoic acid induces TGFbeta-dependent autocrine fibroblast growth. *Oncogene*

C. K. Glass and M. G. Rosenfeld. 2000. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes and Development* 14(2):121-141

C.K. Glass, J.M. Holloway, O.V. Devary and M.G. Rosenfeld. 1988. The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. *Cell* 54(3):313-323

S. Heikkinen, S. Vaisanen, P. Pehkonen, S. Seuter, V. Benes and C. Carlberg. 2011. Nuclear hormone 1alpha,25-dihydroxyvitamin D3 elicits a genome-wide shift in the locations of VDR chromatin occupancy. *Nucleic Acids Research* 39(21):9181-9193

B. Herquel, K. Ouarrarhni, I. Martianov, S. Le Gras, T. Ye, C. Keime, T. Lerouge, B. Jost, F. Cammas, R. Losson and I. Davidson. 2013. Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. *Nature Structural & Molecular Biology* 20(3):339-346

S. Hua, R. Kittler and K. P. White. 2009. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* 137(7):1259-1271

K. Khetchoumian, M. Teletin, J. Tisserand, M. Mark, B. Herquel, M. Ignat, J. Zucman-Rossi, F. Cammas, T. Lerouge, C. Thibault, D. Metzger, P. Chambon and R. Losson. 2007. Loss of Trim24 (Tiflalpha) gene function confers oncogenic activity to retinoic acid receptor alpha. *Nature Genetics* 39(12):1500-1506

S. Khorasanizadeh and F. Rastinejad. 2001. Nuclear-receptor interactions on DNA-response elements. *Trends in Biochemical Sciences* 26(6):384-390

R. Kittler, J. Zhou, S. Hua, L. Ma, Y. Liu, E. Pendleton, C. Cheng, M. Gerstein and K. P. White. 2013. A comprehensive nuclear receptor network for breast cancer cells. *Cell reports* 3(2):538-551

R. Kurokawa, J. DiRenzo, M. Boehm, J. Sugarman, B. Gloss, M. G. Rosenfeld, R. A. Heyman and C. K. Glass. 1994.

Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature* 371(6497):528-531

S. Linder, U. Krondahl, R. Sennerstam and N.R. Ringertz. 1981. Retinoic acid-induced differentiation of F9 embryonal carcinoma cells. *Experimental Cell Research*. 132(453-460

S. Mader, P. Leroy, J.Y. Chen and P. Chambon. 1993. Multiple parameters control the selectivity of nuclear receptors for their response elements. Selectivity and promiscuity in response element recognition by retinoic acid receptors and retinoid X receptors. *The Journal of Biological Chemistry*. 268(591-600

D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon and R. M. Evans. 1995. The nuclear receptor superfamily: the second decade. *Cell* 83(6):835-839

M. Mark, N. B. Ghyselinck and P. Chambon. 2006. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annual Review of Pharmacology and Toxicology* 46(451-480

M. Mark, N. B. Ghyselinck and P. Chambon. 2009. Function of retinoic acid receptors during embryonic development. *Nuclear Receptor Signalling* 7(e002)

J. H. Martens, A. B. Brinkman, F. Simmer, K. J. Francoijs, A. Nebbioso, F. Ferrara, L. Altucci and H. G. Stunnenberg. 2010. PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. *Cancer Cell* 17(2):173-185

S. H. Meijsing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto. 2009. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 324(5925):407-410

E. Moutier, T. Ye, M. A. Choukrallah, S. Urban, J. Osz, A. Chatagnon, L. Delacroix, D. Langer, N. Rochel, D. Moras, G. Benoit and I. Davidson. 2012. Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *The Journal of Biological Chemistry* 287(31):26328-26341

L. Nagy, H. Y. Kao, J. D. Love, C. Li, E. Banayo, J. T. Gooch, V. Krishna, K. Chatterjee, R. M. Evans and J. W. Schwabe. 1999. Mechanism of corepressor binding and release from nuclear hormone receptors. *Genes and Development* 13(24):3209-3216

R. Nielsen, T. A. Pedersen, D. Hagenbeek, P. Moulos, R. Siersbaek, E. Megens, S. Denissov, M. Borgesen, K. J. Francoijs, S. Mandrup and H. G. Stunnenberg. 2008. Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes and Development* 22(21):2953-2967

I. Orlov, N. Rochel, D. Moras and B. P. Klaholz. 2012. Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA. *The EMBO Journal* 31(2):291-300

J. Osz, Y. Brelivet, C. Peluso-Iltis, V. Cura, S. Eiler, M. Ruff, W. Bourguet, N. Rochel and D. Moras. 2012. Structural basis for a molecular allosteric control mechanism of cofactor binding to nuclear receptors. *Proceedings of the National Academy of Sciences of the United States of America* 109(10):E588-594

V. Perissi and M. G. Rosenfeld. 2005. Controlling nuclear receptors: the circular logic of cofactor cycles. *Nature Reviews Molecular and Cellular Biology* 6(7):542-554

S. V. Ramagopalan, A. Heger, A. J. Berlanga, N. J. Maugeri, M. R. Lincoln, A. Burrell, L. Handunnetthi, A. E. Handel, G. Disanto, S. M. Orton, C. T. Watson, J. M. Morahan, G. Giovannoni, C. P. Ponting, G. C. Ebers and J. C. Knight. 2010. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Research* 20(10):1352-1360

N. Ciesielski Rochel, F. Godet, J. Roessle, M. Peluso-Iltis, C. Svergun, DI. Moras, D. . 2011. Common architecture of nuclear receptor heterodimers on DNA direct repeats. *Nature Structural and Molecular Biology* 18(5) 564-571

C. Rochette-Egly and P. Germain. 2009. Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nuclear Receptor Signalling* 7(e005)

D. Schmidt, M. D. Wilson, B. Ballester, P. C. Schwalie, G. D. Brown, A. Marshall, C. Kutter, S. Watt, C. P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek and D. T. Odom. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328(5981):1036-1040

K. Umesono, K. K. Murakami, C. C. Thompson and R. M. Evans. 1991. Direct repeats as selective response elements for

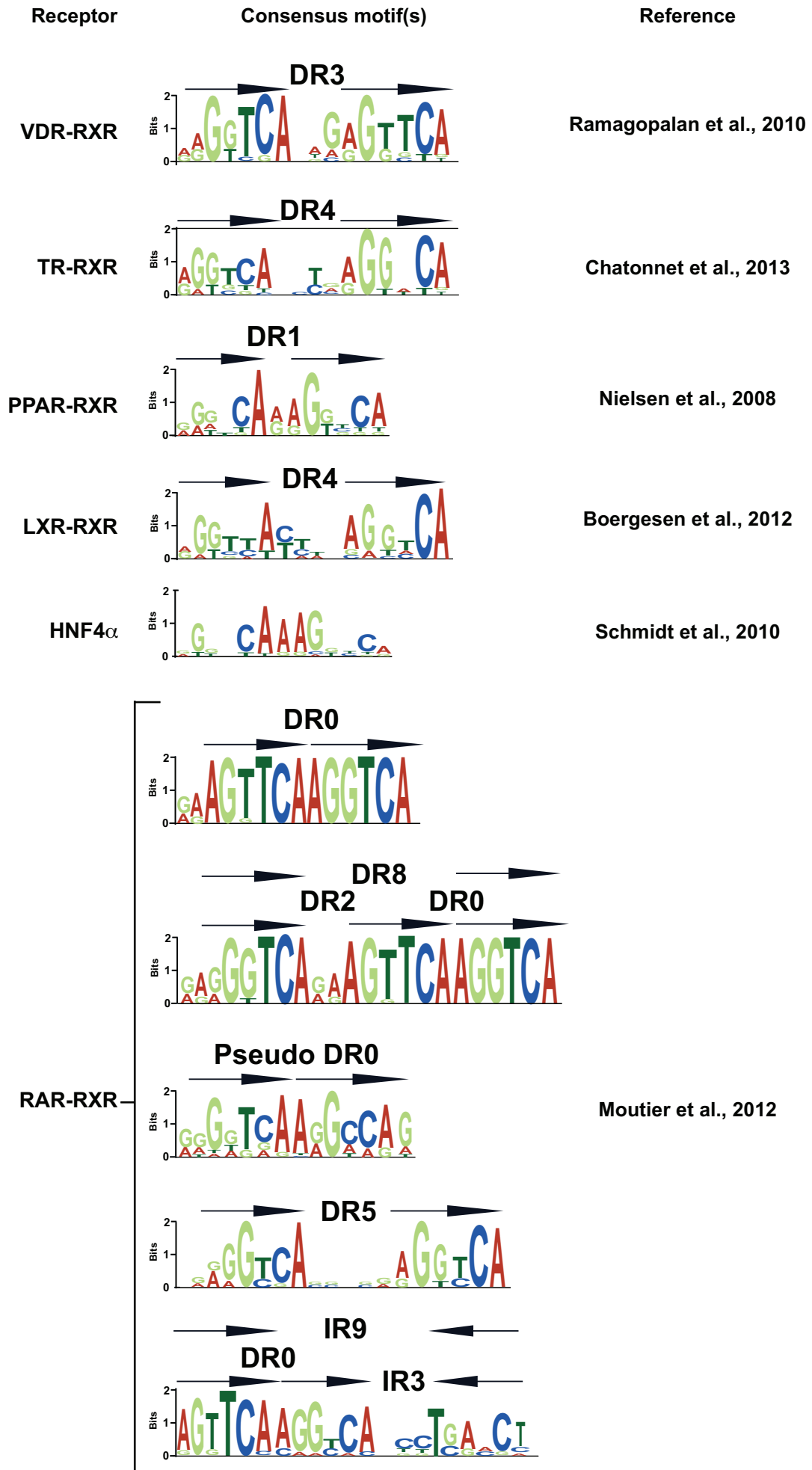
the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* 65(7):1255-1266

N. Vernet, C. Dennefeld, F. Guillou, P. Chambon, N. B. Ghyselinck and M. Mark. 2006. Prepubertal testis development relies on retinoic acid but not rexinoid receptors in Sertoli cells. *The EMBO Journal* 25(24):5816-5825

### **Legend to figure.**

**Figure 1.** Elements occupied by different nuclear receptors.

The figure shows the sequence logo of the most enriched elements occupied by the indicated nuclear receptors. The half-site consensus motifs (initially defined as 5'-RGKTCA-3', with R standing for a purinic [A or G] and K for a keto [G or T] nucleotide) are indicated by arrows. In the case of the elements occupied by RAR-RXR, the consensus sequences for the DR0, DR8 and IR9 elements were derived from the highest occupied sites of each class, while the DR5 sequence was derived from the total data set. As a result, the DR5 element appears to be more degenerate. This illustrates the observation that elements with consensus sequences tend to show higher occupancy than more degenerate elements. Each sequence logo was adapted from results reported in the indicated references.



***Publication annexe***

***n°2***

# **Reninoic Acid Receptors Recognize the Mouse Genome Through Binding Elements with diverse Spacing and Topology**

Emmanuel Moutier\*, Tao Ye\*, Mohamed-Amin Choukrallah<sup>1</sup>, Sylvia Urban, Judit Osz, Amandine Chatagnon<sup>2</sup>, Laurence Delacroix<sup>3</sup>, Diana Langer, Natacha Rochel, Dino Moras, Gerard Benoit<sup>2</sup>, and Irwin Davidson#.

Institut de Génétique et de Biologie Moléculaire et Cellulaire.

CNRS/INSERM/UDS.

1 Rue Laurent Fries,

67404 Illkirch Cédex.

France.

FAX: 33 3 88 65 32 01. TEL: 33 3 88 65 34 40 (45)

1. Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland.

2. Centre de Génétique et de Physiologie Moléculaire et Cellulaire, UMR5534 CNRS, Université de Lyon 1, 43 Boulevard du 11 novembre 1918 F-69622 Villeurbanne Cedex, France.

3. GIGA-Neurosciences, CHU, BAT B36 +1, 1 ave de l'Hôpital, 4000 Liège, Belgium.

\* These authors contributed equally to this study

# Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology<sup>\*[5]</sup>

Received for publication, March 13, 2012, and in revised form, May 15, 2012. Published, JBC Papers in Press, June 1, 2012, DOI 10.1074/jbc.M112.361790

Emmanuel Moutier<sup>‡1,2</sup>, Tao Ye<sup>‡1,2</sup>, Mohamed-Amin Choukallah<sup>§</sup>, Sylvia Urban<sup>‡</sup>, Judit Osz<sup>‡</sup>, Amandine Chatagnon<sup>¶</sup>, Laurence Delacroix<sup>||</sup>, Diana Langer<sup>‡</sup>, Natacha Rochel<sup>‡</sup>, Dino Moras<sup>‡</sup>, Gerard Benoit<sup>¶</sup>, and Irwin Davidson<sup>‡3</sup>

From the <sup>‡</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 1 Rue Laurent Fries, 67404 Illkirch Cédex, France, <sup>§</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland, <sup>¶</sup>Centre de Génétique et de Physiologie Moléculaire et Cellulaire, UMR5534 CNRS, Université de Lyon 1, 43 Boulevard du 11 Novembre 1918 F-69622 Villeurbanne Cedex, France, and <sup>||</sup>GLA-Neurosciences, CHU, BAT B36 + 1, 1 Ave de l'Hôpital, 4000 Liège, Belgium

**Background:** Retinoic acid receptors (RARs) heterodimerize with retinoid X receptors (RXRs) to regulate gene expression.

**Results:** This heterodimer recognizes the genome via a large and diverse repertoire of direct and inverted repeat DNA elements.

**Conclusion:** The observed diversity of binding elements changes the paradigm of how RAR-RXR recognizes the genome.

**Significance:** Half-site spacing in the DNA binding element allosterically regulates RAR function.

Retinoic acid receptors (RARs) heterodimerize with retinoid X receptors (RXRs) and bind to RA response elements (RAREs) in the regulatory regions of their target genes. Although previous studies on limited sets of RA-regulated genes have defined canonical RAREs as direct repeats of the consensus RGKTCA separated by 1, 2, or 5 nucleotides (DR1, DR2, DR5), we show that in mouse embryoid bodies or F9 embryonal carcinoma cells, RARs occupy a large repertoire of sites with DR0, DR8, and IR0 (inverted repeat 0) elements. Recombinant RAR-RXR binds these non-canonical spacings *in vitro* with comparable affinities to DR2 and DR5. Most DR8 elements comprise three half-sites with DR2 and DR0 spacings. This specific half-site organization constitutes a previously unrecognized but frequent signature of RAR binding elements. In functional assays, DR8 and IR0 elements act as independent RAREs, whereas DR0 does not. Our results reveal an unexpected diversity in the spacing and topology of binding elements for the RAR-RXR heterodimer. The differential ability of RAR-RXR bound to DR0 compared to DR2, DR5, and DR8 to mediate RA-dependent transcriptional activation indicates that half-site spacing allosterically regulates RAR function.

All-*trans*-retinoic acid (RA)<sup>4</sup>, the naturally active vitamin A metabolite, exerts a wide range of effects on vertebrate devel-

opment and adult tissue homeostasis by regulating cell proliferation, differentiation, and apoptosis (1–3). RA acts through binding to three members of the nuclear receptor superfamily, RAR $\alpha$ , RAR $\beta$ , and RAR $\gamma$ , that function as ligand-dependent transcriptional regulators. RARs form heterodimers with retinoid receptors (RXRs  $\alpha$ ,  $\beta$ , and  $\gamma$ ) and bind to RA response elements (RAREs) located in the regulatory regions of target genes (4, 5). RAREs are classically described as direct repeats of the consensus sequence 5'-RGKTCA-3' separated by 1, 2, or 5 nucleotides (for review, see Refs. 6–8).

We previously used chromatin immunoprecipitation (ChIP) coupled with array hybridization (ChIP-chip) to identify RAR occupied sites in mouse embryonic fibroblasts and in undifferentiated embryonic stem cells showing that RAR occupancy of target loci is cell type-specific (9). Analysis of these RAR-occupied sites revealed that a majority did not comprise canonical DR1, -2, or -5 elements. This paucity can in part be accounted for by the fact that RAR-RXR binds DR1, DR2, or DR5 elements with non-canonical half-sequences. An additional possibility that we did not address was the existence of DRs with non-canonical spacings. Evidence for the use of alternate spacing has previously been described (10), but their prevalence at RAR-occupied loci has not been described. We have analyzed ChIP-seq data sets from RAR occupancy in embryonic stem cells grown as embryoid body precursors to neuronal differentiation and from F9 embryonal carcinoma cells. We identify DR0 as the most abundant RAR binding element, and we identify a novel composite DR8 as a common but previously unrecognized RARE. Our results reveal an unexpected diversity in the spacing and topology of the DNA binding elements that is unique for the RAR-RXR heterodimer.

<sup>\*</sup> This work was supported by grants from the CNRS, the INSERM, the Université de Strasbourg, the Association pour la Recherche contre le Cancer, the Ligue Nationale contre le Cancer, Institut National du Cancer Grant 2008-037, the EuTRACC program of the European union, and the French Agence Nationale de la Recherche project Rapssodi.

<sup>[5]</sup> This article contains supplemental Experimental Procedures, Figs. 1–7, and Tables 1 and 2.

The data in this paper have been deposited in the GEO Database (<http://www.ncbi.nlm.nih.gov/geo/>) with access number GSE35599.

<sup>1</sup> Both authors contributed equally to this study.

<sup>2</sup> Supported by a fellowship from the Association pour la Recherche contre le Cancer.

<sup>3</sup> An "équipe labellisée" of the Ligue Nationale contre le Cancer. To whom correspondence should be addressed. Tel.: 33-3-88-65-34-40 (ext. 45); Fax: 33-3-88-65-32-01; E-mail: irwin@titus.u-strasbg.fr.

<sup>4</sup> The abbreviations used are: RA, retinoic acid; RAR, RA receptor; RXR, retinoid X receptor; RARE, RA response element; ITC, isothermal titration calorimetry; CAT, chloramphenicol acetyltransferase; ER, estrogen receptor; VDR, vitamin D3 receptor; GDNF, germ cell nuclear factor; EB, embryoid body; DR, direct repeat; IR, inverted repeat.

## EXPERIMENTAL PROCEDURES

**ChIP and ChIP-seq**—ChIP and ChIP-seq experiments were performed on chromatin from murine embryonic stem cells grown as embryoid bodies for 4 days in the absence of RA and then treated for 2 h with all-*trans*-retinoic acid as previously described (9, 11, 12). Further details are provided in the [supplemental Experimental Procedures](#). ChIP-seq was performed using the panRAR antibody (Sc-773, Santa Cruz Biotechnology) and sequenced on an Illumina GAIIx sequencer, and the raw data were analyzed by the Illumina Eland pipeline V1.7 or V1.8 and aligned to the genome with Bowtie. Peak detection was performed using the MACS software (13) under settings where an anti-GFP ChIP from the embryoid bodies was used as a negative control. Peaks were then annotated using GPAT (14) with a window of  $\pm 20$  kb with respect to the coordinates of the beginning and end of Ensembl genes (release 64). Cluster comparison of ChIP-seq data sets was performed with seqMINER (15). F9 embryonal carcinoma cells were grown under standard conditions and treated with  $10^{-6}$  M RA for 2 h before ChIP-seq with the panRAR antibody.<sup>5</sup> ChIP-seq analysis and annotation was performed as described above.

**Electrophoretic Mobility Shift Assay**—EMSA assays were performed essentially as previously described (9) using purified bacterial recombinant RAR $\alpha$  $\Delta$ AB-RXR $\alpha$  $\Delta$ AB as described (16). The RAR binding elements in each oligonucleotide were centered and surrounded by their native flanking sequences. The sequences of the oligonucleotides are available on request. After electrophoresis, the gels were dried and exposed to autoradiographic film or a PhosphorImager plate.

**Isothermal Titration Calorimetry (ITC)**—ITC measurements were performed at 25 °C on a MicroCal ITC<sub>200</sub> (MicroCal). Double-stranded DNA and purified proteins bound to 9-*cis*-RA were dialyzed extensively against the same buffer used in the ITC experiments. The buffer contained 50 mM Hepes, pH 8.0, 100 mM sodium chloride, 2% glycerol, and 1 mM Tris(2-carboxyethyl)phosphine. In a typical experiment 2- $\mu$ l aliquots of DNA at 80–150  $\mu$ M were injected into a 10  $\mu$ M RAR-RXR solution (200- $\mu$ l sample cell). The delay between injections was 120 to 180 s to permit the signal to return to base line before the next injection. ITC titration curves were analyzed using the software Origin 7.0 (OriginLab). Standard free energies of binding and entropic contributions were obtained, respectively, as  $\Delta G = -RT \ln(K_a)$  and  $T\Delta S = \Delta H - \Delta G$  from the  $K_a$  and  $\Delta H$  values derived from ITC curve fitting.

**Bioinformatics Analysis**—The 150 nucleotides surrounding the ChIP-seq peaks were analyzed using a custom JAVA API application to detect perfect consensus 5'-RGKTCA-3' half-sites with the different spacings. Analysis of the 150-bp regions from ChIP-seq peaks where no canonical DR elements were found using the MEME program (17) identified the pseudo-DR0 consensus. MEME analysis was also used to derive the consensus DR0 and DR8 sequences by analysis of 200 ChIP-seq peaks of each class. FIMO was used to search for the FOXA1 consensus motif (JASPAR) and was run in default parameters

with a *p* value cutoff of  $1e-4$ . For SPAMO (spaced motif analysis tool) analysis we used the motifs identified in the MEME analysis as the principle and spaced motifs. DRs and their constituent half-sites were mapped in a 150-bp window using the peak summit from MACS analysis as the central position. An in-house JAVA application was used to align all DRs on the same strand to ensure the sense and antisense matches gave homogeneous positions. The motif position profiles are based on a .wig file-like representation with a step size of 1 bp to calculate the number of overlapping motifs present at each position in the 150-bp window.

**Cell Culture and Transfection Assays**—JI embryonic stem cells (129SV/Jae) were grown on inactivated fibroblast feeder cells in the presence of LIF under standard conditions. Cells were passaged three times in absence of feeders before transfer to bacterial dishes for embryoid body formation. For reporter assays, transfections comprised 1  $\mu$ g of the TATA-chloramphenicol acetyl transferase (CAT) reporters, 1  $\mu$ g of pCH110-expressing  $\beta$ -galactosidase as internal standard, and 1  $\mu$ g of pCMV hRAR $\alpha$  and RXR $\alpha$  expression vectors. Transfections were performed with FuGENE (Roche Applied Science), and 24 h after transfection  $10^{-6}$  M RA was added for an additional 24 h. Extract preparation,  $\beta$ -galactosidase assays, and CAT assays were performed using a Roche Applied Science CAT-ELISA kit as previously described (18). To generate the TATA-CAT reporter plasmids, the various wild-type and repeat elements were generated by DNA synthesis (GeneArt) with flanking BglII and NotI restriction sites. The plasmids containing these regions were amplified, and the inserts were purified and recloned between the equivalent sites in the previously described TATA-CAT plasmid (11, 18).

**mRNA-seq**—Total RNA was extracted from duplicate cultures of embryoid bodies grown for 4 days in the absence of RA and after 24 h of RA treatment. The mRNA-seq libraries were prepared following the Illumina protocol ([supplemental Experimental Procedures](#)). Sequence reads mapped to reference genome mm9/NCBI37 using Tophat (19). Quantification of gene expression was done using Cufflinks (20) and annotations from Ensembl release 62. For each transcript the resulting reads per kilobase of exon model per million mapped reads (RPKM) were converted into raw read counts, and these counts were added for each gene locus. Data normalization was performed as described by Anders and Huber (21) and implemented in the DESeq Bioconductor package. Only regulated transcripts with an RPKM of  $>2$ , an adjusted *p* value of  $<0.1$ , and a log 2-fold change of  $>1$  and  $<-1$  were considered.

## RESULTS

**Representation of DR Spacings at RAR-occupied Loci**—We performed RAR ChIP-seq in mouse embryonic stem cells grown for 4 days as embryoid body (EBs) precursor to neuronal differentiation (22) treated for 2 h with RA. Analysis of this data set revealed 13,385 RAR occupied loci that could be annotated to 12,250 Ensembl and predicted transcripts (equivalent of 6,628 RefSeq genes) ([supplemental Table 1](#)). As seen for other nuclear receptors (23, 24), more than half of the RAR-occupied sites were located greater than 20 kb upstream or downstream of the transcription start site (TSS) with only 10% residing in the

<sup>5</sup> A. Chatagnon, P. Veber, J. Bedo, V. Laudet, A. Kowalczyk, F. d'Alche-Buc, and G. Benoit, manuscript in preparation.

## Characterization of Novel RAREs

proximal promoter region (−5 to +2 kb with respect to the TSS; supplemental Fig. 1).

We analyzed the frequency of all potential DR spacings from DR0 to DR10 (5′-RGKTCA<sub>0–10</sub>RGKTCA-3′) in the 150 bp surrounding the peak summit of the 1000 highest occupied sites based on the number of sequence tags forming the peak. Similarly, as it has been reported that RAR occupied elements are closely associated with estrogen receptor (ER) binding elements (IR3) in MCF7 cells (25), we analyzed the RAR occupied loci for IR0–IR10. This analysis showed that DR0 was by far the most frequently represented spacing followed by DR2, DR8, DR5, and DR1 (Fig. 1A). Similar results were obtained when analyzing 1000 medium and 1000 low occupied sites. In each case the DR0 and DR2 spacings remained the most frequent followed by DR5 and DR8. However, the total number of loci comprising canonical half-sites with these spacings decreases with occupancy. The higher occupied sites comprise a large proportion of elements with canonical half-site repeats, whereas the lower occupied must contain either more elements with degenerate half-sites and/or other spacings (compare the total number of DRs in each class in Fig. 1A).

Analysis of the IR frequencies in each class showed that IR0 is the most frequently found in the highly occupied class, whereas IR9 is also represented in the medium-occupied class (Fig. 1A). Again, the frequency of IR elements is higher in the most occupied class than in the lower occupied classes.

The analysis of the total 13385 sites confirmed that DR0 is the most represented spacing followed by DR2, DR5, and DR8 together with IR0 and IR9 (Fig. 1B). Extending the analysis to DR/IR11–DR/IR20 showed that none of these larger spacings was strongly enriched in the data set (data not shown). DR0, not the canonical DR5, is thus the most highly represented DR element at RAR-occupied loci in EBs irrespective of the degree of occupancy in the cells.

We further analyzed the frequency of the DR and IR elements in the RAR-occupied sites in the proximal promoter region (−5 to +2 kb) to ask if any of these elements are selectively enriched. DR0 remains the most represented element, but the ratios of DR0/DR5 and DR5/DR8 indicate a modest relative enrichment of DR5 at promoters compared with the highest occupied class but not when compared with the total data set (Fig. 1C).

It has been suggested that RAR-occupied motifs in MCF7 cells are closely associated with ER binding sites (IR3) and with sites for the pioneer factor FOXA1 (25, 26). No significant representation of IR3 was seen at RAR-occupied sites in EBs (and F9 cells; Fig. 1, A and D). Analysis of the 150 bp around the top 1000 RAR-occupied peaks in EBs revealed only 47 FOXA1 consensus sites, whereas the same analysis in a 500-bp window revealed 259 potential sites (data not shown). Thus, although no close association of RAR and ER binding elements are seen in EBs, around 25% of RAR-occupied loci show association with FOXA1 when a wider window is used.

To ask if the above DR frequencies are representative of what can be seen in other cell types, we analyzed the RAR ChIP-seq data set obtained from RA-treated F9 embryonal carcinoma cells.<sup>5</sup> Although this data set comprises around 32,000 RAR-occupied loci when analyzed by MACS using the same parameters as for the EB data set, we restricted our analysis to the top

13,385 peaks. Comparison with the EB data set indicated a large fraction of shared sites and potential target genes and a set of sites specific for each cell type (supplemental Fig. 2 and data not shown). Analysis of the F9 data set indicated that DR0 is again the most frequent at the high, medium, or low occupied peaks followed by DR2, DR5, DR8, DR1, and IR0 (Fig. 1D). Overall, the relative DR and IR frequencies are, therefore, similar but not identical to those in the EB data set (Fig. 1B and E) in accordance with the existence of a set of sites specific to F9 cells.

To determine if this spectrum of DR frequencies is found at sites occupied by other nuclear receptors, we analyzed DR and IR usage at sites occupied by the vitamin D3 receptor (VDR) in the data set of Ramagopalan *et al.* (27). In striking contrast to what is observed for RAR, for VDR the DR3 element is strongly represented followed by the DR4, but there is no high frequency of DR0 or IR0 elements at these sites (Fig. 1F). Our analysis is in accord with that reported by the authors of this study who defined DR3 as strongly enriched at these sites. Thus, the RAR-RXR heterodimer binds to DR elements with a much larger variety of spacings than the VDR-RXR heterodimer whose specificity is more restricted.

*In Vitro Binding of RAR-RXR to Non-canonical-spaced DR Motifs*—To determine if DRs with non-canonical spacings can directly bind RAR-RXR, we performed EMSA analysis with oligonucleotide probes derived from sites occupied by RAR and purified bacterially produced recombinant RAR $\alpha$ ΔAB-RXR $\alpha$ ΔAB (Ref. 16; hereafter designated simply RAR-RXR).

We first investigated the ability of DR0 elements with fully canonical half-sites, including that present in the suppressor of cytokine signaling 3 (*Socs3*) gene locus (Fig. 2A) or half-sites with a single base mismatch to compete for RAR-RXR bound to the DR5 from the *Rarb* gene. Several consensus DR0 elements were able to compete in the formation of the RAR-RXR-DR5 complex (see lanes 6–8 in Fig. 2B), whereas DR0s with one or several mismatches were poor competitors or did not compete (lanes 2–4). On the other hand, a DR0 from the AE-binding protein 2 (*Aebp2*) locus with a single mismatch was a good competitor. Thus, as we previously reported (9) in some elements, mismatches are tolerated, whereas at others they are detrimental to *in vitro* RAR-RXR binding.

Competition experiments with increasing quantities of competitor oligonucleotide indicated that the *Rarb* DR5 motif efficiently competes with between 25- and 50-fold of cold competitor (Fig. 2C, lanes 4 and 5). In comparison, the DR0 motif from the *Socs3* gene is less efficient (lanes 7–10), whereas that from the Musashi homolog 2 (*Msi2*) gene is comparable with DR5 (lanes 11–14). Binding of recombinant RAR-RXR to the *Rarb* DR5 and the *Socs3* and *Msi2* DR0 was also assessed by ITC. A dissociation constant ( $K_d$ ) of 73 nM for the *Rarb* DR5 and 79 nM for a canonical DR2 from the *Hoxa10* gene was determined compared with 76 nM for the *Msi2* DR0 and 110 nM for the *Socs3* DR0 (Fig. 3, A and B). Thus, consensus DR0s are *bona fide* RAR-RXR binding elements with *in vitro* affinities comparable with the canonical DR5. Close analysis of thermodynamic parameters observed for the binding of RAR-RXR to DNA suggests that enthalpy solely drives this protein-DNA interaction and that the increase in the binding affinity to DNA correlates with overall favorable enthalpy.

