



**HAL**  
open science

# IDENTIFICATION ET CARACTERISATION MOLECULAIRE D'UNE NOUVELLE FAMILLE DE GENES MARQUEURS DES CELLULES SOUCHES EMBRYONNAIRES DE POULET

Hervé Acloque

► **To cite this version:**

Hervé Acloque. IDENTIFICATION ET CARACTERISATION MOLECULAIRE D'UNE NOUVELLE FAMILLE DE GENES MARQUEURS DES CELLULES SOUCHES EMBRYONNAIRES DE POULET. Sciences du Vivant [q-bio]. Université Claude Bernard Lyon 1, 2003. Français. NNT : . tel-02912760

**HAL Id: tel-02912760**

**<https://hal.inrae.fr/tel-02912760>**

Submitted on 6 Aug 2020

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

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

## **THESE**

Présentée à l'Ecole Normale Supérieure de Lyon  
pour l'obtention du diplôme de Docteur de l'Ecole Normale Supérieure de Lyon  
dans la discipline des Sciences de la vie  
et dans le cadre de l'école doctorale Biologie Moléculaire Intégrée et Cognitive

par

**Hervé ACLOQUE**

---

**IDENTIFICATION ET CARACTERISATION MOLECULAIRE  
D'UNE NOUVELLE FAMILLE DE GENES MARQUEURS DES  
CELLULES SOUCHES EMBRYONNAIRES DE POULET**

---

**Soutenue le 4 juillet 2003**

**Jury composé de :**

**M. Jean Loup DUBAND (rapporteur)**

**M. Jean Paul RENARD (rapporteur)**

**M. Gilbert BRUN**

**M. Pierre COUBLE**

**M. Jacques SAMARUT**

## Remerciements

Je tiens à remercier tout d'abord l'ensemble des personnes avec qui j'ai travaillé durant ces quatre années au laboratoire, que ce soit par le biais d'un service commun ou lors de discussions informelles.

Je remercie aussi également Jacques Samarut pour m'avoir accueilli dans son équipe et pour m'avoir accordé une grande confiance et une grande liberté d'action.

Je remercie les membres du jury d'avoir accepté d'accorder de leur temps à la lecture de ce mémoire.

Je remercie tout particulièrement Edith Legouy et l'ensemble des personnes de l'Unité de Biologie du Développement pour avoir su me donner le goût et l'envie de travailler pour la recherche en biologie.

Je remercie Anne Marie Birot dont j'ai apprécié la compagnie et les qualités durant ces quatre années.

Je remercie toutes les personnes de l'équipe Samarut qui m'ont chacune aidé à réaliser ce travail. Merci également pour les bons moments passés ensemble. Un grand merci plus particulier aux personnes travaillant sur les cellules souches de poulet, Anne, Bertrand et également Fabrice, futur vilain petit thésard. Je remercie également tous les stagiaires qui m'ont aidé durant cette thèse, avec, par ordre chronologique : Abdel, Anne, Bérengère et Amélie.

Je remercie chaleureusement toutes les personnes des équipes voisines, dont les rois du Worms (Olivier, Fred, Guillaume et Laurent) et la reine des poules, j'ai nommé Dominique.

Enfin un grand merci à Florence et à mes proches, qui m'ont toujours encouragé et aussi aidé pour la relecture des manuscrits.

# Résumé

Les cellules souches embryonnaires de poulet (CES) sont des cellules pluripotentes capables de se différencier *in vitro* en de multiples types cellulaires représentant les trois feuilletts embryonnaires. Ces cellules sont également capables de coloniser de nombreux lignages cellulaires d'un embryon hôte, dont la lignée germinale (avec une efficacité très faible).

Afin de préciser la nature moléculaire des cellules CES, nous avons utilisé la stratégie du piégeage de gène pour identifier des gènes exprimés spécifiquement dans ces cellules. Nous avons ainsi isolé une nouvelle famille de gènes, appelés *ENS* pour Embryonic Normal Stem cell genes, exprimés spécifiquement dans les cellules ES et l'embryon précoce de poulet. Trois gènes ont été caractérisés : *cENS-1* code pour une protéine de 490 acides aminés, identique à *cERNI*, dont la séquence codante est insérée entre deux séquences répétées d'environ 500 paires de base. *cENS-2* est une forme tronquée de *cENS-1* et *cENS-3* comporte deux cadres ouverts de lecture correspondant aux gènes rétroviraux *pol* et *env*. Par analogie avec la structure du génome rétroviral, nous avons isolé une région promoteur des gènes *ENS* dont l'activité est spécifique des cellules CES.

Pour comprendre les mécanismes conférant sa tissu-spécificité au promoteur *ENS*, nous avons réalisé un découpage systématique du promoteur et identifié ainsi deux régions, l'une activatrice et importante pour le fonctionnement de ce promoteur, l'autre répressive et restreignant l'activité du promoteur aux cellules CES indifférenciées.

Nous avons caractérisé dans la région activatrice deux éléments de réponse, A et B, fixés par des facteurs présents dans des extraits nucléaires de cellules CES. Nous avons ensuite étudié la fonctionnalité de ces éléments par transfection dans des cellules CES. Nous avons alors montré que seul l'élément B était nécessaire pour activer le promoteur *ENS* dans ces cellules. Par contre, nos résultats suggèrent qu'il n'est pas suffisant pour assurer la tissu-spécificité du promoteur *ENS*. La similitude de la séquence de l'élément B avec celle connue pour fixer le facteur de transcription CP2 (5'-CNRG-N<sub>6</sub>-CNRG) nous a amené à vérifier si le complexe protéique fixé sur l'élément B comprenait le facteur CP2. Par des expériences de supershift avec un antisérum dirigé contre CP2, nous avons montré que le complexe fixé sur l'élément B comprenait un facteur de transcription de la famille CP2. L'analyse du profil d'expression de la protéine CP2 dans les cellules CES confirme cette observation.

La tissu-spécificité du promoteur *ENS* est certainement rendu possible par la présence de ces deux régions, l'une nécessaire à l'activité du promoteur, l'autre bloquant cette activité dans les cellules différenciées. L'analyse fonctionnelle de la région inhibitrice permettra d'identifier les acteurs moléculaires conduisant à l'expression spécifique des gènes *ENS* dans les cellules souches embryonnaires de poulet.

---

# SOMMAIRE

---

<b>SOMMAIRE.....</b>	<b>1</b>
<b>ABREVIATIONS.....</b>	<b>5</b>
<b>PREAMBULE HISTORIQUE .....</b>	<b>7</b>
I UN MONDE SANS CELLULES .....	8
II LA THEORIE CELLULAIRE.....	9
III LA DECOUVERTE DE LA NOTION D'HEREDITE .....	9
IV UNE PORTE OUVERTE POUR L'EXPERIMENTATION.....	11
<b>LA MISE EN EVIDENCE DE LA PLURIPOTENCE CELLULAIRE ET LA DECOUVERTE DES CELLULES SOUCHES EMBRYONNAIRES.....</b>	<b>12</b>
I LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES .....	12
1°) <i>Mise en evidence de la pluripotence des blastomeres chez les animaux</i> .....	12
2°) <i>La pluripotence des cellules embryonnaires de mammiferes</i> .....	13
3°) <i>Les chimeres, veritable demonstration de la pluripotence des cellules embryonnaires</i> .....	14
a) Les chimeres d'agregation.....	14
b) Les chimeres d'injection.....	15
4°) <i>La pluripotence des cellules du blastoderme dans l'embryon de poulet</i> .....	16
II L'ISOLEMENT DES CELLULES SOUCHES EMBRYONNAIRES .....	17
1°) <i>Les cellules embryonnaires carcinomales (EC cells)</i> .....	17
2°) <i>Les cellules souches embryonnaires (cellules ES) de souris</i> .....	19
3°) <i>Les cellules souches embryonnaires de poulet</i> .....	20
<b>CARACTERISATION MOLECULAIRE DES CELLULES SOUCHES EMBRYONNAIRES .....</b>	<b>23</b>
I LE CONTROLE DE L'AUTORENOUVELLEMENT DES CELLULES ES DE SOURIS .....	23
1°) <i>Le LIF et sa voie de signalisation</i> .....	23
a) Le LIF et les cytokines de la famille IL-6.....	23
b) Les proteines STAT et leur activation par le LIF .....	24
2°) <i>Role de STAT3 dans le controle de l'autorenouvellement des cellules ES</i> .....	25
a) STAT3 est necessaire et suffisant au maintien de l'autorenouvellement des cellules ES .....	25
b) Les genes cibles de STAT3 .....	26
3°) <i>La voie ERK bloque l'autorenouvellement des cellules</i> .....	26
a) Presentation rapide de la voie Ras/MAPK.....	26
b) Role de la voie de signalisation dependante de ERKs dans le controle de l'autorenouvellement des cellules ES .....	27
c) Role de SHP2 dans l'activation de la voie de signalisation dependante de ERK.....	27

4°) D'autres voies de régulation contrôlant l'autorenouvellement des cellules ES.....	28
a) La voie LIF/gp130 <i>in vivo</i> .....	29
b) Le facteur ESRF .....	29
c) La thrombopoïétine .....	30
5°) Le cycle cellulaire des cellules ES : sous le contrôle de la voie LIF/STAT3 ? .....	30
a) La transition G1/S dans les cellules différenciées.....	31
b) Le contrôle de la transition G1/S diffère entre cellules ES et cellules différenciées.....	31
<b>II OCT-3/4 : UN REGULATEUR PUISSANT DE LA PLURIPOTENCE DES CELLULES ES.....</b>	<b>33</b>
1°) Le facteur Oct-3/4 appartient à la famille POU.....	33
2°) Importance du gène Oct-3/4 dans le maintien de la pluripotence des cellules embryonnaires.....	34
3°) Les gènes cibles d'Oct-3/4.....	35
4°) La régulation de l'expression du gène Oct-3/4.....	37
<b>III D' AUTRES GENES IMPLIQUES DANS LE MAINTIEN DE LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES ...</b>	<b>38</b>
1°) Le gène FoxD3 .....	39
2°) Le gène Sox2.....	40
3°) Le gène Nanog.....	41
4°) D'autres gènes spécifiques des cellules ES .....	42
5°) Les approches de génomique à grande échelle .....	44
<b>IV CARACTERISATION MOLECULAIRE DES CELLULES DU BLASTODERME DE POULET.....</b>	<b>45</b>
1°) Les cellules germinales primordiales, une population de cellules présentes dans l'épiblaste .....	45
2°) L'induction neurale : un mécanisme déjà actif dans les cellules de l'épiblaste ? .....	46
<b>LE GENOME RETROVIRAL ET SON EXPRESSION DANS LES CELLULES ES .....</b>	<b>49</b>
<b>I PRESENTATION GENERALE DES RETROVIRUS .....</b>	<b>49</b>
1°) Caractéristiques générales des particules rétrovirales .....	49
2°) Le génome des rétrovirus.....	51
3°) Le cycle répliatif des rétrovirus .....	52
4°) Les rétroéléments.....	53
a) Les éléments sans LTR (pseudogènes, rétrogènes et rétroposons) .....	53
b) Les éléments avec LTR (rétrotransposons et rétrovirus endogènes).....	54
<b>II LE CONTROLE DE L'EXPRESSION DU GENOME RETROVIRAL : L'EXEMPLE DES PARTICULES INTRACISTERNALES DE TYPE A.....</b>	<b>55</b>
1°) Présentation générale des particules intracisternales de type A (IAP) .....	56
2°) Présentation des mécanismes régulant l'expression des IAPs .....	56
a) Analyse <i>in vitro</i> des facteurs cellulaires agissant directement sur les LTRs d'IAPs.....	56
b) Mise en évidence de la méthylation dans le contrôle <i>in vivo</i> de l'expression des IAPs .....	57
c) Analyse du rôle de la méthylation dans le contrôle de l'expression des IAPs .....	58
3°) Une régulation à plusieurs niveaux contrôle l'expression des IAPs .....	59
<b>III REGULATION DES GENES RETROVIRAUX DANS LES CELLULES SOUCHES EMBRYONNAIRES .....</b>	<b>60</b>
1°) Expression des génomes rétroviraux dans les cellules ES.....	61
2°) La méthylation du génome viral dans les cellules ES.....	62
3°) D'autres mécanismes régulent l'extinction de l'activité des LTRs dans les cellules ES.....	63

4°) <i>La cellule ES possède un état épigénétique instable</i> .....	65
5°) <i>La régulation des séquences rétrovirales dans les cellules ES et l'embryon précoce de poulet</i> .....	66
<b>LE FACTEUR DE TRANSCRIPTION CP2</b> .....	<b>69</b>
I PRESENTATION GENERAL DU FACTEUR DE TRANSCRIPTION CP2 .....	69
1°) <i>Identification du gène CP2</i> .....	69
2°) <i>Conservation du gène CP2</i> .....	70
3°) <i>Caractéristique de la protéine CP2</i> .....	70
II FONCTIONS DU FACTEUR DE TRANSCRIPTION CP2.....	71
1°) <i>Promoteurs régulés in vitro par CP2</i> .....	71
2°) <i>Partenaires de CP2 identifiés in vitro</i> .....	73
a) <i>Le facteur YY1</i> .....	73
b) <i>D'autres partenaires de CP2</i> .....	73
3°) <i>Fonctions de CP2 in vivo</i> .....	75
<b>RESULTATS ET DISCUSSION</b> .....	<b>77</b>
I IDENTIFICATION D'UNE FAMILLE DE GENES EXPRIMEE SPECIFIQUEMENT DANS LES CELLULES ES ET	
L'EMBRYON PRECOCE DE POULET .....	77
1°) <i>Introduction au travail experimental</i> .....	77
2°) <i>La technique du piégeage de gène (gene-trap)</i> .....	78
3°) <i>Resumé du travail effectué</i> .....	78
a) <i>Identification d'un locus spécifiquement actif dans les cellules ES de poulet</i> .....	78
b) <i>Identification et caractérisation des gènes ENS</i> .....	79
c) <i>Profil d'expression des gènes ENS dans les cellules ES et l'embryon précoce de poulet</i> .....	80
d) <i>Identification du promoteur des gènes ENS</i> .....	81
4°) <i>Article</i> .....	81
5°) <i>Brevet</i> .....	82
6°) <i>Discussion et commentaires sur les résultats présentés</i> .....	83
a) <i>Approche par gene-trap</i> .....	83
b) <i>Identification d'une nouvelle famille de gènes</i> .....	84
c) <i>Profil d'expression des gènes ENS</i> .....	84
d) <i>Conservation des gènes ENS</i> .....	85
e) <i>Origine rétrovirale des gènes ENS</i> .....	86
f) <i>Identification d'une région promoteur</i> .....	87
II UN FACTEUR DE TRANSCRIPTION DE LA FAMILLE CP2 EST NECESSAIRE A L'ACTIVATION DU PROMOTEUR ENS	88
1°) <i>Introduction au travail experimental</i> .....	88
2°) <i>Présentation du travail effectué</i> .....	89
a) <i>Mise en évidence d'une région activatrice dans le promoteur ENS</i> .....	89
b) <i>La région A n'est pas capable de trans-activer le promoteur ENS</i> .....	90
c) <i>Caractérisation de l'élément B</i> .....	91
d) <i>L'élément B est nécessaire à l'activation du promoteur ENS</i> .....	92
e) <i>L'élément B fixe un facteur de transcription appartenant à la famille CP2</i> .....	92
f) <i>Clonage et profil d'expression du facteur CP2 dans les cellules ES de poulet</i> .....	93

3°) <i>Manuscript</i> .....	93
4°) <i>Discussions et Commentaires sur les résultats présentés</i> .....	94
a) Identification d'une région activatrice .....	94
b) L'élément B est nécessaire à l'activation du promoteur ENS.....	94
c) Un facteur de transcription appartenant à la famille CP2 fixe l'élément B.....	95
<b>PERSPECTIVES ET CONCLUSION GENERALE</b> .....	<b>97</b>
I MISE EN EVIDENCE D'UNE REGION REPRESSEUR SPECIFIQUE DES CELLULES DIFFERENCIEES .....	97
II ACTIVITE DU PROMOTEUR <i>ENS</i> DANS L'EMBRYON DE POULET.....	97
III ACTIVITE DU PROMOTEUR <i>ENS</i> DANS D'AUTRES ESPECES ANIMALES .....	98
IV CONCLUSIONS.....	98
<b>REFERENCES</b> .....	<b>100</b>
<b>ANNEXES</b> .....	<b>117</b>



---

## ABREVIATIONS

---

### **A**

aa : acides aminés  
ADN : Acide DésoxyriboNucléique  
ALV : Avian Leukemic Virus  
ARN : Acide RiboNucléique

### **B**

BMP : Bone Morphogenic Protein

### **C**

CDK : Cyclin Dependant Kinase  
CES : Chicken Embryonic Stem  
CNTF : Ciliary Neurotropic Factor  
CT-1 : Cardiotrophin-1

### **D**

Dnmt : DNA methyltransferase

### **E**

EC : Embryonal Carcinoma  
ES : Embryonic Stem  
EG : Eyal-Giladi (stades de développement des embryons de poulet)  
EG : Embryonic Germinal  
EMSA: Electrophoretic Mobility Shift Assay  
ENS : Embryonic Normal Stem cell  
ER : Estrogen Receptor  
ERNI : Embryonic Response to Neural Induction  
ESRF : ES self Renewal Factor

### **F**

bFGF : basic Fibroblast Growth Factor

### **H**

HDAC : Histone Deacetylase

### **I**

IAP : Intra cisternal A Particle  
IGF-1 : Insulin Like Growth Factor 1  
IL-6 : Interleukine 6  
IL-11 : Interleukine 11

### **K**

Kb: kilobase

**L**

LIF : Leukemia Inhibitory Factor  
LINE : Long Interspersed Nuclear Element  
LTR : Long Terminal Repeat

**M**

MCI : Masse Cellulaire Interne  
MEF : Mouse Embryonic Fibroblast  
MMLV : Moloney Murine Leukemia Virus  
MSCV : Mouse Stem Cell Virus

**N**

NTF : Neurogenic element binding Transcription Factor

**O**

OSM : Oncostatin M

**P**

p.c. : post coïtum  
pb : paire de base

**R**

RB : Rétinoblastome  
RT : Reverse Transcriptase  
RSV : Rous Sarcoma Virus

**S**

SCF : Stem Cell Factor  
SINE : Short Interspersed Nuclear Element  
SOCS3 : Suppressor Of Cytokine Signalling 3  
Sox : SRY-related HMG box  
STAT : Signal Transducer and Activator of Transcription

**T**

Tpo : Thrombopoïétine

**V**

VIH : Virus de l'Immunodéficience Humaine

## PREAMBULE HISTORIQUE

De nombreux organismes pluricellulaires dérivent d'une cellule unique, l'œuf ou zygote. Après de multiples divisions cellulaires et au cours de mécanismes complexes qui contrôlent le devenir de ces cellules, cet œuf aura été à l'origine d'un organisme nouveau, génétiquement différent mais très proche cependant des organismes dont il est issu. Cette idée nous est aujourd'hui familière mais pose de nombreuses questions. Comment une cellule peut-elle être à l'origine d'un organisme composé parfois de centaines de milliards de cellules ? Quels sont les déterminants qui confèrent à la cellule œuf sa totipotence ? Comment vont se former les différents organes ? Quels sont les mécanismes qui vont gouverner la différenciation de leurs cellules en leur attribuant une fonction au sein de l'organisme ?

C'est à ces questions que tente de répondre l'étude du développement embryonnaire et en particulier l'étude des cellules souches embryonnaires. En effet, ces cellules, par leurs caractéristiques uniques de prolifération et de pluripotentialité, sont un modèle performant d'étude du développement embryonnaire à la fois *in vivo* par leur capacité à coloniser l'embryon mais aussi *in vitro* par leur capacité à se différencier en culture dans de très nombreux types cellulaires. Ces cellules vous seront présentées plus en détail ultérieurement mais j'aimerais m'attarder dans cette introduction sur les cheminements historiques et culturels qui ont fortement influencé l'étude du développement embryonnaire et, par conséquent, la découverte de la pluripotence des cellules embryonnaires.

Un bref regard en arrière nous montre d'ailleurs que l'émergence des travaux portant sur le développement embryonnaire mais aussi sur l'évolution des espèces animales est récente. La première fécondation observée chez l'oursin par Oskar Hertwig (1849-1922) ne date en effet que de 1875. Et si ces problématiques furent étudiées dès l'Antiquité, les limitations techniques et les concepts philosophiques et religieux ont largement faussé les réponses qui leur furent apportées.

## I UN MONDE SANS CELLULES

Dès l'Antiquité Hippocrate (460-377 av. J.-C.) avançait l'idée que le développement du fœtus résultait de la rencontre des semences du mâle et de la femelle. Selon lui, les matériaux séminaux étaient émis par toutes les parties de l'organisme et transmis par les humeurs jusqu'aux organes reproducteurs. Cette théorie, dite de la pangénèse, a été soutenue jusqu'au 19<sup>ème</sup> siècle, y compris par Darwin.

Deux autres théories furent également avancées durant l'Antiquité pour expliquer les mécanismes aboutissant à la formation d'un organisme complet.

La théorie de la préformation avançait l'hypothèse qu'il existait dans chaque semence un être miniature qui, dans certaines conditions, croissait jusqu'à donner un individu adulte. Il existait déjà dans cette individu miniature tous les membres de sa descendance, emboîtés les uns dans les autres, à l'image des poupées russes.

La seconde théorie, l'épigénèse, développée par Aristote (384-322 av. J.-C.), s'inspire de la philosophie essentialiste.

Celle-ci pose l'hypothèse que chaque forme vivante dérive d'une idée (eidos) ou essence, elle-même d'origine divine. L'épigénèse pose donc l'hypothèse que l'embryon se constitue par l'organisation progressive des éléments de la matière brute. Ainsi la semence apporte l'idée, l'essence de l'individu tandis que l'organisme maternel, qui porte cet individu, pourvoie aux éléments nécessaires à son développement. Aristote la décrit ainsi dans son Histoire des Animaux: « *La femelle fournit toujours le matériau et le mâle apporte ce qui permet de façonner celui-ci, de lui donner forme ; c'est cela qui, selon nous, représente le caractère spécifique de chacun des sexes ; c'est cela que veut dire être mâle ou être femelle* » .

Ces deux théories de l'épigénèse et de la préformation ont fortement influencé le monde occidental et persisteront en l'état jusqu'au 19<sup>ème</sup> siècle. Elles auront pour conséquence de renforcer la conviction sur la fixité des espèces et de favoriser les recherches sur ce qu'est l'homme dans son Essence et non dans sa nature biologique.

Il faudra attendre le 17<sup>ème</sup> et le 18<sup>ème</sup> siècle pour que le débat entre préformationnistes et partisans de l'épigénèse reprenne avec des arguments nouveaux, surtout grâce à des avancées technologiques. Cependant la théorie dominante à l'époque demeure la préformation. En effet sa simplicité est associée à des arguments forts comme la découverte, en 1672, d'œufs chez la

femme par Reinier De Graaf (1641-1673) ou encore, en 1677, d'animalcules spermatiques dans la semence mâle par Antonie Van Leeuwenhoek (1632-1723), inventeur du microscope. Il se développe alors deux théories : les ovistes qui supposent que l'individu miniature est contenu dans l'œuf et les animalculistes qui supposent que l'individu est contenu dans la semence mâle (**Figure 1**).

## **II LA THEORIE CELLULAIRE**

C'est la découverte des cellules et de leur noyau au 19<sup>ème</sup> siècle et la formulation de la théorie cellulaire qui vont véritablement fournir les arguments nécessaires pour mieux comprendre l'origine des embryons. La théorie cellulaire pose en effet le principe que le développement des parties élémentaires des organismes réside dans la formation des cellules. La cellule devient alors le point commun à tous les organismes vivants. L'observation des premières fécondations va confirmer l'existence d'une cellule œuf ou zygote qui résulte de la rencontre de deux autres cellules, l'animalcule mâle (ou spermatozoïdes) et l'œuf femelle (ou ovocytes). On pensait à cette époque que le spermatozoïde n'avait qu'un rôle d'activation de l'ovocyte lors de la fécondation. Mais c'est Oskar Hertwig (1849-1922) qui montre en 1875 l'importance du rôle des noyaux des gamètes dans la fécondation. Hertwig montre ainsi que le noyau de la cellule œuf provient de la fusion des noyaux des gamètes mâle et femelle et que les noyaux de toutes les cellules de l'embryon futur proviennent, par divisions successives, du noyau de cette cellule œuf.

L'ensemble de ces travaux rendit caduque la théorie préformationniste et renforça celle de l'épigénèse. Il restait cependant à découvrir les facteurs conférant à l'embryon ses différentes formes.

## **III LA DECOUVERTE DE LA NOTION D'HEREDITE**

Il était en effet nécessaire de comprendre comment se transmettait aux cellules les caractéristiques propres à l'espèce et à l'individu. C'est ici qu'intervient la nécessité d'aborder le problème de l'hérédité et par la même celui de l'évolution des espèces.

Durant de nombreux siècles, sous l'influence de la pensée Antique puis de la chrétienté, le paradigme général consistait à poser que les espèces préexistaient d'après un

schéma initial. Toutes les espèces étaient fixées et la notion d'évolution n'avait pas lieu d'être.

Ce n'est que relativement tard, au 18<sup>ème</sup> siècle, que le chevalier de Lamarck ( 1744-1829) repris l'idée d'une continuité entre les êtres vivants et l'expliqua par la théorie du transformisme qui regroupait deux lois, celle de la modification de l'organe par la modification des besoins et des habitudes et celle de l'hérédité des caractères acquis. Plus simplement les modifications induites par l'environnement sur l'organisme sont ensuite transmises à la descendance.