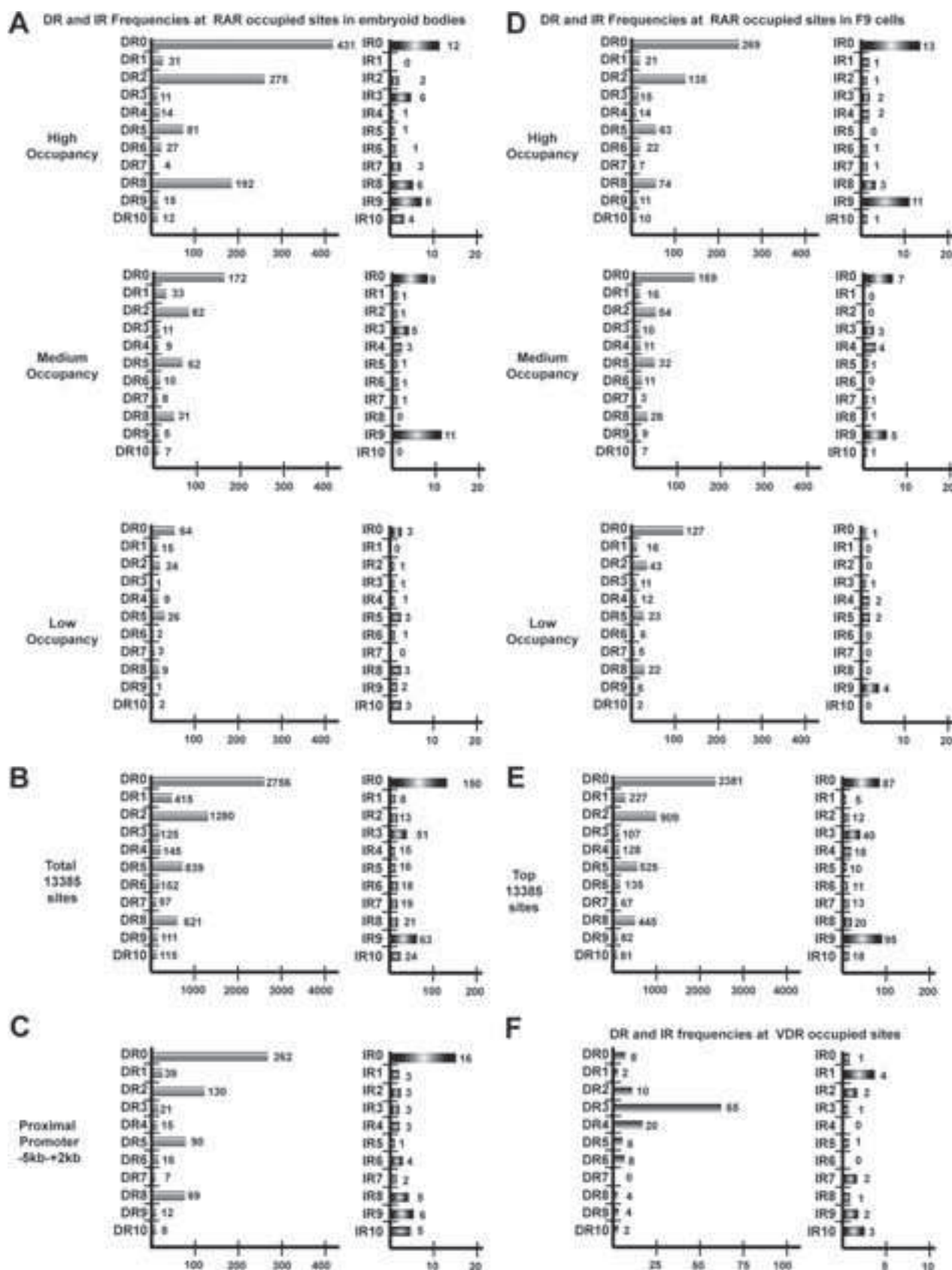
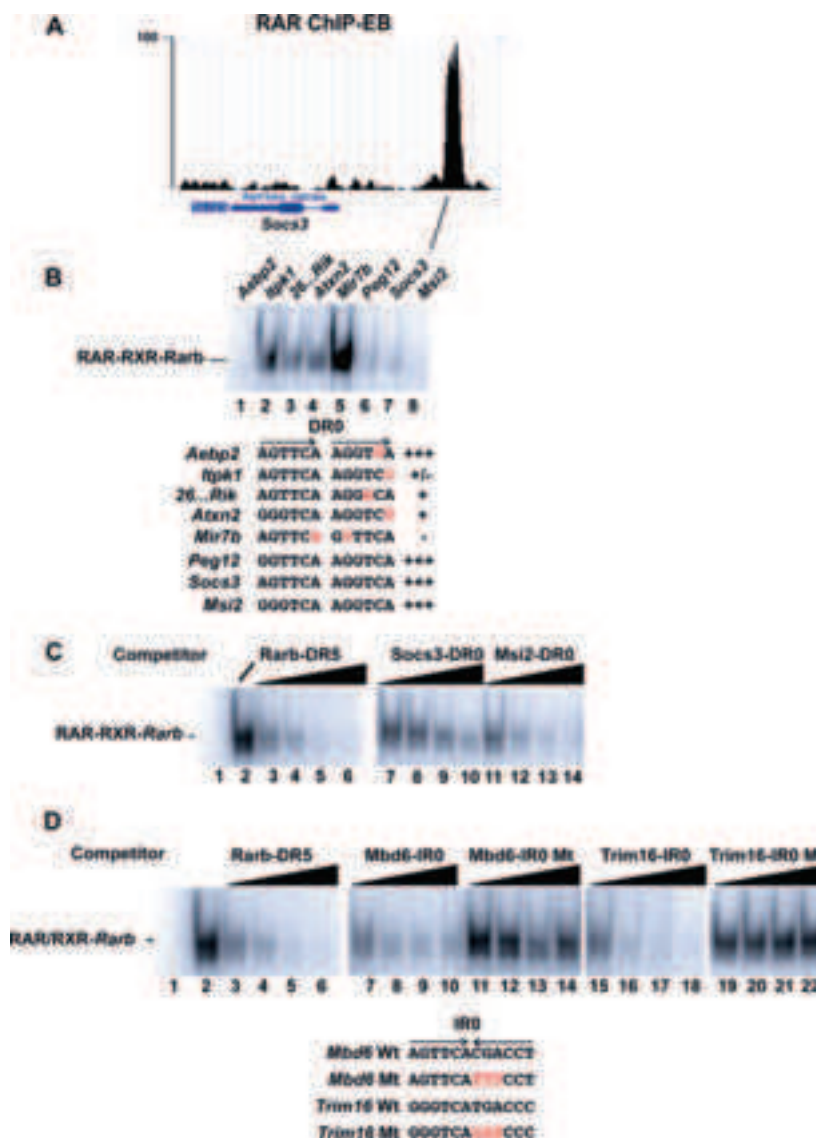


FIGURE 1. DR and IR frequencies at RAR-occupied sites in embryoid bodies and F9 cells. *A*, frequencies of the indicated half-site spacings at the high, middle, and low occupied sites in the embryoid body RAR ChIP-seq data set are shown. *B*, DR/IR frequencies in the total data set of 13385 embryoid body sites are shown. *C*, DR/IR frequencies at promoter proximal sites at the high, middle, and low occupied sites in the RAR F9 cell RAR ChIP-seq data set are shown. *D*, DR/IR frequencies at the top 13385 F9 sites are shown. *E*, DR/IR frequencies at VDR-occupied sites from the data set of (27) are shown.

## Characterization of Novel RAREs



**FIGURE 2. RAR-RXR binding to DR0 elements.** *A*, shown is a University of California at Santa Cruz web browser view of sequence tag density in .wig file format of the RAR-occupied site at the *Socs3* gene comprising a DR0 in EBs. *B*, EMSA analysis shows the ability of the indicated DR0 elements to compete with the labeled *Rarb* DR5 element for RAR-RXR complex formation. The sequences of the DR0 elements within the competing oligonucleotides are shown with the repeated half-sites indicated by arrows. Variations from the consensus half-site sequence are indicated in red. All competitors were used a 100-fold excess. *C*, competition was performed with increasing quantities (10-, 25-, 50-, and 100-fold excess) of the oligonucleotides shown above each lane. Lane 1 is the oligonucleotide probe with no recombinant RAR-RXR, and lane 2 is the oligonucleotide probe with RAR-RXR but no competitor. *D*, EMSA competition analysis of the indicated IR0 elements is shown. The sequences of the IR0 motifs within the competing oligonucleotides are shown with the inverted half-sites indicated by arrows. Mutated nucleotides are indicated in red. Competition was performed with increasing quantities as above.

ChIP-seq identified loci with IR0 elements that efficiently competed for complex formation (supplemental Fig. 3, *A* and *B*, lanes 4, 6, and 8), whereas mutated versions conserving only one half-site showed no competition (lanes 5, 7, and 9). Titration experiments showed that the IR0s from the methyl-CpG binding domain protein 6 (*Mbd6*) and tripartite motif containing 16 (*Trim16*) genes competed with efficiencies comparable with that of the DR5 element (Fig. 2*D*). This was confirmed by ITC showing a  $K_d$  of 58 nM for the *Trim16* IR0 element and 97 nM for the consensus IR0 from the *Vat1* gene locus (Fig. 3, *A* and *B*).

**A Novel Composite DR8 RAR Binding Element**—DR8 is comparable in frequency to DR5 in the EB and F9 data sets. These elements fall into two classes comprising either a simple DR8 with no other potential half-sites in the spacer sequence or

composite half-sites formed by juxtaposition of three half-sites with DR2 and DR0 spacings (Fig. 4*A*). Of the 431 DR0 elements in the top 1000 EB sites (Fig. 1*A*), 183 form DR8 with an additional 5'-RGKTCA-3' half-site. MEME analysis of the remaining DR0 motifs revealed the presence of a variant (5'-RGATCA-3') half-site at a further 61 loci (Fig. 4*A*). This unique and specific topological arrangement of half-sites therefore constitutes a highly represented element at RAR-occupied loci and leaves only 187 DR0s with no recognizable 5' half-site with a DR2 spacing. Similarly in F9 cells, 63 of the 74 DR8s in the highly occupied class have an additional DR2-spaced half-site (supplemental Fig. 4*A*).

We next asked whether RAR bound the DR2, DR0, or DR8 spacings of these elements. We chose two composite DR8 elements identified by ChIP-seq at the v-maf musculoaponeurotic

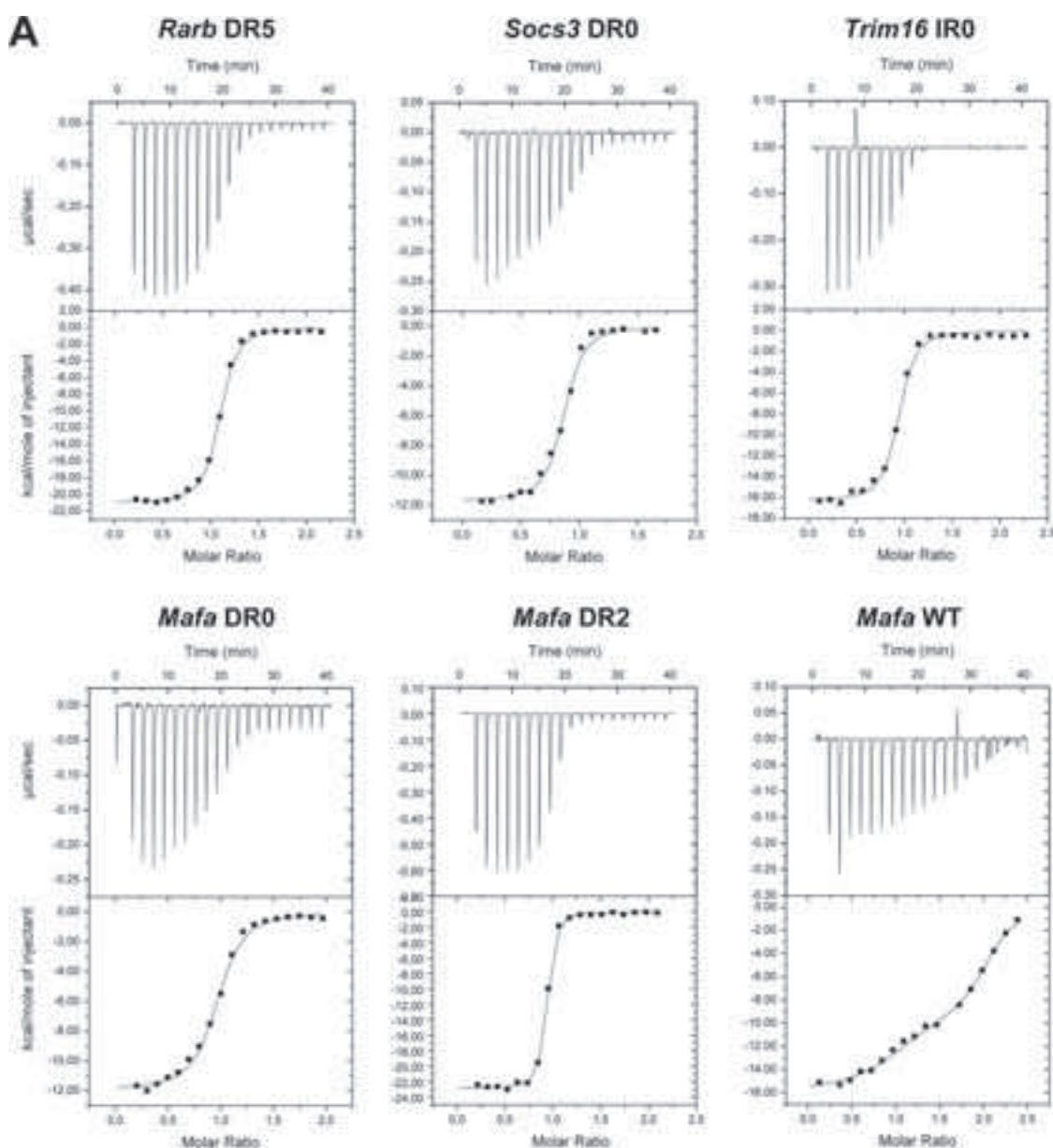
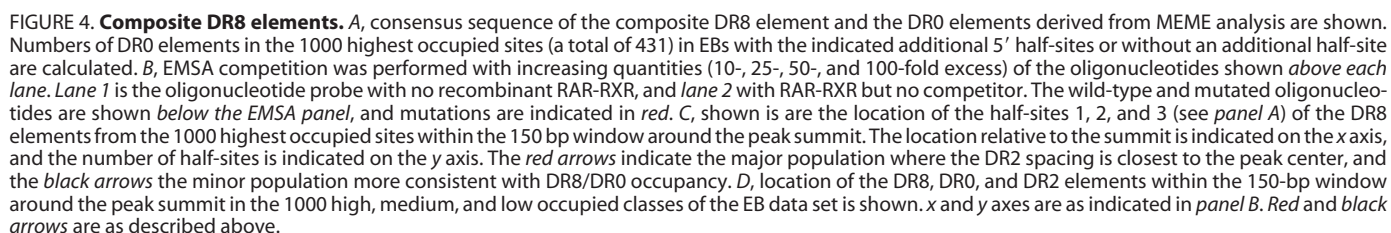


FIGURE 3. **Quantification of the interaction between RAR-RXR and different RAR binding element oligonucleotides by ITC.** A, representative ITC titrations of RAR/RXR to the indicated DR/IR elements are shown. RAR-RXR binds DR5, DR2, DR0, and IR0 with a stoichiometry of 1 DNA per heterodimer. B, thermodynamic parameters (dissociation constant ( $K_d$ ), enthalpy ( $\Delta H$ ), and entropy ( $-T\Delta S$ )) of DNA binding to RAR-RXR was determined by ITC.



fibrosarcoma oncogene family, protein A (*Mafa*) and CD97 antigen (*Cd97*) gene loci (supplemental Fig. 4B). The wild-type *Mafa* and *Cd97* elements both efficiently competed complex formation in EMSA assays (supplemental Fig. 4C, lanes 4 and 9 and Fig. 4B, lanes 7–10). Oligonucleotides mutated in the central half-site leaving the DR8 spacing also competed (supplemental Fig. 4C, lanes 5 and 12 and Fig. 4B, lanes 19–22). Mutation of the 5' half-site leaving only the DR0 resulted in much less efficient competition (supplemental Fig. 4C, lanes 6 and 10 and Fig. 4B, lanes 11–14). In contrast, mutation of the 3' site leaving the DR2 led to efficient competition (supplemental Fig. 4C, lanes 7 and 11 and Fig. 4B, lanes 15–18). Mutation of the 5' and 3' sites leaving only a single half-site essentially abolished competition (supplemental Fig. 4C, lanes 8 and 13 and Fig. 4B, lanes 23–26). Further evidence that DR8 can bind RAR-RXR comes from the *Dedd* gene locus whose simple DR8 is an efficient competitor in EMSA assays (supplemental Fig. 5A, compare lanes 1–4 with 9–12 and the mutated element in lanes 5–8). It is noteworthy that at least *in vitro*, DR8 elements with the variant 5'-RGATCA-3' half-site are much less efficient competitors than those with the 5'-RGKTCA-3' half-sites (supplemental Fig. 5B).

The binding of the RAR-RXR to the *Mafa* element was also analyzed by ITC, showing that the *Mafa* DR2 had a high affinity with a  $K_d$  of 25 nM, whereas this DR0 had a lower affinity than the other DR0s with a  $K_d$  of 180 nM (Fig. 3, A and B). ITC performed with the WT *Mafa* element containing all three half-sites showed bimodal binding. The first phase indicated occupancy of a high affinity site with a  $K_d$  of 13 nM followed by a second phase with much lower affinity of 180 nM. These results are consistent with binding of high affinity DR2 or DR8 and lower affinity to the DR0.

We also labeled oligonucleotides comprising the WT and mutated *Mafa* DR8, the *Msi2* DR0, and the *Dedd* DR8 and tested their ability to form a complex with recombinant RAR-RXR. The WT *Mafa* element formed a complex with mobility identical to *Rarb* DR5 or *Hoxa10* DR2 elements (supplemental Fig. 5C, lanes 4 and 5, 1 and 2, and 14 and 15). Complex formation was also seen with the DR8 and DR2 combinations, whereas binding to the DR0 was less efficient (supplemental Fig. 5C, lanes 8 and 9). The *Msi2* DR0 and the *Dedd* DR8 also form complexes with RAR-RXR (supplemental Fig. 5C, 12 and 13 and 16 and 17). These results confirm those of the EMSA competition and ITC showing that DR0 and DR8 form complexes with RAR-RXR. Furthermore, a single RAR-RXR heterodimer appears to bind to the WT *Mafa* DR8, as no slower migrating complex corresponding to an additional RAR and/or RXR bound to the third half-site was observed.

We analyzed the position of the constitutive half-sites in the 183 DR8s from the top 1000 sites of EB data set with respect to the CHIP-seq peak summit. For each half-site two populations are seen. The first is consistent with occupancy of the DR2 motif localizing the two half-sites of the DR2 close to the summit and the 3' half-site of the DR0 downstream (Fig. 4C). The second population is more consistent with occupancy of the DR0 and localizes the 5' half-site of the DR2 to a more upstream position and the 3' half-site of the DR0 toward the peak center. As a consequence of the fact that the first of these two popula-

tions is more abundant, DR8 elements (in this case all of the DR8s from the data set) show a skewed localization biased toward the 3' side of the peak center (Fig. 4D). Consistent with this, analysis of the total DR2 population shows a bias toward the 5' side as a subpopulation of DR2s are present in DR8 elements where the DR0 is occupied but not the DR2 (Fig. 4D). In addition, analysis of the total DR0 population shows elements localized 3' to the center corresponding to DR8 motifs where the DR2 is occupied and a second population located around the summit corresponding to occupancy of the DR0.

This analysis is consistent with the idea that the DR2 element of the composite DR8 is preferentially occupied with a subpopulation showing occupancy of the DR0 spacing. Our analysis does not reveal whether this is a stochastic effect on the general DR8 population or whether differences in half-site sequence and context favor occupancy of a specific spacing at different subpopulations of sites.

**DR8 Elements Act as RA-responsive Elements**—We next asked if the DR0, IR0, and DR8 elements could act as RAREs. Three copies of the wild-type or mutated elements were inserted upstream of a TATA element in a CAT reporter vector (Fig. 5A). These vectors were transfected into COS1 cells along with vectors expressing RAR and RXR and CAT activity measured in the presence and absence of RA.

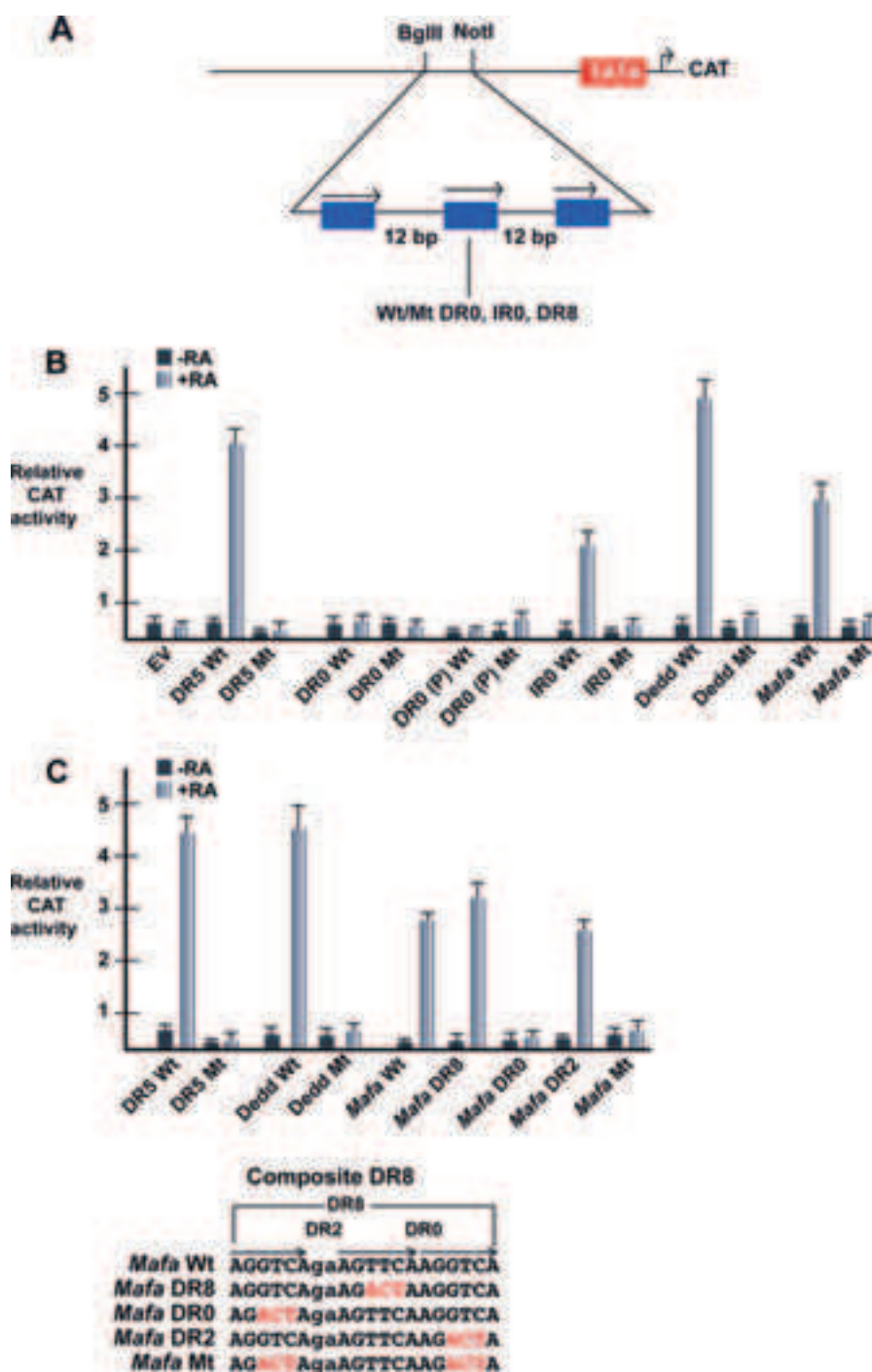
A positive control vector with three copies of the Wt *Rarb* DR5 element strongly responded to RA, whereas mutation of one half-site in each copy abolished the response, and no effect of RA was seen using the empty vector (Fig. 5B). Vectors comprising IR0 and simple DR8 all showed a robust response to RA that was lost when one the half-sites was mutated (Fig. 5B). In contrast, no significant response was seen with vectors carrying DR0 elements. These results show that IR0 and the DR8 elements act as independent RAREs, but the DR0 does not have this activity.

To dissect the activity of the composite DR8 element, we transfected vectors in which mutations leave the DR8, DR2, or DR0 elements intact. As in the first set of experiments, both the simple and composite DR8s show a robust RA response (Fig. 5, B and C). Mutation leaving only the DR8 shows a strong RA response (Fig. 5C) in agreement with the observation that a simple DR8 acts as a RARE. The DR2 is also RA responsive, whereas when only the DR0 or a single half-site are present, no RA response is seen. These results indicate that in the composite DR8, the DR2 and DR8, but not DR0, combinations are RA-responsive.

**Characterization of a Degenerate Pseudo-DR0 Element**—The combination of all the above half-site spacings accounts for only a subset of the RAR-occupied loci. For example, in the EB data set, DR0, DR1, DR2, DR5, DR8, and IR0 with canonical half-site sequences are present at 6,061 of the 13,385 sites, whereas in the F9 data set, they account for 4,574 of the 13,385 selected sites. It should be noted, however, that in EBs there are 1010 elements in the 1000 highest occupied sites, and almost all loci have at least one element with the above spacings.

To determine if there are other sequences that may be recognized by the RAR-RXR, we selected a series of sites with no consensus DR element and performed *de novo* motif detection using the MEME program. One of the motifs generated by this

## Characterization of Novel RAREs



**FIGURE 5. DR8, IR0, but not DR0 act as RAREs.** *A*, shown is a schematic representation of the reporter vector where three copies of the indicated motifs spaced by 12 nucleotides are inserted upstream of a TATA element and the CAT reporter gene. *B*, shown is relative CAT activity of the different reporter vectors 48 h after transfection. EV is empty vector, and the elements in the other vectors are shown below each lane. The DR5 is from the *Rarb* gene, DR0 is from the *Socs3* gene, the IR0 is from the *Trim16* gene, DR0 (P) is the pseudo-DR0 motif from the *Hoxb13* locus, DR8 (C) is the composite DR8 element from the *Mafa* gene, and DR8 is the simple DR8 from the *Dedd* gene. *C*, relative CAT activity from a second series of transfections shows the activity of the mutated composite *Mafa* DR8 element. Wt, wild-type; Mt, mutant.

analysis (pseudo-DR0) resembles a degenerate DR0-type motif (Fig. 6A and supplemental Fig. 6). This motif was present at RAR-occupied sites such as those at the *Hoxb13* and the WD repeat and SOCS box-containing 2 (*Wsb2*) gene loci (Fig. 6, B and C, and supplemental Fig. 6). A more degenerate version of this sequence with a non-consensus 3'-half-site was found at the Rho family GTPase 3 (*Rnd3*) gene locus (Fig. 6C and supplemental Fig. 6).

We performed EMSA competition with wild-type and mutated versions of the *Hoxb13* and *Rnd3* elements. The wild-type elements efficiently competed (Fig. 6C, lanes 5 and 9), whereas versions in which the central 5'-TCAA-3' core is mutated did not compete (lanes 6 and 10). Mutation of single nucleotides in the second pseudo-half-site did not, however, affect competition (lanes 7 and 8 and lanes 11 and 12). ITC measurement of binding to these elements indicated a  $K_d$  for *Hoxb13* of 116 nM and of 93 nM

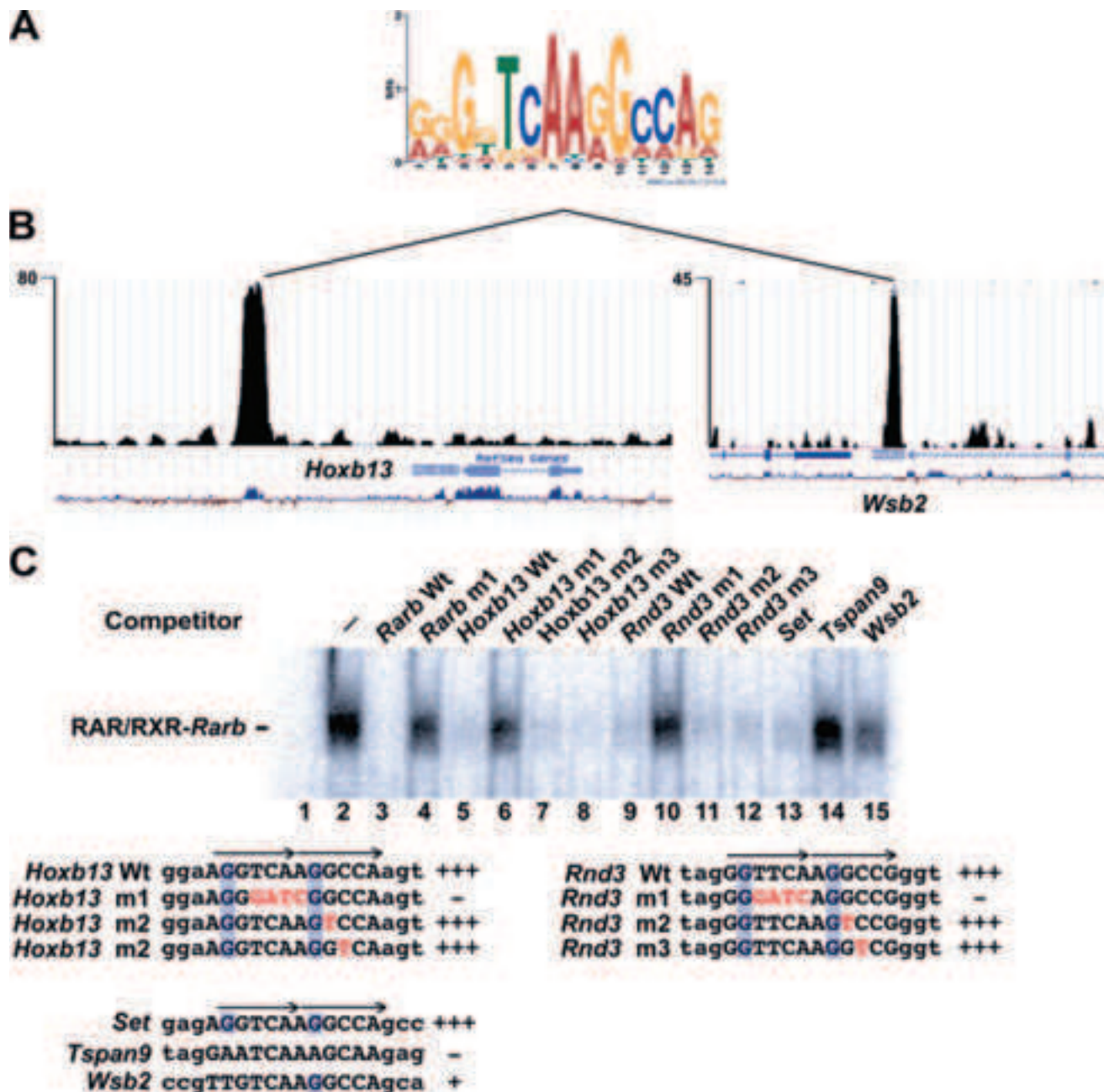


FIGURE 6. Identification of a pseudo-DR0 motif. A, shown is the sequence of the pseudo-DR0 motif identified from the MEME analysis. B, shown is the University of California at Santa Cruz web browser view of sequence tag analysis in .wig file format of the RAR-occupied sites at the *Hoxb13* and *Wsb2* genes comprising the pseudo-DR0 in embryoid bodies. C, EMSA competition analysis of the various indicated pseudo-DR0s to compete for RAR/RXR complex formation. The sequences of the pseudo-DR0 elements within the competing oligonucleotides are shown with the half-sites indicated by arrows. Mutated nucleotides are indicated in red, and the conserved G at position 2 of the half-sites are boxed in blue. All competitors were used a 100-fold excess.

for *Rnd3* (Fig. 3B). Thus, despite its non-consensus 3'-half-site sequence, the *Rnd3* element binds RAR-RXR with an affinity comparable with the canonical DR5.

The pseudo-DR0 from the SET nuclear oncogene (*Set*) has a sequence identical to that of *Hoxb13* and efficiently competes (Fig. 6C, lane 13). In contrast, the tetraspanin 9 (*Tspan9*) gene element does not compete (lanes 14). Similarly, the *Wsb2* gene element also shows almost no ability to compete (lane 15). Consequently at these loci, it is probable that RAR binding is mediated by other as yet unidentified elements. The above elements contain the 5'-TCAA-3' core, but the conserved G at position 2

in canonical half-sites and in the pseudo-half-sites of the *Hoxb13* and *Rnd3* elements is not conserved. Although these motifs do not contain 2 canonical half-sites, their ability to bind RAR-RXR requires conservation of the G residue at position 2 in the first and second half-sites.

We tested the ability of the *Hoxb13* pseudo-DR0 to act as an independent RARE by inserting three copies in the above-described CAT reporter vector. After transfection, no RA response was observed, consistent with the fact that consensus DR0 elements do not show RA responsiveness in this assay (Fig. 5B).

## Characterization of Novel RAREs

**Association of DR0 and DR8 Elements with Other Motifs**—While performing the MEME analysis on the 150 bp at the DR0- and DR8-containing sites, we noted that these were not the only motifs represented at these sites. Both the MEME analysis and a subsequent SPAMO (29) analysis revealed the presence of several other motifs specifically located both 5' and 3' to the DR0 elements often with a rather precise spacing (supplemental Fig. 7A). TOMTOM analysis (30) of these motifs revealed that one of them located 5' to the DR0/8 was a potential PITX2 binding site, with the other motifs showing no significant similarity to known transcription factor binding sites (supplemental Fig. 7B). This analysis indicates that DR0/8 lie within a longer and more complex element made up of several highly represented motifs with specific locations with respect to the DR0/8 (supplemental Fig. 7C). Note, however, that most regions contain a subset of these motifs but only a few comprise all of them. This characteristic is specific for the DR0/8 as no such conservation was seen around the DR5 elements (data not shown).

**DR5 Is Enriched at RAR-occupied Sites Associated with RA-regulated Genes**—We next examined the relationship between RAR occupancy and RA-regulated gene expression. EBs were grown for 4 days in the absence of RA and then treated with RA for 24 h. The transcripts whose expression is induced or repressed by RA compared with the untreated EBs were then assessed by RNA-seq. Transcripts showing a greater than log 2 1-fold change in expression were determined (supplemental Table 2) identifying 824 induced and 379 repressed transcripts. The induced transcripts are strongly enriched in Hox-family and other homeobox-containing transcription factors responsible for subsequent EB patterning. No specific class was represented in the repressed genes.

These data were then compared with those having an RAR-occupied element in the 20-kb upstream/downstream or anywhere within the gene body in the 20h ChIP-seq data set. This analysis does not take into account RAR-occupied sites in far intergenic regions that cannot be readily be assigned to potential target genes. Keeping in mind this caveat, 495 induced and 200 repressed transcripts were identified as potential direct RAR targets (Fig. 7A). A majority of the transcripts whose expression is induced and about half those repressed by RA at 24 h are potentially directly regulated by the RARs.

We next asked whether the RAR-occupied sites associated with RA-regulated transcripts were enriched in a particular class of DR/IR element. The 495 up-regulated transcripts were associated with 943 RAR-occupied sites showing a profile where, although DR0 remained the most common element, the DR5 was relatively enriched (Fig. 7B). At the highly occupied EB sites, the ratio of DR5 to DR0 was 0.18 compared with 0.3 in the total data set, whereas at the up-regulated genes it is enriched to 0.46 (Fig. 7D). Similarly the DR8 element remains prominent in this category. The IR0 element is well represented at the up-regulated transcripts, with 20 elements at the 943 sites compared with the 12 at the 1000 most occupied sites (Figs. 7B and 1A). At RA-repressed transcripts, DR0 and DR2 elements remain the most represented, with a low number of DR5 and DR8 elements, but IR0 elements are strongly depleted (Fig. 7, C and D). Thus, transcripts whose expression is activated by RA in

the EBs are relatively enriched in DR5 and IR0, and DR2 and DR8 elements remain highly represented.

## DISCUSSION

**Flexibility in Half-site Spacing and Topology of Binding Elements Bound by the RAR-RXR Heterodimer**—Here we show that DR0 not DR5 is the most frequent half-site spacing seen at RAR-occupied sites in EBs or F9 cells. DR0 elements can be subdivided into three classes with either an additional canonical half-site or a variant 5'-RGATCA-3' half-site with DR2 spacing forming composite DR8s or DR0s with no additional 5' half-sites. DR0 of the latter class bind RAR-RXR *in vitro* with affinities comparable with those of DR2 and DR5. The location of a subset of DR0 elements at the center of the ChIP-seq peak is also consistent with their occupancy in cells. In addition, we also identified pseudo-DR0 elements with 3' half-sites differing significantly from the consensus sequence. Remarkably, some of these elements have a high affinity *in vitro* for RAR-RXR.

Despite the prevalence of the DR0 element at RAR-occupied loci, it does not act as an independent RARE when placed upstream of a minimal TATA element. It remains to be determined whether the DR0 can confer RA responsiveness in the context of natural promoters acting in combination with other transcription factors. DR0s are components of a larger element composed of several sequence motifs among which are potential PITX2 binding sites. DR0s may function in the context of this larger element in cell types where the appropriate additional factors are present.

The inability of the RAR to activate transcription when bound to DR0 compared with DR2, DR5, and DR8 indicates that half-site spacing exerts allosteric control on RAR activity. We have modeled the structure of the RAR-RXR DNA binding domains bound to DR0 and IR0 spacings based on the known structures of these DNA binding domains bound to DR5 or of the ecdysone receptor heterodimer bound to a IR1 (Ref. 31 and supplemental Fig. 8A). Three-dimensional models indicate that a conformational change of the DNA binding domain may be required to circumvent steric hindrance upon binding to IR0 or to favor dimer interactions on DR0 (supplemental Fig. 8, B and C). These conformational changes likely affect the overall conformation of the heterodimer and co-regulator positioning. Future biophysical and structural experiments will precisely define these changes and why they are compatible with function in the case of IR0 but not in the case of DR0. Nevertheless, these observations reinforce previous examples of allosteric control of RAR-RXR function (32) and are reminiscent of those showing that the DNA binding element can act as an allosteric regulator of glucocorticoid receptor (33).

In the context of embryoid bodies it is interesting to note that RAR-RXR binding to DR0s may antagonize the activity of GCNF (germ cell nuclear factor or NR6A1), a distantly related member of the nuclear receptor superfamily that binds DR0 and appears important for down-regulation of pluripotency genes in RA-induced differentiation (34, 35). RAR-RXR and GCNF may, therefore, compete for binding to DR0 elements and, hence, antagonize each other. It remains to be determined whether GCNF and RAR-RXR compete for binding to DR0s or bind different sets of DR0s.

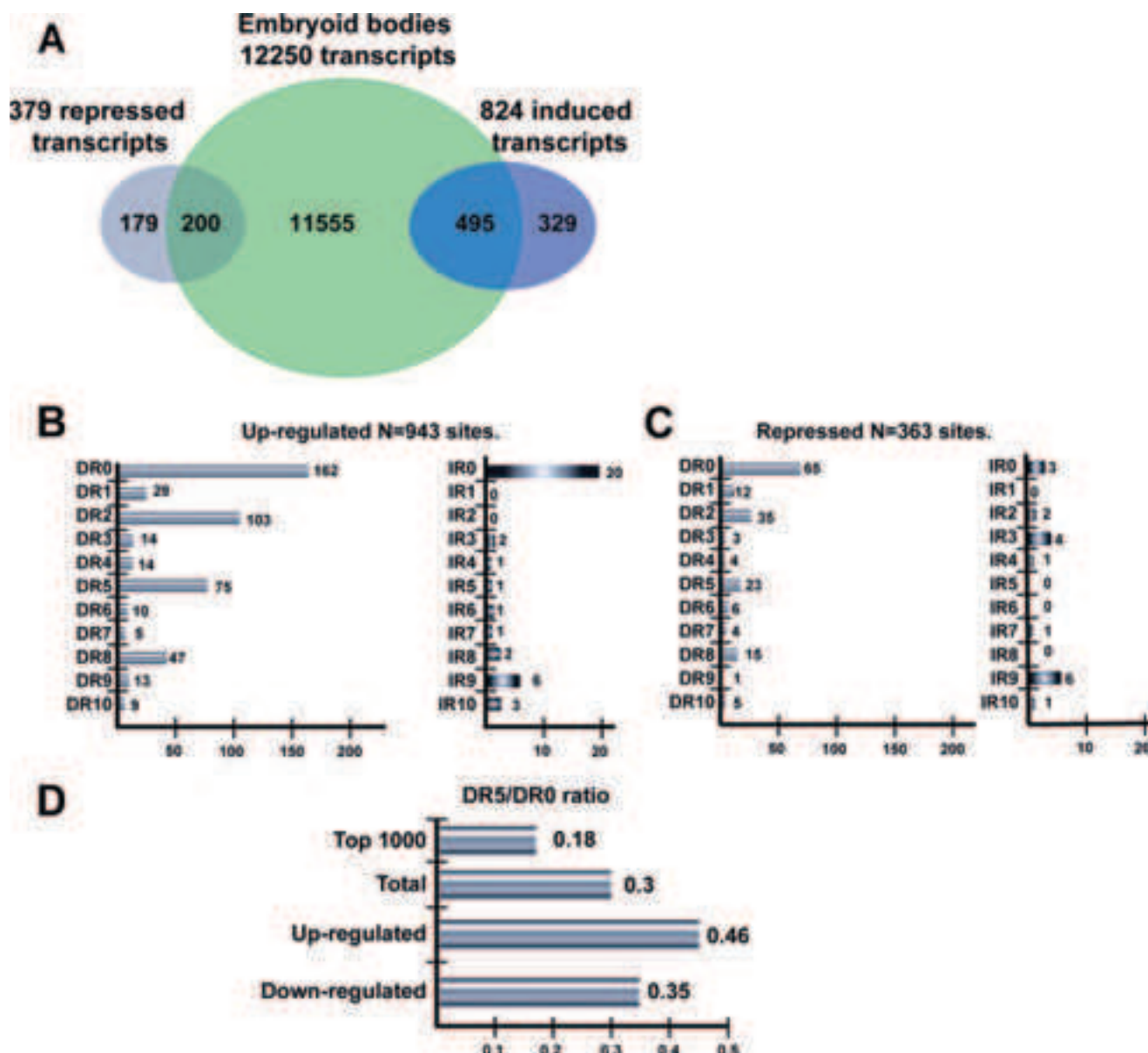


FIGURE 7. **Association of RAR-occupied sites with gene expression.** A, shown is a Venn diagram representing the overlap between the total number of transcripts associated with RAR binding sites and those regulated by RA. B and C, shown are DR and IR frequencies at the RAR-occupied sites associated with RA-induced and repressed transcripts, respectively. D, DR5/DR0 ratios in the indicated classes of RAR-occupied sites.

EMSA and ITC showed that IR0 elements are high affinity RAR-RXR binding sites, and transfection assays showed that they can act as independent RAREs. Previously, an IR0 type RARE has been reported in the mouse gene nuclear receptor subfamily 2, group C, member 1 (*Nr2c1*) gene (36), showing that this element acts as a RARE in natural promoters, and it acts as a RARE when the RAR was expressed in yeast (37). This is consistent with their enrichment at RA-regulated transcripts, suggesting many of these elements act as RAREs. Nevertheless, although IR0 is the most frequent IR element, it is much less abundant than DRs at RAR-occupied sites. We also observed representation of IR9 at RAR occupied sites, but no binding of RAR-RXR to these sites was seen in EMSA (unpublished data).

Together our data indicate that the RAR-RXR heterodimer can bind to a variety of half-site spacings. This property is not seen for VDR-RXR that exhibits a strong specificity for DR3

(27) nor for PPAR-RXR that recognizes DR1 (39). The ability to recognize such diverse spacings can be explained by the lack of defined secondary structure of the hinge connecting the DNA and ligand binding domains of the RAR and RXR (16). This contrasts with the  $\alpha$ -helical structure of the VDR and thyroid hormone receptor hinges that imposes a constraint on half-site spacing (16, 40). For PPAR-RXR, constraint is imposed by the additional interactions of the PPAR hinge with the 5'-AACT-3' motif positioned 5' of the half-site.

**Composite DR8, a Novel Signature of RAR-occupied Loci**—An important observation of our study is the high frequency of composite DR8 elements at RAR-occupied sites shortly after RA treatment. Composite DR8 elements efficiently bind RAR-RXR *in vitro*, and dissection of the DR2, DR8, and DR0 components shows that the DR2 and DR8 combinations both bind efficiently, whereas the tested DR0s are less efficient. It remains

to be determined why the DR0 in this context appears to be less favored. RAR-RXR binds the DR8 spacing as indicated by the fact that DR8 elements with no intervening half-site efficiently compete RAR-RXR in EMSA *in vitro*, and they act as RAREs in cells. Similarly, mutations of the composite DR8 leaving only the DR8 spacing also act as RAREs in cells. Moreover, a simple DR8 element has been previously reported to mediate the response to 9-*cis*-RA in the *IGFBP3* promoter (41) or the human H1(0) histone gene (42).

Despite these observations, our data are more compatible with preferential occupation of the DR2 and the DR0 than of the DR8 in cells. Moreover, as both the DR2 or DR8 spacings act as RAREs, what is the significance of the DR2-DR0 half-site arrangement in the composite DR8? One possibility is that an RAR-RXR heterodimer may bind to the DR2 or DR0 spacings along with an orphan receptor that could bind to the additional 5' or 3' half-site. Although further experiments will be required to determine whether this occurs *in vivo*, our data show that this previously unrecognized half-site organization is a signature of a large number of RAR-occupied sites.

Our results are not the first to report overlapping or composite RAR binding elements. For example, competitive binding to a composite DR3-DR9 arrangement of three half-sites has been shown to mediate RA inhibition of VDR activity at the *Itgb3* promoter (43). Similarly, composite elements mediating both RAR and ER responses have been identified in the lactoferrin and placental lactogen promoters (44, 45), but such composite elements also mediate antagonism between these pathways in breast cancer cells (25). Also, a composite DR/IR sequence forms part of a complex regulatory element involved in regulation of  $\gamma$ F-crystallin expression by RAR, PAX6, and large MAF proteins (46). Although the above illustrate specific examples of overlapping elements, the composite DR8 described here represents a more specific and frequent half-site organization.

Comparison of RAR genomic occupancy and RA-regulated transcription shows that a majority of transcripts regulated by RA at 24 h are potential direct targets showing occupancy by RAR at 2 h. This is consistent with the observation that the major changes in expression of many of the transcripts takes place over the first 12 h (38) and our unpublished data. Although DR5 elements are enriched at sites associated with RA-regulated transcription, consistent with their known role as RAREs, DR2, DR8, and IR0 are also strongly represented at sites associated with RA-regulated transcripts. This together with their ability to act as independent RAREs suggest that each type of element may contribute to the RA response in EBs.

Comparison of our findings with those previously reported highlight similarities but also major differences. Mahony *et al.* (28) have previously reported ChIP-seq data from EBs treated with RA for 8 h. These EBs were not grown under the same conditions as reported here, and their data set comprised many fewer unique sequence reads, resulting in 1924 RAR-occupied sites, far less than identified in our data sets. Mahony *et al.* (28) analyzed their data set for the presence of DR0-DR10 and IR0-IR10 and found DR5 and DR2 as the most highly represented elements with only a small number of DR0 and almost no IR0. In contrast, Mahoney *et al.* (28) did not identify the composite DR8 motif in their data sets.

Similarly, the analysis of Hua *et al.* (25) of RAR occupancy in MCF7 cells identified DR5 as the most frequent element but also noted a low frequency of DR0 elements. Surprisingly, however, they observed very few DR2 and DR8 elements at RAR occupied loci in this cell type. Unlike MCF7 cells, we did not observe a high frequency of IR3 ER binding elements colocalizing with RAR in EBs or F9 cells. Differences in species and/or cell type may explain these contrasting observations.

In summary, our results change the paradigm for how RAR-RXR recognizes the genome from predominantly DR2 and DR5 elements to a more complex situation with a variety of half-site spacings and topologies and an allosteric regulation of RAR function by the half-site spacing of the DNA binding element.

**Acknowledgments**—We thank Carole Peluso-Iltis for technical help in protein purification and all members of the Institut de Génétique et de Biologie Moléculaire et Cellulaire high throughput sequencing platform, in particular C. Keime for analysis of RNA-seq data and S. Le Gras for pipeline analysis.

## REFERENCES

- Blomhoff, R., and Blomhoff, H. K. (2006) Overview of retinoid metabolism and function. *J. Neurobiol.* **66**, 606–630
- Mark, M., Ghyselinck, N. B., and Chambon, P. (2006) Function of retinoid nuclear receptors. Lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu. Rev. Pharmacol. Toxicol.* **46**, 451–480
- Niederreither, K., and Dollé, P. (2008) Retinoic acid in development. Toward an integrated view. *Nat. Rev. Genet.* **9**, 541–553
- Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006) International Union of Pharmacology. LXIII. Retinoid X receptors. *Pharmacol. Rev.* **58**, 760–772
- Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J., and Gronemeyer, H. (2006) International Union of Pharmacology. LX. Retinoic acid receptors. *Pharmacol. Rev.* **58**, 712–725
- Bastien, J., and Rochette-Egly, C. (2004) Nuclear retinoid receptors and the transcription of retinoid target genes. *Gene* **328**, 1–16
- Balmer, J. E., and Blomhoff, R. (2002) Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808
- Balmer, J. E., and Blomhoff, R. (2005) A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.* **96**, 347–354
- Delacroix, L., Moutier, E., Altobelli, G., Legras, S., Poch, O., Choukrallah, M. A., Bertin, I., Jost, B., and Davidson, I. (2010) Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Mol. Cell. Biol.* **30**, 231–244
- Mangelsdorf, D. J., Umesono, K., and Evans, R. M. (1994) in *The Retinoids: Biology, Chemistry, and Medicine* (Sporn, M. B., Roberts, A. B., and Goodman, D. S., eds) pp. 319–349, Raven Press, New York
- Kobi, D., Steunou, A. L., Dembélé, D., Legras, S., Larue, L., Nieto, L., and Davidson, I. (2010) Genome-wide analysis of POU3F2/BRN2 promoter occupancy in human melanoma cells reveals Kitl as a novel regulated target gene. *Pigment Cell Melanoma Res.* **23**, 404–418
- Martianov, I., Choukrallah, M. A., Krebs, A., Ye, T., Legras, S., Rijkers, E., Van Ijcken, W., Jost, B., Sassone-Corsi, P., and Davidson, I. (2010) Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. *BMC Genomics* **11**, 530
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biology* **9**, R137
- Krebs, A., Frontini, M., and Tora, L. (2008) GPAT. Retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* **9**,

533

15. Ye, T., Krebs, A. R., Choukallah, M. A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011) seqMINER. An integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35
16. Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., Svergun, D. I., and Moras, D. (2011) Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* **18**, 564–570
17. Bailey, T. L. a. E., C. (1994) *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36, AAAI Press, Menlo Park, CA
18. Mengus, G., Fadloun, A., Kobi, D., Thibault, C., Perletti, L., Michel, I., and Davidson, I. (2005) TAF4 inactivation in embryonic fibroblasts activates TGF $\beta$  signaling and autocrine growth. *EMBO J.* **24**, 2753–2767
19. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat. Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111
20. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515
21. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology* **11**, R106
22. Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y. A. (2004) Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**, 1003–1009
23. Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., and Stunnenberg, H. G. (2009) ChIP-seq of ER $\alpha$  and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* **28**, 1418–1428
24. Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoutte, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiryanov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S., and Brown, M. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–1297
25. Hua, S., Kittler, R., and White, K. P. (2009) Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137**, 1259–1271
26. Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoutte, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A., and Brown, M. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43
27. Ramagopalan, S. V., Heger, A., Berlanga, A. J., Maugeri, N. J., Lincoln, M. R., Burrell, A., Handunnetthi, L., Handel, A. E., Disanto, G., Orton, S. M., Watson, C. T., Morahan, J. M., Giovannoni, G., Ponting, C. P., Ebers, G. C., and Knight, J. C. (2010) A ChIP-seq defined genome-wide map of vitamin D receptor binding. Associations with disease and evolution. *Genome Res.* **20**, 1352–1360
28. Mahony, S., Mazzoni, E. O., McCuine, S., Young, R. A., Wichterle, H., and Gifford, D. K. (2011) Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12**, R2
29. Whittington, T., Frith, M. C., Johnson, J., and Bailey, T. L. (2011) Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res.* **39**, e98
30. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007) Quantifying similarity between motifs. *Genome Biology* **8**, R24
31. Devarakonda, S., Harp, J. M., Kim, Y., Ozyhar, A., and Rastinejad, F. (2003) Structure of the heterodimeric ecdysone receptor DNA binding complex. *EMBO J.* **22**, 5827–5840
32. Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M. G., Heyman, R. A., and Glass, C. K. (1994) Regulation of retinoid signaling by receptor polarity and allosteric control of ligand binding. *Nature* **371**, 528–531
33. Meijnsing, S. H., Pufall, M. A., So, A. Y., Bates, D. L., Chen, L., and Yamamoto, K. R. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–410
34. Gu, P., LeMenuet, D., Chung, A. C., Mancini, M., Wheeler, D. A., and Cooney, A. J. (2005) Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation. *Mol. Cell. Biol.* **25**, 8507–8519
35. Zechel, C. (2005) The germ cell nuclear factor (GCNF). *Mol. Reprod. Dev.* **72**, 550–556
36. Lee, C. H., and Wei, L. N. (1999) Characterization of an inverted repeat with a zero spacer (IR0)-type retinoic acid response element from the mouse nuclear orphan receptor TR2–11 gene. *Biochemistry* **38**, 8820–8825
37. Heery, D. M., Pierrat, B., Gronemeyer, H., Chambon, P., and Losson, R. (1994) Homo- and heterodimers of the retinoid X receptor (RXR) activated transcription in yeast. *Nucleic Acids Res.* **22**, 726–731
38. Simandi, Z., Balint, B. L., Poliska, S., Ruhl, R., and Nagy, L. (2010) Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. *FEBS Lett.* **584**, 3123–3130
39. Nielsen, R., Pedersen, T. A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., Denissov, S., Børgesen, M., Francoijs, K. J., Mandrup, S., and Stunnenberg, H. G. (2008) Genome-wide profiling of PPAR $\gamma$ :RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev.* **22**, 2953–2967
40. Orlov, I., Rochel, N., Moras, D., and Klaholz, B. P. (2012) Structure of the full human RXR/VDR nuclear receptor heterodimer complex with its DR3 target DNA. *EMBO J.* **31**, 291–300
41. Chang, Y. S., Cho, J. Y., Cho, H. A., Kim, H. J., Chang, J., Ahn, C. M., Kim, S. K., and Kim, S. K. (2006) 9-*cis*-Retinoic acid induces insulin-like growth factor binding protein-3 through DR-8 retinoic acid-responsive elements. *Cancer Biol. Ther.* **5**, 586–592
42. Bouterfa, H. L., Piedrafita, F. J., Doenecke, D., and Pfahl, M. (1995) Regulation of H1(0) gene expression by nuclear receptors through an unusual response element. Implications for regulation of cell proliferation. *DNA Cell Biol.* **14**, 909–919
43. Cao, X., Teitelbaum, S. L., Zhu, H. J., Zhang, L., Feng, X., and Ross, F. P. (1996) Competition for a unique response element mediates retinoic acid inhibition of vitamin D3-stimulated transcription. *J. Biol. Chem.* **271**, 20650–20654
44. Stephanou, A., and Handwerger, S. (1995) Retinoic acid and thyroid hormone regulate placental lactogen expression in human trophoblast cells. *Endocrinology* **136**, 933–938
45. Lee, M. O., Liu, Y., and Zhang, X. K. (1995) A retinoic acid response element that overlaps an estrogen response element mediates multihormonal sensitivity in transcriptional activation of the lactoferrin gene. *Mol. Cell. Biol.* **15**, 4194–4207
46. Yang, Y., Chauhan, B. K., Cveklova, K., and Cvekl, A. (2004) Transcriptional regulation of mouse  $\alpha$ B- and  $\gamma$ F-crystallin genes in lens. Opposite promoter-specific interactions between Pax6 and large Maf transcription factors. *J. Mol. Biol.* **344**, 351–368

## **Supplemental Experimental Procedures.**

### **ChIP.**

RA treated EBs were directly dounced in 1XPBS with 0,4% formaldehyde for 10 min at room temperature and the reaction was stopped by adding glycine to final concentration 0.2M for 10 minutes at room temperature. Fixed cells were rinsed twice with PBS and resuspended in lysis buffer (10 mM EDTA pH 8, 50 mM Tris-HCl pH 8, SDS 1 %). Lysate was sonicated 30 min (30 sec on / 30 sec off) in a Diagenode water bath-sonicator and centrifuged at 14000 rpm for 10 min. The cleared supernatant was used immediately in ChIP experiments or stored at -80°C.

50-150 ug of sonicated chromatin was diluted 10 times in ChIP Dilution Buffer (SDS 0.01%, Triton X-100 1.1%, 1.2 mM EDTA pH 8, 16.7 mM Tris-HCl pH 8 and 167 mM NaCl) and pre-cleared for 1 hour, rotating at 4°C, with 50 ul blocked beads (Protein G Sepharose (PGS) 50% slurry incubated for 4 hours minimum with BSA 0.5 mg/mL and yeast tRNA 0.2 mg/mL) before the overnight incubation with 2-5ug of specific antibody. The bound material was recovered after a 2 hours incubation, rotating at 4°C, with 30 ul blocked PGS. The beads were washed, for 5 minutes, once in Low Salt Buffer (SDS 0.1%, Triton X-100 1%, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8 and 150 mM NaCl), twice in High Salt Buffer (SDS 0.1%, Triton X-100 1%, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8 and 500 mM NaCl), twice in LiCl Buffer (0.25 M LiCl, NP-40 1%, Na Deoxycholate 1%, 1 mM EDTA pH 8 and 10 mM Tris-HCl pH 8) and twice in TE. ChIPed material was eluted by two 15 minute incubations at room temperature with 250 ul Elution Buffer (SDS 1%, 0.1 M NaHCO<sub>3</sub>). Chromatin was reverse-crosslinked by adding 20 ul of NaCl 5M and incubated at 65°C for 4 hours minimum and DNA was submitted to RNase and proteinase K digestion and extracted by phenol-chloroform.

## **RNA-seq.**

The mRNA-seq libraries were prepared following the Illumina protocol with some modifications. Briefly, mRNA was purified from total RNA using oligo-dT magnetic beads and fragmented using divalent cations at 95°C for 5 minutes. The cleaved mRNA fragments were reverse transcribed to cDNA using random primers. This was followed by second strand cDNA synthesis using Polymerase I and RNase H. The double strand cDNA fragments were blunted, phosphorylated and ligated to single-end adapter dimers followed by PCR amplification (30 sec at 98°C; [10 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C] x 13 cycles; 5 min at 72°C). After PCR amplification, surplus PCR primers and dimer adapters were removed by purification using AMPure beads (Agencourt Biosciences Corporation). Size selection was performed by electrophoresis on a 2% agarose gel and DNA fragments in the range of ~250-350bp were excised and purified using QIAquick Gel Extraction Kit (Qiagen). DNA libraries were checked for quality and quantified using 2100 Bioanalyzer (Agilent). The libraries were loaded in the flowcell at 6pM concentration and clusters generated and sequenced on the Illumina Genome Analyzer IIX as single-end 54 base reads. Image analysis and base calling were performed using the Illumina Pipeline version 1.6.

## **Legends to Supplemental Figures and Tables.**

**Supplemental Figure 1.** Schematic representation of the distribution of RAR-occupied site locations with respect to gene annotation. The % of sites at each defined class of locations are indicated.

**Supplemental Figure 2.** RAR genomic occupancy in EBs and F9 cells. **A.** Comparative clustering of the EB and F9 cell 2 hour RAR ChIP-seq data. Tag densities from each dataset were collected in a window of 10 kb around the 13385 EB RAR-occupied sites. The collected

values were subjected to k-means clustering coupled to ranked based-normalization. Examples of read density plots for several classes showing comparable or preferential occupation in either data set are shown to the right of the figure. **B.** Venn diagramme comparison of RAR-occupied genes in embryoid bodies and F9 cells. The total number of annotated Ensembl/Refseq target transcripts/genes in each cell type is indicated along with those that are shared and specific to each data set.

**Supplemental Figure 3.** Binding of RAR-RXR to IR0 elements. **A.** UCSC view of sequence tag density in .wig file format of the RAR-occupied sites at the *Mbd6* and *4930452B06Rik* genes comprising an IR0 in EBs. **B.** EMSA competition analysis of the indicated IR0 elements. The sequences of the IR0 motifs within the competing oligonucleotides are shown with the inverted half sites indicated by arrows. Mutated nucleotides are indicated in red. All competitors were used a 100-fold excess.

**Supplemental Figure 4.** **A.** Examples of simple and composite DR8 elements that are occupied by RAR at the indicated gene loci in embryoid bodies. The frequency of each type of DR8 in the top 1000 sites in EBs and F9 cells is indicated. **B.** UCSC view of sequence tag density in .wig file format of the RAR-occupied sites at the *Mafa* and *Cd97* genes comprising a composite DR8 in embryoid bodies. **C.** EMSA competition analysis of the various indicated IR0 elements to compete for RAR/RXR complex formation. The sequences of the DR8 motifs within the competing oligonucleotides are shown with the half sites indicated by arrows. Mutated nucleotides are indicated in red. All competitors were used a 100-fold excess.

**Supplemental Figure 5.** EMSA competition analysis of simple and composite DR8. **A.** Competition with a simple DR8 from the *Dedd* gene locus. Competition was performed with increasing quantities (10, 25, 50, 100-fold excess) of the oligonucleotides shown above each lane. Lane 1 is the oligonucleotide probe with no recombinant RAR-RXR and lane 2 with RAR-RXR, but no competitor. The wild type and mutated oligonucleotides are shown below

the EMSA panel, and mutations are indicated in red. **B.** Competition with a DR8 comprising a variant 5'RGATCA-3' half site from the Neuropilin 1 (*Nrp1*) gene. The sequences of the *Nrp1* DR8 within the competing oligonucleotides are shown with the half sites indicated by arrows. Mutated nucleotides are indicated in red. **C.** Binding of RAR-RXR to different half sites spacings. The labelled oligonucleotides used are indicated above each lane along with the presence or absence of recombinant RAR-RXR.