Par la suite, Charles Darwin (1809-1882), constatant la variabilité considérable du vivant, montra l'importance de la sélection naturelle dans les mécanismes contrôlant l'évolution des espèces. Il intégra l'hérédité des caractères acquis à sa théorie, et repris une vieille hypothèse d'Hippocrate sur l'hérédité, la pangénèse. En bref, ce ne sont pas les éléments reproducteurs qui engendrent les nouveaux individus, mais toutes les cellules du corps entier : elles émettent des gemmules qui s'agrègent dans les éléments reproducteurs, afin de reconstituer au moment venu des cellules équivalentes à leurs cellules d'origine.

Ce sont en fait les travaux d'August Weisman (1834-1914) qui vont vraiment apporter les concepts d'une nouvelle théorie de l'hérédité. Il introduit l'idée d'une lignée cellulaire immortelle, la lignée germinale (germen) où les cellules se succèdent pour aboutir aux gamètes. C'est elle qui transmet les caractères à travers les générations, par l'intermédiaire de de lignées cellulaires somatiques (soma) qui représentent l'ensemble des cellules de l'organisme non sexuelles. Il propose également que les cellules germinales soient le support de la variabilité observée entre les individus. La découverte des chromosomes (1888), de la réduction chromatique lors de la méiose (1887) l'amène à poser qu' : « *il existe une substance héréditaire , un véhicule matériel des tendances héréditaires et cette substance est contenue dans le noyau de la cellule germinative et dans cette partie du filament nucléaire qui, à certains moments, revêt la forme d'anses ou de baguettes courtes (chromosomes)* »(Weismann A., 1887 cité dans Le Douarin 2000).

Alors que la théorie cellulaire détruisait le mythe de la préformation, la fusion de la théorie de l'hérédité et de la théorie chromosomique a vraiment permis de développer une réflexion nouvelle sur le développement embryonnaire, annonçant la génétique moderne. Il restait alors à identifier les facteurs présents sur les chromosomes qui contrôlaient la transmission des caractères héréditaires mais aussi les déterminants cellulaires.

C'est le rapprochement entre les travaux de Gregor Mendel sur la transmission des caractères héréditaires (1866) et de la théorie chromosomique qui va permettre, en 1902, à Sutton de conclure : « *L'association des chromosomes maternels et paternels en paires suivie de leur séparation pendant la réduction chromatique(...) constituent sans doute les bases physiques des lois mendélienne de l'hérédité.* » (cité dans Le Douarin 2000).

Par la suite beaucoup de travaux dont ceux de Morgan sur la drosophile confirmèrent la théorie chromosomique de l'hérédité et permirent de définir le gène comme unité fondamentale de l'hérédité .

Enfin la découverte de la structure de l'ADN en 1953 par James Watson et Crick puis le déchiffrement du code génétique permirent d'appréhender de façon directe les mécanismes gouvernant le déterminisme cellulaire en démontrant notamment le codage des protéines par les gènes (**Figure 2**).

#### **IV UNE PORTE OUVERTE POUR L'EXPERIMENTATION**

L'histoire de la biologie nous montre donc que les révolutions majeures qui ont bouleversé cette discipline sont relativement récentes et remontent pour la plupart au 19<sup>ème</sup> siècle. La question même de la pluripotence cellulaire ne pouvait émerger en l'absence de la connaissance de l'existence des cellules.

Cette époque est aussi le théâtre de l'émergence d'une nouvelle approche expérimentale qui ouvre un nouveau champ de connaissance. Celle-ci pose comme principe que l'expérience *in vitro* reproduit ce qui se passe au sein même de l'embryon. Il est donc possible de réduire le modèle d'étude de l'organisme entier au tissu voire à la cellule. D'autre part elle pose également le principe que les perturbations induites par l'expérimentateur sur un système vivant fournit des informations sur le fonctionnement normal de ce système.

Cette nouvelle façon d'appréhender le développement embryonnaire développé par Wilhelm Roux en 1894 va permettre d'aborder la question de la pluripotence.

Quel est en effet le devenir et les caractéristiques d'une ou plusieurs cellules embryonnaires sorties de leur contexte embryonnaire ?

---

## PARTIE A

# LA MISE EN EVIDENCE DE LA PLURIPOTENCE CELLULAIRE ET LA DECOUVERTE DES CELLULES SOUCHES EMBRYONNAIRES

---

## I LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES

### 1°) MISE EN EVIDENCE DE LA PLURIPOTENCE DES BLASTOMERES CHEZ LES ANIMAUX

- Totipotence des blastomères d'oursin

Les premiers travaux qui montrèrent la pluripotence des cellules embryonnaires furent publiés par Driesch en 1891. L'auteur a isolé des blastomères d'embryons d'oursin aux stades 2 et 4 cellules. Ceux ci, comme un zygote, vont ensuite se développer normalement pour donner des larves normales mais de tailles réduites. On parle alors de totipotence car une cellule unique est capable de produire un nouvel organisme. A des stades plus tardifs cependant, les blastomères isolés ne sont plus capables de produire une larve viable.

- Totipotence et pluripotence des cellules embryonnaires de triton

Chez les Vertébrés, Speeman reproduisit en 1902 les expériences réalisées par Driesch sur l'embryon d'oursin et montra que deux blastomères isolés à partir d'un embryon de triton au stade 2 cellules étaient totipotents, donc capables de redonner deux têtards viables. Il montra aussi qu'une larve de triton, au stade 4, 8 ou 16 blastomères, scindée en deux, est capable de régénérer deux embryons entiers. Les blastomères à ces stades embryonnaires sont qualifiés de pluripotent, c'est à dire qu'il participe à la formation des différents tissus embryonnaires mais sont incapable, *per se*, de produire un nouvel organisme.

Ces travaux montrent donc que la destinée des cellules embryonnaires précoces aussi bien chez des invertébrés que des vertébrés ne semble pas figée dès les premières divisions cellulaires. Les cellules demeurent pluripotentes et sont capables de compenser partiellement ou totalement l'absence des autres cellules embryonnaires.



## 2°) LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES DE MAMMIFERES

Les cellules embryonnaires de mammifères présentent-elles les mêmes propriétés que celles d'oursin ou de triton ? La question présente un intérêt particulier puisqu'elle permet d'aborder l'étude du développement embryonnaire humain, mais l'accessibilité difficile des embryons de mammifères a certainement handicapé les études menées sur ce modèle. En effet, le développement des embryons de mammifères diffère des autres animaux car il est exclusivement interne, dans l'organisme maternel. Il est divisé en deux phases : la phase préimplantatoire où l'embryon est en « suspension » dans l'oviducte puis l'utérus et la phase d'implantation où l'embryon adhère à la matrice utérine et développe un tissu d'échange entre lui et l'organisme maternel, le placenta.

Après la fécondation, l'embryon de mammifère va donc se diviser jusqu'à former une petite grappe de cellules appelée morula. Ensuite celle-ci va se compacter pour former deux tissus, le trophoblaste, qui est composé des cellules les plus extérieures de la morula et qui donnera par la suite les tissus extra-embryonnaires, et la masse cellulaire interne (MCI), qui dérive des cellules les plus internes de la morula et qui donnera les cellules du futur embryon. Le trophoblaste va sécréter un liquide à l'origine d'une cavité : le blastocoele. La MCI va alors donner deux tissus, l'ectoderme embryonnaire ou épiblaste et l'endoderme primitif qui donnera certains tissus extraembryonnaires (revue dans Lu et *al.* 2001). L'embryon est alors appelé blastocyste. L'embryon va ensuite s'implanter dans la matrice utérine par l'intermédiaire de récepteurs spécifiques et poursuivre son développement (**Figure 3 et 4**).

Les expérimentations menées sur les embryons de mammifères ont vraiment été rendues possibles grâce à la maîtrise de leur culture *in vitro*. C'est en 1965 que sont établies les conditions requises pour que les embryons de souris puissent accomplir *in vitro* les phases du développement allant de l'œuf fécondé jusqu'au blastocyste (Brinster 1965). Cette méthode de culture a ensuite été étendue à d'autres espèces de mammifères comme le lapin, le rat, le mouton, les bovins ou l'homme.

Grâce à ces avancées techniques, les travaux portant sur la pluripotence des cellules de l'embryon des mammifères ont pu être réalisés sur l'embryon précoce de mammifères préimplantatoire, les phases de développement entre l'implantation et la gastrulation demeurant difficile à étudier (entre 4,5 et 6,5 jours post coïtum ou p.c.).

Les premières expériences qui validaient chez les mammifères les résultats de Speeman et de Driesch ont été réalisées avec des embryons de lapins au stade 2 cellules (Seidel 1952). La dissociation des deux blastomères redonnait naissance à deux embryons entiers normaux. D'autre part, la scission d'un blastocyste en deux embryons redonnent deux embryons viables à condition que chacun des embryons scindés contiennent des cellules de la masse cellulaire interne.

Ces résultats montrent que les cellules de la masse cellulaire interne du blastocyste ne sont pas prédéterminées à donner tel ou tel organe ou tissu mais qu'elles sont capables de potentialités de développement très étendues.

### 3°) LES CHIMÈRES, VÉRITABLE DÉMONSTRATION DE LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES

Les chimères sont des animaux composés de cellules provenant de plusieurs organismes différents. Elles possèdent donc des cellules génétiquement différentes que l'on peut reconnaître à l'aide de marqueurs spécifiques. Ces animaux sont créés artificiellement par l'homme et sont un outil très intéressant pour suivre les potentialités de cellules introduites dans un organisme receveur. Les chimères vont se révéler être un outil très puissant pour démontrer la pluripotence des cellules embryonnaires de mammifères.

#### a) Les chimères d'agrégation

Les chimères d'agrégation consistent à fusionner plusieurs embryons. Les premiers travaux menés sur les chimères d'agrégation ont montré la capacité de deux morulas (8 cellules) de souris à fusionner et à produire ensuite des embryons viables, composés de cellules provenant des deux morulas utilisées. Ainsi, à partir d'une morula issue d'une souche de souris albinos et agrégée avec une morula d'une souche de souris à pelage noir, des souris chimères ont été obtenues (**Figure 5**). Ces souris présentent un pelage comportant à la fois des zones blanches et des zones noires. (Tarkowsky 1961, Mintz et *al.* 1962). Ultérieurement des travaux ont montré que des chimères issues de l'agrégation de trois morulas portant chacune des marqueurs différents (pelage blanc; noir et brun) sont capables de transmettre les caractéristiques de chacune des morulas d'origine à leur descendance (Markert et *al.* 1977).

Les différentes cellules ont donc contribué à la formation du pelage mais aussi à la colonisation de la lignée germinale.

La production de chimères est un outil important pour étudier les lignages cellulaires et les contributions respectives d'un type cellulaire par rapport à un autre. Elles nous montrent ici que les cellules issues de la morula sont interchangeables et qu'elles peuvent toutes contribuer à l'ensemble des tissus formant l'embryon. Ces cellules peuvent donc être qualifiées de pluripotentes.

Cependant la technique par agrégation n'étant pas fonctionnelle avec des blastocystes, il était impossible de savoir par cette approche si les cellules de la masse cellulaire interne du blastocyste étaient, comme les cellules issues de la morula, capables de coloniser les différents tissus embryonnaires.

#### b) Les chimères d'injection

Afin de tester la contribution des cellules de la masse cellulaire interne dans un embryon receveur, une technique de microinjection de ces cellules dans un blastocyste hôte a été mise au point (Gardner 1968). Elle consiste à dissocier les cellules de la masse cellulaire interne et à les réinjecter dans le blastocoele d'un embryon receveur (**Figure 6**).

Les cellules injectées possèdent un ou plusieurs marqueurs qui permettent de les différencier des cellules issues de l'embryon hôte. Pour montrer l'efficacité de sa méthode, Gardner a utilisé des blastocystes receveurs issus de souris albinos et des cellules donneuses issues de blastocystes provenant de souris avec une translocation chromosomale et un marqueur de pigmentation. Il obtint de nombreuses souris mosaïques.

Grâce à cette méthode il a démontré que les cellules de la masse cellulaire interne du blastocyste pouvaient contribuer à l'ensemble des tissus dérivés de l'œuf, à l'exception du trophoctoderme.

La colonisation des différents tissus de l'embryon de souris par les cellules de la masse cellulaire interne montre donc que ces cellules sont pluripotentes. Il est intéressant de remarquer que, contrairement aux blastomères d'amphibiens ou d'oursins, les cellules de la masse cellulaire interne forment un groupe d'équivalence, c'est à dire que chaque cellule possède les mêmes potentialités de différenciation. Chez les amphibiens ou chez l'oursin, la

segmentation des blastomères va très vite restreindre leurs potentialités de différenciation et tous les blastomères d'un embryon au stade 8-16 cellules ne sont pas équivalents.

#### 4°) LA PLURIPOTENCE DES CELLULES DU BLASTODERME DANS L'EMBRYON DE POULET

Chez le poulet, l'œuf est pondu 20 à 23 heures après la fécondation. Le développement embryonnaire précoce est divisé en 14 stades, numérotés en chiffres romains (stade I à XIII décrits par Eyal-Giladi et Kochav en 1976). Au cours de cette période l'œuf va subir de nombreuses divisions cellulaires. Au moment de la ponte, l'embryon, au stade X est composé de 40 000 à 80 000 cellules indifférenciées. Il a la forme d'un disque, le blastoderme. Le vitellus est séparé de l'épiblaste dans sa partie centrale par une cavité, le blastocœle et est en contact avec le vitellus dans sa partie périphérique. Ces deux parties correspondent respectivement à l'aire pellucide et à l'aire opaque (**Figure 7**).

Après la ponte, si l'œuf est placé en incubation à 38,5°C, son développement va se poursuivre. L'épiblaste va subir de nombreux remaniements qui vont permettre la formation des trois feuilletts embryonnaires contribuant à l'embryon proprement dit, alors que les cellules de l'aire opaque vont surtout contribuer à la formation des tissus extra-embryonnaires (**Figure 8**). Le développement embryonnaire après la ponte est divisé en 46 stades indiqués en chiffres arabes (stades 1 à 46 d'après Hamburger et Hamilton, 1951).

La pluripotence des cellules du blastoderme de poulet a été montrée dès 1960 par Spratt et Haas. Les auteurs ont montré qu'un blastoderme dissocié en quatre parties pouvait redonner quatre embryons entiers.

Par la suite il a été montré que les cellules du blastoderme de poulet pouvaient coloniser un embryon hôte (Marzullo, 1970). A partir d'amas cellulaires d'embryons issus d'œufs fraîchement pondus, l'auteur identifia 3 chimères sur 239 embryons receveurs injectés (1,25%). Cependant, si les cellules colonisaient l'embryon receveur, aucun poussin n'éclopait.

Afin d'améliorer les expériences de Marzullo, ces travaux ont été reproduits à partir d'embryons identifiés au stade X (EG) et dont les cellules du blastoderme ont été dissociées à la trypsine (Petitte et *al.*, 1990). L'embryon receveur est au même stade que celui dont sont issues les cellules à injecter. Sur 53 embryons injectés avec 200 à 500 cellules, les auteurs obtinrent un poussin chimère (1,9%) capable de transmettre à sa descendance les génotypes des deux embryons dont il était issu (plume blanche pour le dwarf white leghorn et plume grise pour le Barred Plymouth rock) et 6 embryons chimères (11%) morts avant l'éclosion.

Afin d'augmenter ces rendements et de faciliter la colonisation de la lignée germinale de l'embryon receveur par les cellules injectées, la même équipe a utilisé des embryons receveurs préalablement irradiés afin de freiner leur développement (Carsience et *al.*, 1993). Sur 809 embryons injectés avec 100 cellules, 118 chimères furent identifiées (15%) et sur 296 embryons injectés avec 200-400 cellules, 55 chimères furent également identifiées (19%). Sur cette dernière expérience, 3 femelles et 5 mâles (2.7%) transmettaient le génotype des cellules injectées à leur descendance illustrant ainsi la colonisation de la lignée germinale par ces cellules. Un mâle possédait même une colonisation très forte de sa lignée germinale car toute sa descendance possédait le génotype des cellules injectées.

La colonisation des différents tissus de l'embryon de poulet par les cellules de l'épiblaste montre donc que ces cellules sont pluripotentes.

Si la mise en évidence de la pluripotence des cellules de la masse cellulaire interne chez la souris ou de l'épiblaste chez le poulet était acquise, la conservation *in vitro* de la pluripotence de ces cellules mises en culture restait à réaliser

## **II L'ISOLEMENT DES CELLULES SOUCHES EMBRYONNAIRES**

### 1°) LES CELLULES EMBRYONNAIRES CARCINOMALES (EC CELLS)

Les cellules embryonnaires carcinomales sont des cellules dont les caractéristiques sont très proches de celle des cellules de la masse cellulaire interne de l'embryon.

Ces cellules sont dérivées de tumeurs appelées tératocarcinomes. Elles se développent dans le testicule ou l'ovaire à partir d'une cellule germinale qui entre en prolifération. Ces tumeurs sont composées de plusieurs types cellulaires : des cellules différenciées en types cellulaires variés (neurones, muscles, os...) et des cellules épithéliales, prolifératives et indifférenciées, à l'origine des cellules différenciées présentes dans la tumeur. Malheureusement, même si ces tumeurs étaient connues depuis longtemps et semblaient être un modèle intéressant pour l'étude du développement embryonnaire, leur rareté en a durablement limité l'étude.

C'est en 1954 que Stevens et Little publient que 1% des mâles de la lignée 129 de souris développent des tératomes testiculaires. Cette découverte fut le point de départ des travaux qui permirent d'isoler et de caractériser les cellules EC puis les cellules souches embryonnaires (ES cells).

Stevens montra également que la fréquence des tératocarcinomes pouvait être augmentée en greffant des crêtes génitales dans le testicule de souris adulte (Stevens, 1964).

Les premières cultures de cellules EC à partir de tératocarcinomes de souris datent du début des années 1970 (Kahan et Ephrussi 1970, Evans 1972). Ces cellules peuvent être maintenues *in vitro* mais sont aussi capables de reformer des tératocarcinomes quand elles sont réinjectées dans un hôte. Ces cellules ont également tendance à se différencier spontanément en culture et l'utilisation d'une couche cellulaire nourricière composée de fibroblastes embryonnaires de souris (feeder) a permis de maintenir ces cellules dans un état indifférencié (Martin et Evans 1975). Le retrait du feeder induit la différenciation de ces cellules qui, en suspension, s'agrègent et forment des corps embryoïdes dans lesquels les cellules vont se différencier en des types cellulaires très divers (Martin et Evans 1975). Le feeder doit donc produire un ou des facteurs stimulant la prolifération cellulaire et inhibant la différenciation. Ces cellules ont également permis de montrer qu'il existait des molécules capables d'orienter la différenciation. L'acide rétinoïque, par exemple, permet de différencier des cellules EC de la lignée F9 en endoderme pariétal (Strickland et al. 1980), des cellules de la lignée P19 en neurones (Jones-Villeneuve et al. 1982).

Les caractéristiques des cellules EC rappellent fortement celles des cellules embryonnaires précoces. A l'image des cellules de la masse cellulaire interne du blastocyste, elles sont aussi capables de coloniser des tissus embryonnaires lorsqu'elles sont réinjectées dans un blastocyste (Papaioannou et al. 1975). Elles perdent ainsi leur tumorigénicité et, comme le montre les travaux de Beatrice Mintz, même conservées à l'état de tumeurs durant 8 ans dans des générations de souris successives, ces cellules de tératocarcinomes sont capables de coloniser un embryon et de donner des souris mosaïques viables avec une colonisation de la lignée germinale (Mintz et Illmensee 1975). Cela démontrait aussi que le caractère malin des cellules EC pouvait être réversible.

La même expérience fut ensuite réalisée à partir de cellules EC en culture depuis au moins 9 passages. Après injection dans un blastocyste, ces cellules sont capables de coloniser

toutes les lignées cellulaires de l'embryon dont la lignée germinale (Stewart et Mintz, 1981, Bradley et *al.*, 1984). Les auteurs proposaient alors d'utiliser ces cellules pour la production d'animaux transgéniques.

Un inconvénient majeur de ces cellules réside cependant dans l'instabilité de leur information génétique. En effet la plupart des cellules EC présentent des caryotypes anormaux et perdent en culture leur capacité de différenciation et de colonisation de l'embryon (Martin 1980).

## 2°) LES CELLULES SOUCHES EMBRYONNAIRES (CELLULES ES) DE SOURIS

Les résultats obtenus grâce aux chimères d'injection et l'expérience acquise par la culture des cellules EC ont sans doute facilité l'obtention des premières lignées de cellules souches embryonnaires en 1981 (Evans et Kaufman 1981 ; Martin 1981). A partir de souris de la lignée 129 (dont sont issues les cellules EC) et à l'aide des marqueurs connus et spécifiques des cellules EC (antigènes de surface reconnus par le polyclonal M1-22-25 et l'anti-I-Ma) Evans et Kaufman ont défini les stades embryonnaires les plus adéquats à la mise en culture de cellules pluripotentes. Ils ont déduit de leurs observations une fenêtre optimale autour de 5 jours et demi post coïtum, juste après l'implantation. Les embryons étant difficile à prélever dans ces conditions, ils ont induit *in vitro* la maturation de blastocystes, eux-même en diapause par l'ovariectomie des femelles gestantes. Après cette étape de culture, ils obtinrent des embryons avec des cellules morphologiquement proches de celles de l'épiblaste d'un embryon de 5.5 jours. En présence de cellules nourricières (feeder), ces cellules prolifèrent et demeurent indifférenciées. Elles ressemblent fortement aux cellules EC et en présentent d'ailleurs de nombreuses caractéristiques : mêmes marqueurs de surface, même tumorigénicité sous forme de tératocarcinomes lorsqu'elles sont injectées dans le flanc d'une souris adulte, même capacité à former des corps embryoïdes.

La même équipe a montré par la suite la capacité de ces cellules à coloniser tous les lignages cellulaires, dont la lignée germinale d'un embryon hôte, et ce, même après une longue période de culture (Bradley et *al.* 1984).

Ces résultats confirment que les cellules ES sont pluripotentes et qu'elles peuvent répondre de façon appropriées aux facteurs de régulation du développement embryonnaire,

aussi bien en terme de prolifération que de différenciation et migration. Elles conservent ainsi toutes les caractéristiques des cellules dont elles sont issues, les cellules de l'épiblaste.

Par la suite, les conditions de cultures des cellules ES ont été améliorées et l'on a pu remplacer la couche de cellules nourricières par l'emploi de milieux conditionnés (Smith et Hooper 1983). En fait une cytokine purifiée, appelé LIF pour Leukemia inhibitory Factor permet de maintenir l'autorenouvellement des cellules ES indifférenciées (Smith et *al.* 1988). Cet effet est limité au LIF et à d'autres cytokines de la famille de l'interleukine-6 activant le récepteur gp130 (Yoshida et *al.* 1994).

L'ensemble de ces résultats permet donc de donner une définition des cellules souches embryonnaires :

Même après de longue période de culture, ces cellules sont capables :

- De s'autorenouveler indéfiniment, sans altération de leur phénotype, ni de leur caryotype euploïde
- de se différencier *in vitro* dans de nombreux types cellulaires représentatifs des trois feuillets embryonnaires : ectoderme, mésoderme et endoderme
- de participer *in vivo* à la formation de toutes les lignées cellulaires de l'embryon dont la lignée germinale lorsqu'elles sont réinjectées dans un embryon précoce
- de participer à la formation de tumeurs (tératocarcinome) lorsqu'elles sont réinjectées dans un organisme adulte.

Les cellules ES présentent un rapport nucléo-cytoplasmique élevé et sont de taille relativement petites (environ 10 µm). Elles poussent sous la forme d'agrégats plus ou moins réfringents suivant la lignée cellulaire et les conditions de culture utilisées.

### 3°) LES CELLULES SOUCHES EMBRYONNAIRES DE POULET

Comme on l'a vu précédemment, la pluripotence des cellules du blastoderme de l'embryon de poulet a été démontrée dès 1970 (Marzullo 1970). La maîtrise de la culture des cellules souches embryonnaires de souris et la découverte de l'importance du LIF pour



l'autorenouvellement de ces cellules (Smith et *al.*, 1988) a ensuite permis d'étendre ces résultats au poulet.

En 1996 la première culture de cellules ES de poulet était publiée (Pain et *al.*, 1996). Les cellules mises en culture dérivait d'embryons de poulet fraîchement pondus au stade X (EG). Comme pour les cellules ES murines, les cellules de poulet étaient cultivées sur un feeder de cellules STO irradiées et dans un milieu complété en LIF mais aussi en Interleukine 11 (IL-11), Stem Cell Factor (SCF), Insulin like growth factor I (IgF1) et basic fibroblast growth factor (bFGF). Ces différentes cytokines ont été testées pour leur effet sur l'autorenouvellement des cellules embryonnaires de poulet. Il est troublant de constater que ce milieu est finalement plus proche de celui de culture des cellules embryonnaires germinales (cellules EG) de souris que de celui des cellules ES de souris. Les cellules embryonnaires germinales résultent en effet de la mise en culture des cellules germinales primordiales, prélevées dans l'embryon au niveau des crêtes génitales (10.5 à 12.5 jours post coïtum). Ces cellules prolifèrent en présence de LIF, de SCF et de bFGF. Elles présentent aussi tous les critères de pluripotence exposés précédemment : capacité à former des chimères et colonisation de la lignée germinale, autorenouvellement illimité, différenciation *in vitro*, tumorigénicité (Matsui et *al.*, 1992 ; Resnick et *al.* 1992).

A l'image des cellules ES et des cellules EG, les cellules embryonnaires de poulet expriment de nombreux marqueurs spécifiques des cellules pluripotentes : activité de la phosphatase alcaline élevée, activité de la télomérase élevée, expression des antigènes de surfaces reconnus par les anticorps monoclonaux SSEA-1, EMA-1, ECMA-7.