**Supplemental Figure 6.** List of potential pseudo DR0 elements. The conserved G residues at position 2 of the half sites are boxed. Those that do not conserve both of these residues are unlikely to constitute RAR/RXR binding elements as exemplified by the *Tspan9* and *Wsb2* elements in Figure S7.

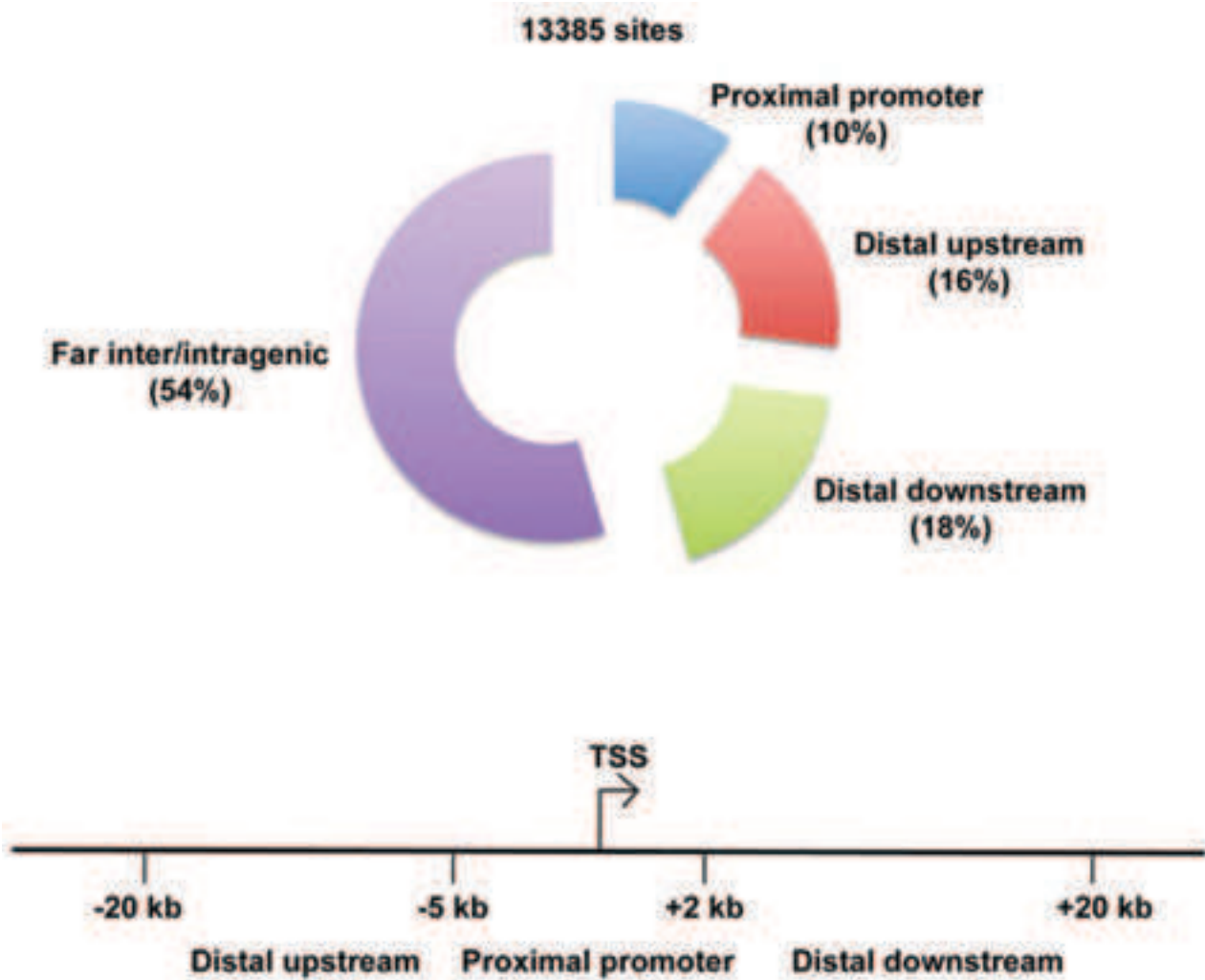
**Supplemental Figure 7.** DR0/8 are embedded in a longer sequence element. **A.** Results of the MEME and SPAMO analysis are summarised. The consensus sequences for each of the identified motifs are shown along with their frequency, location and distance relative to the 200 top DR0 and 186 top DR8 elements in the EB data set. **B.** Results of TOMTOM analysis showing the potential similarity of one of the motifs to the consensus PITX2 binding sites in the Transfac database. **C.** Schematic representation of the locations of the identified motifs relative to the DR0/8 element.

**Supplemental Figure 8.** Model of RAR-RXR binding to IR0 and DR0. **A.** Architecture of RAR/RXR/DR5 complex in solution obtained by small angle X-ray scattering refinement. **B** 3D model of RAR/RXR/IR0 complex based on ecdysone receptor heterodimer bound to an IR1. The region involving a steric clash between the two DBDs is circled. **C.** 3D model of RAR/RXR/DR0 complex.

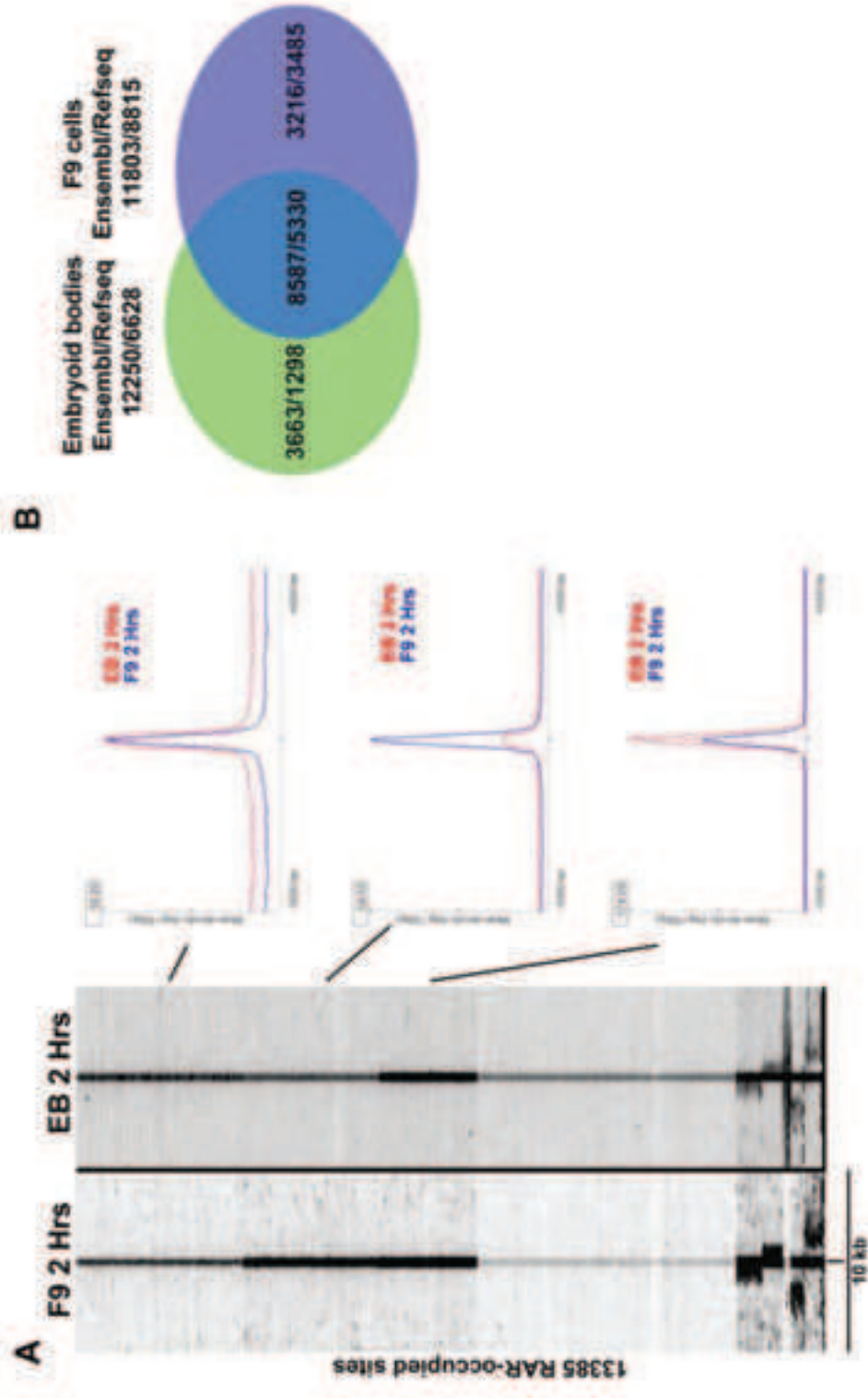
**Supplemental Table 1.** ChIP-seq identification of RAR-occupied loci in EB after 2 hours of RA-treatment. Excel table showing the output from the MACS-GPAT analysis. Page 1 shows the results of the peak detection and annotation. The raw sequence data was analysed first by

the MACs programme to detect peaks. The results of this analysis are indicated in columns A-I. The regions with significant peaks were then annotated using the GPAT programme to detect nearby annotated transcripts. The results of annotation are shown in the remaining columns. For all peaks chromosome location, length, number of reads, pvalue, and local fold enrichment are indicated. GPAT annotation was performed using a window of +/- 20kb with respect to the Ensembl gene coordinates. The identity of the gene closest to the peak and of any secondary genes located within the 20kB window are indicated along with the distance of the peak to the TSS. Page 2 shows genes with an RAR-occupied site.

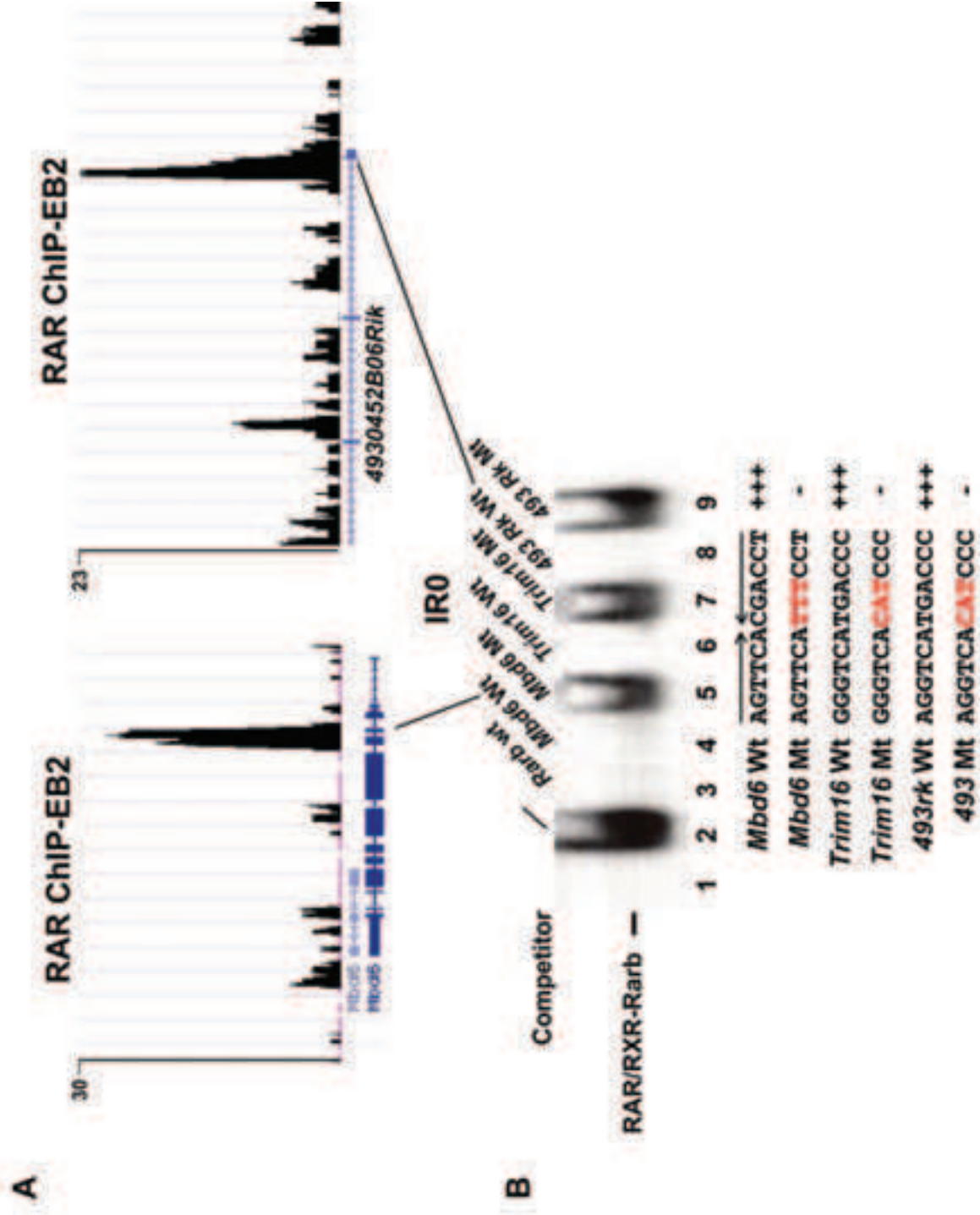
**Supplemental Table 2.** RA-regulated genes in EBs. Pages 1 and 2 of the table show the Ensembl IDs, gene names and descriptions for transcripts that are induced or repressed by RA in EBs, respectively. The normalised expression values for the duplicate samples are indicated along with the mean value, the Log2 change in expression and adjusted pvalue.



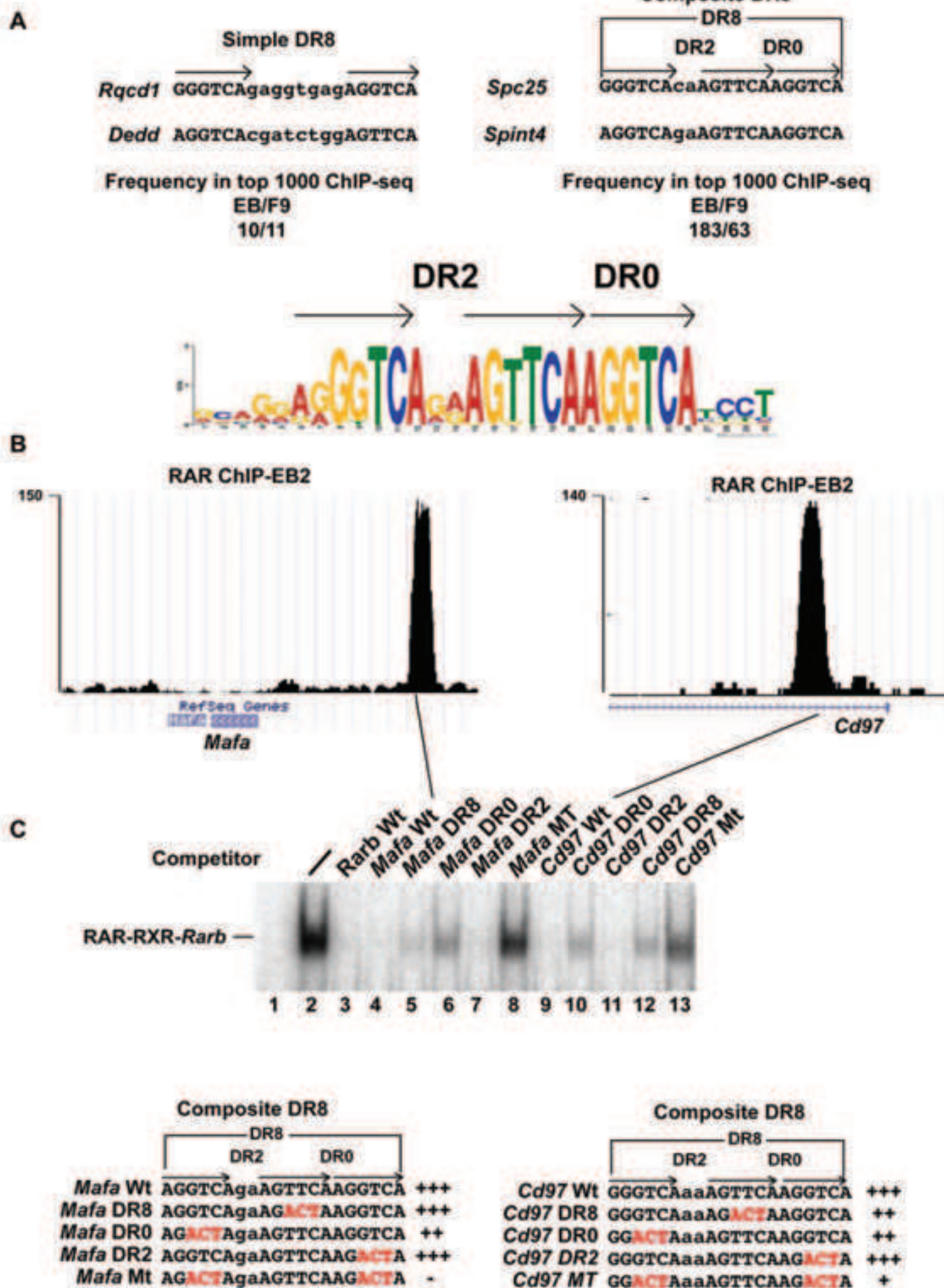
Moutier et al., Supplemental Fig. 1

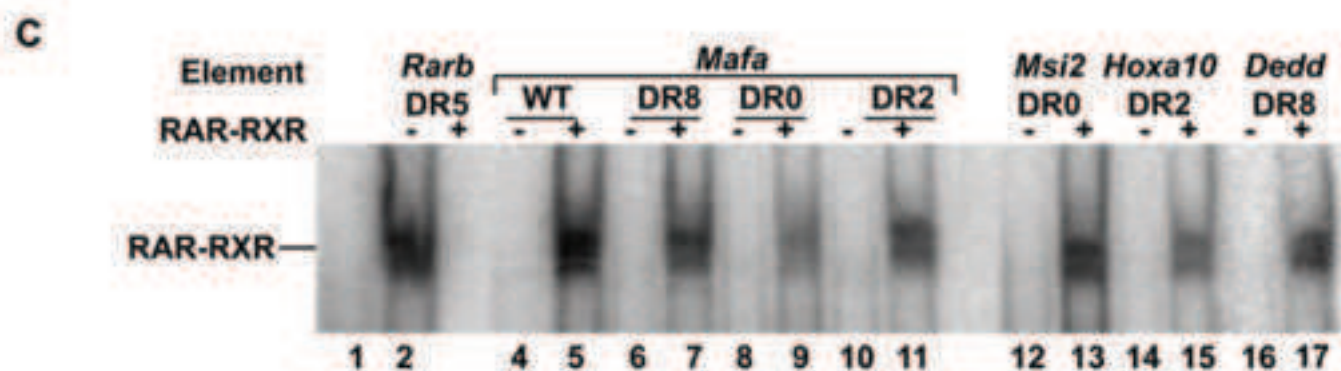
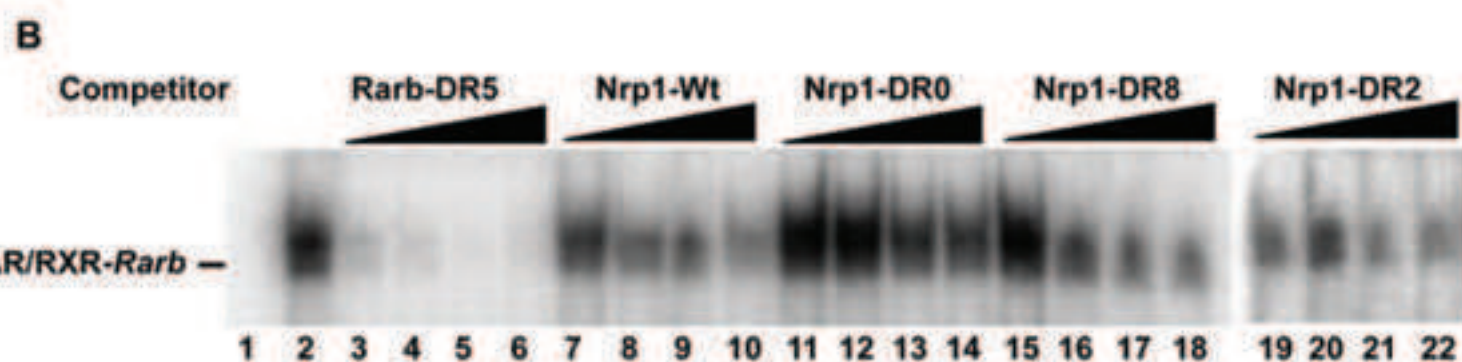
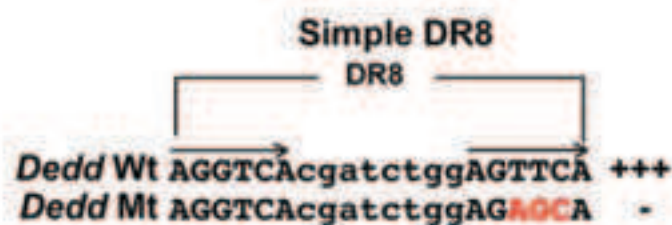
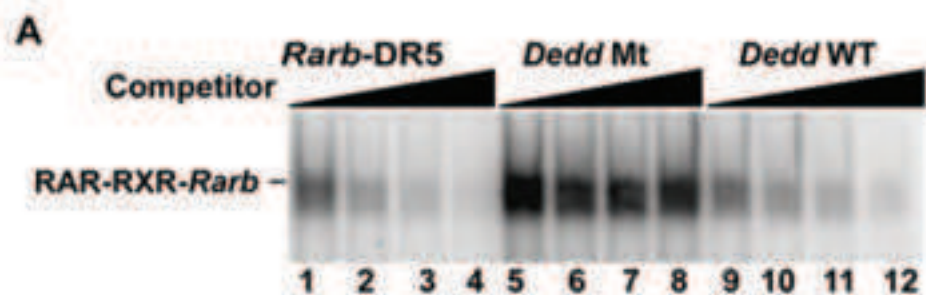


Moutier et al., Supplemental Fig. 2



Moutier et al., Supplemental Fig. 3





# Gene ID and location

## Pseudo DR0 motif

| Name                             | Strand | Start | p-value  | Sites                                                     |
|----------------------------------|--------|-------|----------|-----------------------------------------------------------|
| Cenpm chr15:82077583-82078082    | -      | 246   | 9.99e-09 | AATAATCTCT <b>GGTTCAA</b> <b>CCAG</b> OCTGGTCTAC          |
| Set1 chr2:29916796-29917295      | +      | 118   | 1.93e-08 | GAACCTCTAGA <b>CAGTCA</b> <b>CCAG</b> CCTGGTCTAT          |
| Gx2h1 chr7:54053800-54054299     | +      | 244   | 3.69e-08 | GTGCTACCAA <b>GAATTCAA</b> <b>CCAG</b> TTTTGTAT           |
| Rnd3 chr2:51002808-51003307      | -      | 189   | 2.95e-07 | CAGAGGTCTA <b>GGTTCAA</b> <b>CCAG</b> GTTGTGCT            |
| Hoxb13 chr11:96051518-96052017   | -      | 180   | 3.66e-07 | CACCTCCAGG <b>AAGTCA</b> <b>CCAA</b> GTTGAAAGTT           |
| Myblf chr7:134355912-134356411   | +      | 103   | 4.26e-07 | CAGATCTCT <b>GGTTCAA</b> <b>CCAG</b> GAAGACTGGG           |
| Slc39f2 chr9:53618751-53619250   | +      | 5     | 4.93e-07 | TGGCC <b>CAGTCA</b> <b>CCAG</b> GATAGGTGAC                |
| Lipe chr7:26180609-26181108      | -      | 390   | 8.05e-07 | TGGGCGGGCC <b>AAGTTCAA</b> <b>CCAG</b> GTGAAATGAGA        |
| Capn3 chr7:30981291-30981790     | -      | 455   | 1.87e-06 | CCTGAGTGCT <b>GAATTCAA</b> <b>CCAA</b> GTGCCACCCAC        |
| Wiscr27 chr5:135408686-135409185 | +      | 430   | 2.27e-06 | GCGGNTTCT <b>GAATTC</b> <b>CCAG</b> CCTGGTATAC            |
| Akshf7 chr17:57136318-57136817   | -      | 263   | 2.51e-06 | GAGGTACGT <b>GAATTCAA</b> <b>CCAA</b> GGTTAGAAGT          |
| Wrb2 chr5:117806572-117807071    | +      | 230   | 2.92e-06 | CGCTGGGCC <b>TTTTCAA</b> <b>CCAG</b> CATTTCAGCT           |
| Pbx2 chr17:34729694-34730193     | -      | 148   | 3.47e-06 | TTTAGGGCT <b>AAGTCA</b> <b>CCAG</b> CTTTGAGAGA            |
| Pick1 chr15:79054023-79054522    | -      | 225   | 3.79e-06 | GGTGAGGTCA <b>GAAT</b> <b>CCAA</b> <b>CCAG</b> CAGCAGCCCA |
| Figl1 chr11:11709549-11710048    | -      | 470   | 5.31e-06 | CATGCTTTG <b>AAGTTAA</b> <b>CCAG</b> AAAGGCAGGC           |
| Rf1 chr2:51926638-51927137       | +      | 365   | 6.73e-06 | GGGTGGTGT <b>GAATTCAA</b> <b>CCAG</b> AAGATTCCA           |
| Gli1 chr10:126777917-126778416   | +      | 3     | 7.89e-06 | CAG <b>AAGTTCAA</b> <b>CCAG</b> GTCTCCCCAG                |
| Tspan9 chr6:128092547-128093046  | +      | 41    | 8.55e-06 | CAGGTCTGTA <b>GAATTCAA</b> <b>CCAG</b> AGTTCTTTGT         |
| Mdn1 chr10:79609754-79610253     | -      | 204   | 9.26e-06 | CCAGTCTCC <b>AAGTCA</b> <b>CCAG</b> GCCGAGAGAC            |
| Spv2 chr14:106300009-106300508   | -      | 158   | 1.07e-05 | AGTTTAGTTA <b>GAAT</b> <b>CCAA</b> <b>CCAT</b> CTGGGAAGAA |
| Gna1 chr3:89084900-89085399      | -      | 220   | 1.15e-05 | CTAGGAGGGA <b>AAGTCA</b> <b>CCAG</b> AGCCCTCTC            |
| Nng2 chr2:29100631-29101130      | +      | 437   | 1.43e-05 | AGACTGGAT <b>GAATTCAA</b> <b>CCAG</b> AGTGAAGCAA          |
| Ywhag2 chr5:136410834-136411333  | +      | 211   | 1.43e-05 | AAGGAAAGAA <b>AAGTTCAA</b> <b>CCAT</b> AGATCTGGAC         |
| Dgat1 chr15:76342532-76343031    | -      | 63    | 1.90e-05 | TCTTCCACCG <b>AAGTTCCA</b> <b>CCAA</b> GCTGCACACT         |

DR0

Moutier et al., Supplemental Fig. 6

**A**

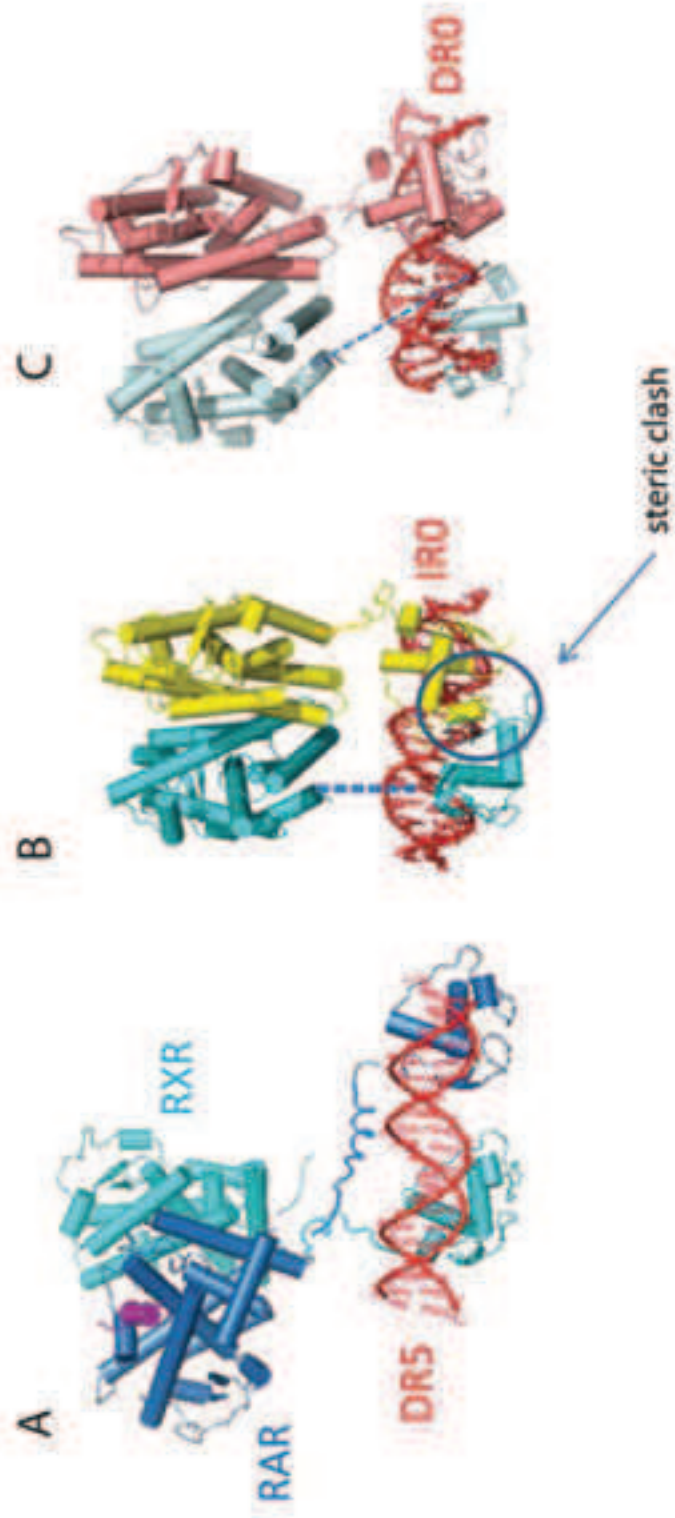
|   | Motif | # Motifs<br>/200 DR0 | # Motifs<br>/186 DR8 | Distance    |
|---|-------|----------------------|----------------------|-------------|
|   |       | 200                  | 186                  | NA          |
| 1 |       | 126                  | 109                  | 5-15 bp 5'  |
| 2 |       | 120                  | 114                  | 19-21 bp 5' |
| 3 |       | 98                   | 56                   | 20-22 bp 5' |
| 4 |       | 48                   | 113                  | 0 bp 3'     |
| 5 |       | 127                  | 132                  | 15-17 bp 3' |
| 6 |       | 57                   | NA                   | 25-30 bp 3' |

**B**



**C**





Moutier et al., Supplemental Fig. 8

***Publication annexe***  
***n°3***

## **Phosphorylation of the retinoic acid receptor RARc2 is crucial for the neuronal differentiation of mouse embryonic stem cells**

Ziad Al Tanoury, Samia Gaouar, Aleksandr Piskunov\*, Tao Ye, Sylvia Urban, Bernard Jost, Céline Keime, Irwin Davidson, André Dierich and Céline Rochette-Egly

## RESEARCH ARTICLE

# Phosphorylation of the retinoic acid receptor RAR $\gamma$ 2 is crucial for the neuronal differentiation of mouse embryonic stem cells

Ziad Al Tanoury, Samia Gaouar, Aleksandr Piskunov\*, Tao Ye, Sylvia Urban, Bernard Jost, Céline Keime, Irwin Davidson, Andrée Dierich and Cécile Rochette-Egly<sup>‡</sup>

### ABSTRACT

Retinoic acid (RA) plays key roles in cell differentiation and growth arrest by activating nuclear RA receptors (RARs) ( $\alpha$ ,  $\beta$  and  $\gamma$ ), which are ligand-dependent transcription factors. RARs are also phosphorylated in response to RA. Here, we investigated the *in vivo* relevance of the phosphorylation of RARs during RA-induced neuronal differentiation of mouse embryonic stem cells (mESCs). Using ESCs where the genes encoding each RAR subtype had been inactivated, and stable rescue lines expressing RARs mutated in phospho-acceptor sites, we show that RA-induced neuronal differentiation involves RAR $\gamma$ 2 and requires RAR $\gamma$ 2 phosphorylation. By gene expression profiling, we found that the phosphorylated form of RAR $\gamma$ 2 regulates a small subset of genes through binding an unusual RA response element consisting of two direct repeats with a seven-base-pair spacer. These new findings suggest an important role for RAR $\gamma$  phosphorylation during cell differentiation and pave the way for further investigations during embryonic development.

**KEY WORDS:** Retinoic acid, RA, RAR, Nuclear receptor, Phosphorylation, Transcription, Differentiation

### INTRODUCTION

Retinoic acid (RA), the active metabolite of vitamin A, regulates multiple biological processes and plays key roles in embryonic development through the regulation of cell growth and differentiation (Samarut and Rochette-Egly, 2012). RA exerts its effects through nuclear RA receptors (RARs), which are ligand-regulated transcription factors with a well-defined domain organization and structure, consisting of a variable N-terminal domain (NTD), and two well-structured and conserved domains, a central DNA-binding domain (DBD) and a C-terminal ligand-binding domain (LBD) (Bastien and Rochette-Egly, 2004; Rochette-Egly and Germain, 2009).

The RA signalling system is highly complex as it comprises three subtypes (RAR $\alpha$ , RAR $\beta$  and RAR $\gamma$ ), and for each subtype there are at least two isoforms, differing in their NTD (Chambon, 1996). Moreover, RARs act as heterodimers with another family of nuclear receptors, the retinoid X receptors (RXRs) (Germain et al., 2006). According to the canonical model, RXR–RAR

heterodimers control the expression of their target genes involved in cell growth and differentiation through binding to specific and polymorphic RA response elements (RAREs) located in their regulatory regions (Moutier et al., 2012) and through the dynamic association and dissociation of coregulator complexes (Rosenfeld et al., 2006).

During the past decade, this scenario became more complicated with the discovery that RA also has non-genomic effects and induces the rapid activation of the MAPK pathway (Piskunov and Rochette-Egly, 2012; Stavridis et al., 2010), and that RARs are phosphoproteins (Al Tanoury et al., 2013b; Rochette-Egly, 2003). Indeed, we have shown that the RA-activated kinases induce RARs phosphorylation at a serine residue located in the NTD (Bastien et al., 2000; Rochette-Egly et al., 1997). Interestingly, this serine residue is conserved between RARs and during evolution (Samarut et al., 2011), emphasizing that phosphorylation of RARs might be important in RA signalling. However, the *in vivo* relevance of the RAR phosphorylation still remains to be defined, especially in the context of development. As cell differentiation is one of the most crucial steps during development, mouse embryonic stem cells (mESCs) (Gudas and Wagner, 2011) provide an interesting model to study the influence of RAR phosphorylation. Indeed, ESCs are pluripotent cells that self-renew indefinitely and can differentiate *in vitro* into a large variety of cell types (Wilson et al., 2009), such as neuronal cells, in response to RA (Bibel et al., 2007).

Here, by using mESCs specifically lacking each RAR subtype and stable rescue cell lines expressing RAR phosphomutants, we demonstrate that their neuronal differentiation crucially involves the phosphorylated form of RAR $\gamma$ 2. Finally, genome-wide RNA-seq experiments identify direct target genes whose expression is controlled by phosphorylated RAR $\gamma$ 2. These results suggest that there is an important role for RAR phosphorylation in RA signalling during cell differentiation, and pave the way for further investigations during embryonic and tissue development.

### RESULTS

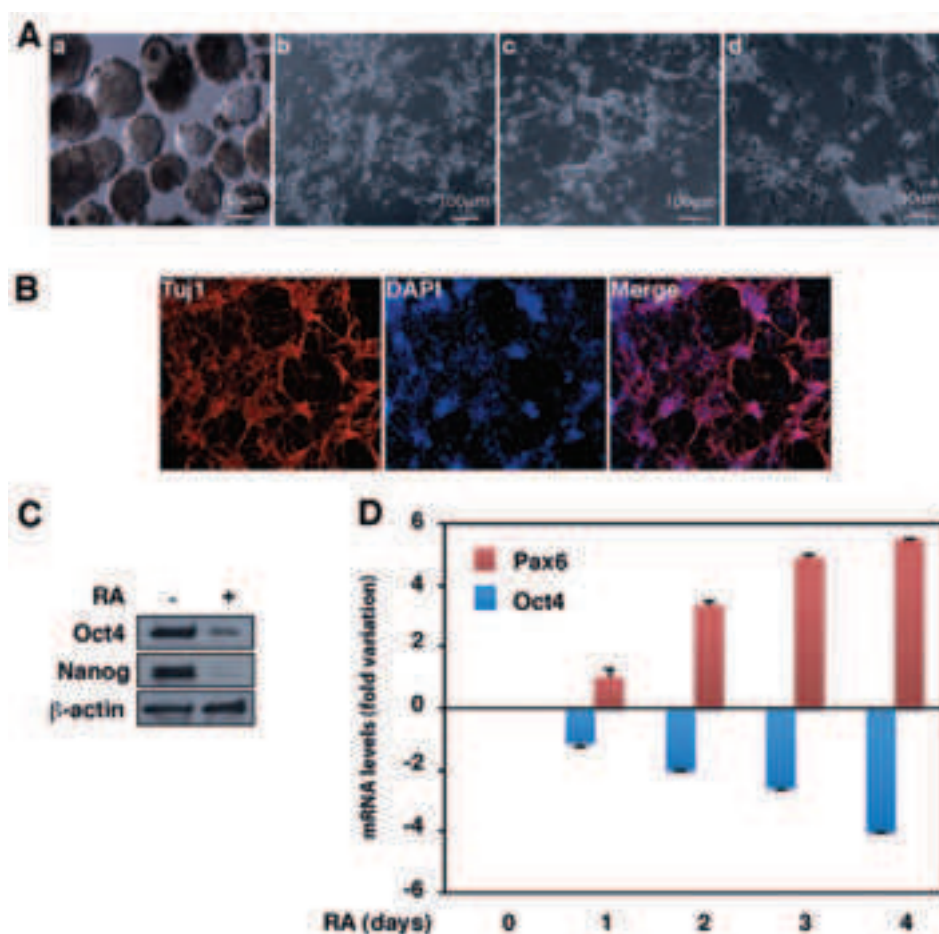
#### RAR $\gamma$ 2 mediates the RA-induced neuronal differentiation of mESCs

When cultured as cellular aggregates, treated with RA and then dissociated and plated according to the protocol of Bibel et al. (Bibel et al., 2007; Bibel et al., 2004), pluripotent mESCs first adopted a spindle-shape morphology characteristic of neural progenitors (Fig. 1A, panel b) (Bibel et al., 2004). Then within 2 days these progenitors gave rise to glutamergic neuronal cells (Fig. 1A, panels c and d) and formed a dense axon network expressing class III  $\beta$ -tubulin (Tuj1) by 7–9 days (Fig. 1B). Concomitantly, the markers for pluripotency [*Oct4* (also known as *Pou5f1*) and *Nanog*] were downregulated, whereas those for neuronal differentiation (*Pax6*) were increased (Fig. 1C,D).

IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), INSERM, U596, CNRS, UMR7104, Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch Cedex, France.

\*Present address: International Biotechnology Center 'Generium' 601125 Volgynsky, Russia.

<sup>‡</sup>Author for correspondence (cegly@igbmc.fr)



**Fig. 1. Neuronal differentiation of mESCs in response to RA.**

(A) Morphology of mESCs at different stages of their differentiation after RA addition. (a) RA-treated cellular aggregates before their dissociation. (b) Progenitors 24 hours after plating dissociated cellular aggregates. (c, d) Neurons, 7–9 days after plating. A dense axon network has formed. (B) Neuronal differentiation of mESCs as assessed by immunofluorescence analysis of class III  $\beta$  tubulin (Tuj1) at 11 days after RA treatment (7 days after plating of the dissociated cellular aggregates). (C) Immunoblotting analysis of the stem cell pluripotency markers (Oct4 and Nanog) 4 days after RA addition. (D) RT-qPCR analysis of the downregulation of Oct4 and upregulation of the differentiation marker Pax6 after RA addition. The results are the mean  $\pm$  s.e.m. of triplicates from two experiments.

As this entire process is initiated by activation of RARs, we investigated which specific RAR subtype ( $\alpha$ ,  $\beta$  or  $\gamma$ ) is required to drive differentiation. ESCs constitutively expressed the RAR $\alpha$  and RAR $\gamma$  proteins and their levels were not significantly affected upon RA treatment (Fig. 2A). In contrast, the RAR $\beta$  protein was hardly detectable but was induced in response to RA (Fig. 2A). Such data suggest that the RAR $\alpha$  and/or RAR $\gamma$  subtypes might be the primary RA targets for mediating the initial phase of neuronal differentiation. To investigate the role of these receptors in the RA-driven neuronal differentiation of mESCs, we used cell lines in which the *Rara* or *Rarg* genes (encoding RAR $\alpha$  and RAR $\gamma$ , respectively) were disrupted by homologous recombination (Lohnes et al., 1993; Lufkin et al., 1993) (Fig. 2B). Our results show that, like wild-type (WT) ESCs, cells lacking RAR $\alpha$  became neuronal progenitors (Fig. 2C) that gave rise to neurons forming a dense axon network (Fig. 2D). In contrast, cells lacking RAR $\gamma$  did not differentiate into cells of the neuronal lineage. They did not become neuronal progenitors (Fig. 2C) and axon structures did not appear even up to 15 days after RA addition (Fig. 2D). These results indicate that RAR $\gamma$  is essential for the RA-induced commitment of mESCs into neuronal precursors that give rise to neurons. They also indicate that RAR $\alpha$  and RAR $\beta$  cannot functionally compensate for loss of RAR $\gamma$  during this process.

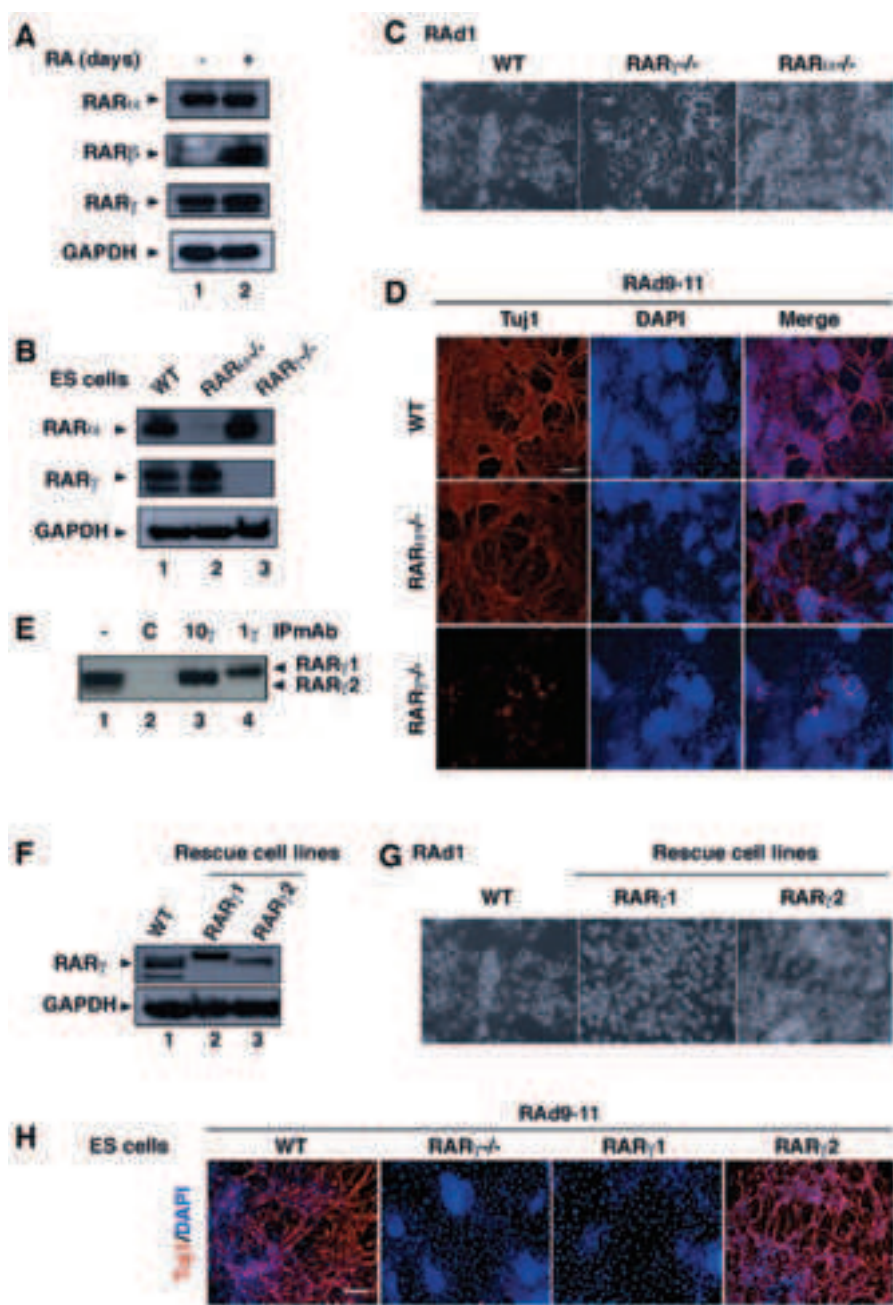
Seven murine RAR $\gamma$  isoforms (mRAR $\gamma$ 1 to mRAR $\gamma$ 7) have been characterized so far, generated by alternative splicing of at least seven exons and differing in their NTD (Kastner et al.,

1990). In mESCs, RAR $\gamma$ 1 and RAR $\gamma$ 2 are the predominant isoforms, whereas the others (RAR $\gamma$ 3 to RAR $\gamma$ 7) have not been detected (Kastner et al., 1990). Moreover, both RAR $\gamma$ 1 and RAR $\gamma$ 2 can be detected at the protein level in mESCs as assessed by immunoblotting after immunoprecipitation with antibodies recognizing specifically each isoform (Bastien et al., 2000) (Fig. 2E).

To investigate the role played by each isoform in the RA-induced neuronal differentiation of mESCs, stable rescue lines expressing the RAR $\gamma$ 2 or RAR $\gamma$ 1 proteins were established from *Rarg*<sup>-/-</sup> cells. Several clones were obtained for each ‘rescue’ transgene, and these expressed RAR $\gamma$ 1 or RAR $\gamma$ 2 at levels similar to the endogenous receptors (Fig. 2F; supplementary material Fig. S1A). We investigated the ability of the RAR $\gamma$ 1 and RAR $\gamma$ 2 rescue lines to differentiate in response to RA. Remarkably, the RAR $\gamma$ 2 rescue lines became neuronal progenitors (Fig. 2G) that gave rise to neurons forming a dense axon network (Fig. 2H). In contrast, the RAR $\gamma$ 1-expressing lines did not give spindle-shaped neuronal progenitors (Fig. 2G) nor any axon structures (Fig. 2H). These results suggest that the RAR $\gamma$ 2 isoform mediates the effects of RA for the commitment of mESCs into the neuronal lineage.

#### **RAR $\gamma$ 2 phosphorylation is required for the neural differentiation of mESCs**

RARs are phosphoproteins and RAR $\gamma$ 2 comprises two serine (S) residues (S66 and S68) that are located in a proline-rich motif of



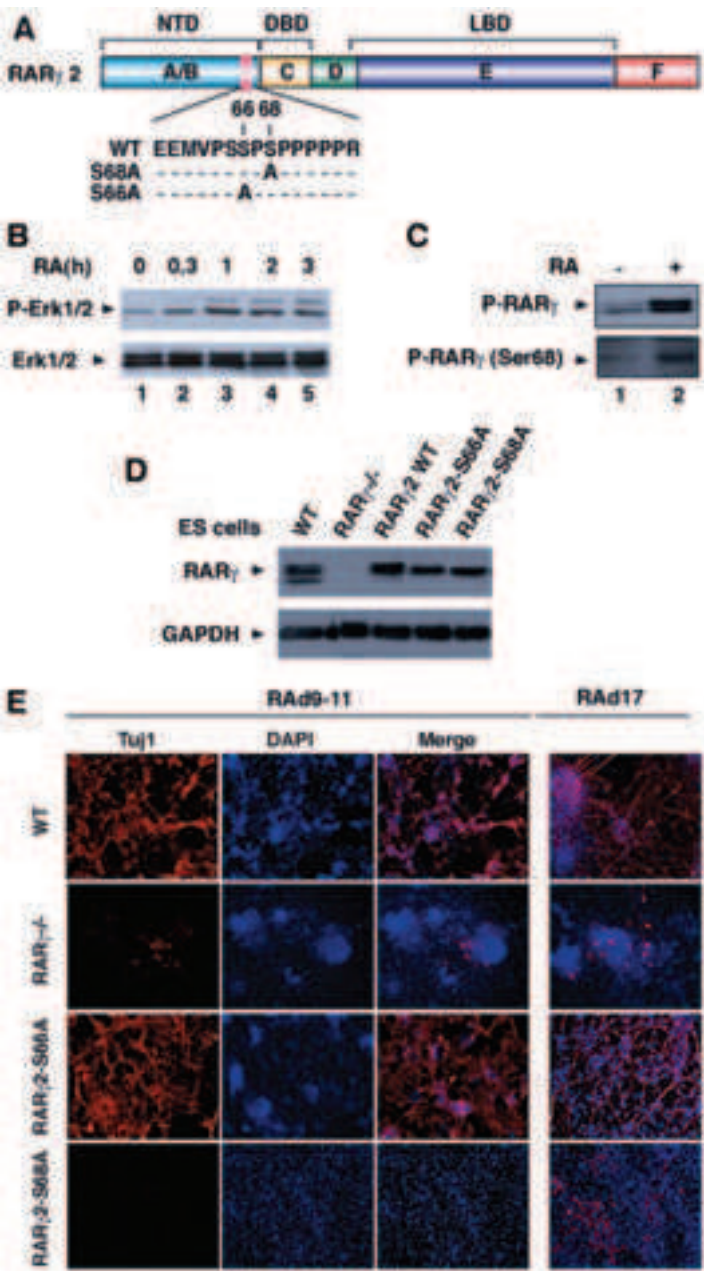
**Fig. 2. The RA-induced neuronal differentiation of mESCs involves RAR $\gamma$ 2.** (A) Immunoblots showing that the RAR $\alpha$  and RAR $\gamma$  subtypes are expressed in mESCs, and that RAR $\beta$ 2 is induced in response to RA. (B) Knockout of RAR $\alpha$  and RAR $\gamma$  in mESCs as assessed by immunoblotting. (C) 24 hours after dissociation and plating of the RA-treated cellular aggregates, *Rarg*<sup>-/-</sup> (RAR $\gamma$ <sup>-/-</sup>) ESCs do not give rise to neuronal progenitors, whereas the *Rara*<sup>-/-</sup> (RAR $\alpha$ <sup>-/-</sup>) counterparts do, as the WT cells. (D) In contrast to the WT and RAR $\alpha$ <sup>-/-</sup> ESCs, the RAR $\gamma$ <sup>-/-</sup> cells do not differentiate into neurons, as assessed by immunofluorescence of Tuj1. Scale bar: 100  $\mu$ m. (E) mESCs express the RAR $\gamma$ 1 and RAR $\gamma$ 2 isoforms as assessed by immunoblotting after immunoprecipitation with antibodies recognizing specifically RAR $\gamma$ 1 (Ab1 $\gamma$ ) or RAR $\gamma$ 2 (Ab10 $\gamma$ ). (F) Immunoblots showing the expression of RAR $\gamma$ 1 and RAR $\gamma$ 2 in the rescue cells. (G) The RAR $\gamma$ 2 but not the RAR $\gamma$ 1 rescue cells become neuronal progenitors 24 hours after dissociation and plating of the RA-treated cellular aggregates. (H) The RAR $\gamma$ 2 but not the RAR $\gamma$ 1 rescue cell lines form a dense axon network by 7–9 days, as assessed by immunofluorescence of Tuj1. Scale bar: 100  $\mu$ m.

the NTD and are substrates for phosphorylation (Bastien et al., 2000) (Fig. 3A). We have previously shown that RARs become rapidly phosphorylated at one of these serine residues (S68 in RAR $\gamma$ 2) subsequent to a kinase cascade initiated by MAPKs after RA treatment (Bruck et al., 2009; Lalevée et al., 2010). In mESCs, Erk1/2 (MAPK3 and MAPK1, respectively) were rapidly (within minutes) activated after RA addition (Fig. 3B). The amount of RAR $\gamma$ 2 phosphorylated at S68 also rapidly increased, as assessed by phosphoprotein affinity purification followed by immunoblotting with antibodies specifically recognizing RAR $\gamma$  phosphorylated at this residue (Fig. 3C).

To investigate whether RAR $\gamma$ 2 phosphorylation plays a role in mESC neuronal differentiation, we generated additional stable rescue lines expressing RAR $\gamma$ 2 with S68 replaced with an alanine

(RAR $\gamma$ 2S68A) in the *Rarg*<sup>-/-</sup> background. Rescue lines expressing RAR $\gamma$ 2 with the other serine residue replaced with an alanine were also established (RAR $\gamma$ 2S66A) as a control. Biophysical studies (nuclear magnetic resonance, circular dichroism and small-angle X-ray scattering) coupled to molecular dynamics simulations performed with synthetic model peptides indicated that replacement of S68 (and S66) by an alanine residue did not affect the global hydrodynamic behaviour of the peptide (B. Kieffer, A. Dejaegere and C.R.E., unpublished data).

Several clones expressing the mutated receptors at levels similar to the endogenous were obtained (Fig. 3D; supplementary material Fig. S1B) and analysed for their differentiation potential. Remarkably, the RAR $\gamma$ 2S68A rescue cell lines did not



**Fig. 3. RAR $\gamma$ 2 phosphorylation at S68 is required for the neuronal differentiation of mESCs.** (A) Schematic representation of the phosphorylation sites located in the NTD of RAR $\gamma$ 2 and of the corresponding mutants. (B) In response to RA, Erk1/2 is rapidly activated as assessed by immunoblotting with antibodies recognizing the active phosphorylated forms (P-Erk1/2). (C) RAR $\gamma$ 2 becomes rapidly phosphorylated at S68 as assessed by immunoblotting after phosphoaffinity purification. (D) Immunoblotting analysis of the expression of WT RAR $\gamma$ 2 and of the phosphomutants in the rescue cells. (E) The RAR $\gamma$ 2S68A rescue line does not form neurons after RA treatment up to 17 days, whereas the RAR $\gamma$ 2S66A rescue line differentiates as efficiently as the WT cells.

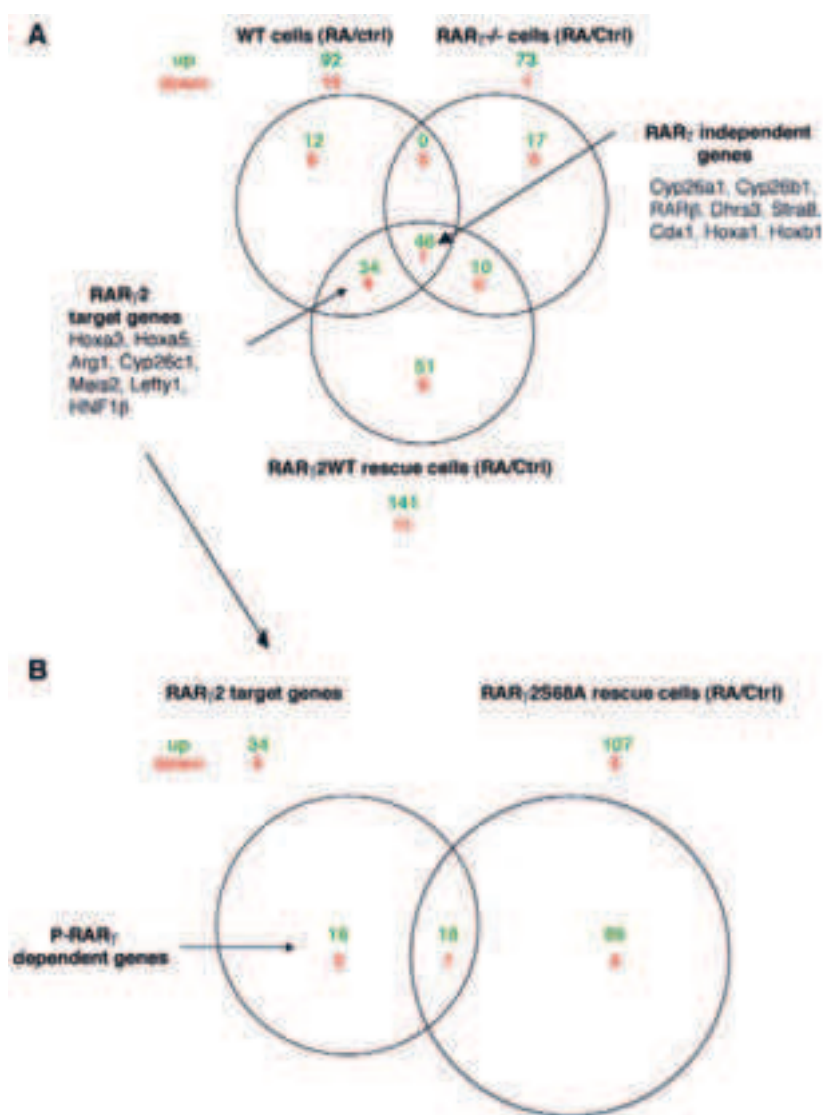
differentiate into neurons. No axon structures were obtained by 7–9 days and even up to 17 days (Fig. 3E) after RA addition. In contrast, the RAR $\gamma$ 2S66A rescue cell line was able to form a dense axon network comparable to the RAR $\gamma$ 2WT rescue line (Fig. 3E). Collectively, these results indicate that RAR $\gamma$ 2 phosphorylation at S68 plays an important role in mediating RA signalling.

**Genome-wide analysis of the early RA target genes that are regulated by RAR $\gamma$ 2**

Then the question was whether the RA-induced neuronal differentiation of mESCs correlates with the activation and/or repression of a subset of genes. Therefore, we profiled the genes that were regulated by RAR $\gamma$ 2 by comparing the WT mESCs to the *Rarg*<sup>-/-</sup> and RAR $\gamma$ 2WT cell lines in the absence and

presence of RA. RNA-seq was performed on cellular aggregates obtained from each cell line with or without a RA treatment of 2 hours, in order to identify the primary (early) RAR $\gamma$ 2 target genes involved in the commitment to the neuronal lineage. Transcripts whose expression was induced or repressed by RA (experiments designated RA/Ctrl) were then assessed (supplementary material Table S1).

The list of genes that were upregulated by RA in the WT mESCs was in agreement with other studies (Mahony et al., 2011; Moutier et al., 2012; Simandi et al., 2010) and included the gene encoding the RAR $\beta$  isoform, the canonical RA target genes involved in RA metabolism (*Cyp26a1*, *Cyp26b1*, *Cyp26c1* and *Dhrs3*), patterning genes exemplified by the Hox genes (*Hoxa1*, *Hoxa3*, *Hoxa5*, *Hoxb1* and *Hoxb4*), other genes encoding homeobox proteins (*Meis2*, *Cdx1*, *Gbx2*, *Dlx3* and *Hnf1b*) and



**Fig. 4. RNA-seq analysis of the RA-regulated genes in WT and *Rarg*<sup>-/-</sup> ESCs, and ESCs expressing RAR $\gamma$ 2WT or RAR $\gamma$ 2 S68A in the *Rarg*<sup>-/-</sup> background.** (A) Venn diagram comparing the genes that are regulated by RA (RA/Ctrl) in mESCs WT, *Rarg*<sup>-/-</sup> (*RAR* $\gamma$ <sup>-/-</sup>) and expressing RAR $\gamma$ 2 in the RAR $\gamma$ -null background (RAR $\gamma$ 2WT rescue cell line). (B) Venn diagram comparing the RAR $\gamma$ 2 target genes selected in A and the genes that are regulated by RA in the RAR $\gamma$ 2S68A rescue cell line.

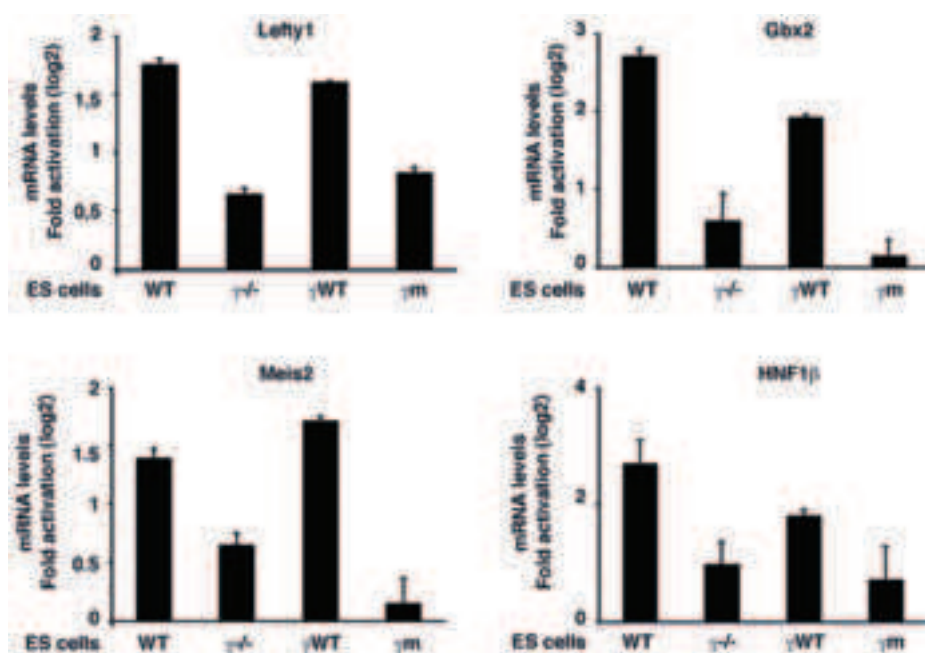
genes with a wide variety of functions such as *Lefty1*, *Arg1* and *Stra8*. Only a few genes were downregulated, and included the *Otx2* gene, also in agreement with other studies.