Elles sont aussi capables de former des corps embryoïdes *in vitro* et de se différencier en de nombreux types cellulaires différents : cellules neuronales, hématopoïétiques ou musculaires. *In vivo*, réinjectées dans une blastula de poulet irradiée (stade X), ces cellules donnent des oiseaux chimères mosaïques, dont certains transmettent à leur descendance le génotype des cellules donneuses.

Ces résultats montrent qu'il est possible d'isoler et de cultiver sur une période de temps élevée des cellules embryonnaires de poulet pluripotentes.

Il est cependant intéressant de s'interroger sur la nature réelle de ces cellules. Sont-elles réellement des cellules ES ou plutôt des cellules EG issues des cellules germinales primordiales présentes dans l'embryon de poulet au stade X ? Plusieurs arguments peuvent être en faveur de la seconde hypothèse : tout d'abord l'approche expérimentale de la culture

des cellules ES de poulet montre la nécessité d'ajouter les facteurs de croissance et cytokines SCF, bFGF et LIF dans le milieu de culture, comme pour la culture des cellules EG de souris (Matsui et *al.* 1992) ou celle des EG de poulet (Park et *al.* 2000). D'autre part, de nombreux travaux avancent l'hypothèse que les cellules germinales primordiales (Primordial germ cells ou PGC) de poulet seraient déjà présentes dans l'embryon de poulet au stade X (EG) (Eyal-Giladi et *al.* 1981, Urven et *al.* 1988, Kagami et *al.* 1997). Urven et *al.* ont ainsi utilisé l'anticorps EMA-1 qui reconnaît une glycoprotéine à la surface des cellules EC murines (Hahnel et Eddy 1983) mais également les cellules PGC de souris (Hahnel et Eddy 1987) pour localiser les cellules PGC de poulet. Ils montrent un marquage très fort des cellules de l'épiblaste. Cependant il est intéressant de souligner que l'anticorps EMA-1 reconnaît également les cellules de la masse cellulaire interne du blastocyste et les cellules ES de souris (Pain et *al.* 1996). Kagami et *al.* apportent une démonstration plus convaincante. En enlevant les cellules du centre du blastoderme d'un embryon de poulet au stade X (EG), ils montrent que le nombre de PGCs présentes dans un embryon au stade 15 est très fortement réduit suggérant ainsi que les PGCs se localiseraient au centre de l'épiblaste et non dans sa partie marginale.

En conclusion, il est difficile de savoir si les cellules en culture issues de blastoderme d'embryon au stade X (EG) sont des cellules ES ou des cellules EG. De toutes façon, il est difficile de comparer le blastoderme de poulet avec les cellules de la masse cellulaire interne du blastocyste. En effet, comme nous le verrons ultérieurement, l'épiblaste de l'embryon de poulet n'est pas une structure uniforme, où les cellules seraient équivalentes et soumises aux mêmes voies de signalisation.

Conscients d'ailleurs de cette ambiguïté, Pain et *al.* ont préféré appeler ces cellules CEC pour chicken embryonic cells, laissant ainsi le champs libre à toute interprétation possible.

---

## **PARTIE B**

# **CARACTERISATION MOLECULAIRE DES CELLULES SOUCHES EMBRYONNAIRES**

---

Un des objectifs de cette thèse est d'identifier des gènes importants pour le maintien de la pluripotence dans les cellules ES de poulet et de souris. Afin d'illustrer le contexte scientifique de cette démarche, un bilan des travaux réalisés sur le maintien de la pluripotence des cellules ES de souris sera présenté en détaillant plus particulièrement le contrôle de l'autorenouvellement des cellules ES par le Leukemia Inhibitory Factor ou LIF et celui de la pluripotence par Oct-3/4. Peu de données étant disponibles sur les cellules ES non murines, l'analyse des données disponibles sur l'embryon de poulet au stade X apportera quelques indications sur le statut moléculaire des cellules embryonnaires dont sont issues les cellules ES de poulet.

## **I LE CONTROLE DE L'AUTORENOUVELLEMENT DES CELLULES ES DE SOURIS**

### **1°) LE LIF ET SA VOIE DE SIGNALISATION**

#### **a) Le LIF et les cytokines de la famille IL-6**

Les cellules ES sont des cellules se divisant de façon symétrique avec un taux de prolifération relativement élevé (8 heures par cycle, Savatier et *al.* 1994). Le maintien de cet autorenouvellement a été montré chez la souris comme étant dépendant d'une cytokine, le LIF pour Leukemia Inhibitory Factor (Smith et *al.* 1988, Williams et *al.* 1988). Le LIF murin est une glycoprotéine de 180 acides aminés qui a été initialement mise en évidence comme bloquant la prolifération et induisant la différenciation en macrophage de cellules leucémiques myéloïdes murines M1, d'où son nom. Il appartient à une famille de cytokines qui regroupent notamment les interleukines IL-11, IL-6, le CNTF (ciliary neurotropic factor), le CT-1 (cardiotrophin-1) ou l'OSM (oncostatin M). Son récepteur, LIFR, appartient à la famille des

récepteurs  $\beta$  aux cytokines et est associé avec la protéine gp130 ( Gearing et *al.*, 1991), qui interagit également avec les autres récepteurs  $\beta$  de la famille de l'IL-6. (**Figure 9**) La redondance observée au niveau des récepteurs explique probablement pourquoi l'inactivation du gène codant pour le LIF n'est pas létale alors que celle de gp130 l'est (Yoshida et *al.* 1996). C'est en fait l'activation de la voie gp130 qui est primordiale pour maintenir l'autorenouvellement des cellules ES et ce, quelle que soit la cytokine et le récepteur associé à gp130 (Yoshida et *al.* 1994, Niwa et *al.* 1998).

#### b) Les protéines STAT et leur activation par le LIF

Les récepteurs  $\beta$  aux cytokines ne possèdent pas d'activité tyrosine kinase dans leur partie intracellulaire. Il existe en fait des tyrosines kinases associées aux récepteurs. Par immunoprécipitation, il a été montré qu'il s'agissait des JAKs pour « Janus kinases » (Stahl et *al.* 1994). C'est la fixation du ligand et l'hétérodimérisation des récepteurs du LIFR avec gp130 qui va entraîner l'activation des JAKs par transphosphorylation (Stahl et *al.* 1994). Celles-ci vont alors phosphoryler les tyrosines présentes dans la région cytoplasmique des récepteurs aux cytokines (Stahl et *al.* 1995). D'autres tyrosine kinases, les HCK pour Hematopoietic Cell Kinase peuvent également transactiver les récepteurs  $\beta$  dans les cellules ES, et ce, de la même façon que les JAKs (Ernst et *al.*, 1996).

Les facteurs de transcription STAT (Signal Transducer and Activator of Transcription) reconnaissent les tyrosines phosphorylées des récepteurs et s'y fixent par leur domaine SH2. Le domaine SH2 (pour Src homology domain 2) confère aux protéines STAT la propriété de transmettre le signal directement de la membrane vers le noyau, sans autre intermédiaire. Dans les cellules ES, ce sont majoritairement les facteurs STAT3 qui vont se fixer aux récepteurs phosphorylés. Ils vont alors être activés par la phosphorylation de leur tyrosine Y705 par les JAKs (Lutticken et *al.* 1994)(**Figure 10**).

Après dissociation des récepteurs, les facteurs STAT3 vont s'homodimériser au niveau de leurs domaines SH2 et de leurs phosphotyrosines. Ces dimères migrent ensuite vers le noyau et vont se fixer sur des séquences cibles pour activer la transcription de certains gènes. Tous les facteurs de transcription STAT peuvent se lier à des variants de la séquence consensus palindromique TTCnnnGAA par l'intermédiaire de leur domaine de fixation à l'ADN (Zhong et *al.*, 1994).

## 2°) ROLE DE STAT3 DANS LE CONTROLE DE L'AUTORENOUVELLEMENT DES CELLULES ES

- a) STAT3 est nécessaire et suffisant au maintien de l'autorenouvellement des cellules ES

Afin de montrer l'importance de STAT3 dans le contrôle de l'autorenouvellement et de la pluripotence des cellules ES, Niwa *et al.* (1998) ont produit des récepteurs gp130 mutés pour différentes tyrosines de leur domaine intracellulaire. L'expression de ces récepteurs mutants montre que l'absence de site de fixation pour STAT3 entraîne un arrêt de l'autorenouvellement et une perte de la pluripotence des cellules.

D'autre part, les auteurs montrent que le facteur STAT3 est absolument nécessaire pour l'autorenouvellement des cellules ES. Ils utilisent dans ce but un mutant dominant négatif de STAT3, STAT3F, où la tyrosine Y705 est remplacée par une phénylalanine, empêchant ainsi la dimérisation et la translocation des facteurs. La surexpression, inductible ou non, de ce dominant négatif, entraîne systématiquement la perte de la pluripotence des cellules ES. D'autre part, ce phénotype peut être compensé par une expression forte de STAT3 qui entre alors en compétition avec le mutant et restaure le phénotype pluripotent. Une surexpression de STAT1 par contre, ne compense pas cette perte de fonction. Ces résultats montrent que le facteur STAT3 est nécessaire à l'autorenouvellement des cellules et que cette fonction lui est spécifique.

Une autre équipe de recherche a montré que le facteur STAT3 était non seulement nécessaire mais aussi suffisant pour maintenir l'autorenouvellement et la pluripotence des cellules ES, en l'absence de LIF. Pour cela, ils ont créé un facteur STAT3 fusionné au domaine de fixation du ligand (LBD) du récepteur à l'oestradiol (ER). Cette protéine chimère présente l'avantage de pouvoir répondre au ligand synthétique 4'-hydroxy-tamoxifène (4'OHT) mais pas aux oestrogènes endogènes, le LBD étant muté pour ne pas les fixer. En présence de 4'OHT, le facteur STAT3 s'homodimérise au niveau des régions LBD, ce qui induit sa translocation nucléaire et l'activation de gènes cibles. Dans ce système, l'activité du facteur STAT3 est donc totalement indépendante de la voie LIF/gp130. Les auteurs montrent ainsi que les cellules ES exprimant le facteur STAT3-ER peuvent être maintenues en autorenouvellement en présence de 4'OHT et en absence de LIF. Ces cellules maintenues en culture en présence de 4'OHT durant 10 passages et injectées dans un blastocyste donnent des souris chimères. L'activité du facteur STAT3 est donc suffisante pour maintenir

l'autorenouvellement et la pluripotence des cellules ES, en l'absence de la voie LIF/gp130 activée par le LIF (Matsuda et *al.* 1999).

#### b) Les gènes cibles de STAT3

Ces résultats suggèrent que les gènes cibles de STAT3 sont responsables du phénotype pluripotent des cellules ES de souris. Cependant peu de gènes régulés par STAT3 ont été identifiés. L'expression des gènes *SOCS3*, *c-fos*, *c-jun* et *junB* est induite par le LIF dans des cellules ES (Duval et *al.* 2000). Le rôle des oncogènes *c-fos*, *c-jun* et *junB* n'a pas encore été analysé mais il est possible que ces trois gènes jouent un rôle dans le contrôle du cycle cellulaire des cellules ES. Le gène *SOCS3* (Suppressor of cytokine signaling-3) réprime l'activité de STAT3 et induit ainsi une boucle d'auto-inactivation de la voie LIF/gp130 dans les cellules ES. Une autre étude a montré que l'expression de la protéine transmembranaire CD9 était également contrôlée par STAT3 (Oka et *al.* 2002) Cette protéine jouerait probablement un rôle en favorisant l'interaction entre les cellules. Une étude à grande échelle des gènes induits par le LIF dans les cellules ES reste à réaliser pour avoir une vision plus exhaustive des facteurs impliqués dans le contrôle de l'autorenouvellement et de la pluripotence des cellules ES.

### 3°) LA VOIE ERK BLOQUE L'AUTORENOUVELLEMENT DES CELLULES ES

L'activation des facteurs STATs n'est pas la seule conséquence de la fixation du LIF sur son récepteur. Une autre voie de signalisation est aussi activée par le LIF au niveau du récepteur gp130 dans les cellules ES, et aboutit à la stimulation de la voie Ras/MAPK (mitogen-activated protein kinase).

#### a) Présentation rapide de la voie Ras/MAPK

La voie Ras/MAPK est activée par le recrutement de Grb2 et du facteur SOS sur des récepteurs membranaires activés. La localisation de Sos à la membrane entraîne l'activation de Ras. Il s'ensuit alors une cascade de transphosphorylation impliquant Raf et la kinase de MAPK, MEK et induisant l'activation de ERK. Cette protéine peut phosphoryler des facteurs de transcription cytoplasmiques comme CREB mais peut aussi agir directement, après

translocation dans le noyau, comme un modulateur de l'activité de régulateurs transcriptionnels comme Elk, Ets, Myc et SRF (serum response factor).

b) Rôle de la voie de signalisation dépendante de ERKs dans le contrôle de l'autorenouvellement des cellules ES

Plusieurs résultats montrent que la voie de signalisation ERK est activée par le LIF dans des cellules ES en culture. Ainsi les kinases ERK1 et ERK2 sont phosphorylées de façon additive dans les cellules ES après ajout du LIF (Bœuf et *al.* 2001). Le facteur de transcription CREB (cAMP Responsive Element Binding) est également phosphorylé après addition du LIF. D'autre part l'emploi d'un inhibiteur de MEK, U0126, bloque l'activité kinase des MEKs et donc la phosphorylation de CREB mais pas la phosphorylation de STAT3 (Bœuf et *al.* 2001). LE LIF peut donc activer deux voies de signalisations bien distinctes dans les cellules ES. Quelle est donc la fonction de cette voie de signalisation dépendante des protéines ERK pour la biologie des cellules ES ?

La fonction de cette voie de signalisation dépendante des protéines ERK a été étudiée à l'aide d'inhibiteurs spécifiques des protéines kinases. L'inhibition de la phosphorylation des protéines ERKs par l'inhibiteur PD098059 qui bloque l'activité kinase de MEK stimule l'autorenouvellement des cellules ES de manière dose dépendante et ne modifie pas la pluripotence de ces cellules. Cet inhibiteur augmente aussi la proportion de cellules indifférenciées dans des corps embryoides en culture (Burdon et *al.* 1999). Ces résultats suggèrent que la voie de signalisation dépendante des ERKs n'est pas impliquée dans le maintien de l'autorenouvellement et de la pluripotence des cellules ES mais aurait plutôt comme rôle d'engager les cellules en différenciation.

c) Rôle de SHP2 dans l'activation de la voie de signalisation dépendante de ERK

On a vu que la voie ERK était activée par le LIF et jouait un rôle dans l'engagement des cellules en différenciation. Comment peut-on relier ces deux mécanismes au niveau moléculaire?

La protéine SHP2 est en fait la charnière moléculaire entre l'activation de gp130 et du LIFR et l'activation de la cascade de phosphorylation qui activent les protéines ERKs.

Il existe au niveau du récepteur gp130 d'autres tyrosines phosphorylées qui ne sont pas utilisées par STAT3 (Niwa et *al.* 1998). Ainsi une tyrosine en position 118 est reconnue grâce à son domaine SH2 par une tyrosine phosphatase cytoplasmique, SHP2 (Stahl 1995). La protéine SHP2 joue un rôle important dans le contrôle de l'autorenouvellement des cellules ES. En effet, si l'on supprime son site de fixation sur le récepteur gp130, on augmente l'autorenouvellement des cellules. De même la suppression de son activité phosphatase (remplacement d'une cystéine de la région catalytique par une sérine) entraîne une augmentation du nombre de colonies indifférenciées (Burdon et *al.* 1999). Ceci peut être expliqué d'une part par l'effet direct de SHP2 qui régule négativement l'activité des JAKs (Symes et *al.* 1997). D'autre part SHP2 active aussi la voie de signalisation dépendante des protéines ERKs, également impliquée dans le contrôle de l'équilibre autorenouvellement/différenciation. En effet, la suppression de l'activité SHP2 soit au niveau du récepteur gp130 soit au niveau de sa région catalytique bloque la voie ERK (Burdon et *al.* 1999).

C'est probablement la liaison d'un complexe Grb2-Sos avec la protéine SHP2 phosphorylée qui, par l'intermédiaire de la protéine Gab1 (Grb2-associated binder protein 1), va activer la voie Ras (Hibi et Hirano 2000). L'invalidation des gènes *SHP-2* et *Grb2* dans les cellules ES montre d'ailleurs une baisse de la capacité des cellules à se différencier (Cheng et *al.* 1998, Qu et *al.* 1998).

En conclusion, la voie de signalisation dépendante des ERK et activée par SHP2 favorise la différenciation des cellules ES. Deux signaux contradictoires, l'un favorisant l'autorenouvellement (STAT3) des cellules et l'autre favorisant la différenciation (SHP2) sont donc simultanément activés lors de la transduction du signal par gp130 (**Figure 11**).

#### 4°) D'AUTRES VOIES DE REGULATION CONTROLANT L'AUTORENOUVELLEMENT DES CELLULES ES

Des données obtenues *in vivo* suggèrent que la voie d'activation par le LIF ne peut pas être la seule impliquée dans le contrôle de l'autorenouvellement des cellules ES.



#### a) La voie LIF/gp130 *in vivo*

L'invalidation des gènes codant pour le LIF (Stewart et *al.* 1992), le LIFR (Li et *al.* 1995), gp130 (Yoshida et *al.* 1996) ou même STAT3 (Takeda et *al.* 1997) n'altère pas le développement embryonnaire durant la période péri-implantatoire ni la formation de l'épiblaste de ces embryons, ce qui prouve que les cellules de la masse cellulaire interne se sont multipliées activement. Il semblerait en fait que la voie LIF/gp130 soit activée *in vivo* dans le cas bien particulier de la diapause embryonnaire. Il s'agit d'un retard d'implantation des blastocystes dans la souris gestante, dû à des conditions physiologiques défavorables à la nidation et à la poursuite du développement embryonnaire. Cet état peut être induit par un déficit en œstrogène obtenu par l'ovariectomie des femelles porteuses. A l'inverse des blastocystes sauvages, des blastocystes *gp130*<sup>-/-</sup> induits en diapause sont incapables de se réimplanter et de se développer lorsqu'ils sont réintroduits dans une femelle normale (Nichols et *al.* 2001). La voie gp130 est donc impliquée *in vivo* dans le maintien de la pluripotence des cellules au moment de la diapause embryonnaire. Il est intéressant de constater que les premières cellules ES ont été obtenues à partir de blastocystes induits en diapause par ovariectomie des femelles à 2.5 jours pc (Evans et Kauffman 1981). On peut penser que ce procédé a permis à l'époque d'activer la voie gp130 et donc de faciliter l'obtention en culture de cellules ES.

Comme on vient de le voir, la voie d'activation par gp130 semble jouer un rôle très marginal pour l'établissement d'une population de cellules pluripotentes dans l'embryon de souris. Cette observation pose de nombreuses questions. Une voie aussi importante *in vitro* pour le maintien de la pluripotence peut-elle vraiment être aussi « accessoire » au cours du développement embryonnaire ? N'existe-t-il pas une redondance entre ces facteurs qui peuvent être substitués *in vivo* par d'autres facteurs de la même famille et qui aboutissent finalement à l'activation des mêmes cibles ? Il se peut aussi que d'autres voies régulent l'autorenouvellement des cellules ES.

#### b) Le facteur ESRF

Le laboratoire d'Austin Smith a ainsi identifié un facteur soluble n'appartenant pas aux cytokines apparentées à l'IL6 mais permettant de maintenir le caractère indifférencié des cellules ES. Ce facteur a été appelé ESRF pour ES cell Renewal Factor. ESRF agit sur une

autre voie de régulation car il fonctionne aussi bien sur des cellules ne possédant plus le LIFR que sur des cellules pour lesquelles le récepteur gp130 est inactivé par un anticorps. D'autre part il n'active pas le facteur STAT3. Des cellules ES cultivées avec un milieu conditionné contenant ce facteur conservent leur pluripotence durant au moins une semaine en culture et peuvent contribuer, quand elles sont injectées dans un embryon hôte, à la formation de chimères dont la lignée germinale a été colonisée par ces cellules (Dani et *al.* 1998). Cette nouvelle voie, dont la nature reste à élucider, pourrait aussi contrôler l'autorenouvellement des cellules de la masse cellulaire interne et de l'épiblaste *in vivo*.

### c) La thrombopoïétine

D'autres résultats suggèrent également que la thrombopoïétine (Tpo) pourrait agir sur l'autorenouvellement des cellules ES par l'intermédiaire d'une autre voie de signalisation que la voie STAT3. En effet une étude montre que la Tpo et le LIF agissent en synergie pour maintenir un phénotype indifférencié aux cellules ES, mutantes pour le facteur SHP-2. Cependant, dans ces cellules, si le facteur STAT3 est activé par Tpo, il ne l'est pas davantage en présence de Tpo plus LIF qu'en présence de LIF seul (Xie et *al.* 2002). Il est donc possible que la Tpo agisse par l'intermédiaire d'une autre voie de signalisation que celle activant STAT3.

### 5°) LE CYCLE CELLULAIRE DES CELLULES ES : SOUS LE CONTROLE DE LA VOIE LIF/STAT3 ?

Le contrôle de l'autorenouvellement des cellules ES peut être associé à celui du contrôle du cycle cellulaire. Il semble donc logique de supposer que la voie gp130 régule le cycle cellulaire des cellules ES.

Les cellules ES présentent un cycle cellulaire très particulier. Tout d'abord, ces cellules hautement prolifératives doublent leur population toutes les 8-12 heures. En comparaison, les fibroblastes embryonnaires de souris (MEF) en culture doublent leur population toutes les 20 heures. Les cellules ES présentent également une phase G1 très brève qui ne dépasse pas 2 heures. La transition G1/S est une étape clé du contrôle de la prolifération cellulaire. Quels sont donc les acteurs moléculaires qui confèrent aux cellules ES un cycle cellulaire si particulier ?

#### a) La transition G1/S dans les cellules différenciées

La prolifération des cellules différenciées est placée sous le contrôle du passage de la phase G1 à la phase de synthèse S. Cette transition est contrôlée par deux mécanismes différents, l'un impliquant la protéine du rétinoblastome RB via son niveau de phosphorylation, l'autre le protooncogène *c-myc*.

- Durant la phase G1, la protéine RB est hypophosphorylée et va alors séquestrer les facteurs de transcription E2F. Elle inhibe ainsi l'expression des gènes nécessaires pour l'entrée de la cellule en phase S. Cependant la protéine RB va également être phosphorylée séquentiellement durant la phase G1 par des complexes comprenant les cyclines et les kinases dépendantes des cyclines (CDK). Ainsi la phosphorylation de RB par le complexe cycline D/CDK4 ou CDK6 réduit sa potentialité à séquestrer E2F, qui active alors la transcription de la cycline E et de la phosphatase *cdc25A*. *cdc25A* va ensuite déphosphoryler CDK2 qui, en interagissant avec la cycline E, va phosphoryler à nouveau RB, empêchant complètement la séquestration de E2F et induisant l'entrée en phase S de la cellule.
- La seconde voie de régulation implique le protooncogène *c-myc* qui active directement la transcription des gènes codant pour la cycline E et *cdc25A*, pour former un complexe cycline E/CDK2 fonctionnel (revue dans Burdon et *al.* 2002).

Ces deux voies fonctionnent en parallèle et modulent l'activité du complexe cycline E/CDK2, du niveau duquel dépend l'entrée en phase S (**Figure 12**).

#### b) Le contrôle de la transition G1/S diffère entre cellules ES et cellules différenciées

- RB ne contrôle pas le cycle cellulaire des cellules ES en G1

Dans les cellules ES, les acteurs contrôlant la transition G1/S sont différents de ceux des cellules différenciées. La forme hypophosphorylée de la protéine RB est ainsi indécélable au cours de la phase G1 et une forme hyperphosphorylée semble majoritairement présente (Savatier et *al.* 1994). D'autres observations suggèrent également que la protéine RB n'est pas

impliquée dans le contrôle du cycle cellulaire des cellules ES. Comme les cellules tumorales dont le cycle cellulaire n'est plus contrôlé par la protéine RB, les cellules ES ne répondent pas à la protéine inhibitrice du cycle cellulaire p16<sup>ink4a</sup>. D'autre part la triple invalidation des gènes *Rb* et de ces partenaires *p107* et *p130* ne perturbe pas la prolifération des cellules ES (Dannenberget al. 2000, Sage et al. 2000) mais confère aux fibroblastes embryonnaires de souris (MEF) *Rb*<sup>-/-</sup>, *p107*<sup>-/-</sup>, *p130*<sup>-/-</sup> un cycle cellulaire très proche de celui des cellules ES. Ces MEFs ne se bloquent plus en phase G1 à confluence ou lors de l'altération de leur ADN et échappent ainsi au contrôle du cycle par p16<sup>ink4a</sup> et p21<sup>cip1</sup> (Brown et al. 1997).

Les MEFs modifiés pour les protéines de la famille RB sont immortelles, non soumises à l'inhibition de contact et sans les systèmes de contrôle normalement présents en phase G1. L'ensemble de ces données suggèrent donc que le contrôle du cycle cellulaire des cellules ES est indépendant, en G1, de la protéine RB.

- Le contrôle de la transition G1/S par le complexe cycline E/CDK2

Les cellules ES expriment les cyclines D1 et D3. Cependant l'activité kinase de CDK4 est indécélable dans les cellules ES. Le complexe cycline D/CDK4 pourrait agir dans les cellules ES en séquestrant p27<sup>kip1</sup> et limiter ainsi son activité inhibitrice sur le complexe cycline E/CDK2 (Sherr et Roberts 1999). Il est probable aussi que le complexe cycline E/CDK2 soit constitutivement actif dans les cellules ES (Savatier et al. 1996) et que ce soit lui qui contrôle la transition G1/S dans les cellules ES (Burton et al. 2002) (**Figure 13**).

Si le contrôle de la transition G1/S semble être sous le contrôle du complexe CyclineE/CDK2, son lien avec la voie gp130 est loin d'être évident. Plusieurs hypothèses sont avancées pour corréler les deux mécanismes (Burton et al. 2002) :

- des cibles de Stat3 comme les gènes *myc* et *pim-1* activeraient le complexe CyclineE/CDK2
- la fixation de SHP2 sur gp130 en interagissant avec le complexe Gab1 recruterait PI3K et entraînerait alors une répression de p27<sup>kip1</sup>, un inhibiteur du cycle cellulaire.
- Les complexes CyclineE/Cdk2 et CyclineD/Cdk4 de la drosophile régulent l'activité de la protéine STAT92E au cours du développement embryonnaire (Chen et al. 2003)

## II *OCT-3/4* : UN REGULATEUR PUISSANT DE LA PLURIPOTENCE DES CELLULES ES

### 1°) LE FACTEUR *OCT-3/4* APPARTIENT A LA FAMILLE POU

*Oct-3/4*, également appelé *Pou5fl*, appartient à la famille des facteurs de transcription POU qui regroupe les gènes *pit* (P), *oct* (O) et *unc* (U). C'est une protéine de 352 acides aminés (Schöler et al. 1990) qui comprend une région conservée d'environ 150 acides aminés codant pour le domaine POU de liaison à l'ADN et deux régions en position N et C-terminal comprenant des domaines de transactivation (**Figure 14**). Le domaine POU est constitué de deux parties distinctes, l'une de 75 acides aminés spécifique des régions POU (POUs) et l'autre de 60 acides aminés composée de la partie C-terminal d'un homéodomaine (POUh). Il reconnaît spécifiquement une séquence de 8 nucléotides, AGTCAAAT (Saijoh et al. 1996) présente dans les promoteurs et les régions activatrices de gènes cibles. Les régions transactivatrices sont nécessaires pour maintenir le caractère pluripotent des cellules ES mais sont en fait redondantes et une seule région en fusion avec un domaine POU suffit à substituer *Oct-3/4* dans les cellules ES (Niwa et al. 2002).

Le gène *Oct-3/4* n'est apparemment conservé que chez les mammifères et n'a jamais été identifié dans d'autres espèces animales, notamment chez le poulet (Soodeen-Karamath et Verrinder-Gibbins 2001).

Le gène *Oct-3/4* a été cloné à partir d'une banque d'ADNc de cellules embryonnaires carcinoïdes (Rosner et al. 1990, Schöler et al. 1990) mais aussi à partir de l'analyse d'une séquence régulatrice spécifiquement active dans les cellules ES (Okamoto et al. 1990). Son profil d'expression est particulièrement intéressant car il est restreint aux cellules totipotentes et pluripotentes de la souris. En effet seul l'ovocyte, les premières cellules embryonnaires, la masse cellulaire interne du blastocyste, l'épiblaste d'embryon au stade 6,5 jours pc (egg cylinder) et les cellules germinales primordiales expriment *Oct-3/4* (Schöler et al. 1990). *In vitro*, seules les cellules ES, les cellules EG et les cellules EC l'expriment, ce qui fait de ce gène un excellent marqueur de la pluripotence cellulaire.

## 2°) IMPORTANCE DU GÈNE *OCT-3/4* DANS LE MAINTIEN DE LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES

Les résultats de plusieurs équipes s'accordent pour donner à *Oct-3/4* un rôle essentiel dans le maintien de la pluripotence des cellules embryonnaires de souris, aussi bien *in vivo* qu'*in vitro*.

*In vivo*, l'invalidation du gène *Oct-3/4* (Nichols et al. 1998) montre que ce gène est nécessaire au maintien de la pluripotence des cellules de la masse cellulaire interne (MCI) du blastocyste. En son absence, le développement embryonnaire s'arrête car les cellules de la MCI sont différenciées en cellules du trophoctoderme. D'autre part, l'étude de l'expression de *Oct-3/4* dans des blastocystes obtenus par clonage de noyaux somatiques montre que la viabilité des embryons est très fortement corrélée à la fiabilité du profil d'expression d'*Oct-3/4* (Boiani et al. 2002)

*In vitro*, une étude a également montré de façon remarquable que le niveau d'expression du gène *Oct-3/4* était crucial pour le contrôle de la pluripotence des cellules ES (Niwa et al. 2000). Les auteurs ont placé le gène *Oct-3/4* sous le contrôle d'un promoteur réprimé de façon dose-dépendante par la tétracycline. Suivant le niveau d'expression d'*Oct-3/4* dans les cellules ES, les auteurs ont observé que le devenir de ces cellules n'étaient pas le même (**Figure 15**). En absence ou à un faible niveau d'expression d'*Oct-3/4*, ils observent une dédifférenciation des cellules ES en cellules du trophoctoderme. Les auteurs parlent de dédifférenciation car comme on l'a vu précédemment, les cellules de la masse cellulaire interne du blastocyste et les cellules ES sont restreintes au niveau de leurs potentialités de différenciation à ne donner que des tissus fœtaux. *Oct-3/4* agirait donc comme un verrou en bloquant la différenciation par défaut en trophoctoderme et en maintenant ainsi la pluripotence de ces cellules. Cependant, s'il s'agissait bien de dédifférenciation, alors les cellules souches du trophoctoderme devraient avoir une potentialité de différenciation plus étendue que les cellules ES, or celle-ci est restreinte aux tissus extra-embryonnaires (Kunath et al. 2001). D'autre part l'expression de *Oct-3/4* dans ces cellules devrait restaurer le phénotype ES mais à ce jour cela n'a pas encore été montré.

Un niveau fort d'expression d'*Oct-3/4* (au moins deux fois supérieur au niveau endogène) entraîne la différenciation des cellules en endoderme primitif et en mésoderme. Ces résultats s'accordent avec ceux obtenus *in vivo* montrant que *Oct-3/4* est surexprimé dans l'endoderme primitif et réprimé dans le trophoctoderme (Palmieri et al. 1994) mais aussi avec

le phénotype des embryons dont le gène *Oct-3/4* a été invalidé. Les auteurs insistent aussi sur le fait que cette activité est spécifique à *Oct-3/4* et non à la présence du domaine POU. La surexpression d'*Oct-1*, *Oct-2* ou *Oct-6* n'a aucun effet sur la différenciation des cellules ES.

L'ensemble de ces résultats suggère un rôle primordial d'*Oct-3/4* pour le maintien de la pluripotence des cellules embryonnaires. *Oct-3/4* agirait grâce à un mécanisme subtil de régulation qui en fonction de son niveau d'expression modulerait l'expression de gènes cibles dont la combinaison donnerait les phénotypes observés. Quels sont donc les mécanismes sous jacents à l'expression d'*Oct-3/4* ? Existe-t'il un lien entre *Oct-3/4* et STAT3 ?

### 3°) LES GENES CIBLES D'*OCT-3/4*

Une manière de répondre aux questions posées précédemment est de travailler en aval du gène *Oct-3/4* et de déterminer les gènes cibles qu'il active ou réprime.

Les gènes cibles de *Oct-3/4* identifiés à ce jour sont résumés dans le tableau suivant :

Gène	Activation par <i>Oct-3/4</i>	Répression par <i>Oct-3/4</i>	Profil d'expression		
			Références	MCI	Epiblaste Cellules ES
<i>Fgf-4</i>	Oui <sup>1</sup>		2, 5	+	+ +
<i>Utf-1</i>	Oui <sup>3</sup>		4, 5	+	+ +
<i>Tera/226</i>	Oui <sup>5,6</sup>		5, 6	+	+ +
<i>Lefty-1</i>	Oui <sup>5,6</sup>		5, 7	+	- +
<i>Otx-2</i>	Oui <sup>5,6</sup>		5, 7	+	+ +
<i>Zfp42/Rex-1</i>	Oui <sup>8</sup>		5, 9	+	- +
<i>PDGF<math>\alpha</math>R</i>	Oui <sup>10</sup>			?	? ?
<i>Osteopontin</i>	Oui <sup>11</sup>			?	? ?
<i>hCG<math>\beta</math></i>		Oui <sup>12</sup>	5	-	- -

<i>Cdx2</i>		Oui <sup>5</sup>	5, 13	-	-	-
<i>Upp/383</i>	Oui <sup>5,6</sup>		5, 6	+	-	+
<i>Slc2a3/Glut-3</i>	Oui <sup>5,6</sup>		5, 6	+	+	+
<i>Sox-2</i>	Oui <sup>14</sup>		5, 7, 15	+	+	+
<i>Hand-1</i>		Oui <sup>5</sup>	5	-	-	-
<i>INF-<math>\tau</math></i>		Oui <sup>19</sup>	19	-	-	-
<i>Fbx15</i>	Oui <sup>18</sup>		18	+	Nd	+
<i>ELA</i>	nd	Nd	5, 16 17	+	-	+

**Tableau I** : Gènes cibles d'*Oct-3/4* et leur profil d'expression dans les cellules pluripotentes (d'après Niwa 2001)

Tous ces gènes exceptés ELA (E1A like activity) sont des candidats comme cibles d'*Oct-3/4*. Les références se rapportant aux gènes cités dans le tableau sont les suivantes : <sup>1</sup>(Yuan et al. 1995), <sup>2</sup>(Rappolee et al. 1994), <sup>3</sup>(Nishimoto et al. 1999), <sup>4</sup>(Okuda et al. 1998), <sup>5</sup>(Niwa et al. 2000), <sup>6</sup>(Saijoh et al. 1996), <sup>7</sup>(Matsui et Niwa, unpublished datas), <sup>8</sup>(Ben-Shushan et al. 1998), <sup>9</sup>(Rogers et al. 1991), <sup>10</sup>(Kraft et al. 1996), <sup>11</sup>(Botquin et al. 1998), <sup>12</sup>(Liu and Roberts 1996), <sup>13</sup>(Beck et al. 1995), <sup>14</sup>(Tomioka et al. 2002), <sup>15</sup>(Collignon et al. 1996), <sup>16</sup>(Suemori et al. 1988), <sup>17</sup>(Dooley et al. 1989), <sup>18</sup>(Tokuzawa et al. 2003), <sup>19</sup>(Ezashi et al. 2001)

On connaît aujourd'hui peu de choses sur la fonction de ces gènes cibles. *Fgf-4*, dont l'étude a été la plus approfondie, est essentiel lors de la période péri-implantatoire (Feldman et al. 1995). Il a en effet un rôle paracrine dans la stimulation de la prolifération des cellules du trophoctoderme qui expriment d'ailleurs le récepteur FGFR2 (Rossant et Cross 2001). Cependant *Fgf-4* n'a pas de rôle évident dans le maintien de l'autorenouvellement des cellules ES (Tanaka et al. 1998). L'inactivation des gènes *Otx-2* (Matsuo et al. 1995), *Lefty-1* (Meno et al. 1998) et *Opn* (Hynes 1996) a montré que la fonction de ces gènes n'était pas primordiale pour la formation des cellules pluripotentes de l'embryon.

Au contraire, les gènes cibles réprimés par *Oct-3/4* et impliqués dans le développement des cellules du trophoctoderme semblent jouer un rôle majeur au cours du développement précoce de l'embryon. L'inactivation du gène *Cdx-2* (Chawengsaksophak et al. 1997) entraîne ainsi une létalité embryonnaire entre 3.5 et 5.5 jours post coïtum et l'inactivation du gène *Hand-1*



(Firulli et al. 1998, Riley et al. 1998) bloque la différenciation des cellules géantes du trophoblaste en cellules placentaires. Ces données suggèrent que la fonction du gène *Oct-3/4* semble également liée à l'apparition de la viviparité chez les mammifères.

Ces résultats montrent aussi que les cibles d'*Oct-3/4* ne sont pas les mêmes que celles de *STAT3*. Ces observations sont d'ailleurs confirmées par des résultats expérimentaux. En effet, l'expression constitutive d'*Oct-3/4* n'empêche pas la différenciation des cellules ES en absence de LIF mais l'ajout du LIF ne change en rien les résultats des expériences lorsque l'expression d'*Oct-3/4* est perturbée (Niwa et al. 2000).

#### 4°) LA REGULATION DE L'EXPRESSION DU GENE *OCT-3/4*

Si l'étude de la fonction d'*Oct-3/4* a apporté des informations sur son rôle dans la maintenance de l'état pluripotent, l'étude de sa régulation permettra peut-être de comprendre comment cette fonction est contrôlée dans l'espace et dans le temps et quels en sont les régulateurs. En effet, l'expression d'*Oct-3/4* débute dès l'activation du génome du zygote et s'éteint au moment de la gastrulation, sauf dans les cellules germinales primordiales.

Le contrôle du niveau d'expression d'*Oct-3/4* est important pour le maintien du phénotype pluripotent. Ce type de régulation suggère qu'il existe une boucle d'autorégulation, mais il n'y a actuellement aucun résultat en prouvant l'existence (Niwa 2001).

L'analyse du promoteur d'*Oct-3/4* a montré l'existence de 2 éléments impliqués dans l'expression spécifique d'*Oct-3/4* dans les cellules pluripotentes. Ces deux éléments n'ont pas la même fonction apparente. L'un, proximal, est nécessaire pour activer l'expression d'*Oct-3/4* dans les cellules de l'épiblaste et dans les cellules EC de la lignée P19. L'autre, distal, active l'expression d'*Oct-3/4* dans l'embryon préimplantatoire, dans les cellules germinales primordiales et dans les cellules EG et ES en culture (Yeom et al. 1996). Ces résultats sont un peu surprenants car les caractéristiques des cellules de l'épiblaste sont proches de celles des cellules ES, même si l'environnement *in vitro* diffère de l'*in vivo*, et que certains mécanismes de régulation peuvent être différents, comme la signalisation par le LIF, par exemple.

*In vitro*, l'induction de la différenciation des cellules ES ou EC par l'acide rétinoïque induit une répression indirecte de l'expression d'*Oct-3/4* (Schoorlemmer et al., 1994, Minucci et al. 1996). Apparemment cette répression est due à une séquence présente dans la région proximale du promoteur. Ce domaine contient notamment plusieurs éléments de réponse se

chevauchant, composés de trois répétitions directes avec un espacement de 1 à 0 base. Plusieurs récepteurs nucléaires orphelins comme COUP-TFI et COUP-TFII se fixent sur cette région et sont impliqués dans le contrôle de l'expression d'*Oct-3/4*. Cependant la répression de l'expression d'*Oct-3/4* apparaît avant la fixation de ces deux facteurs sur le promoteur (Fuhrmann et al. 1999). Deux autres récepteurs nucléaires orphelins peuvent également se fixer sur la région proximale du promoteur, SF-1 (steroidogenic factor-1) et GCNF (germ cell nuclear factor). SF-1 a été montré comme activant le promoteur d'*Oct-3/4* en synergie avec le récepteur de l'acide rétinoïque RAR (Barnea et Bergman 2000) alors que GCNF réprime l'expression d'*Oct-3/4* aussi bien *in vitro* dans les cellules EC qu'*in vivo* dans l'embryon de souris (Fuhrmann et al. 2001).

Un contrôle épigénétique de l'activité du promoteur d'*Oct-3/4* semble également important au cours du développement embryonnaire. Une étude a ainsi montré que la méthylation du promoteur débutait autour de 6,5 jours post-coïtum et ce dans tous les tissus somatiques testés. Elle a aussi montré qu'il existait dans le promoteur un élément permettant la déméthylation des séquences et notamment de l'activateur proximal dans les cellules EC (Gidekel et Bergman 2002).

Le gène *Oct-3/4* est certainement, avec la voie LIF/STAT3, le gène qui a été le plus étudié pour comprendre le mécanisme du maintien de la pluripotence. On sait encore malgré tout peu de chose sur la fonction de ce gène dans la cellule et dans l'embryon, même si les données suggèrent qu'il joue un rôle important, à la fois dans le maintien de la pluripotence des cellules embryonnaires mais aussi probablement dans la mise en place des annexes extra-embryonnaires.

### **III D'AUTRES GENES IMPLIQUES DANS LE MAINTIEN DE LA PLURIPOTENCE DES CELLULES EMBRYONNAIRES**

Si le gène *Oct-3/4* et la voie d'activation de STAT3 par le LIF sont certainement les mécanismes les plus étudiés pour comprendre le maintien du phénotype pluripotent, de nombreuses équipes ont cherché à identifier d'autres gènes impliqués dans ces mécanismes. Plusieurs arguments sont en faveur de l'implication d'autres voies de signalisation pour le maintien de la pluripotence des cellules embryonnaires : *Oct-3/4* n'existe apparemment que

chez les mammifères, or des cellules souches embryonnaires ont été obtenues chez d'autres espèces comme le poulet (Pain *et al.* 1996) ou le poisson médaka (Hong *et al.* 1998). La voie de signalisation par le LIF ne semble pas directement impliqué *in vivo* dans le maintien de l'autorenouvellement des cellules embryonnaires pluripotentes. Il est donc légitime de rechercher d'autres acteurs impliqués à la fois *in vivo* et *in vitro* dans le maintien de la pluripotence des cellules embryonnaires. Quelques gènes importants pour le maintien de la pluripotence des cellules embryonnaires vont être présentés dans les paragraphes suivants.

### 1°) LE GENE *FOXD3*

Le gène *FoxD3* appartient à la famille des facteurs de transcription Forkhead/HNF-3 qui comporte un domaine de fixation à l'ADN en « winged helix » (hélices ailées) du fait de la structure des domaines qui le composent. Ce gène a été identifié à partir d'une banque d'ADNc de cellules EC de la lignée PCC4. On le pensait exprimé spécifiquement dans les cellules ES et les cellules EC comme un répresseur de la transcription dans ces cellules (Sutton *et al.* 1996). Cependant d'autres travaux ont montré qu'il avait un profil d'expression plus large, notamment dans les cellules de la crête neurale (Labosky et Kaestner 1998). De façon intéressante, il se fixe sur l'une des régions activatrices spécifiques des cellules ES isolées par le laboratoire d'Hiroshi Hamada, et qui a également permis d'identifier le facteur *Oct-3/4* (Okamoto *et al.* 1990).

L'étude de l'inactivation du gène *FoxD3* de souris a montré son importance pour le maintien *in vivo* et *in vitro* de la pluripotence des cellules embryonnaires (Hanna *et al.* 2002). En effet des embryons *FoxD3*<sup>-/-</sup> ne se développent pas plus tard que 9,5 jours post coïtum et des anomalies de développement apparaissent dès 6,5 jours p.c. (on peut signaler notamment l'absence de ligne primitive). Si ce gène ne semble pas avoir de fonction lors des premiers stades du développement son rôle semble primordial pour le maintien de la prolifération des cellules de l'épiblaste. En effet, si les blastocystes d'embryons *FoxD3*<sup>-/-</sup> ressemblent au niveau morphologique et moléculaire aux embryons contrôles, des embryons *FoxD3*<sup>-/-</sup> à 6,5 jours p.c. possèdent beaucoup moins de cellules en prolifération, beaucoup plus de cellules en apoptose que des embryons contrôles. L'analyse moléculaire confirme de nombreuses variations par rapport au contrôle : absence des marqueurs du mésoderme, absence des marqueurs de l'épiblaste comme *Oct-3/4* ou *Otx2* et de la gastrulation comme nodal, expression ectopique des marqueurs de l'endoderme et de l'ectoderme extraembryonnaire. Les auteurs ont également montré qu'il semblait impossible d'obtenir des cellules ES *FoxD3*<sup>-/-</sup>, soit à partir

de blastocystes mis en culture (les cellules de la masse cellulaire interne ne proliférant pas), soit à partir de cellules ES *FoxD3*<sup>+/-</sup> dont le deuxième allèle est invalidé. D'autre part, l'injection de cellules ES *FoxD3*<sup>+/+</sup> restaure complètement le phénotype dans 80% des cas avec des cellules *FoxD3*<sup>-/-</sup> capables de contribuer à la formation des différents tissus embryonnaires. Cette observation suggère que *FoxD3* agirait de manière cellule non-autonome (non-cell autonomous) en favorisant l'expression de facteurs (diffusibles ou de contacts) capables d'agir sur les cellules environnantes (dont les cellules *FoxD3*<sup>-/-</sup>).

L'ensemble de ces résultats suggère que le gène *FoxD3* est crucial pour la survie des cellules de l'épiblaste et des cellules ES et qu'il agirait sur les cellules environnantes en activant l'expression de facteurs diffusibles ou de contacts entre cellules. En effet, le fait de pouvoir sauver le phénotype par des expériences de chimères (où les niveaux de colonisation peuvent se situer autour de 25%) suggèrent que peu de cellules *FoxD3*<sup>+/+</sup> sont nécessaires pour maintenir la survie des cellules de l'épiblaste.

## 2°) LE GENE *SOX2*

Le gène *Sox2* appartient à la famille des facteurs de transcription Sox (SRY-related HMG box) qui ne possède qu'un domaine HMG (High Mobility Group) de fixation à l'ADN. *Sox2* a été montré comme important pour le maintien de la prolifération des cellules souches du système nerveux central (Zappone et al. 2000). *Sox2* est également un marqueur des cellules embryonnaires pluripotentes. Comme le gène *Oct-3/4* dont il est un coactivateur (Yuan et al. 1995) *Sox2* est exprimé dans les cellules de l'épiblaste, de la masse cellulaire interne du blastocyste ainsi que dans les cellules germinales primordiales. A l'inverse d'*Oct-3/4*, il est aussi exprimé dans le trophoctoderme et dans l'ectoderme extraembryonnaire qui dérive des cellules du trophoctoderme.

Par une approche très analogue à celle présentée pour le gène *FoxD3*, l'étude de l'invalidation du gène *Sox2* a montré qu'il était primordial pour le développement des cellules de l'épiblaste mais aussi de l'ectoderme extraembryonnaire (Avilion et al. 2003). L'analyse des embryons *Sox2*<sup>-/-</sup> montre des développements anormaux dès 6 jours p.c. qui conduisent à une létalité embryonnaire. Les embryons *Sox2*<sup>-/-</sup> ont une structure très altérée, sans épiblaste mais avec de nombreuses cellules géantes du trophoctoderme et de l'endoderme viscérale. La culture de blastocystes *Sox2*<sup>-/-</sup> ne conduit pas à une propagation des cellules de la MCI mais au contraire à de nombreuses cellules du trophoctoderme et de l'endoderme viscérale. Pour

expliquer le fait que les blastocystes *Sox2*<sup>-/-</sup> soient viables, les auteurs montrent que la protéine maternelle persiste dans l'embryon jusqu'à ce stade. Par la suite, ils montrent également que des injections de cellules ES *Sox2*<sup>+/+</sup> dans des blastocystes *Sox2*<sup>-/-</sup> compensent la perte de fonction jusqu'à 8 jours p.c. mais, par la suite, le développement des chimères *Sox2*<sup>-/-</sup> s'arrête. L'étude approfondie des chimères permet de penser que les cellules de l'épiblaste nécessitent *Sox2* pour survivre (très peu de cellules *Sox2*<sup>-/-</sup> sont présentes dans l'épiblaste des chimères) mais que la létalité embryonnaire des chimères est due à l'absence de *Sox2* dans les tissus extraembryonnaires (les cellules ES injectées ne colonisent pas les tissus extraembryonnaires).

Le rôle de *Sox2*, par son implication dans le développement des annexes embryonnaires et le maintien de la pluripotence des cellules embryonnaires, semble très proche de celui d'*Oct-3/4*. Le fait que ce gène soit un coactivateur d'*Oct-3/4* va d'ailleurs dans ce sens et il est possible que la coopération des deux facteurs soit indispensable pour maintenir le phénotype pluripotent. Les caractéristiques des différents Knock Out des gènes *Oct3/4*, *Sox2* et *FoxD3* sont résumés sur la **figure 16**.

### 3°) LE GENE NANOG

Identifié récemment par deux équipes japonaises et écossaises, le gène *Nanog* est certainement un élément clef pour la compréhension des mécanismes contrôlant la pluripotence dans les cellules ES (Mitsui et al. 2003, Chambers et al. 2003). En effet, la protéine NANOG, qui est un facteur de transcription à homéodomain proche de la famille NK-2, est nécessaire et suffisante pour maintenir les cellules ES dans un état différencié. Son expression est restreinte aux cellules ES indifférenciées et, *in vivo*, aux cellules de la masse cellulaire interne du blastocyste et aux cellules germinales présentes dans les crêtes génitales. D'autre part, l'invalidation de ce gène ne permet plus la formation de l'épiblaste dans un embryon précoce ni la dérivation de cellules ES pluripotentes, mais entraîne une différenciation des cellules en endoderme extraembryonnaire. D'autre part, son expression et son activité sont totalement indépendantes de la voie LIF/Stat3 et de *Oct-3/4*. *Oct-3/4* demeure malgré tout indispensable au maintien de la pluripotence, le gène *Nanog* ne pouvant compléter l'absence d'*Oct-3/4*. L'ensemble de ces résultats montrent que *Nanog* est un élément fondamental du contrôle de la pluripotence aussi bien dans les cellules ES que dans la MCI mais aussi du maintien de l'autorenouvellement des cellules ES en culture.

#### 4°) D'AUTRES GENES SPECIFIQUES DES CELLULES ES

Afin de ne pas être rébarbatif, les gènes définis par la littérature comme spécifiques ou importants pour la biologie des cellules ES de souris vont être récapitulés dans le tableau suivant. Cette énumération ne se veut pas exhaustive.