To assess which of these genes are specifically regulated by RAR $\gamma$ 2 in the presence of RA, the results obtained with the WT cells were cross-referenced with those obtained with the *Rarg*<sup>-/-</sup> cell line and the RAR $\gamma$ 2 WT rescue cell line. The Venn diagram (Fig. 4A) shows that 37% of the genes upregulated in RA-treated WT cells were attenuated in the *Rarg*<sup>-/-</sup> cells and fully restored in the RAR $\gamma$ 2WT rescue line, showing that they are RAR $\gamma$ 2 target genes. These genes are exemplified by the *Meis2*, *Lefty1*, *Hnf1b*, *Arg1*, *Hoxa3* and *Hoxa5* genes, and their regulation by RAR $\gamma$ 2 was corroborated in RT-qPCR experiments (Fig. 5). The *Gbx2* gene can also be added to this list because its activation was significantly decreased (but not below the selected cut-off of the RNA-seq analysis) in the *Rarg*<sup>-/-</sup> cell line and fully restored in the RAR $\gamma$ 2WT rescue cell line (supplementary material Table S1; Fig. 5), indicating that it is a RAR $\gamma$ 2 target.

It is worth noting that 13% of the upregulated genes were also attenuated in the *Rarg*<sup>-/-</sup> cells but not restored in the RAR $\gamma$ 2 rescue line, suggesting that these genes might be regulated by the

RAR $\gamma$ 1 isoform. Given that RAR $\gamma$ 1 does not restore differentiation of mESCs, these genes were considered not to be essential in our model. Finally, 50% of the upregulated genes were not grossly affected in absence of RAR $\gamma$ , suggesting that they are also not essential for the differentiation of mESCs. Among these genes are the canonical RA target genes involved in RA metabolism (*Cyp26a1*, *Cyp26b1* and *Dhrs3*). There is also the gene encoding the RAR $\beta$ 2 isoform, corroborating our first hypothesis that its expression is not essential for neuronal differentiation of ESCs. Remarkably, additional studies performed with mESCs lacking RAR $\alpha$  indicated that the activation of RAR $\beta$ 2 is mediated by the RAR $\alpha$  subtype (supplementary material Fig. S2). In contrast, activation of *Cyp26a1* involves the two receptors (supplementary material Fig. S2) and, hence, RAR $\alpha$  might functionally compensate for RAR $\gamma$  loss. Interestingly, expression of several genes was upregulated in the *Rarg*<sup>-/-</sup> cells compared to WT cells, suggesting a repressive effect of RAR $\gamma$ . Several others were activated only in the RAR $\gamma$ 2 rescue cell line, suggesting that in the WT cells, the function of RAR $\gamma$ 2 might be antagonized by RAR $\gamma$ 1.

In conclusion, these experiments revealed a repertoire of genes that are specifically regulated by RAR $\gamma$ 2 in mESCs. It also



**Fig. 5.** RT-qPCR analysis of the different mESC lines for the RA regulation of the genes selected in the RNA-seq experiments. WT mESCs (WT), *Rarg*<sup>-/-</sup> mESCs ( $\gamma^{-/-}$ ), mESCs expressing RAR $\gamma$ 2WT in the RAR $\gamma$ -null background ( $\gamma$ WT) and mESCs expressing RAR $\gamma$ 2S68A in the RAR $\gamma$ -null background ( $\gamma$ m) were assessed for expression of *Lefty1*, *Gbx2*, *Meis* and *Hnf1b* (HNF1 $\beta$ ) mRNAs. Values are the mean  $\pm$  s.e.m. of triplicates from five to seven separate experiments.

highlighted the complexity of the regulation of the other RA target genes.

#### The phosphorylated form of RAR $\gamma$ 2 regulates a subset of genes

Given that RAR $\gamma$ 2 phosphorylation is required for the neuronal differentiation of mESCs, we investigated the genes that are controlled by the phosphorylated form of this receptor. Cellular aggregates obtained from mESCs expressing RAR $\gamma$ 2S68A were treated with RA for 2 hours as above and gene expression assessed by RNA-seq (supplementary material Table S1). The list of genes with altered expression was compared with the list of RAR $\gamma$ 2 target genes generated above. The Venn diagram (Fig. 4B) shows that  $\sim$ 50% of RAR $\gamma$ 2-target genes are not expressed in the RAR $\gamma$ 2S68A rescue cell line, suggesting that their activation is controlled by the phosphorylation of RAR $\gamma$ 2. These genes include *Lefty1* (left right determination factor 1) and the homeobox gene *Hnf1b* (hepatocyte nuclear factor 1 homeobox B) and their phospho-RAR $\gamma$ 2 dependency was corroborated in RT-qPCR experiments (Fig. 5). The homeobox genes *Gbx2* (gastrulation brain homeobox 2) and *Meis2* can be added to this list. Indeed *Gbx2* activation, which was decreased in the *Rarg*<sup>-/-</sup> cells (but not abolished), was well restored in the RAR $\gamma$ 2WT rescue line but not in the RAR $\gamma$ 2S68A line. Concerning *Meis2*, the values obtained with the RAR $\gamma$ 2S68A rescue line were at the limit of the threshold, suggesting that this gene might be also an interesting candidate. RT-qPCR experiments corroborated that both *Gbx2* and *Meis2* are regulated by the phosphorylation of RAR $\gamma$ 2 (Fig. 5). Taken together, these results suggest that the phosphoRAR $\gamma$ 2 programme would trigger the neuronal differentiation of mESCs via the activation of a particular set of key genes including at least *Lefty1*, *Gbx2*, *Meis2* and *Hnf1b*.

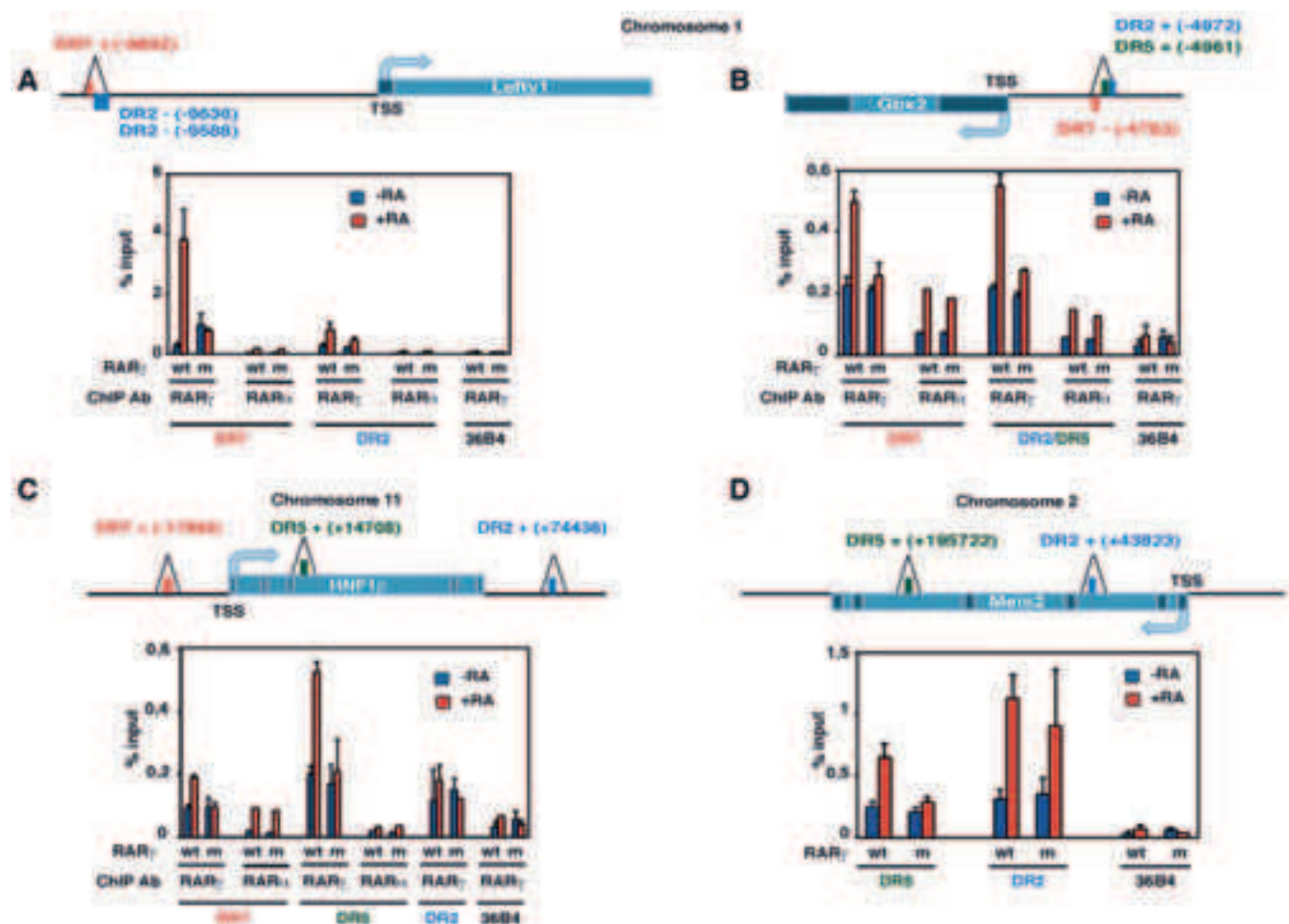
#### Phosphorylation controls the recruitment of RAR $\gamma$ 2 to gene promoters

Next, we asked whether phosphorylation controls the recruitment of RAR $\gamma$ 2 to the promoters of the selected genes. We analysed

the RAR-binding sites mapped in the *Lefty1*, *Meis2*, *Gbx2* and *Hnf1b* genes in ChIP-seq experiments performed with cellular aggregates treated or not with RA for 2 hours and with a pan-RAR antibody (Moutier et al., 2012). We identified one or several major RAR-binding sites that were enriched after RA addition (supplementary material Table S2; Fig. 6). The sequences under these peaks were analysed to detect consensus 5'-RGKTS-3' half sites (with R=A/G, K=G/T and S=C/G, according to the IUPAC convention) with either no (DR0) or 1- to 8-base-pair (DR1–DR8) spacers (Moutier et al., 2012). Then primers specifically surrounding the RAREs were designed and used in ChIP-qPCR experiments performed with cellular aggregates from the RAR $\gamma$ 2WT and RAR $\gamma$ 2S68A rescue cell lines and with our purified specific RAR $\gamma$  antibodies (Lalevée et al., 2010; Mendoza-Parra et al., 2011).

One major RA-regulated RAR-binding site was found in the promoter of the *Lefty1* gene in the  $-10$  kb region flanking the transcription start site (TSS). This site comprises a consensus DR7 element and two DR2s with one mismatch (supplementary material Table S2). Our ChIP experiments revealed that, after RA addition, occupancy of the DR7 sequence was markedly enriched with RAR $\gamma$ 2WT, but not with the phosphomutant (Fig. 6A). Occupancy of the DR2 sequences was also enriched, but less markedly and without significant difference between RAR $\gamma$ 2WT and RAR $\gamma$ 2S68A. ChIP experiments were also performed with RAR $\alpha$  antibodies, using the same chromatin preparations and the same primers. However, none of the sequences were significantly occupied in both the RAR $\gamma$ 2WT and RAR $\gamma$ 2S68A cell lines (Fig. 6A). Collectively, these results indicate that the activation of *Lefty1* requires the binding of the phosphorylated form of RAR $\gamma$ 2 to a DR7 element.

A major RAR-binding site was also identified in the  $-10$  kb region flanking the TSS of the *Gbx2* gene. It contains a DR7, a DR5 and a DR2 element, all with one mismatch (supplementary material Table S2). As for the *Lefty1* gene, occupancy of the DR7 element was enriched with RAR $\gamma$ 2WT, but not with the phosphomutant after RA addition (Fig. 6B). Similar results



**Fig. 6.** ChIP experiments showing that phosphorylation controls the recruitment of RAR $\gamma$ 2 to the promoters of the *Lefty1*, *Gbx2*, *Hnf1b* and *Meis2* genes. Each gene is schematically represented with the location of the RAR-binding peaks identified in ChIP-seq experiments. The location of the RAREs is also indicated. ChIP experiments were performed with RAR $\gamma$ - or RAR $\alpha$ -specific purified antibodies and the RAR-bound DNA fragments were amplified with primers surrounding the different RAREs. wt, wild-type RAR $\gamma$ 2; m, RAR $\gamma$ 2S68A.

were obtained with the sequence encompassing the DR2 and DR5 elements. ChIP experiments performed with RAR $\alpha$  antibodies showed that, in both the RAR $\gamma$ 2WT and RAR $\gamma$ 2S68A cell lines, all elements were also enriched with RAR $\alpha$ , but less markedly (Fig. 6B), explaining at least in part why the activation of the *Gbx2* gene was not fully abrogated upon RAR $\gamma$  invalidation in our RNA-seq experiments.

Concerning the *Hnf1b* gene, three RAR-binding sites were found, one upstream of the TSS with a consensus DR7 sequence and two downstream of the TSS (one intragenic with a DR5 and one downstream of the gene end with a DR2) (supplementary material Table S2; Fig. 6C). Occupancy of the DR2 element was not significantly enriched with RAR $\gamma$ 2 either WT or S68A. However, occupancy of the DR7 and DR5 elements was enriched by RAR $\gamma$ 2WT but not by the phosphomutant after RA addition (Fig. 6C). It is worth noting that, compared to the DR7 element, the DR5 was more markedly enriched by RAR $\gamma$ 2 and was not occupied by RAR $\alpha$ , suggesting that the regulation of the *Hnf1b* gene by the phosphorylated form of RAR $\gamma$ 2 concerns this DR5 RARE. In contrast, the *Meis2* locus comprises two major RAR-binding sites located downstream of the TSS in intragenic regions. One peak contains a consensus DR5 and the other a

consensus DR2 element (supplementary material Table S2; Fig. 6D). Both elements were enriched with RAR $\gamma$ 2WT after RA addition, but only the DR5 element was not enriched with the RAR $\gamma$ 2 phosphomutant (Fig. 6D).

Taken together, these results highlight novel RAR $\gamma$ 2-binding elements with 5- or 7-base-pair spacers, which specifically recruit the phosphorylated form of RAR $\gamma$ 2 in response to RA.

## DISCUSSION

RA-regulated differentiation, proliferation, survival or death is at the basis of complex physiological processes such as development. During the past decade, numerous studies have provided an enormous gain of knowledge into the molecular and structural features of RARs. However, the complexity of the scenarios increased with the discovery that RA also has non-genomic effects and activates several signalling pathways that are integrated in the nucleus and target several proteins, including RARs, for phosphorylation processes (Al Tanoury et al., 2013b). Thus, we are far from understanding how the non-genomic and genomic effects of RA coordinate in embryonic development and organogenesis.

As cell differentiation is one of the most crucial steps during development, we have used mESCs, which can differentiate into

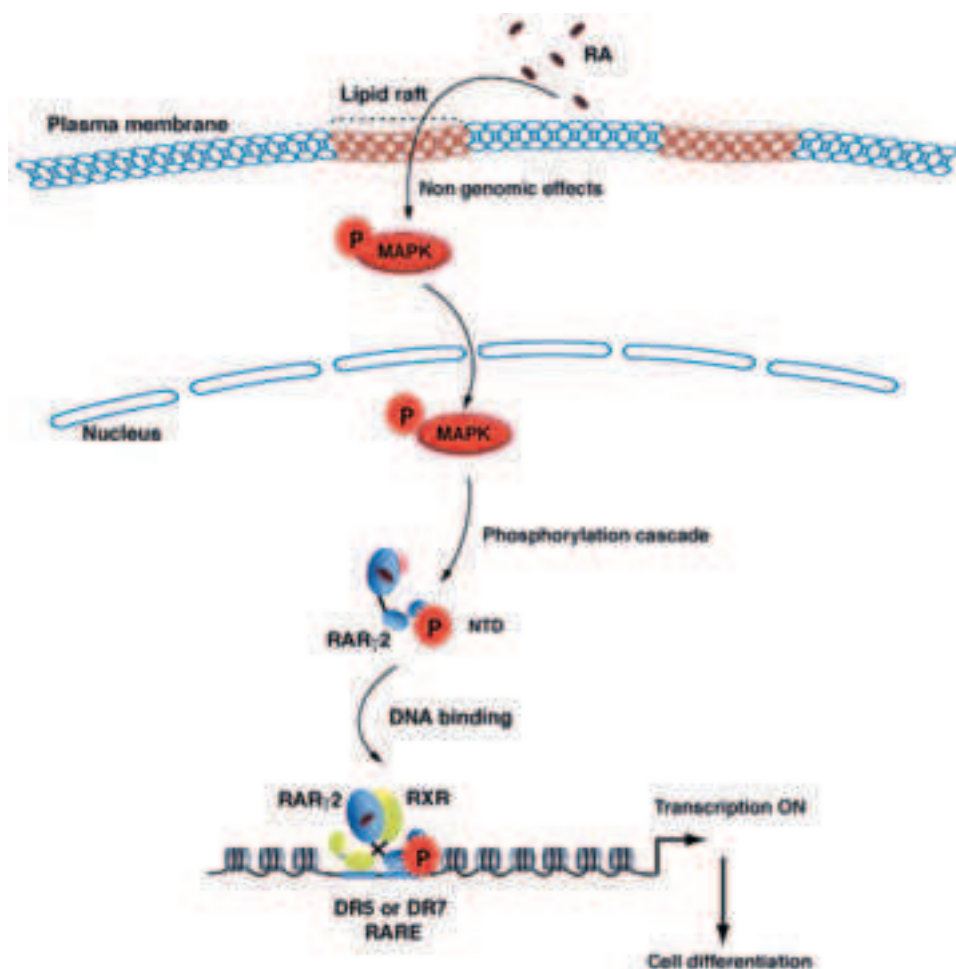
a large variety of cell types, such as neuronal cells, in response to RA. However, whether a particular RAR subtype was involved in the differentiation of these cells was still undefined. Moreover, the relevance of RAR phosphorylation had also not been addressed. Therefore, we investigated which RAR subtype is involved in neuronal differentiation of mESCs and whether their phosphorylation is required. With that aim, we took advantage of mESCs in which the different RARs have been inactivated and used them to generate rescue lines expressing RAR phosphomutants.

The first novel finding of this study is that RAR $\gamma$ , and more precisely the RAR $\gamma$ 2 isoform, which is the predominant RAR $\gamma$  isoform in embryonic stem cells (Kastner et al., 1990), is required for the RA-induced generation of neuronal progenitors, which then give rise to neurons forming a dense axon network. An additional important observation is the identification of a subset of early RA target genes involved in development whose regulation is lost in the RAR $\gamma$ -knockout cells, but well re-established in the RAR $\gamma$ 2 rescue line. Because these genes are activated as early as 2 hours after RA addition to the cell aggregates, one can hypothesize that they are involved in the commitment of ESC to the neuronal lineage. A recent study conducted by Kashyap et al. (Kashyap et al., 2013) has also shown that inactivation of RAR $\gamma$  is associated with a reduced expression of several genes. However, the cells were not analysed

under the same conditions as here and there was no correlation with neuronal differentiation.

The second novel finding is the role of the phosphorylation state of RAR $\gamma$ 2 (Fig. 7). We show that ERKs are rapidly activated after RA addition in mESCs, corroborating other studies (Stavridis et al., 2010). Today, it is known that the rapid activation of the kinase signalling pathways by RA is mediated by non-genomic effects, and that the effects of these kinases are integrated in the nucleus where they phosphorylate several factors involved in the expression of the RA target genes (Al Tanoury et al., 2013b). Here, the important point is that RAR $\gamma$ 2 becomes phosphorylated at its N-terminal serine residue, and that this phosphorylation process is required for the neuronal differentiation of mESCs. Indeed our present data demonstrate that mESCs expressing RAR $\gamma$ 2 mutated at the N-terminal serine residue are not able to enter the neuronal lineage. They also reveal a small set of target genes involved in neuronal development and exemplified by *Lefty1*, *Gbx2*, *Meis2* and *Hnf1b*, which require the phosphorylation of RAR $\gamma$ 2 to be activated. This is the first report showing that RA target genes involved in development are regulated by the phosphorylation of RAR $\gamma$ 2.

The third novel finding is the characterization of RAREs with specific spacings, which recruit the phosphorylated form of RAR $\gamma$ 2 in response to RA. Indeed, we found that the promoters of



**Fig. 7. Working model for the role of RAR $\gamma$ 2 phosphorylation in the expression of RA target genes and neuronal differentiation of mESCs.** In response to RA, the MAPK signalling pathway is activated via non-genomic effects. Then RAR $\gamma$ 2 becomes phosphorylated at the N-terminal serine residue and can be recruited to genes promoters with DR5 or DR7 RAREs. Finally, the corresponding genes are expressed and mESCs differentiate into neurons.

the above genes targeted by phosphoRAR $\gamma$ 2 depict not only canonical DR2 and DR5 response elements but also atypical DR7 elements, and that only the DR5 and DR7 elements recruit specifically the phosphorylated form of RAR $\gamma$ 2 in response to RA. The next challenge will be to determine how phosphorylation of the N-terminal serine residue, which is located in the vicinity of the DBD, regulates the binding of RAR $\gamma$ 2 only to response elements with DR5 or DR7 spacing. However, it is worth pointing out that the spacing of the response elements directs the architecture of the DNA-bound RXR–RAR heterodimers (Br  livet et al., 2012; Rochel et al., 2011). Because phosphorylation of the N-terminal serine residue induces the dissociation of partners with SH3 domains (Lalev  e et al., 2010), preliminary studies have suggested that the interaction of the non-phosphorylated form with SH3 proteins would impede the binding of RAR $\gamma$ 2 only to DRs with specific spacing (DR5 and DR7), whereas the phosphorylated form (without SH3 partners) would be recruited to these DRs in response to RA (B. Kieffer and C.R.-E., unpublished results).

It is important to note that the DR7 elements show a rather low frequency (Moutier et al., 2012) among the RAR-occupied sites, suggesting that they would control the expression of a small set of genes involved in the commitment of pluripotent cells to specific lineages. Here, we have defined some of these genes, i.e. the *Lefty1*, *Gbx2* and *Hnf1b* genes, and we show that they belong to an early phosphoRAR $\gamma$ 2-regulated gene programme and are crucial for the loss of pluripotency and for triggering the neuronal differentiation of mESCs. Remarkably, comparison of our findings with those reported for other cell types highlights that *Lefty1* and *Gbx2*, which are also known as the ‘stimulated by RA’ *Stra3* and *Stra7* genes, respectively (Chazaud et al., 1996; Oulad-Abdelghani et al., 1998), are involved in early neural development (Li et al., 2009; Smith et al., 2008; Sunmonu et al., 2011) and, thus, are also activated in the P19 cell line that differentiates into neurons in response to RA. However, they are absent from the lists of RA-activated genes in cells with other features, such as mouse embryo carcinoma cells (F9 cell line) (Lalev  e et al., 2011; Mendoza-Parra et al., 2011; Su and Gudas, 2008), MCF7 cells (Hua et al., 2009) and mouse embryonic fibroblasts (Al Tanoury et al., 2013a). Collectively such data highlight the importance of these genes for commitment to the neuronal lineage. Concerning the *Hnf1b* and *Meis2* (also known as *Stra10*; Oulad-Abdelghani et al., 1997) genes, they are also involved in neuronal differentiation, but are not restricted to this cell type because they are expressed in F9 cells and other developing tissues (Mendoza-Parra et al., 2011; Oulad-Abdelghani et al., 1997). Nevertheless, all these data suggest that RAR $\gamma$ 2 phosphorylation would control neuronal differentiation via the activation of target genes with specific DR7 and/or DR5 element combinations.

In attempts to validate the role of the early phosphoRAR $\gamma$ 2-regulated genes in the neuronal differentiation of mESCs, we monitored the effects of shRNA-mediated *Gbx2* and *Meis2* knockdown on the appearance of axons after RA treatment. In fact, despite the efficiency of the knockdown, the neuronal differentiation of the cells was not significantly affected. Thus knockdown of one of these transcription factors alone is not sufficient to abolish the differentiation programme. Therefore, either the residual levels of expression are sufficient to sustain differentiation, or the genes act together to establish a sub-programme of interconnected regulatory networks that are not strongly perturbed by the loss of only one of these genes. Further

studies involving simultaneous knockdown of several or all of these genes will be required to answer this question. It is also important to note that almost 50% of the RA-induced genes are still normally activated in the RAR $\gamma$ -knockout cells although these cells do not differentiate into neurons. This again highlights the crucial role played by the small subset of genes specifically induced by the RAR $\gamma$ 2 isoform.

In conclusion, we provide several lines of evidence for a function of RAR $\gamma$ 2 phosphorylation in positively regulating neuronal differentiation (Fig. 7). It would be interesting to extrapolate this work to other RA responsive systems. Ultimately, one can predict specific sets of genes with DR5 and/or DR7 elements that could be controlled by phosphorylation of RARs, depending on the feature of the cells.

## MATERIALS AND METHODS

### Plasmids

All constructs containing the RAR $\gamma$  receptors were cloned into the pCA vector, which is driven by a CAG early promoter, coupled to hygromycin/neomycin resistance. mRAR $\gamma$ 2 and mRAR $\gamma$ 1 were isolated as *XhoI/BamHI* fragments from the pSG5 constructs (Bastien et al., 2000) and subcloned in the same sites of pCA. mRAR $\gamma$ 2S66A and mRAR $\gamma$ 2S68A in pCA were constructed by PCR amplification reactions. Internal oligonucleotides used in the PCR reaction encoded alanine (A) instead of serine (S) residues at positions 66 or 68.

### Antibodies

Rabbit polyclonal antibodies recognizing RAR $\alpha$  [RP $\alpha$ (F)], RAR $\beta$  [RP $\beta$ (F)] and RAR $\gamma$  [RP $\gamma$ (F)] were as described previously (Bastien et al., 2000; Bruck et al., 2009; Rochette-Egly et al., 1992). RP $\gamma$ (F) was purified on sulfoLink gel columns (Pierce Chemical) coupled to the corresponding immunizing peptide (Lalev  e et al., 2010). Mouse monoclonal antibodies specifically recognizing RAR $\gamma$  phosphorylated at position S68 were as previously described (Lalev  e et al., 2010), as were mouse monoclonal antibodies recognizing specifically the RAR $\gamma$ 1 [Ab1 $\gamma$ (A1)] and RAR $\gamma$ 2 [Ab10 $\gamma$ (A2)] isoforms (Bastien et al., 2000). Rabbit polyclonal antibodies against GAPDH were from Sigma-Aldrich and those against Pax6, Oct4 and Nanog were from Abcam Ltd. Mouse monoclonal antibodies recognizing neuronal class III  $\beta$ -tubulin (Tuj1) were from Eurogentec France. The antibodies against ERKs and their active phosphorylated forms were from Santa Cruz Biotechnology.

### Mouse ESC lines and culture conditions

Mouse ESCs (clone D4), derived from the 129 sub-strain were as previously described, as were the *Rarg*<sup>−/−</sup> (clone AA71) and *Rara*<sup>−/−</sup> (clone KC25) cell lines (Lohnes et al., 1993; Lufkin et al., 1993). To establish the rescue lines, the RAR $\gamma$ 2WT, RAR $\gamma$ 2S66A or RAR $\gamma$ 2S68A constructs were introduced into the *Rarg*<sup>−/−</sup> cells by Lipofectamine<sup>®</sup> 2000 transfection (Invitrogen). The stable rescue lines were selected with G418 (350  $\mu$ g/ml) or hygromycin B (100  $\mu$ g/ml) for 1 week and analysed for the presence of the transgene by qPCR and western blotting.

All cell lines were grown on inactivated mouse embryonic fibroblast feeder cells in the presence of LIF under standard conditions [DMEM supplemented with Glut Amax<sup>TM</sup>-I (Fischer Scientific SAS), 15% FCS, non essential amino acids and  $\beta$ -mercaptoethanol]. Then cells were differentiated into neurons according to the protocol of Bibel et al. (Bibel et al., 2007). In brief, cells were trypsinized, plated on non-adhesive bacteriological Greiner Petri dishes (4 $\times$ 10<sup>6</sup> cells) in 15 ml CA (cellular aggregates) medium (DMEM supplemented with GlutaMAX<sup>TM</sup>-I, 10% FCS, non essential amino acids and  $\beta$ -mercaptoethanol) and aggregated for 8 days with a medium change every 2 days. At day four, all-trans RA (2  $\mu$ M) (Sigma-Aldrich Chimie SARL) was added and 4 days later, the cellular aggregates were washed with PBS, dissociated with trypsin, suspended in N2 medium [DMEM/Ham-F12, BSA and N2<sup>®</sup> (Fischer Scientific SAS)], and plated on culture dishes precoated with PORN (Poly-DL-ornithine hydrobromide, Sigma-Aldrich, Chimie SARL) and laminin (Roche Diagnostics). After 1 day, the N2 medium was changed

and 2 days later replaced by neurobasal medium supplemented with B27® (Fisher Scientific).

### Immunoblotting, immunoprecipitation and immunofluorescence assays

Extract preparation and immunoblotting were as described previously (Bour et al., 2005). Immunoprecipitation was performed with mouse monoclonal antibodies immobilized on Dynabeads® Protein A/G (Invitrogen). For immunofluorescence assays, cells were grown on Lab-Tek® glass chamber slides (Thermoscientific), fixed in 4% formaldehyde (PFA)-PBS (20 min), permeabilized with 0.1% Triton X-100 (15 min) and blocked with 3% non-immune serum in PBS (30 min). Then, cells were incubated with primary antibodies, followed by Alexa-Fluor-448- or Alexa-Fluor-555-conjugated secondary antibodies (Invitrogen). Nuclei were counterstained with DAPI (Sigma-Aldrich). Cells were analysed by fluorescence microscopy using a LEICA DMRX microscope equipped with a LEICA True Confocal Scanner TCS SP.

### Detection of phosphorylated RAR $\gamma$

Cell extracts were applied to phosphoprotein purification columns (Qiagen). Column eluates containing protein peaks were concentrated and analysed by immunoblotting as previously described (Bruck et al., 2009).

### RNA extraction and RT-qPCR

Total RNA was extracted from cellular aggregates grown for 4 days in the absence of RA and then treated with RA for 2–6 hours. Aliquots were subjected to RT-qPCR as described (Bruck et al., 2009). Transcripts were normalized according to the housekeeping gene *GAPDH*. Primer amplification and specificity were verified on DNA serial dilutions. Primer sequences are available from the corresponding author upon request.

### High-throughput mRNA sequencing

After isolation of total RNA, a library of template molecules suitable for high-throughput DNA sequencing (RNA-Seq) was created following the Illumina ‘Truseq RNA sample prep v2’ protocol with some modifications. Briefly, mRNA was purified from 2  $\mu$ g total RNA using oligo-dT magnetic beads and fragmented using divalent cations (94°C, 8 minutes). The cleaved mRNA fragments were reverse-transcribed to cDNA using random primers and then the second cDNA strand was synthesized using Polymerase I and RNase H. The next steps of RNA-Seq library preparation were performed in a fully automated system using SPRIworks Fragment Library System I kit (ref A84803, Beckman Coulter, Inc.) with the SPRI-TE instrument (Beckman Coulter, Inc.). Briefly, in this system, double-stranded cDNA fragments were blunted, phosphorylated and ligated to indexed adapter dimers, and fragments in the range of ~200–400 bp were selected. Finally, the library was amplified by PCR [30 s at 98°C (10 s at 98°C, 30 s at 60°C, 30 s at 72°C)  $\times$  12 cycles; 5 min at 72°C] and the surplus PCR primers were removed using AMPure beads (Agencourt Biosciences Corporation) with the Biomek 3000 instrument (Beckman Coulter, Inc.). DNA libraries were checked for quality and quantified using 2100 Bioanalyzer (Agilent). The libraries were loaded in the flow cell at 11 pM concentration and clusters were generated and sequenced in the Illumina HiSeq2000 as single-end 50-base reads.