Gène	Nature	Phénotype KO	Fonction	Profil d'expression
Nanog <sup>14,15</sup>	Facteur de transcription à homéodomaine	Pas d'épiblaste . Léthalité embryonnaire à E5.5 Pas de dérivation d'ES	Maintien de la pluripotence  Maintien de l'autorenouvellement	ES, EC, MCI, PGC
Oct-3/4 <sup>1</sup>	Facteur de transcription de la famille POU	Pas de cellules de la MCI Léthalité embryonnaire à E4.5 Pas de dérivation d'ES	Maintien de la pluripotence.	ES, EC, MCI, endoderme primitif, épiblaste, PGC
Sox2 <sup>2</sup>	Facteur de transcription à boîte HMG	Léthalité embryonnaire à E6.5. Pas de cellules de l'épiblaste ni de l'ectoderme extraembryonnaire. Pas de dérivation d'ES	Maintien de la pluripotence des cellules de l'épiblaste. Formation des cellules de l'ectoderme extraembryonnaire (dépendance autonome des cellules)	ES, EC, nucléaire dans la MCI et cytoplasmique dans le trophectoderme, épiblaste, PGC, ectoderme extraembryonnaire
FGF4 <sup>3</sup> et FGFR2 <sup>4</sup>	Cytokine et son récepteur	Pas de cellules de l'épiblaste  Léthalité embryonnaire à E5.5. Pas de dérivation d'ES FGF4 <sup>-/-</sup> sauf en présence de FGF4 dans le milieu de culture	Activation de la prolifération des cellules souches du trophectoderme et maintien de la pluripotence	ES, EC, embryon préimplantatoire, MCI et endoderme primitif , épiblaste pour FGF4  Embryon préimplantatoire, trophectoderme et ectoderme extraembryonnaire pour FGFR2
FoxD3 <sup>5</sup>	Facteur de transcription de la famille Forkhead	Léthalité embryonnaire à E9.5 Absence des cellules de l'épiblaste. Extension des tissus proximaux comme l'endoderme et l'ectoderme extraembryonnaire  Pas de dérivation d'ES	Maintenance de la prolifération et protection contre l'apoptose des cellules de l'épiblaste.	ES, EC, blastocyste, épiblaste et ectoderme extraembryonnaire.
CRTR1 <sup>6</sup>	Facteur de transcription de la famille CP2	Nd	Répresseur de la transcription...Nd	ES, rein, testicules, intestin grêle...
Taubenuss <sup>7</sup>	Facteur de transcription ???	Léthalité embryonnaire à E4	Apoptose des cellules de la MCI	ES, morula, MCI, embryon 6,5-7,5-8,5-12.5jours...
Pumilio 2 <sup>8</sup>	Interagit avec l'ARN	Nd	Interagit avec l'ARN et DAZ	ES, EG, ovaire et testicule

PGC <sup>9</sup>	Protéine nucléaire ?	Nd		Oocytes, PGC, Embryon preimplantatoire
UTF1 <sup>10</sup>	Facteur de transcription à leucine zipper	Nd	Interagit avec ATF2 et le complexe TFIID comme coactivateur de la transcription. Interagit également avec Oct-3/4 et Sox2p	EC, ES, EG, MCI, épiblaste et ectoderme extraembryonnaire, ovaire, testicule
Rex1 <sup>11</sup> (Zfp-42)	Facteur de transcription à doigt de zinc apparenté à YY1	Nd	Interagit avec Oct-3/4, Oct6 et Rox-1	ES, EC, MCI et trophectoderme, cellules germinales du testicule
Tcl1 <sup>12</sup>		Perte de la fertilité chez les femelles mutantes. Arrêt du développement au stade morula	Protéine qui transloque entre le noyau et le cytoplasme. Interagit avec Akt	Ovaire, testicule, embryon préimplantatoire, thymus et moelle osseuse
Icn <sup>13</sup>		Létalité embryonnaire détecté à E7.5. Pas de dérivation d'ES.	Pas défini	ubiquitaire

**Tableau 2** : tableau récapitulatif de gènes importants pour la biologie des cellules ES ou de gènes marqueurs de ces cellules.

Références dont sont issues les informations : <sup>1</sup> revue dans Niwa et al. 2001, <sup>2</sup> Avilion et al. 2003, <sup>3</sup> Feldman et al. 1995, <sup>4</sup> Arman et al. 1998, <sup>5</sup> Hanna et al. 2002, <sup>6</sup> Rodda et al. 2001, <sup>7</sup> Voss et al. 2000, <sup>8</sup> Moore et al. 2003, <sup>9</sup> Sato et al. 2002, <sup>10</sup> Okuda et al. 1998, <sup>11</sup> Ben-Shushan et al. 1998, <sup>12</sup> Narducci et al. 2002, <sup>13</sup> Pu et al. 2000, <sup>14</sup> Mitsui et al. 2003, <sup>15</sup> Chambers et al. 2003

Il est encore difficile de comprendre les mécanismes contrôlant la pluripotence des cellules dans leur globalité. Si de nouvelles pistes émergent, il est encore difficile de tracer des liens entre les différents acteurs, comme on a pu le voir à travers les derniers paragraphes, aussi bien par l'exemple de la voie LIF/STAT3 que par les fonctions des gènes *Nanog*, *Oct-3/4* ou encore *FoxD3*.

Des recherches récentes sur les cellules souches adultes ont montré que l'on pouvait obtenir des cellules souches pluripotentes à partir d'organes de souris adulte comme le cerveau (Clarke et al. 2000), le sang (Labarge et al. 2002), le muscle (Seale et al. 2001)... Ces cellules prolifèrent et sont capables de coloniser d'autres tissus que celui dont elles sont issues, lorsqu'elles sont réinjectées dans un organisme receveur. Quels sont donc les

mécanismes moléculaires qui confèrent leur pluripotence à toutes ces cellules souches, aussi bien embryonnaires qu'adultes ?

#### 5°) LES APPROCHES DE GENOMIQUE A GRANDE ECHELLE

Afin de répondre à ces questions, il fallait pouvoir comparer le transcriptome (ensemble des transcrits d'une ou d'un ensemble de cellules) de ces différentes cellules pluripotentes. Récemment, grâce au développement de la technologie des puces à ADN qui permet d'analyser un nombre très important de gènes en une seule expérimentation, de telles expérimentations deviennent réalisables et les premiers résultats commencent à être publiés.

Deux études ont ainsi comparé le transcriptome de cellules ES, de cellules souches neurales et de cellules souches hématopoïétiques (Ramalho-Santos et *al.* 2002, Ivanova et *al.* 2002). Il ressort de ces deux études que 216 à 283 gènes seraient amplifiés dans les trois types cellulaires, sur un criblage d'environ 12 000 gènes. L'une des études montre qu'à part 4 gènes qui semblent réellement spécifiques des cellules souches (*Uridine phosphorylase*, *suppressor of Lec15* et 2 EST), les autres gènes existent dans divers types cellulaires. Leurs fonctions dans la biologie des cellules souches seraient liées à leur enrichissement et leur combinaison simultanés (Ramalho-Santos et *al.* 2002). Ils définissent aussi de grandes catégories de gènes apparemment nécessaires aux cellules souches :

- voie de signalisation JAK/STAT et Notch
- sensibilité à l'hormone de croissance et à la thrombine
- interaction avec la matrice extra cellulaire par l'intégrine  $\alpha6/\beta1$  et les protéines ADAM9 et bystine
- forte résistance au stress avec une surexpression des systèmes de réparation de l'ADN, de détoxification, de séquestration et dégradation par les ubiquitines
- remodelage de la chromatine par le biais d'hélicases, de méthylase et de déacétylases des histones
- régulation de la traduction par des hélicases de l'ARN proche de *Vasa*

Il semble donc se dessiner petit à petit un cadre moléculaire nécessaire et suffisant pour le maintien de la pluripotence des cellules souches aussi bien embryonnaires, qu'adultes. Il est évident que les approches à grande échelle nécessitent une validation expérimentale, les



outils pour la réaliser étant de plus en plus nombreux et devraient bientôt permettre d'y parvenir.

## **IV CARACTERISATION MOLECULAIRE DES CELLULES DU BLASTODERME DE POULET**

Les travaux sur les mécanismes moléculaires contrôlant la pluripotence des cellules ES de souris sont de plus en plus nombreux. Chez le poulet, par contre, très peu de données sont disponibles et il est difficile de savoir si ces mécanismes sont conservés.

Par exemple, le gène *Oct-3/4* n'existe pas chez le poulet mais il est possible que ce gène se soit substitué chez les mammifères à d'autres gènes ayant conservé leur fonction chez le poulet. D'autre part, le maintien de l'autorenouvellement en culture des cellules ES de poulet nécessite l'usage de facteurs de croissance activant la voie gp130/STAT3, ce qui suggère que cette voie de signalisation est conservée chez le poulet.

Afin de mieux comprendre les mécanismes contrôlant la pluripotence des cellules ES de poulet, il pourrait être intéressant de se pencher sur les connaissances acquises *in vivo*, concernant l'embryon de poulet dont sont issues les cellules ES.

### **1°) LES CELLULES GERMINALES PRIMORDIALES, UNE POPULATION DE CELLULES PRESENTES DANS L'EPIBLASTE**

Comme on l'a vu précédemment, l'épiblaste de l'embryon de poulet au stade X est composé de plusieurs milliers de cellules n'étant pas, en apparence, toutes équivalentes. Il semble en effet que le devenir de certaines cellules soit déjà spécifié, comme le suggère la présence des cellules germinales primordiales ou de leur précurseurs dans la partie centrale de l'épiblaste de l'embryon au stade X. Ainsi une vingtaine de cellules de l'épiblaste localisées dans la partie ventrale du feuillet (en contact avec le blastocoele) sont immunomarquées avec les anticorps EMA-1 (Urven et *al.* 1988) ou SSEA-1 (Karagenç et *al.* 1996). Cependant, avec les données obtenues à ce jour, il est difficile de savoir si des cellules germinales primordiales sont présentes à des stades du développement plus précoces ou si leur présence au stade X est due à la spécification puis à la différenciation de précurseurs sous l'effet d'une induction

épigénétique, rendue possible par la mise en place à ce stade de l'aire opaque et de l'aire pellucide.

## 2°) L'INDUCTION NEURALE : UN MECANISME DEJA ACTIF DANS LES CELLULES DE L'EPIBLASTE ?

Un autre exemple montre que l'hypothèse de l'existence de populations de cellules bien distinctes dans l'épiblaste n'est pas irréaliste. Des travaux récents ont en effet montré que l'induction neurale chez le poulet avait lieu avant la gastrulation, très précocement même, puisque les cellules pourraient être spécifiées dans certaines voies de différenciation dès le stade IX (EG). Deux équipes ont particulièrement étudié cette induction à des stades très précoces : celle de Thomas Edlund en Suède et celle de Claudio Stern en Angleterre. Inspirée par les travaux réalisés sur la blastula de Xénope où l'induction neurale est présentée comme un équilibre entre les BMP (bone morphological protein) et leurs inhibiteurs, les uns induisant la différenciation en ectoderme épidermal, les autres en ectoderme neural, l'équipe de Thomas Edlund a cherché à mettre en évidence cet équilibre dans les blastulas de poulet. Pour cela ces chercheurs ont mis en culture des explants d'embryons de poulet de différents stades (VIII-IX et XII) dans un milieu asérique et sans facteurs de croissance et ils ont analysé le devenir de ces explants avec des marqueurs spécifiques de l'ectoderme neural (Sox2, Sox3, Pax6 et Otx2) et de l'ectoderme épidermal (Msx1/2, Gata2). De ces expériences, ils concluent que l'induction neurale a lieu peu après le stade VIII puisqu'au stade IX et XII ils détectent déjà en culture deux types de populations de cellules : l'une, issue d'un explant au centre de l'épiblaste, exprime les marqueurs de l'ectoderme neural. L'autre, issue d'explants situés à la périphérie de l'épiblaste, exprime les marqueurs de l'ectoderme épidermal (**Figure 17A<sub>1-2</sub>**). Ensuite, afin de déterminer les facteurs impliqués dans cette délimitation spatiale et cette spécification, ils ajoutent des BMP, des inhibiteurs de BMP, du FGF et des inhibiteurs du FGF dans le milieu de culture des explants (**Figure 17A<sub>3</sub>**). Ils montrent ainsi que chez le poulet, BMP4 induit le caractère épidermal alors que FGF est nécessaire pour l'acquisition du caractère neural. FGF agirait en fait en réprimant l'action de BMP4 dans les cellules amenées à acquérir un caractère neural (**Figure 17B**)(Wilson et *al.* 2000).

L'équipe de Claudio Stern a eu une démarche différente. Après avoir identifié un gène marqueur, cERNI pour Early Response to Neural Induction, dont l'expression est inductible par le nœud de Hensen, elle a étudié l'expression de ce gène en réponse à différents inducteurs greffés sur l'embryon. Ils ont ainsi montré que l'induction neurale avait lieu avant

la gastrulation, au stade XII et que le FGF était nécessaire à cette induction (Streit et *al.* 2000).

Par la suite, l'équipe de Thomas Edlund a affiné sa démonstration. Pour cela elle va s'appuyer sur les travaux de Roeser et *al.* (1999) qui montrent l'hétérogénéité de la localisation de la  $\beta$ -caténine dans l'épiblaste de poulet. Aux stades X et XII, la  $\beta$ -caténine est localisée dans le cytoplasme des cellules de la partie centrale de l'épiblaste mais dans le noyau des cellules de la zone marginale (région de l'épiblaste qui borde l'aire opaque) et de l'aire opaque. La localisation de la  $\beta$ -caténine dépend de son activation par les facteurs Wnt. La présence de ceux-ci induit une délocalisation de la  $\beta$ -caténine qui va alors être transloquée dans le noyau et recrutée par un complexe transcriptionnel pour activer des gènes cibles.

Wilson et *al.* (2002) reprennent la culture d'explants d'embryon aux stades X-XII et ajoutent différents facteurs (BMP4, FGF8, Wnt et leurs inhibiteurs respectifs). Ils déduisent de cette manière que la voie Wnt bloque l'action du FGF dans les cellules de la zone marginale de l'épiblaste. FGF ne peut donc plus inhiber BMP et les cellules acquièrent un caractère épidermal. Par contre dans les cellules du centre de l'épiblaste, la voie Wnt n'est pas activée (la  $\beta$ -caténine est cytoplasmique), FGF réprime BMP et permet la différenciation par défaut de ces cellules en ectoderme neural (**Figure 18 A**).

Ces expériences montrent que le modèle par défaut de l'induction neural s'applique également au poulet (**Figure 18 B**). D'autre part, ces expériences suggèrent qu'il existe déjà des spécifications cellulaires dans l'épiblaste de l'embryon au stade X mais que celles-ci semblent réversibles, au moins partiellement.

Il est malgré tout intéressant de remarquer que, si ces mécanismes ne reflètent pas la réalité *in vivo*, ils semblent importants pour le contrôle des différenciations *in vitro*. Les mêmes observations ont pu être ainsi réalisées *in vitro* sur des cellules ES de souris. Il a été montré que la voie Wnt dirigeait la différenciation des cellules ES dans des lignées cellulaires non neurales (Aubert et *al.* 2002) alors que le FGF était nécessaire à la différenciation *in vitro* de ces cellules en cellules neurales et que BMP bloquait *in vitro* cette voie de différenciation (Ying et *al.* 2003).

L'ensemble de ces données, obtenu à la fois sur l'embryon de poulet et sur les cellules ES de souris, suggère que les cellules ES de poulet sont également sensibles à la voie Wnt/ $\beta$ -

caténine, aux BMP et aux FGF. Dans les cellules ES de poulet (**Figure 19**) comme dans les cellules ES de souris, la  $\beta$ -caténine a une localisation cytoplasmique.

Si actuellement les données moléculaires manquent pour comprendre les mécanismes généraux contrôlant la pluripotence des cellules ES de poulet, l'analyse du transcriptome et du protéome des cellules ES de poulet permettra peut être de rapprocher les différents modèles animaux pour identifier des gènes conservés et importants dans le contrôle de la pluripotence et de l'autorenouvellement des cellules ES.

---

## PARTIE C

# LE GENOME RETROVIRAL ET SON EXPRESSION DANS LES CELLULES ES

---

Les travaux présentés dans cette thèse montrent qu'un gène d'origine rétrovirale est exprimé spécifiquement dans les cellules ES. Le but de cette troisième partie est de présenter brièvement les rétrovirus tout en s'attardant davantage sur la régulation de l'expression de leurs gènes de façon générale et plus spécifiquement dans les cellules ES.

## I PRESENTATION GENERALE DES RETROVIRUS

Les premiers rétrovirus ont été identifiés chez le poulet en 1908 par Ellerman et Bang et en 1911 par Rous. Ces virus étaient capables d'induire des leucémies lorsqu'ils étaient inoculés dans un poulet. Ces travaux n'ont pas été reconnus à leur juste valeur à l'époque, peut-être parce qu'il était alors difficile d'admettre l'idée que les cancers pouvaient être transmis par voie virale. Il fallu donc attendre plus de quarante ans pour qu'un autre rétrovirus tumorigène soit également isolé chez la souris par Gross en 1951 et que l'intérêt porté aux rétrovirus s'accrût. Par la suite, la découverte de la transcriptase inverse des rétrovirus (Baltimore 1970) révolutionna le dogme de la biologie moléculaire, qui posait que l'ARN était produit à partir de l'ADN et non l'inverse. Elle expliqua aussi comment le génome du virus contenu dans le virion sous la forme de deux ARN pouvait être rétrotranscrit en ADN. La découverte par la suite du virus de l'immunodéficience humaine (VIH) et de sa pathologie chez l'homme a sans doute fortement dynamisée les recherches menées sur les rétrovirus.

### 1°) CARACTERISTIQUES GENERALES DES PARTICULES RETROVIRALES

Les rétrovirus appartiennent à une vaste famille de virus à ARN et à enveloppe qui a la particularité de pouvoir s'intégrer dans le génome de leur hôte. La particule rétrovirale typique est représentée sur la **Figure 20**. Le virion d'un diamètre de 80-100nm intègre deux molécules d'ARN identiques de 7 à 12 kilobases, simple brin, linéaires et orientés dans le

sens codant. Son enveloppe est constituée d'une bicouche phospholipidique dans laquelle s'insèrent les sous-unités de la glycoprotéine ENV, composée d'une protéine transmembranaire et d'une protéine de surface SU. A l'intérieur de l'enveloppe on retrouve les protéines structurales du virus et les enzymes rétrovirales :

- la matrice (MA) se localise sur la face interne de l'enveloppe
- la capsid (CA) protège l'ARN du virus
- la nucléocapsid (NC) se fixe sur la molécule d'ARN et la stabilise
- la protéase (PR) qui permet de cliver les précurseurs protéiques, la transcriptase inverse (RT) qui permet d'obtenir une molécule d'ADN à partir de l'ARN du virus et une intégrase (IN) nécessaire à l'intégration de l'ADN du virus dans le génôme de l'hôte.

Les virus ont tout d'abord été classés en quatre catégories A, B, C et D selon la structure de leur virion. La catégorie A ne correspond pas en fait à des rétrovirus fonctionnels. Par la suite, en fonction de l'homologie de leurs séquences génomiques, les rétrovirus ont été classés dans sept genres, résumés dans le tableau suivant :

Genre	Virus typique	Morphologie du virion
1 Virus des sarcomes et leucoses aviaires	RSV ( Rous Sarcoma Virus)	Type C
2 Virus mammaliens de type B	MMTV (Mouse Mammary Tumor Virus)	Type B
3 Groupe du virus de la leucémie murine	MoMLV (Moloney Murine Leukemia Virus)	Type C
4 Groupe des virus de la leucémie humaine et bovine	HTLV (Human T-cell Leukemia Virus)	Type C
5 Groupe des virus de type D	MPMV (Mason-Pfizer Monkey Virus)	
6 Lentivirus	VIH (Virus de l'Immunodéficience Humaine)	Type C avec cœur en cône tronqué
7 Spumavirus	HFV (human foamy virus)	Proche du type C

**Tableau 3** : classification des rétrovirus (d'après Vogt 1997)

## 2°) LE GENOME DES RETROVIRUS

Le génome des rétrovirus doit non seulement comporter les informations nécessaires à la synthèse des protéines du virus mais aussi celles qui permettent sa transcription inverse, son intégration dans le génome de l'hôte et l'expression de ses gènes.

Le génome viral est présent sous deux formes différentes au cours de son cycle répliatif. Dans le virion, le génome viral se compose de deux ARN simple brin. Ces ARNs vont ensuite être convertis en ADN double brin dans la cellule par le mécanisme de transcription inverse. Cet ADN proviral va alors être la source des ARN viraux constituant les nouvelles particules virales (**Figure 21-A**).

Le génome rétroviral contient trois grandes régions (quatre avec *pro* qui est plus petite) codantes pour les protéines du virion décrites précédemment :

- *gag* code pour les protéines du virion qui forment la matrice, la capsid et les nucléoprotéines
- *pro* code pour la protéase du virion
- *pol* code pour la transcriptase inverse, la RnaseH et l'intégrase
- *env* code pour les protéines constituant l'enveloppe du virion

Certains génomes de rétrovirus dits complexes, codent également pour d'autres gènes comme *tat*, *rev*, *vif*, *vpr*, *vpu* et *nef* pour le virus du VIH.

Ces régions codantes sont flanquées par de longues séquences répétées (LTR pour Long Terminal Repeat) localisées aux extrémités du génome (**Figure 21-B**). Celles-ci sont divisées en trois régions : U3, R et U5 et sont importantes à la fois pour la transcription des gènes viraux mais aussi pour l'étape de transcription inverse.

- la région U3 est définie comme la région intercalée entre R et le site d'initiation de la synthèse du brin positif. Elle comprend l'un des deux sites *att* nécessaires à l'intégration de l'ADN proviral. Elle contient également le promoteur nécessaire à l'expression des transcrits rétroviraux. Elle comporte ainsi une boîte CAAT et une boîte TATA permettant la fixation des facteurs basaux de transcription. D'autres

sites de fixations pour des facteurs de transcription sont en général présents dans cette région.

- La région R, répétée aux deux extrémités de l'ARN viral est essentielle pour la mécanique de la transcription inverse. Elle contient également le site d'initiation de la transcription et la séquence signal AAUAAA pour le clivage et la polyadénylation du transcrit primaire.
- La région U5 est la région intercalée entre le site d'initiation de la transcription inverse (PBS) et la région R. U5 contient en 3' le second site *att* requis pour l'intégration. La conservation de la région U5 est crucial pour l'initiation de la transcription inverse.

### 3°) LE CYCLE REPLICATIF DES RETROVIRUS

Le cycle de réplication d'un rétrovirus comprend une succession d'étapes depuis la reconnaissance de la cellule cible jusqu'à la production de nouvelles particules infectieuses (**Figure 22**).

La pénétration du cœur de la particule virale dans la cellule est rendu possible par la fusion des membranes cellulaire et virale. Ce mécanisme met en jeu la reconnaissance de la glycoprotéine d'enveloppe SU par un récepteur cellulaire. L'interaction entre ces deux protéines entraîne un changement conformationnel qui révèle le domaine de fusion de la protéine d'enveloppe TM et permet la fusion des membranes cellulaires et virales.

La capsidie pénètre alors dans le cytoplasme où elle est partiellement dégradée et libère l'ARN et les enzymes virales. L'ARN viral va alors être converti en ADN proviral au cours d'un mécanisme complexe : la transcription inverse. Celle ci a lieu dans le cytoplasme cellulaire et fait intervenir un ARNt cellulaire. En se fixant sur une séquence nommée PBS (Primer Binding Site), localisée dans la région 5' de l'ARN viral, cet ARNt va agir comme un site d'initiation de la transcription inverse et permettre l'élongation de l'ADN proviral. Au cours de la transcription inverse, les séquences situées en 5' et en 3' de l'ARN viral sont dupliquées et génèrent ainsi les LTR flanquantes. L'ARN viral simple brin de type 5' R-U5-*gag-pol-env*-U3-R 3' est donc converti en ADN double brin de type 5' U3-R-U5-*gag-pol-env*-U3-R-U5 3' (pour revue Cristofari 2002)



L'ADN viral est alors transporté dans le noyau et s'intègre de façon aléatoire dans le génome de la cellule hôte. Une fois intégré, l'ADN rétroviral persiste dans le génome de la cellule hôte et se réplique comme n'importe quel gène cellulaire.

Ensuite, c'est en utilisant les machineries cellulaires de transcription, de maturation des ARNm et de traduction que seront produites les nouvelles particules virales. Chez les virus simples, des facteurs cellulaires interagissant avec les LTR interviennent pour la transcription des gènes viraux alors que chez les virus complexes ce sont également des facteurs de trans-activation produits par les virus qui sont impliqués dans l'activation de cette transcription. Les transcrits ARN rétroviraux sont ensuite soumis aux mêmes événements de maturation que les ARN cellulaires : addition d'une coiffe à l'extrémité 5', clivage et polyadénylation à l'extrémité 3'.

Les protéines virales matures sont obtenues par clivage, phosphorylation et glycosylation des précurseurs protéiques produits lors de la traduction des différents ARNm viraux. Elles s'accumulent sous la membrane cellulaire. L'assemblage suit un scénario différent suivant le type de rétrovirus. L'ARN génomique est encapsidé au moment du bourgeonnement grâce à un signal spécifique d'encapsidation ( $\psi$ ) et certainement à une séquence située sur *gag* ou *pol* (Linial et Miller 1990).

#### 4°) LES RETROELEMENTS

Il existe dans le génome des organismes vivants des éléments génétiques qui se rapprochent des rétrovirus. Ces rétroéléments sont des éléments transposables ou éléments génétiques mobiles qui utilisent un intermédiaire ARN au cours de leur transposition. Ils sont classés dans deux grands groupes, différenciés par la liaison ou non de l'élément transposable avec une LTR.

##### a) Les éléments sans LTR (pseudogènes, rétrogènes et rétroposons)

Les pseudogènes sont les rétroéléments les plus « simples » ; ils sont la manifestation de la réintégration d'ARNm cellulaires. Lorsqu'ils ont acquis un promoteur et peuvent donc être transcrits, ils sont aussi appelés rétrogènes. Les SINEs (Short Interspersed Nuclear Elements) appartiennent à cette famille d'éléments transposables mais ne possèdent pas de cadre ouvert de lecture. Ils sont de petites tailles, en général inférieur à 500 pb. La plupart des SINEs dérivent d'ARNt à l'exception des gènes *alu* chez les primates et *BI* chez les rongeurs.

Les rétroposons possèdent un promoteur interne riche en GC, un gène codant pour une protéine non rétrovirale et un gène de type *pol* présentant des homologies avec la transcriptase inverse. Les LINEs (Long Interspersed Nuclear Elements) en sont les prototypes.

Ces deux types d'éléments sont présents en grande quantité dans le génome des plantes et des animaux.

#### b) Les éléments avec LTR (rétrotransposons et rétrovirus endogènes)

Les rétrotransposons existent dans de très nombreux organismes, aussi bien chez les plantes, les champignons que chez les animaux, et ont été plus particulièrement étudiés chez la levure, la drosophile et la souris. Ils présentent une structure générale très proche de celle des rétrovirus. D'une taille de 5 à 9 kb ils comportent deux LTR encadrant les gènes correspondant aux gènes *gag* et *pol* des rétrovirus. Leur structure est donc la suivante : LTR- $\psi$ -*gag-pol*-LTR mais la majorité présente des cadres ouverts de lecture altérés. Cependant il existe parfois des cadres ouverts de lecture complets pour *gag* qui permettent aux rétrotransposons de former des particules, l'existence du signal d'encapsidation  $\psi$  permettant à leur ARN d'être incorporé dans ces particules. Chez la souris, les particules IAP (intracisternal A particles) en sont un exemple. Dans la mesure où il s'agit de particules intracellulaires, la transposition intracellulaire est possible. (Heidmann et Heidmann, 1991).

Les rétrovirus endogènes présentent globalement la même structure génétique que les rétrovirus exogènes simples : des régions répétées non traduites encadrant les gènes structuraux. La structure typique est donc : LTR- $\psi$ -*gag-pol-env*-LTR. Les rétrovirus endogènes sont présents dans le génome sous forme provirale et sont transmis à la descendance par les cellules germinales. Cependant la transmission verticale des génomes rétroviraux endogènes n'impose pas de pression de sélection pour maintenir des cadres de lecture fonctionnels pour la réplication virale ; tout ou partie des gènes structuraux présente ainsi des mutations ponctuelles à l'origine de codons stop ou de décalages de trame.

Chez le poulet, il a été décrit deux grands types de rétrovirus endogènes. Les rétrovirus endogènes classiques possèdent une structure identique à celle des rétrovirus exogènes de type ALV (Avian Leukosis Virus), et d'autres séquences plus éloignées appelées EAV (Endogenous Avian Virus) et E51 (revue dans Allioli 1995).

Certains loci sont capables de produire une forme virale complète et infectieuse comme le virus RAV-0 produit par le locus *ev2* (Astrin et *al.* 1980). Les autres par contre, s'ils ne sont pas infectieux, peuvent parfois produire des protéines virales. Par sélection, il a été possible d'obtenir des poulets viables (lignée 0) dont le génome ne renferme aucun virus endogène apparenté aux ALV (Astrin et *al.* 1979).

Les séquences endogènes de type EAV ont été identifiées chez des animaux de la lignée 0 qui ne renferment pas les structures endogènes proches des ALV (Dunwiddie et Faras 1985). Ces rétrovirus endogènes conservent la structure caractéristique d'éléments répétés encadrant les gènes *gag*, *pol* et *env*. Ceux-ci présentent une homologie de séquence de l'ordre de 60% avec les gènes viraux du virus PR-RSV-C (Dunwiddie et *al.* 1986). La LTR de ces gènes présente tous les éléments nécessaires à son fonctionnement et est capable d'assurer un taux correct de transcription. Une quatrième famille de rétrovirus endogènes a également été découverte et nommée ART-CH (Avian RetroTransposon from CHicken genome) (Nikiforov et Gudkov 1994). D'une longueur de 3,3kb ils possèdent deux LTR encadrant des séquences similaires des gènes *gag* et *pol* des virus ALV et EAV. Ils semblent cependant qu'ils aient perdu la capacité de produire ces protéines.

## **II LE CONTROLE DE L'EXPRESSION DU GENOME RETROVIRAL : L'EXEMPLE DES PARTICULES INTRACISTERNALES DE TYPE A**

Comme on l'a vu précédemment, le contrôle de la transcription du génome virale est assuré par les LTRs au niveau de la région U3. L'analyse des promoteurs de différents rétrovirus a conduit à l'identification de nombreux facteurs cellulaires activant l'expression des gènes viraux. L'étude du contrôle de l'expression des gènes des particules intracisternales de type A (IAP) est particulièrement intéressante car il a été montré que l'activité des LTR dans des souris transgéniques était restreinte aux cellules de la lignée germinale (Dupressoir et Heidmann 1996). Il était donc intéressant de comprendre comment cette expression tissu-spécifique était contrôlée.

## 1°) PRESENTATION GENERALE DES PARTICULES INTRACISTERNALES DE TYPE A (IAP)

Les particules intracisternales de type A ont été observées initialement dans des tumeurs (Dalton 1961). Il existe environ 1000 copies de ces éléments dans le génome de la souris, distribuées sur la plupart des chromosomes (Lueders et Kuff 1977). Ces éléments peuvent s'intégrer par transposition dans d'autres régions du génome. Ils sont très proches des rétrovirus et possèdent deux LTRs flanquant les gènes *gag*, *pol* et *env*. Le gène *env* n'est cependant pas fonctionnel car il comporte plusieurs codons stop dans les trois cadres de lecture et ne permet pas la synthèse d'une protéine d'enveloppe fonctionnelle. Plusieurs types d'IAPs ont été identifiées à ce jour et la variation de la taille de leurs transcrits est due à de multiples délétions internes.

Le profil d'expression des IAPs a été réalisé initialement par Northern Blot et a montré que ces gènes étaient exprimés dans l'embryon préimplantatoire (Poznanski et Calarco 1991) mais aussi chez l'adulte dans de multiples tissus (revue dans Kuff et Lueders 1988). Cependant plusieurs résultats montrent que ces profils d'expression ne seraient pas caractéristiques de l'ensemble des IAP mais seraient plutôt dus à un effet de position sur certains de ces gènes (Lueders et *al.* 1993). L'expression de la majeure partie des IAPs est en effet réprimée dans le génome de la souris.

Pour comprendre la régulation de l'expression des gènes des IAPs, deux approches différentes ont été menées. La première s'est focalisée sur l'étude des facteurs cellulaires interagissant directement avec les LTRs. Il a ainsi été possible de délimiter des régions importantes pour l'expression de ces gènes et de caractériser les facteurs qui y interagissent. La deuxième approche s'est davantage polarisée sur l'étude de la régulation de l'expression des IAPs par la méthylation de l'ADN.

## 2°) PRESENTATION DES MECANISMES REGULANT L'EXPRESSION DES IAPS

### a) Analyse *in vitro* des facteurs cellulaires agissant directement sur les LTRs d'IAPs

La première étape de la caractérisation des LTRs a été de démontrer l'activité promotrice de ces régions. Placées en amont d'un gène rapporteur, les LTRs d'IAPs sont en effet capables d'assurer une fonction de promoteur dans des cellules eukaryotes (Lueders et

*al.* 1984). Il a été montré par la suite que c'était la région U3 qui était plus particulièrement impliquée dans l'activité promotrice de ces LTRs (Christy et Huang 1988). L'analyse fine de cette région U3 a ensuite permis d'identifier dans ses 139 premières paires de base plusieurs sites de fixation pour des facteurs nucléaires. Certains de ces éléments de réponse étaient connus pour fixer les facteurs de transcription Sp-1, Ap-1 ou le récepteur aux glucocorticoïdes (Falzon et Kuff 1988). D'autres au contraire ont été identifiés par la suite, comme le facteur EBP-80 qui se fixe sur le domaine Enh2 pour stimuler la transcription (Falzon et Kuff 1989) ou encore le facteur YY1 qui interagit plus en amont du promoteur pour activer la transcription (Satyamoorthy et *al.* 1993). Une autre séquence activatrice du promoteur qualifiée de proximal (IPE pour IAP proximal enhancer) car localisée près de la boîte TATA, régule également spécifiquement l'expression des IAPs dans des cellules EC de la lignée F9 différenciées en endoderme pariétale (Lamb et *al.* 1992). Cette région fixe une protéine, homologue à la protéine humaine p54<sup>nrb</sup>/NonO et est capable de se fixer à l'ADN et l'ARN (Basu et *al.* 1997).

L'ensemble de ces études a permis d'identifier *in vitro* un certain nombre d'acteurs régulant l'expression des gènes des IAPs dans différents types cellulaires. La structure de cette région promoteur et des éléments qui s'y fixent est résumée sur la **figure 23**.

b) Mise en évidence de la méthylation dans le contrôle *in vivo* de l'expression des IAPs

Si l'acquisition de ces données était nécessaire pour comprendre les mécanismes de base de l'activation des gènes des IAPs, elle n'expliquait pas comment l'expression des IAPs était régulée *in vivo*. En effet les IAPs sont présents en quantité importante dans le génome de la souris et leur niveau d'expression varie d'un locus à l'autre. Certains IAPs ne sont pas exprimés alors que d'autres ont acquis une certaine tissu-spécificité. La localisation génomique des IAPs peut être à l'origine de modifications de leur profil d'expression car la présence d'activateurs, situés près du site d'insertion, peut influencer directement sur la régulation de l'expression des gènes des IAPs et en modifier considérablement le profil d'expression.

Afin de mieux comprendre l'implication directe des LTR des IAPs dans la régulation du profil d'expression de ces gènes *in vivo*, des souris transgéniques ont été produites pour l'expression du gène rapporteur LacZ sous le contrôle d'une LTR d'IAP (Dupressoir et

Heidmann 1996). L'analyse de l'expression du transgène dans les différentes lignées créées a été réalisée soit par Northern Blot et RT-PCR pour détecter les transcrits, soit par coloration des tissus au X-gal pour détecter le gène rapporteur LacZ. Cette analyse montre que le transgène n'est détectable que dans les testicules et ce, sur 5 générations de souris. Une expression faible dans le cerveau a également été détectée mais seulement dans la première génération de souris.

L'analyse plus fine du profil d'expression dans les testicules montre que cette expression est restreinte aux cellules germinales, comme les gonocytes embryonnaires et les spermatogonies undifférenciées. Par contre, aucune expression n'est détectée dans les stades embryonnaires précoces ou dans les gonocytes femelles.

Afin de comprendre l'origine de l'extinction de l'expression des LTRs dans les autres tissus les auteurs ont analysé le niveau de méthylation du transgène dans ces tissus versus le testicule. Ils ont ainsi montré que le transgène était méthylé dans ces tissus mais pas dans le testicule, ce qui permettait le maintien de son expression. D'autre part, la même démarche expérimentale appliquée aux IAPs endogènes montre que ceux ci sont soumis à la même régulation et sont hypométhylés dans le testicule.

Une autre équipe a également analysé en détail le profil de méthylation *in vivo* de ces gènes dans des souris sauvages et obtient des résultats proches, même s'ils ne sont pas tout à fait identiques (Walsh et al. 1998). Ils observent ainsi une déméthylation uniquement dans les cellules germinales primordiales mâles et femelles et dans l'embryon préimplantatoire mais pas dans les ovocytes adultes bloqués en prophase de méiose ou dans les spermatogonies.

Ces résultats montrent que l'activité des LTRs des IAPs chez la souris est très restreinte et semble se limiter aux cellules germinales primordiales et, dans certaines conditions, aux cellules souches de l'ovaire et du testicule. Ces travaux montrent également que la méthylation joue un rôle important dans la régulation de l'expression des IAP.

#### c) Analyse du rôle de la méthylation dans le contrôle de l'expression des IAPs

D'autres études réalisées *in vitro* et *in vivo* confirment d'ailleurs l'importance de la méthylation pour la régulation de l'expression des IAPs.

La méthylation de la région *enh2* inhibe la fixation du facteur EBP-80 et bloque ainsi son activité activatrice (Falzon et Kuff 1991). L'inactivation de la *méthyltransférase-1* (*Dnmt1*), gène nécessaire au maintien du niveau de méthylation de l'ADN, réactive fortement l'expression des gènes des IAPs (50 à 100 fois) et entraîne une déméthylation des LTRs, et ce,

quelle que soit le type d'IAP et le tissu étudié (Walsh et al. 1998). Une autre étude récente montre que l'insertion du promoteur d'IAP dans un locus fortement sollicité dans certains organes permet de lever l'extinction de l'expression par la méthylation dans ces organes (Barbot et al. 2002). Les auteurs montrent ainsi que l'insertion d'un rétrotransposon IAP dans le locus d'un gène soumis à l'horloge circadienne (le gène *m.nocturnin*) entraîne dans le foie une déméthylation progressive de l'élément IAP et aboutit chez la souris âgée à une levée importante de l'inhibition de la transcription du facteur IAP. Les auteurs supposent que la réactivation répétée par les éléments régulateurs de l'horloge au niveau du locus *m.nocturnin* est à l'origine de cette déméthylation progressive et cumulative.

### 3°) UNE REGULATION A PLUSIEURS NIVEAUX CONTROLE L'EXPRESSION DES IAPS

L'ensemble de ces résultats suggère qu'il existe plusieurs niveaux de régulation de l'expression des IAPs. Tout d'abord un système d'expression localisé dans la région U3 permet l'activation de la transcription dans les cellules par le biais d'un promoteur basal et de régions activatrices. Un système épigénétique va aussi limiter la fonctionnalité de ce système d'expression dans un type cellulaire donné, ici les cellules souches germinales primordiales. Comment sont régulées ces méthylations ? Plusieurs hypothèses peuvent être avancées :

- Les facteurs induisant la méthylation du promoteur sont recrutés au niveau de séquences cibles présentes sur le promoteur et entraînent rapidement l'extinction de son activité. Le récepteur nucléaire orphelin GCNF (Germ Cell Nuclear Factor) pourrait être un bon candidat à ce type de régulation. En effet ce facteur restreint l'expression du gène *Oct-3/4* dans les cellules germinales primordiales par l'intermédiaire d'un élément de réponse localisé dans le promoteur d'*Oct-3/4*.
- Il existe dans la région U3 du promoteur une séquence de liaison pour un facteur nucléaire spécifique des cellules germinales primordiales. Celui-ci, en activant spécifiquement l'expression du gène dans les cellules germinales primordiales le protégerait contre les événements de méthylation. C'est l'hypothèse avancée par Barbot et al. (2002) pour le contrôle de l'expression des IAPs par les gènes de l'horloge circadienne dans le foie.
- Toutes les cellules de l'organisme vont spontanément méthyler les îlots CpG présents dans les séquences IAPs par le biais des méthyltransférases. Il est alors possible que ces gènes restent actifs dans les cellules germinales primordiales car, tout au long du développement embryonnaire, ces cellules sont protégées de la méthylation (soit par l'absence des méthyltransférase soit par l'inhibition de celle-ci) ou alors car une

reprogrammation épigénétique survient dans ces cellules et déméthyle globalement leur génome. Plusieurs arguments sont en faveur de cette hypothèse. En effet le génome des cellules germinales primordiales (PGC) subit une première méthylation avant d'être à nouveau déméthylé au cours du développement embryonnaire comme le montre la **figure 24** (pour revue Reik et *al.* 2001). Une étude montre ainsi que le niveau de méthylation de différents gènes (dont les éléments IAPs) diminue dans des cellules germinales primordiales isolées de 11.5 jours pc à 13.5 jours pc (Hajkova et *al.* 2002).

Même si les approches utilisées pour expliquer le contrôle de l'expression des IAPs apportent parfois des résultats en légères contradictions concernant le type de cellules méthylées ou non, l'ensemble des données actuellement disponibles suggère fortement que la méthylation est un régulateur fort de l'expression des éléments IAPs au cours du développement embryonnaire. L'expression spécifique des IAPs dans les cellules germinales primordiales serait en effet due à une déméthylation générale du génome dans ces cellules.

L'exemple des IAPs montre combien l'étude de la régulation des séquences rétrovirales est délicate. Alors que l'on pourrait spontanément supposer que toute l'information est facilement accessible par l'analyse de la région U3 on se rend compte, qu'*in vivo*, un grand nombre de facteurs intervient aussi bien en *cis* qu'en *trans* et induit des modifications épigénétiques importantes. Si ces modifications sont peut-être la conséquence d'un mécanisme de défense contre l'insertion de séquences d'ADN exogènes dans l'organisme (Yoder et *al.* 1997), la façon dont ces mécanismes sont ciblés sur ces séquences demeure encore méconnue.

### **III REGULATION DES GENES RETROVIRAUX DANS LES CELLULES SOUCHES EMBRYONNAIRES**

L'idée d'utiliser les rétrovirus pour faire de la transgénèse à partir des cellules ES ou des cellules EC était particulièrement intéressante. En infectant les cellules et en les réinjectant dans un embryon receveur, il était possible d'obtenir des animaux transgéniques dont la lignée germinale était capable de transmettre le transgène à leurs descendance.



## 1°) EXPRESSION DES GENOMES RETROVIRAUX DANS LES CELLULES ES

Afin de montrer la faisabilité d'une stratégie de transgénèse par des vecteurs rétroviraux, de nombreuses équipes ont infecté des cellules EC et ES à partir de vecteurs rétroviraux dérivés du virus MMLV (Moloney murine Leukemia Virus). Les résultats ont été relativement décevants puisque les auteurs observaient rapidement une extinction de l'activité des LTRs dans ces cellules ( Linney et *al.* 1984, Stewart et *al.* 1982, Tsukiyama et *al.* 1989). Un problème majeur a été rapidement mis en évidence : les séquences régulatrices des rétrovirus étaient très rapidement inactivées. Cette restriction de l'activité des LTRs des rétrovirus dans les cellules ES ne semble pas limitée aujourd'hui au virus MMLV. D'autres travaux utilisant le virus VIH-1 ou bien encore le virus MSCV (mouse stem cell virus) ont également montré qu'une extinction de l'expression du génome viral survenait à plus ou moins long terme (Cherry et *al.* 2000, Pannell et *al.* 2000).

Il est intéressant de constater le parallèle entre cette inactivation des LTRs dans les cellules ES et celle qui se produit également chez l'embryon précoce de souris aux stades embryonnaires d'où sont dérivées les cellules ES. Des infections d'embryons préimplantatoires par le virus MMLV montrent ainsi que ces stades embryonnaires ne sont pas permissifs pour la réplication du virus (Jaenisch et *al.* 1975). D'autres travaux montrent également que les LTRs du virus MMLV sont inactivées dans l'embryon précoce (notamment au stade blastocyste) et qu'ultérieurement, à 8-9 jours post coïtum cette activité est restreinte aux cellules du mésoderme et de l'ectoderme (Savatier et *al.* 1990). Par contre, l'activité d'un promoteur ubiquitaire (comme celui du gène  *$\beta$ -actin*) placé dans la même structure rétrovirale, n'est pas affectée lorsque ce promoteur est introduit dans l'embryon. Ces observations ont été également constatées dans les cellules ES (Soriano et *al.* 1991) et suggèrent donc que l'inactivité du virus MMLV dans les cellules serait due à une régulation spécifique aux niveaux de ses LTRs.

Néanmoins cette hypothèse ne peut expliquer totalement pourquoi différents vecteurs rétroviraux ne fonctionnent pas à long terme dans les cellules ES. D'autres mécanismes d'extinction de l'expression sont certainement impliqués, comme le suggère la corrélation entre la répression de l'expression du génome proviral et la méthylation des LTRs, aussi bien dans l'embryon pré et périimplantatoire (Jähner et *al.* 1982) que dans les cellules ES ou les cellules souches hématopoïétiques (Challita et Kohn 1994).

L'ensemble de ces résultats suggère que les LTRs des rétrovirus dans les cellules ES et dans l'embryon précoce sont inactivées par plusieurs mécanismes faisant intervenir

simultanément des méthyltransférases mais aussi des facteurs agissant au niveau des LTR et induisant une extinction rapide de l'expression du génome viral.

## 2°) LA METHYLATION DU GENOME VIRAL DANS LES CELLULES ES

La méthylation de l'ADN a été proposée comme un mécanisme de défense contre la présence d'ADN étranger dans la cellule et notamment contre les rétrovirus et les éléments transposables (Yoder *et al.* 1997). L'extinction de l'activité des LTRs des rétrovirus dans les cellules ES et EC peut ainsi être corrélée avec une méthylation de l'ADN de ces régions et peut être levée par l'usage de traitements chimiques (5-azacytidine) entraînant une déméthylation générale du génome (Stewart *et al.* 1982). La présence d'une forte activité de méthylation *de novo* de l'ADN dans les cellules ES conforte cette hypothèse en comparaison des cellules différenciées où cette activité est beaucoup plus faible et où les infections fonctionnent mieux (Lei *et al.* 1996).

La découverte chez les mammifères d'une première méthyltransférase a conforté l'idée selon laquelle la méthylation était directement impliquée dans le contrôle de l'expression des rétrovirus. L'invalidation de la première méthyltransférase de souris (*Dnmt-1*) a ainsi montré que les rétrovirus endogènes présents dans le génome des cellules ES *dnmt1<sup>-/-</sup>* étaient beaucoup moins méthylés que ceux des cellules ES *+/+* (Li *et al.* 1992). L'expression de *Dnmt-1* dans des cellules ES *dnmt<sup>-/-</sup>* compense la perte de fonction et restaure la méthylation des séquences IAP endogènes (Biniszkiewicz *et al.* 2002) démontrant ainsi l'importance de cette méthyltransférase dans le maintien du niveau de méthylation de l'ADN dans le génome des cellules ES mais aussi de l'embryon.

Cependant, l'invalidation du gène *Dnmt-1* n'a pas permis de comprendre totalement comment était méthylé *de novo* le génome des cellules ES. En effet, dans les cellules ES *dnmt<sup>-/-</sup>* le niveau de méthylation du génome est faible mais stable dans le temps. D'autre part il n'y a pas de différences significatives entre le niveau de méthylation d'un ADN proviral après infection d'une cellule ES normale ou *dnmt<sup>-/-</sup>* infectées par le virus MMLV. Ces résultats suggèrent ainsi la présence d'autres méthyltransférases chez la souris, capables de méthyler *de novo* l'ADN cellulaire et viral (Lei *et al.* 1996).

La découverte d'une deuxième méthyltransférase codée par le gène *Dnmt-2* et son invalidation dans les cellules ES n'a pas permis de mieux comprendre ces mécanismes car cette enzyme ne présentait ni activité de méthylation *de novo* des séquences rétrovirales

intégrées dans des cellules ES *dnmt2<sup>-/-</sup>* infectées, ni maintien du niveau de méthylation dans ces cellules (Okano et al. 1998).

C'est par contre l'identification de deux nouvelles méthyltransférases chez la souris, Dnmt3a et Dnmt3b, qui va permettre d'élucider le mécanisme de méthylation *de novo* dans l'embryon de souris et dans les cellules ES. L'invalidation de ces gènes a ainsi démontré qu'ils étaient requis pour la méthylation *de novo* du génome de l'embryon postimplantatoire et des cellules ES (Okano et al. 1999). Contrairement à Dnmt1 qui méthyle préférentiellement des molécules d'ADN déjà hémiméthylées, Dnmt3a et Dnmt3b méthylent préférentiellement un ADN non méthylé et ce aussi bien sur des CpG que sur des CpA (Dodge et al. 2002). Ces résultats suggèrent donc que la régulation de la méthylation se fait en deux temps : tout d'abord une méthylation *de novo* du génome embryonnaire au cours du développement péri et postimplantatoire assurée par les méthyltransférases Dnmt3a et Dnmt3b puis une maintenance de la méthylation assurée par la méthyltransférase Dnmt1 qui reconnaît les substrats hémiméthylés et permet la maintenance de la méthylation au cours des divisions cellulaires successives (**Figure 25**).

Si ces expériences ont partiellement élucidé le mécanisme contrôlant la méthylation du génome embryonnaire, elles ne répondent pas à la problématique concernant la place de la méthylation dans les mécanismes de régulation de l'expression des rétrovirus dans les cellules ES. En effet la méthylation de l'ADN est-elle à l'origine ou au contraire la conséquence de cette répression de l'activité des LTRs rétrovirales ? Plusieurs travaux montrent en effet que les mécanismes d'extinction par la méthylation ne sont pas forcément directement impliqués dans le contrôle de l'activité des LTRs rétrovirales puisque l'extinction de l'expression du virus MMLV est constatée 2 jours après l'infection de cellules EC alors que l'évènement de méthylation n'est détecté que 8 à 16 jours après l'infection (Stewart et al. 1982).