Image analysis and base calling were performed using CASAVA v1.8.2 sequence reads were mapped onto the mm9 assembly of the mouse genome by using Tophat v1.4.1 (Trapnell et al., 2009) and the bowtie v0.12.7 aligner. Only uniquely aligned reads have been retained for further analyses. Gene expression was quantified using HTSeq v0.5.3p3 (Anders and Huber, 2010) and gene annotations from Ensembl release 66. Read counts were normalized across libraries with the method proposed by Anders and Huber (Anders and Huber, 2010). Comparisons of interest were performed using the statistical method proposed by Anders and Huber (Anders and Huber, 2010) implemented in the DESeq v1.6.1 Bioconductor package. *P*-values are adjusted for multiple testing by using the Benjamini and Hochberg (Benjamini and Hochberg, 1995) method. Only genes with  $|\log_2 \text{fold-change}| > 1$  or  $< -1$  and an adjusted

*P* < 0.05 were considered. Functional analyses of these genes were performed using the Manteia program (<http://manteia.igbmc.fr>).

### Motif research

The gene regions located  $\pm 10$  kb from gene limits (Ensembl release 63) were analysed using regular expression search to detect perfect consensus 5'-RGKTS-3' half sites with different spacing. The potential RAR-binding elements were aligned on the same strand to ensure the sense and antisense matches gave homogeneous positions.

### Chromatin immunoprecipitation experiments

Cellular aggregates were treated with RA for 45 minutes and chromatin immunoprecipitation (ChIP) experiments were performed as previously described (Bruck et al., 2009). Control ChIP were performed without antibodies, and RAR $\gamma$  was immunoprecipitated with purified RP $\gamma$ (F) immobilized on Dynabeads® Protein A (Invitrogen). RAR $\alpha$  was also immunoprecipitated as previously described (Bruck et al., 2009). Immunoprecipitated DNA was amplified by PCR primers designed using Primer3 software (Rozen and Skaletsky, 2000), which are available upon request. Occupancy of the promoters was calculated by normalizing the PCR signals from the immunoprecipitated samples to the signals obtained from the input DNA.

### Acknowledgements

We are grateful to Jean-Marie Garnier for the RAR $\gamma$  constructs and to M. Oulad Abdelghani (IGBMC) for the mouse monoclonal antibodies. Special thanks to Marie Hestin, Regis Lutz and the cell culture facilities for help.

### Competing interests

The authors declare no competing interests.

### Author contributions

Z.A.T. and A.P. devised and undertook all experimental work and analyzed the data. A.D. generated the KO cell lines. S.G. performed the RT-qPCR and ChIP experiments. S.U. and I.D. performed and analyzed the ChIP-seq experiments. B.J. performed the RNA-seq experiments; T.Y. and C.K. performed the bioinformatic analysis of the results. C.R.E. analyzed the data and wrote the paper.

### Funding

This work was supported by the Centre national de la recherche scientifique (CNRS); the Institut national de la santé et de la recherche médicale (INSERM); the Agence Nationale pour la Recherche [grant numbers ANR-05-BLAN-0390-02, ANR-09-BLAN-0297-01, ANR-SVS8-11-Rarescales]; the Association pour la recherche sur le Cancer [grant numbers ARC-07-1-3169 and SL220110603474]; the Fondation pour la Recherche Médicale (FRM) [grant number DEQ20090515423], and the Institut National du Cancer [grant numbers INCa-PL09-194, PL07-96099]. A.P. was supported by FRM and the Lady TATA Memorial Trust, Z.A.T. by INCA, and S.U. by the Ministère de la Recherche. I.D. is an ‘équipe labélisée’ of the Ligue Nationale contre le Cancer. The Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) sequencing facility is a member of the ‘France Génomique’ consortium [grant number ANR10-INBS-09-08].

### Supplementary material

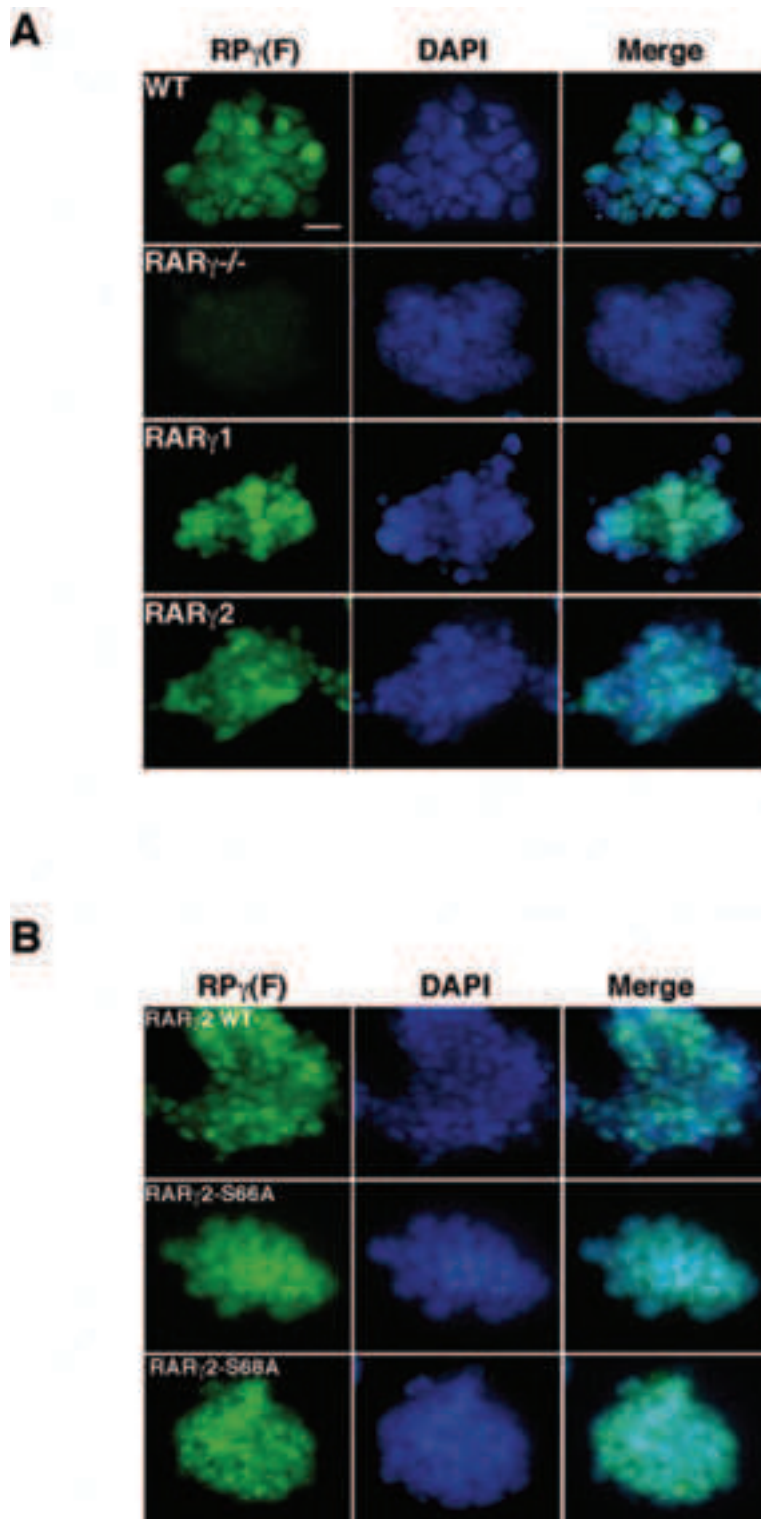
Supplementary material available online at <http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.145979/-/DC1>

### References

- Al Tanoury, Z., Piskunov, A. and Rochette-Egly, C. (2013). Vitamin A and retinoid signaling: genomic and nongenomic effects. *J. Lipid Res.* **54**, 1761–1775.
- Al Tanoury, Z., Piskunov, A., Andriamorasiresy, D., Gaouar, S., Lutz, R., Ye, T., Jost, B., Keime, C. and Rochette-Egly, C. (2014). Genes involved in cell adhesion and signaling: A new repertoire of Retinoic Acid Receptors target genes in mouse embryonic fibroblasts. *J. Cell Sci.* **127**, 521–533.
- Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106.
- Bastien, J. and Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **328**, 1–16.
- Bastien, J., Adam-Stitah, S., Riedl, T., Egly, J. M., Chambon, P. and Rochette-Egly, C. (2000). TFIIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J. Biol. Chem.* **275**, 21896–21904.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57B**, 289–300.

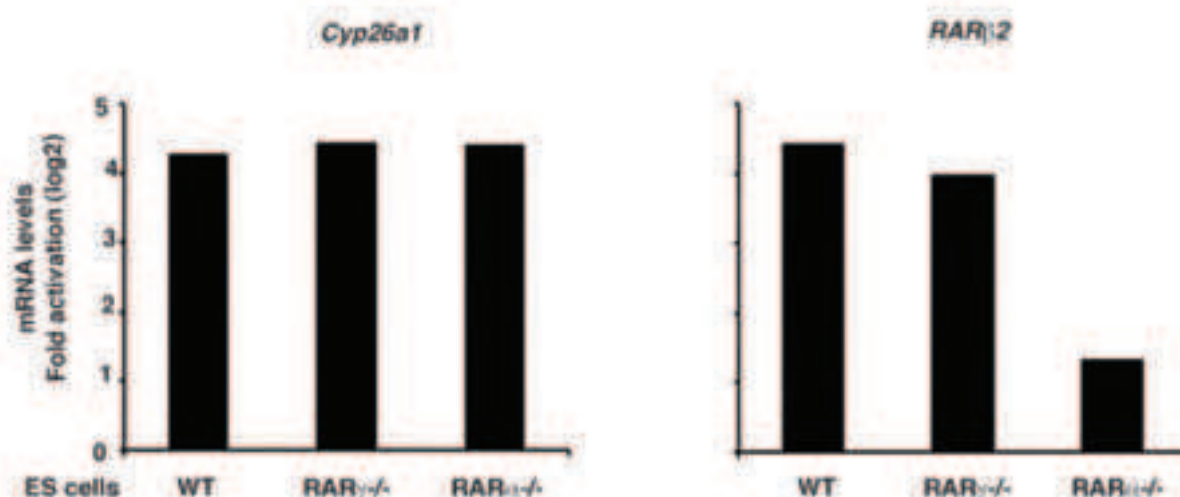
- Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M. and Barde, Y. A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**, 1003–1009.
- Bibel, M., Richter, J., Lacroix, E. and Barde, Y. A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* **2**, 1034–1043.
- Bour, G., Plassat, J. L., Bauer, A., Lalevée, S. and Rochette-Egly, C. (2005). Vinexin beta interacts with the non-phosphorylated AF-1 domain of retinoid receptor gamma (RARgamma) and represses RARgamma-mediated transcription. *J. Biol. Chem.* **280**, 17027–17037.
- Brélivet, Y., Rochel, N. and Moras, D. (2012). Structural analysis of nuclear receptors: from isolated domains to integral proteins. *Mol. Cell. Endocrinol.* **348**, 466–473.
- Bruck, N., Vitoux, D., Ferry, C., Duong, V., Bauer, A., de Thé, H. and Rochette-Egly, C. (2009). A coordinated phosphorylation cascade initiated by p38MAPK/MSK1 directs RARalpha to target promoters. *EMBO J.* **28**, 34–47.
- Chambon, P. (1996). A decade of molecular biology of retinoic acid receptors. *FASEB J.* **10**, 940–954.
- Chazaud, C., Bouillet, P., Oulad-Abdelghani, M. and Dollé, P. (1996). Restricted expression of a novel retinoic acid responsive gene during limb bud dorsoventral patterning and endochondral ossification. *Dev. Genet.* **19**, 66–73.
- Germain, P., Chambon, P., Eichele, G., Evans, R. M., Lazar, M. A., Leid, M., De Lera, A. R., Lotan, R., Mangelsdorf, D. J. and Gronemeyer, H. (2006). International Union of Pharmacology. LXIII. Retinoid X receptors. *Pharmacol. Rev.* **58**, 760–772.
- Gudas, L. J. and Wagner, J. A. (2011). Retinoids regulate stem cell differentiation. *J. Cell. Physiol.* **226**, 322–330.
- Hua, S., Kittler, R. and White, K. P. (2009). Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137**, 1259–1271.
- Kashyap, V., Laursen, K. B., Brenet, F., Viale, A. J., Scandura, J. M. and Gudas, L. J. (2013). RARγ is essential for retinoic acid induced chromatin remodeling and transcriptional activation in embryonic stem cells. *J. Cell Sci.* **126**, 999–1008.
- Kastner, P., Krust, A., Mendelsohn, C., Garnier, J. M., Zelent, A., Leroy, P., Staub, A. and Chambon, P. (1990). Murine isoforms of retinoic acid receptor gamma with specific patterns of expression. *Proc. Natl. Acad. Sci. USA* **87**, 2700–2704.
- Lalevée, S., Bour, G., Quinternet, M., Samarut, E., Kessler, P., Vitorino, M., Bruck, N., Delsuc, M. A., Vonesch, J. L., Kieffer, B. et al. (2010). Vinexinβ, an atypical “sensor” of retinoic acid receptor gamma signaling: union and sequestration, separation, and phosphorylation. *FASEB J.* **24**, 4523–4534.
- Lalevée, S., Anno, Y. N., Chatagnon, A., Samarut, E., Poch, O., Laudet, V., Benoit, G., Lecompte, O. and Rochette-Egly, C. (2011). Genome-wide in silico identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp). *J. Biol. Chem.* **286**, 33322–33334.
- Li, B., Kuriyama, S., Moreno, M. and Mayor, R. (2009). The posteriorizing gene Gbx2 is a direct target of Wnt signalling and the earliest factor in neural crest induction. *Development* **136**, 3267–3278.
- Lohnes, D., Kastner, P., Dierich, A., Mark, M., LeMour, M. and Chambon, P. (1993). Function of retinoic acid receptor gamma in the mouse. *Cell* **73**, 643–658.
- Lufkin, T., Lohnes, D., Mark, M., Dierich, A., Gorry, P., Gaub, M. P., LeMour, M. and Chambon, P. (1993). High postnatal lethality and testis degeneration in retinoic acid receptor alpha mutant mice. *Proc. Natl. Acad. Sci. USA* **90**, 7225–7229.
- Mahony, S., Mazzoni, E. O., McCuine, S., Young, R. A., Wichterle, H. and Gifford, D. K. (2011). Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12**, R2.
- Mendoza-Parra, M. A., Walia, M., Sankar, M. and Gronemeyer, H. (2011). Dissecting the retinoid-induced differentiation of F9 embryonal stem cells by integrative genomics. *Mol. Syst. Biol.* **7**, 538.
- Moutier, E., Ye, T., Choukrallah, M. A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D. et al. (2012). Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *J. Biol. Chem.* **287**, 26328–26341.
- Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P. and Dollé, P. (1997). Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. *Dev. Dyn.* **210**, 173–183.
- Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Mattei, M. G., Dollé, P. and Chambon, P. (1998). Stra3/lefty, a retinoic acid-inducible novel member of the transforming growth factor-beta superfamily. *Int. J. Dev. Biol.* **42**, 23–32.
- Piskunov, A. and Rochette-Egly, C. (2012). A retinoic acid receptor RARα pool present in membrane lipid rafts forms complexes with G protein αQ to activate p38MAPK. *Oncogene* **31**, 3333–3345.
- Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Ittis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y. et al. (2011). Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. *Nat. Struct. Mol. Biol.* **18**, 564–570.
- Rochette-Egly, C. (2003). Nuclear receptors: integration of multiple signalling pathways through phosphorylation. *Cell. Signal.* **15**, 355–366.
- Rochette-Egly, C. and Germain, P. (2009). Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7**, e005.
- Rochette-Egly, C., Gaub, M. P., Lutz, Y., Ali, S., Scheuer, I. and Chambon, P. (1992). Retinoic acid receptor-beta: immunodetection and phosphorylation on tyrosine residues. *Mol. Endocrinol.* **6**, 2197–2209.
- Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J. M. and Chambon, P. (1997). Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIID and phosphorylation by CDK7. *Cell* **90**, 97–107.
- Rosenfeld, M. G., Lunyak, V. V. and Glass, C. K. (2006). Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.* **20**, 1405–1428.
- Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386.
- Samarut, E. and Rochette-Egly, C. (2012). Nuclear retinoic acid receptors: conductors of the retinoic acid symphony during development. *Mol. Cell. Endocrinol.* **348**, 348–360.
- Samarut, E., Amal, I., Markov, G. V., Stote, R., Dejaegere, A., Laudet, V. and Rochette-Egly, C. (2011). Evolution of nuclear retinoic acid receptor alpha (RARα) phosphorylation sites. Serine gain provides fine-tuned regulation. *Mol. Biol. Evol.* **28**, 2125–2137.
- Simandi, Z., Balint, B. L., Poliska, S., Ruhl, R. and Nagy, L. (2010). Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. *FEBS Lett.* **584**, 3123–3130.
- Smith, J. R., Vallier, L., Lupo, G., Alexander, M., Harris, W. A. and Pedersen, R. A. (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* **313**, 107–117.
- Stavridis, M. P., Collins, B. J. and Storey, K. G. (2010). Retinoic acid orchestrates fibroblast growth factor signalling to drive embryonic stem cell differentiation. *Development* **137**, 881–890.
- Su, D. and Gudas, L. J. (2008). Gene expression profiling elucidates a specific role for RARgamma in the retinoic acid-induced differentiation of F9 teratocarcinoma stem cells. *Biochem. Pharmacol.* **75**, 1129–1160.
- Sunmonu, N. A., Li, K., Guo, Q. and Li, J. Y. (2011). Gbx2 and Fgf8 are sequentially required for formation of the midbrain-hindbrain compartment boundary. *Development* **138**, 725–734.
- Trapnell, C., Pachter, L. and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.
- Wilson, V., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* **136**, 1591–1604.

## Supplementary Material



**Figure S1: Immunofluorescence experiments showing the expression of RAR $\gamma$  in the different rescue lines**

A. Expression of RAR $\gamma$ 1 and RAR $\gamma$ 2. B. Expression of the RAR $\gamma$ 2 phosphomutants.



**Figure S2: RA regulation of the *Cyp26a1* and *RARβ2* genes in mES cells WT, *RARγ*<sup>-/-</sup> and *RARα*<sup>-/-</sup>.**

The results have been extracted from the RNA-seq data. Adjusted p values are <0,05.

**Table S1: List of the genes that are regulated by RA in mES cells WT, *RARγ*<sup>-/-</sup> and in the different rescue cell lines.** Ensembl IDs, gene names, descriptions and normalized expression values for transcripts that are induced or repressed by RA in the different cell lines are shown. The log2 change in expression and adjusted p value are also indicated.

[Download Table S1](#)

**Table S2: List of the RAR binding peaks at the *Lefty1*, *Gbx2*, *Meis2* and *HNF1β* genes**

The position of the genes in the genome is indicated as well as the sequence of the RAREs with their location compared to the TSS.

[Download Table S2](#)

# ***Bibliographie***

# Bibliographie

- Aberdam, D., K. Gambaro, et al. (2007). "Embryonic stem cells as a cellular model for neuroectodermal commitment and skin formation." Comptes rendus biologies **330**(6-7): 479-484.
- Akbari, O. S., A. Bousum, et al. (2006). "Unraveling cis-regulatory mechanisms at the abdominal-A and Abdominal-B genes in the Drosophila bithorax complex." Developmental biology **293**(2): 294-304.
- Alazard, R., M. Blaud, et al. (2005). "Identification of the 'NORE' (N-Oct-3 responsive element), a novel structural motif and composite element." Nucleic acids research **33**(5): 1513-1523.
- Albert, T. K., K. Grote, et al. (2010). "Basal core promoters control the equilibrium between negative cofactor 2 and preinitiation complexes in human cells." Genome biology **11**(3): R33.
- Alison, M. R., S. Islam, et al. (2009). "Cell therapy for liver disease." Current opinion in molecular therapeutics **11**(4): 364-374.
- Andersen, B. and M. G. Rosenfeld (2001). "POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease." Endocrine reviews **22**(1): 2-35.
- Andrews, P. W. (2002). "From teratocarcinomas to embryonic stem cells." Philosophical transactions of the Royal Society of London. Series B, Biological sciences **357**(1420): 405-417.
- Armache, K. J., H. Kettenberger, et al. (2003). "Architecture of initiation-competent 12-subunit RNA polymerase II." Proceedings of the National Academy of Sciences of the United States of America **100**(12): 6964-6968.
- Aruga, J., A. Kamiya, et al. (2006). "A wide-range phylogenetic analysis of Zic proteins: implications for correlations between protein structure conservation and body plan complexity." Genomics **87**(6): 783-792.
- Aruga, J., O. Minowa, et al. (1998). "Mouse Zic1 is involved in cerebellar development." The Journal of neuroscience : the official journal of the Society for Neuroscience **18**(1): 284-293.
- Avilion, A. A., S. K. Nicolis, et al. (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes & development **17**(1): 126-140.
- Bailey, P. J., J. M. Klos, et al. (2006). "A global genomic transcriptional code associated with CNS-expressed genes." Experimental cell research **312**(16): 3108-3119.
- Bain, D. L., Aaron F. Heneghan, et al. (2007). "Nuclear Receptor Structure: Implications for Function." Annual Review of Physiology **69**: 201-220.
- Bain, D. L., A. F. Heneghan, et al. (2007). "Nuclear receptor structure: implications for function." Annual review of physiology **69**: 201-220.
- Bain, G., D. Kitchens, et al. (1995). "Embryonic stem cells express neuronal properties in vitro." Developmental biology **168**(2): 342-357.
- Baranek, C., E. Sock, et al. (2005). "The POU protein Oct-6 is a nucleocytoplasmic shuttling protein." Nucleic acids research **33**(19): 6277-6286.
- Bastien, J., S. Adam-Stitah, et al. (2000). "TFIIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7." The Journal of biological chemistry **275**(29): 21896-21904.
- Bastien, J. and C. Rochette-Egly (2004). "Nuclear retinoid receptors and the transcription of retinoid-target genes." Gene **328**.
- Becker, K. A., P. N. Ghule, et al. (2006). "Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase." Journal of cellular physiology **209**(3): 883-893.
- Bell, A. C., A. G. West, et al. (2001). "Insulators and boundaries: versatile regulatory elements in the eukaryotic genome." Science **291**(5503): 447-450.

- Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**: 407-412.
- Berlin, I., L. Denat, et al. (2012). "Phosphorylation of BRN2 modulates its interaction with the Pax3 promoter to control melanocyte migration and proliferation." Molecular and cellular biology **32**(7): 1237-1247.
- Berman, B. P., Y. Nibu, et al. (2002). "Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome." Proceedings of the National Academy of Sciences of the United States of America **99**(2): 757-762.
- Bernstein, B. E., M. Kamal, et al. (2005). "Genomic maps and comparative analysis of histone modifications in human and mouse." Cell **120**(2): 169-181.
- Bertrand, N., D. S. Castro, et al. (2002). "Proneural genes and the specification of neural cell types." Nature reviews. Neuroscience **3**(7): 517-530.
- Bibel, M., J. Richter, et al. (2007). "Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells." Nature protocols **2**(5): 1034-1043.
- Bibel, M., J. Richter, et al. (2004). "Differentiation of mouse embryonic stem cells into a defined neuronal lineage." Nature neuroscience **7**(9): 1003-1009.
- Blaud, M., C. Vossen, et al. (2004). "Characteristic patterns of N Oct-3 binding to a set of neuronal promoters." Journal of molecular biology **339**(5): 1049-1058.
- Boheler, K. R., J. Czyz, et al. (2002). "Differentiation of pluripotent embryonic stem cells into cardiomyocytes." Circulation research **91**(3): 189-201.
- Botquin, V., H. Hess, et al. (1998). "New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2." Genes & development **12**(13): 2073-2090.
- Bourguet, W., V. Vivat, et al. (2000). "Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains." Molecular cell **5**(2): 289-298.
- Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-256.
- Breathnach, R. and P. Chambon (1981). "Organization and expression of eucaryotic split genes coding for proteins." Annual review of biochemistry **50**: 349-383.
- Bruck, N., D. Vitoux, et al. (2009). "A coordinated phosphorylation cascade initiated by p38MAPK/MSK1 directs RARalpha to target promoters." The EMBO journal **28**(1): 34-47.
- Buratowski, S., S. Hahn, et al. (1989). "Five intermediate complexes in transcription initiation by RNA polymerase II." Cell **56**(4): 549-561.
- Burdon, T., I. Chambers, et al. (1999). "Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells." Cells, tissues, organs **165**(3-4): 131-143.
- Burdon, T., C. Stracey, et al. (1999). "Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells." Developmental biology **210**(1): 30-43.
- Burke, T. W. and J. T. Kadonaga (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." Genes & development **10**(6): 711-724.
- Buttery, L. D., S. Bourne, et al. (2001). "Differentiation of osteoblasts and in vitro bone formation from murine embryonic stem cells." Tissue engineering **7**(1): 89-99.
- Campbell, K. and M. Gotz (2002). "Radial glia: multi-purpose cells for vertebrate brain development." Trends in neurosciences **25**(5): 235-238.
- Carninci, P., A. Sandelin, et al. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." Nature genetics **38**(6): 626-635.
- Castro, D. S., D. Skowronska-Krawczyk, et al. (2006). "Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif." Developmental cell **11**(6): 831-844.

- Cazillis, M., V. Lelievre, et al. (2005). "[Neural differentiation of murine embryonic stem cells ES]." Medecine sciences : M/S **21**(5): 484-490.
- Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-655.
- Chambon, P. (1996). "A decade of molecular biology of retinoic acid receptors." FASEB journal : official publication of the Federation of American Societies for Experimental Biology **10**(9): 940-954.
- Chan, R. J., S. A. Johnson, et al. (2003). "A definitive role of Shp-2 tyrosine phosphatase in mediating embryonic stem cell differentiation and hematopoiesis." Blood **102**(6): 2074-2080.
- Chen, C., D. Daugherty, et al. (2012). "Oligodendrocyte progenitor cells derived from mouse embryonic stem cells give rise to type-1 and type-2 astrocytes in vitro." Neuroscience letters **523**(2): 180-185.
- Cheung, W. L., K. Ajiro, et al. (2003). "Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase." Cell **113**(4): 507-517.
- Chuang, J. H., L. C. Tung, et al. (2011). "An approach for differentiating uniform glutamatergic neurons from mouse embryonic stem cells." Analytical biochemistry **410**(1): 149-151.
- Chung, S., J. I. Moon, et al. (2011). "ES cell-derived renewable and functional midbrain dopaminergic progenitors." Proceedings of the National Academy of Sciences of the United States of America **108**(23): 9703-9708.
- Cook, A. L. and R. A. Sturm (2008). "POU domain transcription factors: BRN2 as a regulator of melanocytic growth and tumorigenesis." Pigment cell & melanoma research **21**(6): 611-626.
- Coraux, C., C. Hilmi, et al. (2003). "Reconstituted skin from murine embryonic stem cells." Current biology : CB **13**(10): 849-853.
- Core, L. J. and J. T. Lis (2008). "Transcription Regulation Through Promoter-Proximal Pausing of RNA Polymerase II." Science **319**: 1791-1792.
- Cornish, E. J., S. M. Hassan, et al. (2009). "A microarray screen for direct targets of Zic1 identifies an aquaporin gene, aqp-3b, expressed in the neural folds." Developmental dynamics : an official publication of the American Association of Anatomists **238**(5): 1179-1194.
- Cui, H. and R. F. Bulleit (1998). "Expression of the POU transcription factor Brn-5 is an early event in the terminal differentiation of CNS neurons." Journal of neuroscience research **52**(6): 625-632.
- Davis, S., T. H. Aldrich, et al. (1993). "LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor." Science **260**(5115): 1805-1808.
- Delacroix, L., E. Moutier, et al. (2010). "Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells." Molecular and cellular biology **30**(1): 231-244.
- Deng, W. and S. G. Roberts (2005). "A core promoter element downstream of the TATA box that is recognized by TFIIB." Genes & development **19**(20): 2418-2423.
- Deng, W. and S. G. Roberts (2007). "TFIIB and the regulation of transcription by RNA polymerase II." Chromosoma **116**(5): 417-429.
- Doetschman, T. C., H. Eistetter, et al. (1985). "The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium." Journal of embryology and experimental morphology **87**: 27-45.
- Dominguez, M. H., A. E. Ayoub, et al. (2013). "POU-III transcription factors (Brn1, Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex." Cerebral cortex **23**(11): 2632-2643.
- Doyon, Y., C. Cayrou, et al. (2006). "ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation." Molecular cell **21**(1): 51-64.
- Eisen, T. G. (1996). "The control of gene expression in melanocytes and melanomas." Melanoma research **6**(4): 277-284.

- Elms, P., A. Scurry, et al. (2004). "Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation." Gene expression patterns : GEP **4**(5): 505-511.
- Englund, C., A. Fink, et al. (2005). "Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex." The Journal of neuroscience : the official journal of the Society for Neuroscience **25**(1): 247-251.
- Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.
- Farboud, B. and M. L. Privalsky (2004). "Retinoic acid receptor-alpha is stabilized in a repressive state by its C-terminal, isotype-specific F domain." Molecular endocrinology **18**(12): 2839-2853.
- Fico, A., G. Manganelli, et al. (2008). "High-throughput screening-compatible single-step protocol to differentiate embryonic stem cells in neurons." Stem cells and development **17**(3): 573-584.
- Finlay, B. L. and R. B. Darlington (1995). "Linked regularities in the development and evolution of mammalian brains." Science **268**(5217): 1578-1584.
- Fraichard, A., O. Chassande, et al. (1995). "In vitro differentiation of embryonic stem cells into glial cells and functional neurons." Journal of cell science **108 ( Pt 10)**: 3181-3188.
- Fuda, N. J., M. B. Ardehali, et al. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature **461**(7261): 186-192.
- Fujii, H. and H. Hamada (1993). "A CNS-specific POU transcription factor, Brn-2, is required for establishing mammalian neural cell lineages." Neuron **11**(6): 1197-1206.
- Fujimi, T. J., M. Hatayama, et al. (2012). "Xenopus Zic3 controls notochord and organizer development through suppression of the Wnt/beta-catenin signaling pathway." Developmental biology **361**(2): 220-231.
- Furushima, K., T. Murata, et al. (2000). "A new murine zinc finger gene, Opr." Mechanisms of development **98**(1-2): 161-164.
- Gadisseux, J. F. and P. Evrard (1985). "Glial-neuronal relationship in the developing central nervous system. A histochemical-electron microscope study of radial glial cell particulate glycogen in normal and reeler mice and the human fetus." Developmental neuroscience **7**(1): 12-32.
- Gaj, T., C. A. Gersbach, et al. (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." Trends in biotechnology **31**(7): 397-405.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." Journal of molecular biology **196**(2): 261-282.
- Gaston-Massuet, C., D. J. Henderson, et al. (2005). "Zic4, a zinc-finger transcription factor, is expressed in the developing mouse nervous system." Developmental dynamics : an official publication of the American Association of Anatomists **233**(3): 1110-1115.
- Ge, W., F. He, et al. (2006). "Coupling of cell migration with neurogenesis by proneural bHLH factors." Proceedings of the National Academy of Sciences of the United States of America **103**(5): 1319-1324.
- Geijsen, N., M. Horoschak, et al. (2004). "Derivation of embryonic germ cells and male gametes from embryonic stem cells." Nature **427**(6970): 148-154.
- Gerrard, L., L. Rodgers, et al. (2005). "Differentiation of human embryonic stem cells to neural lineages in adherent culture by blocking bone morphogenetic protein signaling." Stem cells **23**(9): 1234-1241.
- Ghyselinck, N., B. Chapellier, et al. (2002). "Dissection génétique de la fonction de l'acide rétinoïque dans la physiologie de l'épiderme." Annales de Dermatologie et de Vénéréologie **129**: 793-799.
- Gotz, M. and Y. A. Barde (2005). "Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons." Neuron **46**(3): 369-372.
- Graf, U., E. A. Casanova, et al. (2011). "The Role of the Leukemia Inhibitory Factor (LIF) - Pathway in Derivation and Maintenance of Murine Pluripotent Stem Cells." Genes **2**(1): 280-297.
- Grenfell, S. J., D. S. Latchman, et al. (1996). "Oct-1 [corrected] and Oct-2 DNA-binding site specificity is regulated in vitro by different kinases." The Biochemical journal **315 ( Pt 3)**: 889-893.