### 3°) D'AUTRES MECANISMES REGULENT L'EXTINCTION DE L'ACTIVITE DES LTRs DANS LES CELLULES ES

Plusieurs équipes ont mis en évidence que des facteurs cellulaires agissant en *trans* régulaient négativement l'expression de l'ADN rétroviral dans les cellules EC et ES. Le facteur YY1, le facteur A, la protéine ELP (embryonic LTR-binding protein) ont ainsi été identifiées comme répresseurs de l'expression des LTRs rétrovirales (revue dans Pannell et Ellis 2001). L'analyse des LTRs du virus MMLV a ainsi montré qu'il était possible de les modifier de façon à augmenter significativement leur activité dans les cellules ES, en

favorisant la fixation des éléments activateurs et en supprimant la fixation d'éléments répresseurs (revue dans Challita et al. 1995). D'autre part l'utilisation de LTRs tronquées d'un autre rétrovirus PCMV, dérivé du virus du sarcome myeloprolifératif (MPSV) a permis de construire des vecteurs rétroviraux, comme le MESV (mouse embryonic stem cell virus) ou le MSCV (mouse stem cell virus) avec une efficacité d'expression bien plus forte dans les cellules EC et les cellules ES (Osborne et al. 1999, Grez et al. 1990, Cherry et al. 2000). Cependant, les études prouvent que même ces vecteurs modifiés sont soumis dans les cellules ES à des mécanismes d'extinction d'expression, indépendant de la méthylation. Ainsi l'infection de cellules ES *dnmt1*<sup>-/-</sup>, *dnmt3*<sup>-/-</sup> ou normales par le virus MSCV (mouse stem cell virus) montre que son niveau basal d'expression n'est pas modifié quel que soit le type cellulaire testé et qu'il reste beaucoup plus faible que dans d'autres cellules comme les fibroblastes transformés de la lignée cellulaire 3T3 (Cherry et al. 2000, Pannell et al. 2000). Ces résultats suggèrent donc qu'il existe dans les cellules ES un mécanisme d'extinction rapide et indépendant des méthyltransférases dont le rôle plus tardif serait de maintenir dans le temps cette extinction. Cependant la façon dont est ciblée cette extinction dans les cellules ES est encore mal comprise car la plupart des facteurs cellulaires agissant en *trans* sont exprimés de façon ubiquitaire.

Il semble aujourd'hui que les facteurs agissant sur la structure de la chromatine soient impliqués dans ce mécanisme d'extinction rapide. L'étude de la structure de la chromatine par immunoprécipitation des histones H3 acétylées et phosphorylées, H4 acétylées et H1 dans des lignées de souris transgéniques pour un vecteur rétroviral de type MSCV montre ainsi que la chromatine est dans un état répressif pour la transcription de l'ADN proviral (Pannell et al. 2000).

Afin de comprendre pourquoi les LTRs rétrovirales étaient inactivées dans les cellules ES mais pas dans les cellules différenciées, les auteurs supposent l'existence d'un mécanisme qu'ils nomment somno-complexe et qui fait intervenir des complexes protéiques encore inconnus et agissant spécifiquement dans les cellules souches. Ils proposent donc un modèle en trois étapes pour expliquer le mécanismes d'extinction de l'activité des LTRs rétrovirales dans les cellules ES et l'embryon précoce (**Figure 26**, Pannell et Ellis 2001) :

- Fixation de facteurs cellulaires *trans*-répresseurs sur les LTRs rétrovirales
- Recrutement de facteurs de remodelage de la chromatine
- Fermeture de la chromatine à la transcription par la fixation de protéines associées à la chromatine

- Méthylation *de novo* par Dnmt3a et Dnmt3b
- ADN proviral maintenu silencieux par Dnmt1

Ce modèle demeure encore très spéculatif car peu d'acteurs impliqués dans ces différentes étapes ont été identifiés, notamment concernant le remodelage de la chromatine. Il reste également à définir la dynamique exacte de ces différents événements mais surtout à identifier les moyens de reconnaissance des LTRs viraux par la cellule, puisque les mutations produites dans les LTRs n'empêchent jamais totalement l'extinction de l'expression. Il est ainsi intéressant de s'intéresser aux mécanismes de reconnaissance de l'ADN exogène par les cellules ES et de l'embryon précoce, ou encore aux mécanismes de reconnaissance et de protection de l'ADN endogène.

La deuxième hypothèse est particulièrement attrayante puisque de nombreux gènes endogènes ubiquitaires sont protégés des mécanismes d'extinction d'expression au cours du développement. Le développement des stratégies de gene-trap ou de Knock-In vise d'ailleurs à utiliser ces loci protégés pour placer des transgènes dans des régions où ils ne seront pas réprimés.

#### 4°) LA CELLULE ES POSSEDE UN ETAT EPIGENETIQUE INSTABLE

La forte activité des méthyltransférases dans les cellules ES suggère que le génome de ces cellules subit des remaniements importants au cours du temps et que les populations de cellules ES sont probablement différentes entre le moment de leur mise en culture et à des passages plus tardifs.

Les travaux de Humpherys et coll. (2001) sur l'expression des gènes soumis à l'empreinte dans les cellules ES confirment cette hypothèse. Les gènes soumis à l'empreinte sont des gènes exprimés préférentiellement selon leur position sur le génome d'origine maternel ou paternel. Il a été montré que cette régulation de l'expression était due notamment à une méthylation différentielle des loci situés sur le génome d'origine paternel ou maternel. L'étude de l'expression du gène H19 dans les cellules ES montrent de très fortes variations selon la lignée cellulaire étudiée mais aussi selon les clones cellulaires dérivés d'une même lignée. Ces variations peuvent être corrélées avec le niveau de méthylation du locus. Ces résultats montrent que le contrôle de l'expression de certains gènes endogènes est particulièrement instable d'une cellule ES à l'autre, ce qui pourrait être étendu sans doute aux gènes exogènes comme l'ADN proviral.

Pour expliquer cette instabilité on pourrait mettre en avant le vieillissement des cellules après de nombreux passages en culture, la présence d'aberrations chromosomiques ou bien encore une structure chromatinienne plus permissive dans ces cellules, qui permettrait la mise en place de régulations aberrantes au cours des passages en culture. Si cet état chromatinien permissif existe dans l'embryon, il est beaucoup plus restreint dans le temps.

#### 5°) LA REGULATION DES SEQUENCES RETROVIRALES DANS LES CELLULES ES ET L'EMBRYON PRECOCE DE POULET

Peu de données sont disponibles concernant la régulation de l'expression de l'ADN proviral des rétrovirus dans l'embryon précoce de poulet et les cellules issues du blastoderme. Il est donc difficile de généraliser et de comparer avec les travaux réalisés chez la souris. Néanmoins la plupart des résultats publiés ne vont pas dans le sens d'une conservation des mécanismes d'extinction des séquences rétrovirales observés chez la souris.

La plupart des études initiées chez le poulet sur l'expression des séquences rétrovirales ont été menées pour montrer la faisabilité de la transgénèse chez cet oiseau par le biais de vecteurs rétroviraux.

Certains auteurs ont ainsi montré qu'il était possible d'infecter des embryons de poulet au stade X (EG) et de suivre la transmission de l'ADN proviral dans le génome de leur descendance (Salter et *al.* 1986, Crittendenn et *al.* 1990, Bosselman et *al.* 1989, Thoraval et *al.* 1995). Cependant leurs analyses ont porté sur la présence de l'ADN proviral dans l'ADN génomique des oiseaux et non sur son niveau d'expression au fil des générations. Chez la caille, une étude s'est penchée sur le niveau d'expression d'un gène rapporteur *cat* inséré dans un vecteur rétroviral (Lee et Shuman 1990). Un oiseau mâle transmettait le vecteur à sa descendance où l'expression du gène *cat* variait suivant le tissu étudié (très élevée dans l'intestin et le thymus, faible dans le rein et les gonades, nulle dans le sang, le cœur et le foie). S'il est possible que ces régulations soient dues aux LTRs, il ne faut pas non plus exclure un effet de position qui dirigerait l'expression du transgène dans ces tissus, étant donné que la descendance d'un seul individu a pu être étudiée. Une autre étude réalisée chez le poulet semble tout de même confirmer que l'ADN proviral n'est pas éteint au cours du développement embryonnaire après l'infection des cellules du blastoderme d'un embryon de poulet au stade X. L'expression du gène rapporteur LacZ sous le contrôle des LTRs

rétrovirales est ainsi détectée à la fois dans le mésoderme et l'ectoderme de nombreux embryons (14 observés) (Reddy et al. 1991).

Si les premières études se sont portées sur l'infection des embryons *in ovo*, quelques publications portent sur l'infectabilité et l'activité du génome rétrovirale dans des cellules du blastoderme en culture. Les résultats sont, par contre, contradictoires. Une première étude semble en effet s'accorder avec les observations réalisées sur les cellules ES de souris. A partir de culture de cellules du blastoderme d'embryon de poulet au stade X, Mitrani et al. (1987) avaient montré que l'infection de ces cellules était possible mais que l'expression provirale restait très faible et insuffisante pour produire des particules virales. D'autres auteurs, par contre, ont montré que les vecteurs rétroviraux dérivés du RSV sont capables de s'intégrer et de s'exprimer dans ces cellules du blastoderme de la même façon que dans des fibroblastes embryonnaires (Reddy et al. 1991).

Les résultats obtenus au laboratoire s'accordent avec ces derniers résultats et suggèrent que certaines séquences rétrovirales ne sont pas soumises aux mêmes mécanismes de régulation dans les cellules ES de poulet et de souris. En effet contrairement aux cellules ES de souris, nous avons montré que des séquences d'origine rétrovirale, appelées *ENS* pour Embryonic Normal Stem cell genes, sont fortement exprimées dans les cellules ES et dans l'embryon précoce de poulet puis s'éteignent, que ce soit au cours du développement embryonnaire ou de la différenciation *in vitro* de ces cellules (Acloque et al. 2001). Les cellules ES de poulet seraient donc permissives pour l'expression de ces séquences d'origine rétrovirale. Malgré tout, il semble que des mécanismes de régulation de l'expression des gènes par la méthylation existe. Il a ainsi été montré que le génome de l'embryon de poulet était soumis précocément à une régulation par la méthylation et qu'une déméthylation générale du génome par la 5-Azacytidine pouvait perturber fortement la gastrulation avec une ligne primitive très réduite (Zagris et Podimatas 1994).

Pour expliquer l'expression de ces rétrovirus endogènes dans les cellules ES de poulet, on peut avancer plusieurs hypothèses:

- les cellules ES de poulet sont issues d'embryons où la méthylation du génome n'a pas encore eu lieu. Leur génome est donc largement déméthylé et le demeure en l'absence de méthylation *de novo*. Plus tard, au cours du développement

embryonnaire ou lors de l'induction *in vitro* de la différenciation, le génome des cellules est à nouveau méthylé et les séquences rétrovirales endogènes sont éteintes.

- Les cellules ES sont en fait des cellules EG dont le génome est maintenu dans un état d'hypométhylation (comme les cellules germinales primordiales embryonnaires) et où les séquences rétrovirales endogènes (comme les IAPs chez la souris) sont exprimées
- Les séquences rétrovirales ENS exprimées bénéficient d'un effet de position dans les cellules ES qui favorisent leur expression
- Les séquences rétrovirales ENS ne sont pas éteintes dans les cellules ES et sont protégées de la méthylation dans ces cellules car la fonction du gène *ENS-1* est devenue cruciale pour la survie embryonnaire dans cette espèce. On peut alors supposer qu'au cours de l'évolution les individus conservant l'expression de ce gène aux stades embryonnaires précoces ont été avantagés.

Comme on l'a vu au cours de ce chapitre, la régulation de l'expression des gènes rétroviraux dans les cellules ES est relativement complexe et dépend de nombreux facteurs aussi bien endogènes qu'exogènes. L'étude des mécanismes régulant l'expression d'un ADN proviral doit donc intégrer à la fois l'analyse des facteurs agissant directement sur le promoteur de ce gène mais aussi l'étude des facteurs épigénétiques contrôlant son environnement chromatinien, comme la méthylation ou l'acétylation des histones par exemple.



---

## PARTIE D

### LE FACTEUR DE TRANSCRIPTION CP2

---

L'analyse du promoteur des gènes *ENS* a montré qu'un élément de réponse nécessaire à l'activité du promoteur était reconnu par le facteur de transcription CP2. L'objectif de cette quatrième partie est donc de présenter l'essentiel des connaissances acquises sur ce facteur de transcription.

#### I PRESENTATION GENERAL DU FACTEUR DE TRANSCRIPTION CP2

##### 1°) IDENTIFICATION DU GENE *CP2*

Le facteur de transcription CP2 a été identifié à la fin des années 80 par des équipes travaillant sur différents systèmes de régulation de l'expression des gènes. Tout d'abord identifié comme un facteur cellulaire se fixant sur la région CP2 du promoteur de la globine  $\alpha$  de souris (Barnhart et al. 1988), sur le promoteur tardif de SV40 (Kim et al. 1987) ou bien encore sur la séquence leader du promoteur du VIH (Wu et al. 1988), les facteurs de transcription appelés respectivement CP2, LSF (Late SV40 transcription Factor) ou LBP-1 (Leader Binding Protein-1) ont été clonés en 1992 par Lim et al. (pour CP2), en 1994 par Shirra et al. (pour LSF) et par Yoon et al. (pour LBP-1). Chez le poulet, le gène *CP2* a été identifié comme un activateur du promoteur du gène de la *crystalline  $\alpha A$*  (Murata et al. 1998).

Chez la souris, le gène est structuré en 16 exons (15 pour l'homme), s'étend sur 30 kb d'ADN génomique et se localise à l'extrémité distale du chromosome 15 (chromosome 12 chez l'homme), proche des gènes *Gdc-1*, *Wnt-1* et *Rarg*, une région synténique du chromosome 12 humain (Swendeman et al. 1994).

## 2°) CONSERVATION DU GENE *CP2*

Le gène *CP2* appartient à la famille des facteurs de transcription de type NTF dont l'archétype est le facteur NTF-1 (*Neurogenic element binding Transcription Factor*) de drosophile. *NTF-1*, encore appelé *grainyhead* ou *Elf-1*, est important pour la mise en place de l'axe dorso-ventral de l'embryon de drosophile et régule directement l'expression de nombreux gènes comme *engrailed* ou *ultrabithorax*. Chez les mammifères, deux homologues de *NTF-1* ont été initialement identifiés et caractérisés, *CP2* et *LBP-1a* (*NF2d9* chez la souris) (Lim et al. 1992, Yoon et al. 1994). Ces deux gènes sont relativement proches puisque leur séquence peptidique est identique à 72%. *LBP-1d* est une forme tronquée de *CP2* résultant d'un épissage alternatif. *LBP-1c* est incapable de fixer l'ADN (Yoon et al. 1994). D'autres gènes de la même famille ont été identifiés par la suite dont le gène CRTR-1 de souris (pour CP2-Related Transcriptional Repressor également appelé LBP-9 chez l'homme) dont la séquence peptidique est identique à 70% avec celle de *CP2* (Rodda et al. 2001). Une étude récente a également identifié d'autres gènes de mammifères plus apparentés à *NTF-1* que *CP2* comme *MGR* (Mammalian grainyhead), *BOM* (Brother of MGR) ou *SOM* (Sister of MGR) mais également un nouveau gène de drosophile *dCP2* plus proche de *CP2* que de *NTF-1* (Wilanowski et al. 2002). L'analyse phylogénétique de ces différents gènes, représentée sur la **figure 27**, met en évidence deux grandes branches regroupant *grainyhead* et ses homologues mammaliens sur la première ramification et *dCP2* et ses homologues *CP2* (*LBP-1c*, *LSF*), *LBP-1a* (*NF2d9*) et *LBP-9* (*CRTR-1*) sur l'autre.

## 3°) CARACTERISTIQUE DE LA PROTEINE *CP2*

La protéine *CP2* est codée par un ARNm d'une longueur de 4kb environ qui comporte un cadre ouvert de lecture de 1506 pb. La protéine *CP2* se compose donc de 502 acides aminés chez la souris. Sa séquence peptidique est très conservée entre la souris et l'homme avec 96% d'homologie de séquence (Lim et al. 1992). Chez le poulet elle est légèrement plus longue avec une taille de 506 acides aminés mais reste très homologue aux gènes mammaliens avec une identité de séquence proche de 90% (Murata et al. 1998).

La taille de la protéine révélée par immunoblot est de 64 kD. Elle comprend un grand domaine de fixation de l'ADN de 318 acides aminés (acides aminés 65 à 383) dont une région indispensable à la fixation (acides aminés 189 à 239), est hautement conservée entre les

différentes espèces (100% d'homologie avec le poulet, la souris et l'homme) et est proche de celle du gène *NTF-1* de drosophile (avec 40% d'homologie). Ce domaine de fixation à l'ADN reconnaît un site de fixation constitué de deux répétitions directes du type : CNRG N<sub>6</sub> CNRG (Lim et *al.* 1993). Chacune des deux répétitions est indispensable à l'interaction, la protéine CP2 ne reconnaissant pas une répétition seule, au contraire de NTF-1. La protéine CP2 se fixe sur cet élément de réponse sous la forme de tétramères grâce à son domaine d'oligomérisation composé de 137 acides aminés (acides aminés 266 à 403) (Shirra et *al.* 1998, Murata et *al.* 1998). Ces données montrent que le domaine de fixation à l'ADN et le domaine d'oligomérisation se chevauchent, suggérant que ces deux régions sont nécessaires pour conserver l'interaction (**Figure 28**). Le site de fixation à l'ADN au sens strict doit être beaucoup plus restreint en taille et correspond sans doute à la région très conservée entre CP2 et NTF-1. D'autres régulations peuvent également influencer sur la capacité de CP2 à se fixer sur l'ADN. Il a ainsi été montré que la phosphorylation de CP2 sur la sérine 291 (conservée chez le poulet) augmente la capacité de la protéine à se lier à l'ADN. Une mutation de cet acide aminé ou le traitement par des phosphatases réduit d'ailleurs la fixation de CP2 sur l'ADN, confirmant ainsi la régulation par des kinases de l'activité du facteur CP2 (Volker et *al.* 1997).

Le profil d'expression de la protéine CP2 est relativement large puisqu'elle semble exprimée dans tous les tissus testés aussi bien embryonnaires qu'adultes, chez la souris et le poulet (Swendeman et *al.* 1994, Murata et *al.* 1998). Cependant il n'y a pas de données disponibles sur le profil d'expression de ce gène dans l'embryon précoce, les études précédentes ayant portées sur des embryons de plus de 9.5 jours post coïtum pour la souris et de 14 jours d'incubation pour le poulet. Par contre un gène proche de CP2, le gène CRTR-1 (LBP-9), est exprimé fortement dans les cellules souches embryonnaires de souris et dans certains tissus comme le rein (Rodda et *al.* 2001).

## II FONCTIONS DU FACTEUR DE TRANSCRIPTION CP2

### 1°) PROMOTEURS REGULES *IN VITRO* PAR CP2

De nombreuses études de promoteur *in vitro* ont montré que le facteur de transcription CP2 pouvait se fixer sur ces promoteurs et en activer ou réprimer l'activité transcriptionnelle. Ces différents promoteurs sont présentés dans le tableau suivant :

Nom du promoteur	Activation ou répression	Références
$\alpha$ A-crystalline de poulet	Activation	Murata <i>et al.</i> 1998
promoteur tardif SV40	Activation	Huang <i>et al.</i> 1990
$\alpha$ -globine de souris	Activation	Lim <i>et al.</i> 1992
VIH-1	Répression	Yoon <i>et al.</i> 1994
$\gamma$ -fibrinogène de rat	nd	Lim <i>et al.</i> 1993
c-Fos humain	nd	Shirra et Hansen 1998
Thymidilate synthase murine et humaine	Activation	Powell <i>et al.</i> 2000
Sérum amyloïde A humain	Activation	Bing <i>et al.</i> 1999
Interleukine 2 humaine	Répression	Casolaro <i>et al.</i> 2000
Interleukine 4 humaine	Activation	Casolaro <i>et al.</i> 2000
Synthase de l'uroporphyrinogène III	Nd	Solis <i>et al.</i> 2001
Collagène de type III du poulet	Répression	Zhang <i>et al.</i> 1999
Phospholipase C- $\beta$ 3 de rat	nd	Kang <i>et al.</i> 1997
Complexe majeur d'histocompatibilité classe II Ea	nd	Bellorini <i>et al.</i> 1996

**Tableau 4** : Différents promoteurs régulés par CP2

Comme on le voit ici, l'expression de nombreux gènes est régulée par CP2, confirmant ainsi son rôle assez général dans le contrôle de l'expression des gènes. Son activité peut cependant être modulée *in vitro* par des protéines partenaires qui, suivant les promoteurs étudiées, contrôlent son activité soit comme répresseur ou activateur de la transcription.

## 2°) PARTENAIRES DE CP2 IDENTIFIES *IN VITRO*

### a) Le facteur YY1

L'interaction entre CP2 et YY1 a été particulièrement bien étudiée dans le cadre de l'étude du promoteur du VIH-1.

YY1 est un facteur de transcription de 68 kDa exprimé dans de nombreux types cellulaires et interagissant avec de nombreux autres facteurs nucléaires aussi bien viraux que cellulaires. YY1 a été montré comme réprimant l'activité des LTR du virus VIH-1. Une première étude a ainsi mis en évidence que l'interaction entre CP2 et YY1 était nécessaire pour induire cette répression (Romerio et *al.* 1997). Il a été montré par la suite que l'interaction entre CP2 et YY1 nécessitait d'une part, les domaines de fixation à l'ADN et d'oligomérisation de CP2, et les deux premiers doigts de zinc (sur 4 au total) de YY1 d'autre part. La fixation du complexe CP2/YY1 sur la LTR du VIH-1 va alors permettre le recrutement par YY1 d'une déacétylase d'histone (HDAC) nécessaire à la répression de la LTR virale (Coull et *al.* 2000). Il semble que la fixation de CP2 sur l'ADN soit indispensable pour assurer cette répression transcriptionnelle. La surexpression d'un dominant négatif de CP2 capable de s'associer avec les CP2 non mutés mais incapable de fixer l'ADN induit ainsi une augmentation du niveau d'acétylation des histones et une diminution du niveau de recrutement de HDAC1 par YY1, limitant alors la répression des LTR virales (He et Margolis 2002). Une expérience identique, basée sur l'utilisation de polyamides capables d'entrer en compétition avec CP2 pour son site de fixation à l'ADN, montre également que la non fixation de CP2 sur la LTR rétrovirale lève l'inhibition induite par le complexe CP2/YY1 (Coull et *al.* 2002).

Comme on l'a vu dans cet exemple, l'association de CP2 (ou de multimères de CP2) avec YY1 permet la modulation de l'activité d'un promoteur, ici sa répression, par le recrutement d'un troisième partenaire induisant un état transcriptionnel répressif de la chromatine.

### b) D'autres partenaires de CP2

D'autres partenaires de CP2 ont également été identifiés dans des contextes différents. Un exemple intéressant est l'identification de NF-E4 comme un partenaire de CP2 pour assurer la régulation temps et tissu spécifique de l'expression de la globine  $\gamma$  fœtale. Les gènes

codant pour les cinq globines fœtales et adultes sont situés dans un même locus mais sont exprimés successivement au cours du développement embryonnaire. Ils dépendent d'une même région de régulation (LCR pour Locus Control Region) pour laquelle leurs différents promoteurs sont en compétition. Certaines régions des promoteurs des gènes de globine permettent de capter l'activité de la LCR plus fortement à certains stades du développement. Une région du promoteur de la globine- $\gamma$  (SSE pour stage selector element) est ainsi connue pour fixer un complexe protéique nommée SSP (pour stage selector protein), important pour bloquer la compétition avec le promoteur de la globine  $\beta$ . Il a été montré que CP2 appartenait à ce complexe protéique (Jane et *al.* 1995). Cependant CP2 étant exprimé de façon ubiquitaire au cours du développement, d'autres protéines du complexe devaient en assurer la spécificité dans l'espace et dans le temps. La recherche de partenaires de CP2 par double hybride a montré que le domaine d'oligomérisation de CP2 pouvait fixer le facteur NF-E4 et que cette protéine était nécessaire pour permettre au complexe SSP de se fixer sur l'ADN. NF-E4 possède un profil d'expression restreint au foie fœtal, au cordon embryonnaire et à la moelle osseuse. Sa surexpression dans des lignées de cellules hématopoïétiques conduit à la surexpression des globines embryonnaires  $\epsilon$  et  $\gamma$  (Zhou et *al.* 2000). L'ensemble de ces données suggèrent que le facteur NF-E4 est bien un élément clé de la régulation de l'expression des globines embryonnaires et confère, en association avec CP2, la spécificité temporelle et spatiale de l'expression des globines embryonnaires. Il est intéressant de se poser la question du rôle de CP2 dans ce complexe. Apparemment, il n'est pas nécessaire pour fixer l'ADN de l'élément SSE, même si un site de fixation pour CP2 (CNRG-N<sub>6</sub>-CNRG) est présent dans cette séquence. Il est possible que NF-E4 agisse comme un compétiteur de CP2, l'empêchant d'adopter sa structure tétramérique fonctionnelle. En présence de forte quantité de NF-E4, CP2 ne se fixerait donc pas sur son élément de réponse mais serait recruté par NF-E4 et induirait l'expression des gènes des globines embryonnaires. En absence de NF-E4, l'homotétramère CP2 peut fixer son élément de réponse. On peut alors supposer qu'il puisse recruter YY1 et HDAC1 et réprimer ainsi l'expression des globines embryonnaires, favorisant alors l'expression de la globine  $\beta$ .

Un autre partenaire de CP2, la protéine dinG, appartient à la famille *Polycomp*. Cette famille de gène est connue pour réprimer l'expression des gènes homéotiques au cours du développement embryonnaire de la drosophile. Elle est aussi impliquée dans les mécanismes de maintien de la chromatine dans un état transcriptionnellement répressif. Cependant, peu de

membres de ce complexe ont la capacité de se lier à l'ADN. Une étude a montré que CP2 interagissait avec la protéine dinG aussi bien *in vitro* que *in vivo* dans des cellules en culture. CP2 pourrait donc agir comme un moyen d'ancrage à l'ADN à l'intérieur du complexe (Tuckfield et al. 2002).

Comme on vient de le voir, CP2 est capable d'interagir avec des partenaires différents, suivant le contexte cellulaire dans lequel il est étudié. Ces partenaires protéiques lui confèrent alors des fonctions distinctes contribuant à l'expression tissu-spécifique de certains gènes au cours du développement embryonnaire.