- Grinberg, I. and K. J. Millen (2005). "The ZIC gene family in development and disease." Clinical genetics **67**(4): 290-296.
- Grinberg, I., H. Northrup, et al. (2004). "Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation." Nature genetics **36**(10): 1053-1055.
- Gronemeyer, H. and V. Laudet (1995). "Transcription factors 3: nuclear receptors." Protein profile **2**(11): 1173-1308.
- Guerout, N., X. Li, et al. (2014). "Cell fate control in the developing central nervous system." Experimental cell research **321**(1): 77-83.
- Hamazaki, T., Y. Itoh, et al. (2001). "Hepatic maturation in differentiating embryonic stem cells in vitro." FEBS letters **497**(1): 15-19.
- Hartfuss, E., R. Galli, et al. (2001). "Characterization of CNS precursor subtypes and radial glia." Developmental biology **229**(1): 15-30.
- Hashimoto-Torii, K., M. Torii, et al. (2008). "Interaction between Reelin and Notch signaling regulates neuronal migration in the cerebral cortex." Neuron **60**(2): 273-284.
- Hatano, S. Y., M. Tada, et al. (2005). "Pluripotential competence of cells associated with Nanog activity." Mechanisms of development **122**(1): 67-79.
- Haubensak, W., A. Attardo, et al. (2004). "Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis." Proceedings of the National Academy of Sciences of the United States of America **101**(9): 3196-3201.
- He, X., M. N. Treacy, et al. (1989). "Expression of a large family of POU-domain regulatory genes in mammalian brain development." Nature **340**(6228): 35-41.
- Herr, W. and M. A. Cleary (1995). "The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain." Genes & development **9**(14): 1679-1693.
- Herr, W., R. A. Sturm, 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 & development **2**(12A): 1513-1516.
- Hodawadekar, S. C. and R. Marmorstein (2007). "Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design." Oncogene **26**(37): 5528-5540.
- Hubner, K., G. Fuhrmann, et al. (2003). "Derivation of oocytes from mouse embryonic stem cells." Science **300**(5623): 1251-1256.
- Huttner, W. B. and M. Brand (1997). "Asymmetric division and polarity of neuroepithelial cells." Current opinion in neurobiology **7**(1): 29-39.
- Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." FEBS letters **583**(11): 1713-1720.
- Illingworth, R. S., U. Gruenewald-Schneider, et al. (2010). "Orphan CpG islands identify numerous conserved promoters in the mammalian genome." PLoS genetics **6**(9): e1001134.
- Iso, T., L. Kedes, et al. (2003). "HES and HERP families: multiple effectors of the Notch signaling pathway." Journal of cellular physiology **194**(3): 237-255.
- Ivanova, N., R. Dobrin, et al. (2006). "Dissecting self-renewal in stem cells with RNA interference." Nature **442**(7102): 533-538.
- Kageyama, R., J. Hatakeyama, et al. (2006). Roles of Hes bHLH Factors in Neural Development. Transcription Factors. G. Thiel: 3-22.
- Kageyama, R., T. Ohtsuka, et al. (2005). "Roles of bHLH genes in neural stem cell differentiation." Experimental Cell Research **306**(2): 343-348.
- Kageyama, R., T. Ohtsuka, et al. (2008). "Roles of Hes genes in neural development." Development, Growth & Differentiation **50**: 97-103.
- Kageyama, R., T. Ohtsuka, et al. (2008). "Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition." Nature neuroscience **11**(11): 1247-1251.
- Kapiloff, M. S., Y. Farkash, et al. (1991). "Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements." Science **253**(5021): 786-789.

- Kasibhatla, S., P. Taylor, et al. (1999). "Jun kinase phosphorylates and regulates the DNA binding activity of an octamer binding protein, T-cell factor beta1." Molecular and cellular biology **19**(3): 2021-2031.
- Keller, G. (2005). "Embryonic stem cell differentiation: emergence of a new era in biology and medicine." Genes & development **19**(10): 1129-1155.
- Khorasanizadeh, S. and F. Rastinejad (2001). "Nuclear-receptor interactions on DNA-response elements." Trends in biochemical sciences **26**(6): 384-390.
- Klemm, J. D., M. A. Rould, et al. (1994). "Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules." Cell **77**(1): 21-32.
- Kolch, W. (2000). "Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions." The Biochemical journal **351 Pt 2**: 289-305.
- Koleske, A. J. and R. A. Young (1994). "An RNA polymerase II holoenzyme responsive to activators." Nature **368**(6470): 466-469.
- Kopan, R. and M. X. Ilagan (2009). "The canonical Notch signaling pathway: unfolding the activation mechanism." Cell **137**(2): 216-233.
- Kopp, J. L., B. D. Ormsbee, et al. (2008). "Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells." Stem cells **26**(4): 903-911.
- Kriegstein, A. R. and M. Gotz (2003). "Radial glia diversity: a matter of cell fate." Glia **43**(1): 37-43.
- Kuhlbrodt, K., B. Herbarth, et al. (1998). "Cooperative function of POU proteins and SOX proteins in glial cells." The Journal of biological chemistry **273**(26): 16050-16057.
- Kuhlbrodt, K., B. Herbarth, et al. (1998). "Sox10, a novel transcriptional modulator in glial cells." The Journal of neuroscience : the official journal of the Society for Neuroscience **18**(1): 237-250.
- Kuo, J. S., M. Patel, et al. (1998). "Opl: a zinc finger protein that regulates neural determination and patterning in *Xenopus*." Development **125**(15): 2867-2882.
- Kuroda, T., M. Tada, et al. (2005). "Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression." Molecular and cellular biology **25**(6): 2475-2485.
- Kutach, A. K. and J. T. Kadonaga (2000). "The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters." Molecular and cellular biology **20**(13): 4754-4764.
- Lagrange, T., A. N. Kapanidis, et al. (1998). "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB." Genes & development **12**(1): 34-44.
- Ledermann, B. and K. Burki (1991). "Establishment of a germ-line competent C57BL/6 embryonic stem cell line." Experimental cell research **197**(2): 254-258.
- Lee, S. H., N. Lumelsky, et al. (2000). "Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells." Nature biotechnology **18**(6): 675-679.
- Li, B., M. Carey, et al. (2007). "The Role of Chromatin during Transcription." Cell **128**(4): 707-719.
- Lim, L. S., F. H. Hong, et al. (2010). "The pluripotency regulator Zic3 is a direct activator of the Nanog promoter in ESCs." Stem cells **28**(11): 1961-1969.
- Lim, L. S., Y. H. Loh, et al. (2007). "Zic3 is required for maintenance of pluripotency in embryonic stem cells." Molecular biology of the cell **18**(4): 1348-1358.
- Liu, A. and L. A. Niswander (2005). "Bone morphogenetic protein signalling and vertebrate nervous system development." Nature reviews. Neuroscience **6**(12): 945-954.
- Liu, N., M. Lu, et al. (2007). "Molecular mechanisms involved in self-renewal and pluripotency of embryonic stem cells." Journal of cellular physiology **211**(2): 279-286.
- Lodato, M. A., C. W. Ng, et al. (2013). "SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state." PLoS genetics **9**(2): e1003288.
- Loh, K. M. and B. Lim (2011). "A precarious balance: pluripotency factors as lineage specifiers." Cell stem cell **8**(4): 363-369.
- Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nature genetics **38**(4): 431-440.

- Louvi, A. and S. Artavanis-Tsakonas (2006). "Notch signalling in vertebrate neural development." Nature reviews. Neuroscience **7**(2): 93-102.
- Lu, R., A. Yang, et al. (2011). "Dual functions of T-box 3 (Tbx3) in the control of self-renewal and extraembryonic endoderm differentiation in mouse embryonic stem cells." The Journal of biological chemistry **286**(10): 8425-8436.
- Lumsden, A. and R. Krumlauf (1996). "Patterning the vertebrate neuraxis." Science **274**(5290): 1109-1115.
- Maden, M. (2007). "Retinoic acid in the development, regeneration and maintenance of the nervous system." Nature reviews neuroscience **8**: 755-765.
- Malatesta, P., M. A. Hack, et al. (2003). "Neuronal or glial progeny: regional differences in radial glia fate." Neuron **37**(5): 751-764.
- Mangelsdorf, D. J. and R. M. Evans (1995). "The RXR heterodimers and orphan receptors." Cell **83**(6): 841-850.
- Mark, M., N. B. Ghyselinck, et al. (2006). "Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis." Annual review of pharmacology and toxicology **46**: 451-480.
- Markstein, M., P. Markstein, et al. (2002). "Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo." Proceedings of the National Academy of Sciences of the United States of America **99**(2): 763-768.
- Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nature reviews. Molecular cell biology **6**(11): 838-849.
- Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proceedings of the National Academy of Sciences of the United States of America **78**(12): 7634-7638.
- Masui, S. (2010). "Pluripotency maintenance mechanism of embryonic stem cells and reprogramming." International journal of hematology **91**(3): 360-372.
- Matsuda, T., T. Nakamura, et al. (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." The EMBO journal **18**(15): 4261-4269.
- Maunakea, A. K., R. P. Nagarajan, et al. (2010). "Conserved role of intragenic DNA methylation in regulating alternative promoters." Nature **466**(7303): 253-257.
- McEvilly, R. J., M. O. de Diaz, et al. (2002). "Transcriptional regulation of cortical neuron migration by POU domain factors." Science **295**(5559): 1528-1532.
- Merzdorf, C. S. (2007). "Emerging roles for zic genes in early development." Developmental dynamics : an official publication of the American Association of Anatomists **236**(4): 922-940.
- Meyuhas, O. (2000). "Synthesis of the translational apparatus is regulated at the translational level." European journal of biochemistry / FEBS **267**(21): 6321-6330.
- Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.
- Miura, T., M. P. Mattson, et al. (2004). "Cellular lifespan and senescence signaling in embryonic stem cells." Aging cell **3**(6): 333-343.
- Miyata, T., A. Kawaguchi, et al. (2004). "Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells." Development **131**(13): 3133-3145.
- Monsoro-Burq, A. H., E. Wang, et al. (2005). "Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction." Developmental cell **8**(2): 167-178.
- Moore, L. D., T. Le, et al. (2013). "DNA methylation and its basic function." Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology **38**(1): 23-38.
- Moutier, E., T. Ye, et al. (2012). "Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology." The Journal of biological chemistry **287**(31): 26328-26341.

- Mummery, C. L., C. E. van den Brink, et al. (1987). "Commitment to differentiation induced by retinoic acid in P19 embryonal carcinoma cells is cell cycle dependent." Developmental biology **121**(1): 10-19.
- Nagpal, S., S. Friant, et al. (1993). "RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization in vivo." The EMBO journal **12**(6): 2349-2360.
- Nagpal, S., M. Saunders, et al. (1992). "Promoter context- and response element-dependent specificity of the transcriptional activation and modulating functions of retinoic acid receptors." Cell **70**(6): 1007-1019.
- Nagy, L. and J. W. Schwabe (2004). "Mechanism of the nuclear receptor molecular switch." Trends in biochemical sciences **29**(6): 317-324.
- Nagy, L. and J. W. R. Schwabe (2004). "Mechanism of the nuclear receptor molecular switch." Trends in Biochemical Science **29**(6): 317-324.
- Nakada, Y., T. L. Hunsaker, et al. (2004). "Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification." Development **131**(6): 1319-1330.
- Nakai, S., H. Kawano, et al. (1995). "The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse." Genes & development **9**(24): 3109-3121.
- Nakano, T., H. Kodama, et al. (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science **265**(5175): 1098-1101.
- Nakata, K., T. Nagai, et al. (1998). "Xenopus Zic family and its role in neural and neural crest development." Mechanisms of development **75**(1-2): 43-51.
- Nayernia, K., J. Nolte, et al. (2006). "In vitro-differentiated embryonic stem cells give rise to male gametes that can generate offspring mice." Developmental cell **11**(1): 125-132.
- Niederreither, K. and P. Dollé (2008). "Retinoic acid in development: towards an integrated view." Nature Reviews Genetics **9**: 541-553.
- Nieto, L., G. Joseph, et al. (2007). "Differential effects of phosphorylation on DNA binding properties of N Oct-3 are dictated by protein/DNA complex structures." Journal of molecular biology **370**(4): 687-700.
- Nieto, M., E. S. Monuki, et al. (2004). "Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex." The Journal of comparative neurology **479**(2): 168-180.
- Nieto, M., C. Schuurmans, et al. (2001). "Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors." Neuron **29**(2): 401-413.
- Nishikawa, S. I., S. Nishikawa, et al. (1998). "Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages." Development **125**(9): 1747-1757.
- Niwa, H., T. Burdon, et al. (1998). "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." Genes & development **12**(13): 2048-2060.
- Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nature genetics **24**(4): 372-376.
- Niwa, H., K. Ogawa, et al. (2009). "A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells." Nature **460**(7251): 118-122.
- Noctor, S. C., V. Martinez-Cerdeno, et al. (2004). "Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases." Nature neuroscience **7**(2): 136-144.
- Ogbourne, S. and T. M. Antalis (1998). "Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes." The Biochemical journal **331** ( Pt 1): 1-14.
- Oki, M., H. Aihara, et al. (2007). "Role of histone phosphorylation in chromatin dynamics and its implications in diseases." Sub-cellular biochemistry **41**: 319-336.

- Park, I. H., N. Arora, et al. (2008). "Disease-specific induced pluripotent stem cells." Cell **134**(5): 877-886.
- Pavri, R., B. Lewis, et al. (2005). "PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator." Molecular cell **18**(1): 83-96.
- Perez, E., W. Bourguet, et al. (2012). "Modulation of RXR function through ligand design." Biochimica et biophysica acta **1821**(1): 57-69.
- Perissi, V. and M. G. Rosenfeld (2005). "Controlling nuclear receptors: the circular logic of cofactor cycles." Nature Reviews Molecular Cell Biology **6**: 542-554.
- Perry, M. W., A. N. Boettiger, et al. (2010). "Shadow enhancers foster robustness of Drosophila gastrulation." Current biology : CB **20**(17): 1562-1567.
- Pevzner, V., R. Kraft, et al. (2000). "Phosphorylation of Oct-2 at sites located in the POU domain induces differential down-regulation of Oct-2 DNA-binding ability." The Biochemical journal **347 Pt 1**: 29-35.
- Ponjavic, J., B. Lenhard, et al. (2006). "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters." Genome biology **7**(8): R78.
- Portela, A. and M. Esteller (2010). "Epigenetic modifications and human disease." Nature biotechnology **28**(10): 1057-1068.
- Pourebahim, R., R. Houtmeyers, et al. (2011). "Transcription factor Zic2 inhibits Wnt/beta-catenin protein signaling." The Journal of biological chemistry **286**(43): 37732-37740.
- Puelles, E. (2007). "Genetic control of basal midbrain development." Journal of neuroscience research **85**(16): 3530-3534.
- Puelles, L. and J. L. Rubenstein (2003). "Forebrain gene expression domains and the evolving prosomeric model." Trends in neurosciences **26**(9): 469-476.
- Qi, X., T. G. Li, et al. (2004). "BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways." Proceedings of the National Academy of Sciences of the United States of America **101**(16): 6027-6032.
- Ramsahoye, B. H., D. Biniszkiwicz, et al. (2000). "Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a." Proceedings of the National Academy of Sciences of the United States of America **97**(10): 5237-5242.
- Remenyi, A., K. Lins, et al. (2003). "Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers." Genes & development **17**(16): 2048-2059.
- Remenyi, A., H. R. Scholer, et al. (2004). "Combinatorial control of gene expression." Nature structural & molecular biology **11**(9): 812-815.
- Remenyi, A., A. Tomilin, et al. (2001). "Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping." Molecular cell **8**(3): 569-580.
- Revazova, E. S., N. A. Turovets, et al. (2007). "Patient-specific stem cell lines derived from human parthenogenetic blastocysts." Cloning and stem cells **9**(3): 432-449.
- Rhee, J. M., C. A. Gruber, et al. (1998). "Highly cooperative homodimerization is a conserved property of neural POU proteins." The Journal of biological chemistry **273**(51): 34196-34205.
- Rice, D. S. and T. Curran (2001). "Role of the reelin signaling pathway in central nervous system development." Annual review of neuroscience **24**: 1005-1039.
- Rochel, N., F. Ciesielski, et al. (2011). "Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings." Nature structural & molecular biology **18**(5): 564-570.
- Rochette-Egly, C. and P. Germain (2009). "Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs)." Nuclear receptor signaling **7**: e005.
- Rohwedel, J., K. Guan, et al. (1999). "Induction of cellular differentiation by retinoic acid in vitro." Cells, tissues, organs **165**(3-4): 190-202.
- Rohwedel, J., V. Maltsev, et al. (1994). "Muscle cell differentiation of embryonic stem cells reflects myogenesis in vivo: developmentally regulated expression of myogenic determination genes and functional expression of ionic currents." Developmental biology **164**(1): 87-101.

- Rohwedel, J., U. Sehlmeier, et al. (1996). "Primordial germ cell-derived mouse embryonic germ (EG) cells in vitro resemble undifferentiated stem cells with respect to differentiation capacity and cell cycle distribution." Cell biology international **20**(8): 579-587.
- Roth, S. Y., J. M. Denu, et al. (2001). "Histone acetyltransferases." Annual review of biochemistry **70**: 81-120.
- Ryan, A. K. and M. G. Rosenfeld (1997). "POU domain family values: flexibility, partnerships, and developmental codes." Genes & development **11**(10): 1207-1225.
- Sandelin, A., P. Carninci, et al. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nature reviews. Genetics **8**(6): 424-436.
- Sasai, N., K. Mizuseki, et al. (2001). "Requirement of FoxD3-class signaling for neural crest determination in *Xenopus*." Development **128**(13): 2525-2536.
- Sato, N., L. Meijer, et al. (2004). "Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor." Nature medicine **10**(1): 55-63.
- Sato, T., N. Sasai, et al. (2005). "Neural crest determination by co-activation of Pax3 and Zic1 genes in *Xenopus* ectoderm." Development **132**(10): 2355-2363.
- Sawadogo, M. and R. G. Roeder (1985). "Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay." Proceedings of the National Academy of Sciences of the United States of America **82**(13): 4394-4398.
- Saxonov, S., P. Berg, et al. (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proceedings of the National Academy of Sciences of the United States of America **103**(5): 1412-1417.
- Scholer, H. R. (1991). "Octamania: the POU factors in murine development." Trends in genetics : TIG **7**(10): 323-329.
- Scholer, H. R., R. Balling, et al. (1989). "Octamer binding proteins confer transcriptional activity in early mouse embryogenesis." The EMBO journal **8**(9): 2551-2557.
- Schonemann, M. D., A. K. Ryan, et al. (1995). "Development and survival of the endocrine hypothalamus and posterior pituitary gland requires the neuronal POU domain factor Brn-2." Genes & development **9**(24): 3122-3135.
- Schoonjans, L., V. Kreemers, et al. (2003). "Improved generation of germline-competent embryonic stem cell lines from inbred mouse strains." Stem cells **21**(1): 90-97.
- Schubeler, D., D. M. MacAlpine, et al. (2004). "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote." Genes & development **18**(11): 1263-1271.
- Scully, K. M., E. M. Jacobson, et al. (2000). "Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification." Science **290**(5494): 1127-1131.
- Shimada, M., H. Niida, et al. (2008). "Chk1 is a histone H3 threonine 11 kinase that regulates DNA damage-induced transcriptional repression." Cell **132**(2): 221-232.
- Shogren-Knaak, M., H. Ishii, et al. (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions." Science **311**(5762): 844-847.
- Smale, S. T. and D. Baltimore (1989). "The "initiator" as a transcription control element." Cell **57**(1): 103-113.
- Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." Annual review of biochemistry **72**: 449-479.
- Smart, I. H., C. Dehay, et al. (2002). "Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey." Cerebral cortex **12**(1): 37-53.
- Smit, D. J., A. G. Smith, et al. (2000). "Domains of Brn-2 that mediate homodimerization and interaction with general and melanocytic transcription factors." European journal of biochemistry / FEBS **267**(21): 6413-6422.
- Smith, A. G. (2001). "Embryo-derived stem cells: of mice and men." Annual review of cell and developmental biology **17**: 435-462.

- Smith, A. G., G. Brightwell, et al. (1998). "Redox regulation of Brn-2/N-Oct-3 POU domain DNA binding activity and proteolytic formation of N-Oct-5 during melanoma cell nuclear extraction." Melanoma research **8**(1): 2-10.
- Smith, A. G., J. K. Heath, et al. (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.
- Smith, D. S., P. L. Greer, et al. (2001). "Cdk5 on the brain." Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research **12**(6): 277-283.
- Smukler, S. R., S. B. Runciman, et al. (2006). "Embryonic stem cells assume a primitive neural stem cell fate in the absence of extrinsic influences." The Journal of cell biology **172**(1): 79-90.
- Solter, D. and B. B. Knowles (1978). "Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1)." Proceedings of the National Academy of Sciences of the United States of America **75**(11): 5565-5569.
- Soria, B. (2001). "In-vitro differentiation of pancreatic beta-cells." Differentiation; research in biological diversity **68**(4-5): 205-219.
- Stead, E., J. White, et al. (2002). "Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities." Oncogene **21**(54): 8320-8333.
- Stoffel, M., L. Vallier, et al. (2004). "Navigating the pathway from embryonic stem cells to beta cells." Seminars in cell & developmental biology **15**(3): 327-336.
- Strubing, C., G. Ahnert-Hilger, et al. (1995). "Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature inhibitory and excitatory neurons." Mechanisms of development **53**(2): 275-287.
- Sturm, R. A. and W. Herr (1988). "The POU domain is a bipartite DNA-binding structure." Nature **336**(6199): 601-604.
- Sugitani, Y., S. Nakai, et al. (2002). "Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons." Genes & development **16**(14): 1760-1765.
- Sun, C., Y. Nakatake, et al. (2009). "Dax1 binds to Oct3/4 and inhibits its transcriptional activity in embryonic stem cells." Molecular and cellular biology **29**(16): 4574-4583.
- Sunabori, T., A. Tokunaga, et al. (2008). "Cell-cycle-specific nestin expression coordinates with morphological changes in embryonic cortical neural progenitors." Journal of cell science **121**(Pt 8): 1204-1212.
- Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.
- Takahashi, T., B. Lord, et al. (2003). "Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes." Circulation **107**(14): 1912-1916.
- Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proceedings of the National Academy of Sciences of the United States of America **99**(6): 3740-3745.
- Tanaka, S., Y. Kamachi, et al. (2004). "Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells." Molecular and cellular biology **24**(20): 8834-8846.
- Tang, F., K. Shang, et al. (2002). "Differentiation of embryonic stem cell to astrocytes visualized by green fluorescent protein." Cellular and molecular neurobiology **22**(1): 95-101.
- Tang, H., E. Hornstein, et al. (2001). "Amino acid-induced translation of TOP mRNAs is fully dependent on phosphatidylinositol 3-kinase-mediated signaling, is partially inhibited by rapamycin, and is independent of S6K1 and rpS6 phosphorylation." Molecular and cellular biology **21**(24): 8671-8683.
- Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nature structural & molecular biology **14**(11): 1025-1040.
- Thomas, M. C. and C. M. Chiang (2006). "The general transcription machinery and general cofactors." Critical reviews in biochemistry and molecular biology **41**(3): 105-178.
- Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.

- Tomilin, A., A. Remenyi, et al. (2000). "Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration." *Cell* **103**(6): 853-864.
- Toyooka, Y., N. Tsunekawa, et al. (2003). "Embryonic stem cells can form germ cells in vitro." *Proceedings of the National Academy of Sciences of the United States of America* **100**(20): 11457-11462.
- Tropepe, V., S. Li, et al. (2006). "Identification of a BMP inhibitor-responsive promoter module required for expression of the early neural gene *zic1*." *Developmental biology* **289**(2): 517-529.
- Uranishi, K., T. Akagi, et al. (2013). "Dax1 associates with Esrrb and regulates its function in embryonic stem cells." *Molecular and cellular biology* **33**(10): 2056-2066.
- Valdimarsdottir, G. and C. Mummery (2005). "Functions of the TGFbeta superfamily in human embryonic stem cells." *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* **113**(11-12): 773-789.
- Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." *Science* **291**(5507): 1304-1351.
- Vierbuchen, T., A. Ostermeier, et al. (2010). "Direct conversion of fibroblasts to functional neurons by defined factors." *Nature* **463**(7284): 1035-1041.
- Wang, D., D. A. Bushnell, et al. (2009). "Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution." *Science* **324**(5931): 1203-1206.
- Wang, H., H. Yang, et al. (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." *Cell* **153**(4): 910-918.
- Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." *Nature genetics* **40**(7): 897-903.
- Wapinski, O. L., T. Vierbuchen, et al. (2013). "Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons." *Cell* **155**(3): 621-635.
- Warr, N., N. Powles-Glover, et al. (2008). "Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation." *Human molecular genetics* **17**(19): 2986-2996.
- Wegner, M. (2005). "Secrets to a healthy Sox life: lessons for melanocytes." *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **18**(2): 74-85.
- Willert, K. and K. A. Jones (2006). "Wnt signaling: is the party in the nucleus?" *Genes & development* **20**(11): 1394-1404.
- Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." *Nature* **336**(6200): 684-687.
- Wobus, A. M. and K. R. Boheler (2005). "Embryonic stem cells: prospects for developmental biology and cell therapy." *Physiological reviews* **85**(2): 635-678.
- Wobus, A. M., H. Holzhausen, et al. (1984). "Characterization of a pluripotent stem cell line derived from a mouse embryo." *Experimental cell research* **152**(1): 212-219.
- Wolffe, A. P. and D. Pruss (1996). "Targeting chromatin disruption: Transcription regulators that acetylate histones." *Cell* **84**(6): 817-819.
- Wu, H. and Y. Zhang (2011). "Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells." *Cell cycle* **10**(15): 2428-2436.
- Xiao, Q., L. Zeng, et al. (2007). "Stem cell-derived Sca-1+ progenitors differentiate into smooth muscle cells, which is mediated by collagen IV-integrin alpha1/beta1/alpha5 and PDGF receptor pathways." *American journal of physiology. Cell physiology* **292**(1): C342-352.
- Yagi, T., D. Ito, et al. (2011). "Modeling familial Alzheimer's disease with induced pluripotent stem cells." *Human molecular genetics* **20**(23): 4530-4539.
- Yamada, T., M. Yoshikawa, et al. (2002). "In vitro functional gut-like organ formation from mouse embryonic stem cells." *Stem cells* **20**(1): 41-49.
- Ying, Q. L., J. Nichols, et al. (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." *Cell* **115**(3): 281-292.

- Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nature biotechnology **21**(2): 183-186.
- Young, R. A. (1991). "RNA polymerase II." Annual review of biochemistry **60**: 689-715.
- Yu, P., G. Pan, et al. (2011). "FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation." Cell stem cell **8**(3): 326-334.
- Zhang, J., W. L. Tam, et al. (2006). "Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1." Nature cell biology **8**(10): 1114-1123.
- Zhang, X., Y. V. Bernatavichute, et al. (2009). "Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana." Genome biology **10**(6): R62.
- Zhang, Y. and D. Reinberg (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." Genes & development **15**(18): 2343-2360.
- Zimmer, C., M. C. Tiveron, et al. (2004). "Dynamics of Cux2 expression suggests that an early pool of SVZ precursors is fated to become upper cortical layer neurons." Cerebral cortex **14**(12): 1408-1420.
- Zwilling, S., A. Annweiler, et al. (1994). "The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP." Nucleic acids research **22**(9): 1655-1662.



Sylvia URBAN



# **Brn2 et Zic1 spécifient l'identité neuronale des cellules souches embryonnaires murines lors de la différenciation induite par l'acide rétinoïque**

## **Résumé**

Les cellules souches embryonnaires (ES) murines peuvent être différenciées *in vitro* en une population homogène de neurones glutamatergiques semblables aux neurones présents dans le cortex *in vivo*, suite à un traitement par l'acide rétinoïque (AR). Bien que le rôle de l'AR soit bien étudié, les facteurs qui spécifient le destin neuronal ne sont pas connus. Nous montrons ici que Pou3f2 (Brn2) est un facteur essentiel à la différenciation neuronale des cellules ES *in vitro*. L'utilisation de l'approche de différenciation *in vitro* associée à des techniques de génomique à haut débit (RNA-seq, ChIP-seq) a permis d'identifier des gènes régulés directement ou indirectement par Brn2. Parmi ces gènes se trouvent *Ascl1*, *Hes5* ou *Pou6f1*, qui sont des gènes clés dans la neurogenèse. La comparaison de nos données avec des expériences précédemment publiées nous a permis d'identifier un nombre restreint de gènes cibles de Brn2 quelque soit le protocole de différenciation utilisé. Parmi ces gènes se trouve Zic1. Nous montrons que Zic1 coopère avec Brn2 pour spécifier le destin neuronal des cellules ES *in vitro*.

Mots clés : Différenciation neuronale, Acide Rétinoïque, Brn2, Zic1

## **Résumé en anglais**

Mouse embryonic stem (ES) cells can be differentiated *in vitro* into a highly homogenous population of glutamatergic neurons, similar to those present in the cerebellar cortex by treatment with retinoic acid (RA). While the role of RA in differentiation is well studied, the downstream factors that specify the neural fate of the ES cells are not known. Here we show that Pou3f2 (Brn2), with a known role in neuronal differentiation *in vivo*, is essential for neuronal differentiation of ES cells *in vitro*. Using our *in vitro* differentiation protocol combined with high throughput techniques (RNA-seq, ChIP-seq) we show that Brn2 directly and indirectly regulates a set of target genes with essential roles in neurogenesis such as *Ascl1*, *Hes5* or *Pou6f1*. Integration of these results with previously published datasets allowed us to identify a core set of Brn2 target genes common to each differentiation model. Amongst these is transcription factor Zic1. We show that Zic1 and Brn2 cooperate to specify the neural fate of RA-treated ES cells *in vitro*.

Keywords : Neuronal differentiation, Retinoic Acid, Brn2, Zic1