### 3°) FONCTIONS DE *CP2 IN VIVO*

Très peu d'études ont été réalisées sur la fonction du gène *CP2 in vivo*. Les résultats de l'invalidation du gène *CP2* ont été relativement décevants. En effet le phénotype des souris *CP2<sup>-/-</sup>* est quasiment normal et les animaux sont fertiles (Ramamurthy et al. 2000). Aucune perturbation de l'expression des gènes de la globine n'a été décelée alors qu'il avait été montré *in vitro* l'importance de *CP2* dans le contrôle de la transition pour l'expression des globines embryonnaires et adultes. Il semble donc, soit que les observations réalisées *in vitro* ne reflètent pas les régulations *in vivo*, soit qu'il existe un phénomène de compensation de *CP2* par une protéine homologue. Cet argument est avancé par les auteurs de l'invalidation du gène *CP2* qui montrent que l'absence de *CP2* est compensé *in vivo* par le facteur NF2d9 (LBP-1a chez l'homme), qui est une protéine très proche de *CP2* (Ramamurthy et al. 2000). La double invalidation des gènes *CP2* et *Nf2d9* permettra peut-être de mieux comprendre leurs fonctions *in vivo* chez la souris.

Chez l'homme, le polymorphisme du gène *CP2* a été corrélé avec le développement chez des patients de la maladie d'alzheimer (Luedeking-Zimmer et al. 2003). Des arguments obtenus *in vitro* suggèrent que la protéine *CP2* pourrait être impliquée dans cette pathologie car elle interagit avec Fe65, une protéine importante pour la translocation de la protéine APP ( $\beta$ -amyloid precursor protein), dont la protéolyse conduit à l'accumulation du peptide  $\beta$ -amyloïde à l'origine de la maladie (Bruni et al. 2002, Minopoli et al. 2001).

CP2 est un facteur de transcription ubiquitaire capable de se tétramériser pour activer ou réprimer la transcription de nombreux gènes. Sa capacité à s'associer avec d'autres protéines nucléaires lui confère une modularité importante dans le contrôle de l'expression des gènes au cours du développement embryonnaire.



---

## **PARTIE E**

# **RESULTATS ET DISCUSSION**

---

Les résultats présentés dans cette thèse portent sur la caractérisation d'une famille de gènes exprimés spécifiquement dans les cellules ES et l'embryon précoce de poulet. Un premier article présente la méthodologie qui nous a permis d'identifier ces gènes et les différentes caractéristiques de ces gènes. Ces travaux ont également permis le dépôt d'un brevet pour la France et l'Union Européenne et en cours d'extension pour la Norvège, le Canada et les Etats-Unis. Un deuxième article relate les travaux réalisés sur les mécanismes contrôlant l'activité du promoteur de ces gènes. Nous avons ainsi identifié une région activatrice contrôlée, en partie au moins, par le facteur de transcription CP2. Une troisième étude, qui n'est pas encore achevée, se consacre à la caractérisation d'une région inhibitrice localisée dans le promoteur et particulièrement active dans les cellules différenciées. Cette région pourrait conférer au promoteur sa spécificité de fonction dans les cellules souches embryonnaires pluripotentes.

## **I IDENTIFICATION D'UNE FAMILLE DE GENES EXPRIMEE SPECIFIQUEMENT DANS LES CELLULES ES ET L'EMBRYON PRECOCE DE POULET**

### **1°) INTRODUCTION AU TRAVAIL EXPERIMENTAL**

L'objectif initial du projet était d'obtenir l'expression ubiquitaire et stable d'un gène rapporteur dans les cellules ES de poulet. L'idée était ensuite d'injecter les cellules ES génétiquement modifiées dans un embryon receveur et d'évaluer ainsi la capacité de ces cellules à coloniser cet embryon receveur. La nature réelle de ces cellules n'étant pas encore connue avec précision, l'étude du chimérisme avec des gènes marqueurs permet de définir le potentiel réel de différenciation de ces cellules dans un contexte embryonnaire. Plusieurs stratégies peuvent être envisagées pour exprimer un transgène dans les cellules ES.

L'expérience acquise chez la souris suggérait néanmoins d'utiliser préférentiellement une technique de knock-in ou une technique de *gene-trap* plutôt qu'une approche par un vecteur d'expression classique. En effet l'utilisation de loci endogènes pour l'expression de transgènes est beaucoup plus efficace dans ces cellules que l'utilisation de promoteurs exogènes, souvent rendus silencieux au cours des passages en culture.

## 2°) LA TECHNIQUE DU PIEGEAGE DE GENE (*GENE-TRAP*)

Cette technique consiste à insérer, dans le génome des cellules, un vecteur ne possédant qu'un site accepteur d'épissage et un gène de sélection couplé à un gène rapporteur (**Figure 29-A**). La sélection des cellules permet d'isoler les clones ayant intégrés le transgène en aval d'un promoteur endogène actif dans ces cellules. Il en résulte l'expression d'un transcrite chimère, fusion des séquences endogènes du locus piégé et du gène rapporteur. Il est alors facile d'identifier par PCR le gène endogène piégé. Cette technique permet également d'obtenir des mutants par insertion et d'identifier ensuite *in vivo* (chez la souris) la fonction des gènes piégés.

Le vecteur choisi pour le *gene-trap*, ROSA  $\beta$ -geo, a été développé par Friedrich et Soriano (1991) et se compose de deux LTR de MMLV mutées encadrant un site accepteur d'épissage et une fusion du gène LacZ avec le gène de résistance à la néomycine (**Figure 29-B**).

L'ensemble des travaux expérimentaux concernant le *gene-trap* a été réalisé par Valérie Risson, qui était alors en stage post-doctoral au laboratoire et travaillait avec Bertrand Pain. Elle a également réalisé les hybridations *in situ* et les chimères présentées dans l'article.

## 3°) RESUME DU TRAVAIL EFFECTUE

### a) Identification d'un locus spécifiquement actif dans les cellules ES de poulet

Après transfection du vecteur ROSA- $\beta$ geo dans les cellules ES de poulet, nous avons isolé un clone cellulaire, appelé 9N2-5, exprimant spécifiquement le gène rapporteur LacZ dans les cellules ES indifférenciées. L'expression du transgène est forte dans les cellules ES mais s'éteint au cours de la différenciation induite par l'acide rétinoïque (**cf article Figure 1**). Lorsque ces cellules sont réinjectées dans un embryon hôte au stage X, l'expression du gène

LacZ est détectée dans les chimères lors des premiers stades du développement puis s'éteint, alors qu'il est toujours possible de détecter la présence du transgène dans l'ADN génomique de ces chimères (**cf article Figure 2**). L'ensemble de ces données confirme que le transgène ROSA- $\beta$ geo inséré dans les cellules 9N2-5 s'exprime de façon prépondérante dans les cellules ES indifférenciées et l'embryon précoce de poulet.

#### b) Identification et caractérisation des gènes ENS

Afin de connaître le locus dans lequel s'était inséré le transgène ROSA- $\beta$ geo, nous avons utilisé la technique du 5'RACE. Nous avons ainsi identifié une petite séquence appartenant au locus endogène. Le criblage d'une banque d'ADNc de cellules ES de poulet avec une sonde correspondant à cette petite séquence, nous a permis d'identifier une nouvelle famille de gènes appelés *ENS* (pour Embryonic Normal Stem cell). Trois gènes *ENS* ont été caractérisés (**cf article Figure 7A**): *cENS-1* se compose d'un cadre de lecture inséré entre deux séquences répétées (dont une au moins est présente dans chacun des gènes *ENS*). Son transcrit, d'une taille de 4,2 kb, comprend une séquence codante pour une protéine de 490 acides aminés (**cf article Figure 3**), identique à *cERNI* (Early Response to Neural Induction). Le gène *cERNI*, isolé par Streit et *al.* (2000), a été identifié comme un marqueur précoce de l'induction neurale, induit *in vivo* par le nœud de Hensen ou le FGF-8. La protéine ENS-1 ne présente aucune homologie avec des protéines décrites mais comprend un domaine « coiled-coil » du résidu 80 au résidu 137. Les domaines coiled-coil sont souvent impliqués dans les interactions entre protéines. La séquence codante du gène *ENS-1* a été retrouvée par PCR (polymerase chain reaction) ou southern blot chez de nombreux oiseaux appartenant à la famille des gallinacées (faisan, caille, perdrix rouge et grise, dinde pintade) mais semble absente du génome d'autres oiseaux (comme les canards ou les pigeons), de vertébrés (saumon, xénope, souris, singe et homme) et d'invertébrés (drosophile ou ascidie)(**cf article Figure 4**). L'absence d'homologue dans le génome de la souris et de l'homme a pu être confirmée par la suite grâce à la connaissance de la séquence complète de ces génomes.

D'une taille de 2.5 kb, le transcrit du gène *cENS-2* représente une forme courte de *cENS-1*. Encadré par les mêmes séquences répétées que *cENS-1*, il comporte une séquence codante tronquée pour 83 acides aminés (la délétion représente 1415pb). Si ce transcrit peut résulter d'un épissage alternatif, il existe également sous cette forme dans l'ADN génomique.

*cENS-3* est relativement différent des deux premiers gènes décrits mais il comporte une extrémité 3' identique à *cENS-1* et *cENS-2* sur environ 500pb. Il se compose de deux cadres de lecture adjacents, codant successivement pour des gènes homologues aux gènes rétroviraux *pol* et *env*. Nous avons montré que, de façon surprenante, la séquence codante pour *cENS-1* était liée avec celle du gène *pol* sur l'ADN génomique mais aussi sur des ADNc produits à partir d'ARN totaux de cellules ES de poulet (**Figure 7B**). La séquence codante d'*ENS-1* ne présente cependant aucune homologie avec celle des gènes *gag* connus.

Il existe dans le génome du poulet de nombreuses copies de ces trois gènes, et notamment du gène *cENS-1*, pour lequel nous avons détecté au moins onze loci différents. La connaissance de la séquence complète du génome du poulet, qui devrait être bientôt disponible, apportera beaucoup de réponses sur la structure génomique de cette famille de gènes, aussi bien concernant le nombre de copies que la localisation chromosomique de ces 3 gènes.

c) Profil d'expression des gènes *ENS* dans les cellules ES et l'embryon précoce de poulet

Le profil d'expression des gènes *ENS* a été établi par Northern Blot et hybridation *in situ*. Nous avons ainsi détecté trois transcrits différents, exprimés fortement dans les cellules ES de poulet, puis de plus en plus faiblement au cours de la différenciation cellulaire ou du développement embryonnaire (**cf article Figure 5**). Les tailles estimées des trois transcrits pourraient correspondre aux trois gènes *ENS*. La première bande, à environ 2.5kb, représenterait le transcrit *cENS-2*, la deuxième, majoritaire, d'une taille avoisinant les 4.5kb, correspondrait au gène *cENS-1* et la plus grande, à environ 10kb, pourrait être un transcrit comprenant les trois séquences codantes pour *ENS-1*, *pol* et *env*. *In vivo*, dans l'embryon de poulet, les gènes *ENS* sont exprimés fortement dans l'épiblaste de la blastula de poulet (stade X H&H). Lors de la gastrulation, l'expression est restreinte aux cellules de la future plaque neurale (stade 3-4 EG), puis plus tardivement, à la frontière entre les cellules de l'ectoderme épidermal et neural (stade 5 EG). Au stade 12 (EG), l'expression des gènes *ENS* se restreint au sinus rhomboïdal, à quelques cellules dans l'endoderme et l'intestin antérieur (**cf article Figure 6**) puis devient indétectable ultérieurement.

#### d) Identification du promoteur des gènes *ENS*

La présence de séquences répétées aux extrémités des transcrits, l'absence d'introns et l'existence de séquences codantes pour *pol* et *env* dans le gène *cENS-3* suggèrent une origine rétrovirale des gènes *ENS*. Par analogie avec la structure des LTR rétrovirales, nous avons délimité des régions U3, R et U5 potentielles (**cf article Figure 8**). La région R était la plus facile à caractériser car elle correspond à la région répétée à chaque extrémité du transcrit. La région U5 commence là où s'achève R et se termine par la présence d'une répétition inversée (*att*) et du PBS (Primer Binding Site) permettant la fixation de l'ARNt nécessaire à la transcription inverse dans les génomes rétroviraux. L'extrémité 3' de la région U5 correspond également à la jonction entre l'ADN endogène et l'extrémité 3' de l'ADN proviral. N'ayant pas trouvé de PBS dans la séquence *ENS-1* à l'aide des logiciels d'analyse de séquence, l'extrémité 3' de la région U5 reste indéfinie. La région U3 se termine là où commence R et son extrémité 5' correspond normalement à la jonction entre l'ADN génomique endogène et l'ADN proviral. Elle est bordée également par la répétition inversée (*att*). Ne disposant pas de la séquence génomique complète des gènes *ENS*, cette limite est, elle aussi, encore inconnue, même si pour les besoins de l'étude nous l'avons fixée au début de l'homologie entre les séquences de l'extrémité 3' des gènes *ENS-1* et *ENS-3* (**Figure 30**). La région U3 des rétrovirus est connue pour avoir une activité promoteur qui peut être tissu-spécifique (Seifarth et al. 1998). Afin de tester si la région U3 que nous avons définie possédait, elle aussi, cette activité transcriptionnelle, nous l'avons placée, ainsi que la région R, en amont du gène rapporteur pour la luciférase. Ces constructions ont ensuite été transfectées dans différentes lignées cellulaires dont des cellules ES de poulet, des fibroblastes de caille ou des cellules HeLa. Nous avons alors observé que non seulement cette région U3 possédait une activité promoteur mais que celle-ci était particulièrement active dans les cellules ES de poulet (**cf article Figure 8**). D'autre part, l'activité de ce promoteur décroît lorsque les cellules ES de poulet sont induites à se différencier par l'acide rétinoïque. L'ensemble de ces résultats montre que nous avons identifié un promoteur fonctionnel dans la région U3 potentielle des gènes *ENS* et que l'activité de ce promoteur est particulièrement forte dans les cellules ES de poulet indifférenciées.

#### 4°) ARTICLE

# Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo

Hervé Acloque<sup>1</sup>, Valérie Risson<sup>1,2</sup>, Anne-Marie Birot, Riyota Kunita<sup>3</sup>,  
Bertrand Pain<sup>4</sup>, Jacques Samarut\*

Laboratoire de Biologie Moléculaire et Cellulaire de l'Ecole Normale Supérieure de Lyon, CNRS UMR5665, INRA LA 913, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Received 10 January 2001; received in revised form 15 February 2001; accepted 21 February 2001

## Abstract

Chicken embryonic stem (CES) cells are pluripotent cells derived from chicken early blastoderm. In order to identify new genes specifically expressed in these pluripotent cells, we have used a gene trap strategy and cloned a novel gene family called *cENS* for chicken Embryonic Normal Stem cell gene. The *cENS* genes expression decreases after induction of CES cells differentiation in culture and is restricted in vivo to the very early embryo. We have characterized three different *cENS* genes. One, *cENS-1*, is composed of an open reading frame inserted between two terminal direct repeats which are the common point of the *cENS* genes. *cENS-1* encodes a protein identical to cERNI, a recently described protein. *cENS-2* is a truncated form of *cENS-1*. *cENS-3* presents two adjacent open reading frames coding respectively for env and pol related proteins. The presence of conserved direct repeats, of retrovirus related genes and the absence of introns argue in favor of a retroviral origin of the *cENS* genes. In the *cENS* we identified a promoter region whose activity is strong in CES cells and decreases after induced differentiation showing a highly specific transcriptional activity specific of undifferentiated chicken embryonic stem cells. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

**Keywords:** Gene trap; Embryonic stem cells; Chicken; Retrovirus; Long terminal repeats; Promoter

## 1. Introduction

Embryonic stem (ES) cells were first derived from the inner cell mass of mouse blastocysts (Evans and Kaufman, 1981; Martin, 1981). These cells are pluripotent as shown by their ability to differentiate in all embryonic tissues including germ line upon grafting into a host blastocyst. ES-related cells so called embryonic germ (EG) cells have been derived from cultivated mouse primordial germ cells (Matsui et al., 1992). Human ES and EG cells have been recently isolated (Thomson et al., 1998; Shambloott et al., 1998). Pluripotent cells with ES-like activity as assessed in in vitro culture were also isolated from various animal

species such as rhesus monkey (Thomson et al., 1995), mink (Sukoyan et al., 1993), hamster (Doetschman et al., 1988), pig (Wheeler, 1994), chicken (Pain et al., 1996) and zebrafish (Sun et al., 1995). Among these, only the chicken ES cells (CES) have been demonstrated to be able to generate germ line transmission although with a low efficiency (Pain et al., 1996). These pluripotent cells are of most interest for two reasons. First they provide invaluable tools to develop a powerful genetics in vertebrates through manipulating ES genome by gene targeted mutagenesis. This application has been so far developed only in the mouse. Second these cells offer a unique model to investigate the molecular mechanisms underlying the control of self-renewing and differentiation commitment together with the molecular bases of pluripotentiality. These mechanisms have been yet investigated only in the mouse ES cells and are still poorly understood. The growth of undifferentiated mouse ES cells in culture is highly dependent upon stimulation by the leukemia inhibitory factor (LIF) (Smith and Hooper, 1987; Niwa et al., 1998; Matsuda et al., 1999) but LIF may not be the only required factor for sustaining pluripotentiality (Dani et al., 1998).

\* Corresponding author. Tel.: +33-4-7272-8536; fax: +33-4-7272-8080.

E-mail address: jacques.samarut@ens-lyon.fr (J. Samarut).

<sup>1</sup> These two authors contributed equally to this work.

<sup>2</sup> Present address: Laboratoire de Physiologie et Pharmacologie Clinique, CNRS-UMR 5014, Faculté de Pharmacie, 8 avenue Rockefeller, 69373 Lyon Cedex 08, France.

<sup>3</sup> Present address: Department of Applied Biological Chemistry, Faculty of Agriculture, Tohoku University, Sendai, Japan.

<sup>4</sup> Present address: Vivalis Company, CHU-Hôtel-Dieu, Place A. Ricordeau, HNB Aile Nord, 44093 Nantes cedex 1, France.

Mouse ES cells show a special cell replication control with a short cell cycle (6–7 h) and a very reduced G1 phase (1 h) (Savatier et al., 1994). Although they express very low levels of D-type cyclins, the hypophosphorylated RB protein remains undetectable during the M-G1-S transition indicating that the RB protein is rapidly re-phosphorylated in G1 (Savatier et al., 1996). Moreover the mouse ES cells do not express the p21 cell cycle inhibitor (Savatier et al., 1996) and do not seem to have a p53-mediated cell cycle check point control (Aladjem et al., 1998). Upon differentiation the cells recover features of somatic cells, i.e. longer cell cycle, cycling phosphorylation and dephosphorylation of the RB protein along the cell cycle, expression of p21 and p53 controlled arrest in response to DNA damage.

The regulation of pluripotency of mouse ES cells seems to be tightly controlled by the POU transcription factor Oct3/4. Indeed this transcription factor is essential for the initial formation of pluripotent founder cell population in the mouse embryo (Nichols et al., 1998). The level of expression of Oct3/4 is also critical for the maintenance of the mouse stem cells pluripotency and self-renewal (Niwa et al., 2000).

Chicken ES cells (CES) have been isolated from stage X chicken embryos and could be grown for long-term in vitro culture. These cells are able to differentiate in vitro in many cell lineages. Upon grafting into recipient early chicken embryos these cells can participate in the development of the host tissues and colonize the germ lineage (Pain et al., 1996). These cells likely represent cells of the early blastoderm epiblast, the equivalent of the primitive ectoderm in the mouse inner cell mass. The CES are of great potential interest for the production of transgenic chickens for further agronomic developments. They also provide helpful tools for experimental avian embryology. In order to better understand the molecular mechanisms which govern pluripotency in these cells, and by the way in the early epiblast, we searched for genes specifically expressed in these pluripotent cells. Using a gene trap strategy we cloned a gene family which is specifically expressed in the undifferentiated CES and we were able to identify a promoter of these genes which is active specifically in undifferentiated chicken ES cells.

## 2. Results

### 2.1. Isolation of the 9N2-5 CES clone

The aim of the procedure was to insert the pROSA  $\beta$ -geo gene trap vector downstream of a cellular gene promoter that is specifically expressed in undifferentiated CES cells. This vector contains two inverted LTR (from MMLV) with mutated enhancer activity, flanking a splice acceptor site upstream of a LacZ-Neomycin resistance fusion gene followed by a polyadenylation signal (Friedrich and Soriano, 1991). Normal undifferentiated CES cells were electroporated with the linearized plasmid and then selected

with G418. The number of drug resistant clones observed 7–10 days later was about 1 for  $10^6$  electroporated cells. Nine clones containing cells with the typical morphology of undifferentiated chicken ES cells (Pain et al., 1996) were isolated and further expanded. All these clones were positive for the  $\beta$ -galactosidase activity (not shown).

To further test for the LacZ expression upon differentiation a quick screening was set up. Samples of each clone were seeded in culture without feeder cells and without LIF and FGF, and in the presence of retinoic acid. One clone named 9N2.5 was retained as it quickly became negative for the LacZ activity upon differentiation. The morphology of the undifferentiated 9N2.5 cells was typical of that of undifferentiated CES cells (Pain et al., 1996). These cells expressed epitopes specific of chicken ES cells (not shown), and were positive for the telomerase activity (not shown). When the 9N2.5 cells were induced to make embryoid bodies, they gave rise to a wide spectrum of cell types including muscle, nerve, blood, fibroblasts (not shown) similarly to normal parental CES cells (Pain et al., 1996).

In a first step toward identifying the integration site of the ROSA $\beta$ -geo insert, a Southern blot analysis was performed on the 9N2.5 genomic DNA. Using EcoRI, which cuts only once in the ROSA $\beta$ -geo vector, a LacZ probe revealed a single integration locus for the transgene (not shown). To further assert that the pROSA $\beta$ -geo vector has inserted into a gene specifically expressed in undifferentiated CES cells, we analyzed the expression of the LacZ transcripts in undifferentiated 9N2.5 cells and in these cells after differentiation induced by retinoic acid. As shown in Fig. 1 from Northern blot analysis the amount of LacZ RNA strongly decreased as soon as 1 day after RA addition. One single LacZ transcript of 4.7 kb was detected in the 9N2.5 cells. In the culture conditions used in Fig. 1, overcrowding of the culture and spontaneous differentiation of CES cells occurred which explains the slow decrease of the lacZ transcript in cultures without RA.

### 2.2. Analysis of expression of the ROSA $\beta$ -geo in 9N2.5 cells in vivo

To further investigate the expression of the ROSA $\beta$ -geo vector in cells derived from the 9N2.5 cells upon normal

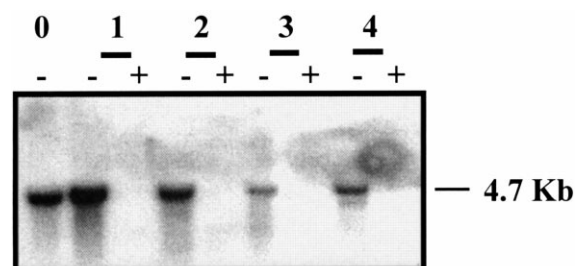


Fig. 1. Lac Z mRNA expression in 9N2-5 cells. Northern blot was performed with total RNAs of 9N2-5 cells treated (+ lanes) or not (- lanes) with retinoic acid for 1–4 days (lanes 1–4). Filters were hybridized with a Lac Z probe.

differentiation *in vivo*, the 9N2.5 cells were grafted into slightly irradiated stage X (EG) recipient chicken embryos. As shown previously, in these conditions the CES cells participate in host tissue formation as illustrated by the development of chimerical plumage (Pain et al., 1996).

At various times after grafting, the recipient embryos were collected and stained *in toto* to reveal LacZ expression (Fig. 2). Clusters of blue cells were detected in the epiblast of the area pellucida in stage XIII (EG) embryo, i.e. 16–18 h after cells injection in recipient embryo (Fig. 2A). LacZ positive cells were found only in the epiblast layer and not in the newly formed hypoblast (direct observation with dissecting microscope). This observation suggests either that the grafting cells do not participate to hypoblast formation or that, when contributing to hypoblast they loose expression of the LacZ transgene. At later stage, during gastrulation (stage 5 H&H), LacZ expression was restricted within the primitive streak and the extraembryonic area corresponding to the prospective neural plate (Streit and Stern, 1999) (Fig. 2B). In the primitive streak, the  $\beta$ -galactosidase activity was localized in dots, mainly in the Hensen's node and the regressing streak. At stage 13 (H&H), the LacZ positive spots were detected in the caudal part of the embryo comprising the open neural plate (Fig. 2C).

At 5 days of embryonic development, the distribution of the  $\beta$ -galactosidase activity was followed on serial semi thin paraffin sections from several embryos. Sparse small clusters of LacZ positive cells were seen in the ependymal layer of the neural tube, in branchial arches, notochord, epithelium of ears, wall of the heart, and in the kidney tubules. Isolated rare LacZ positive cells were also detected within tissues of the digestive system and in the genital buds.

It is clear that the number of LacZ positive cells strongly and quickly decreased as the recipient embryos developed. From two chimeric embryos that hatched we could detect the ROSA $\beta$ -geo transgene in genomic DNA isolated from the skin, the gizzard and the liver whereas no LacZ activity could be detected in these tissues (not shown). These observations strongly suggest that descendants of the grafted 9N2.5 CES cells persist in the chimeric chicks although they no longer express the transgene. All these data confirm that the ROSA $\beta$ -geo transgene inserted into the 9N2.5 is transcribed mostly in undifferentiated CES cells.

### 2.3. Identification of the ROSA $\beta$ -geo integration locus

A 5' RACE method was applied on 9N2-5 cells total mRNAs to identify the 5' sequence upstream to LacZ.

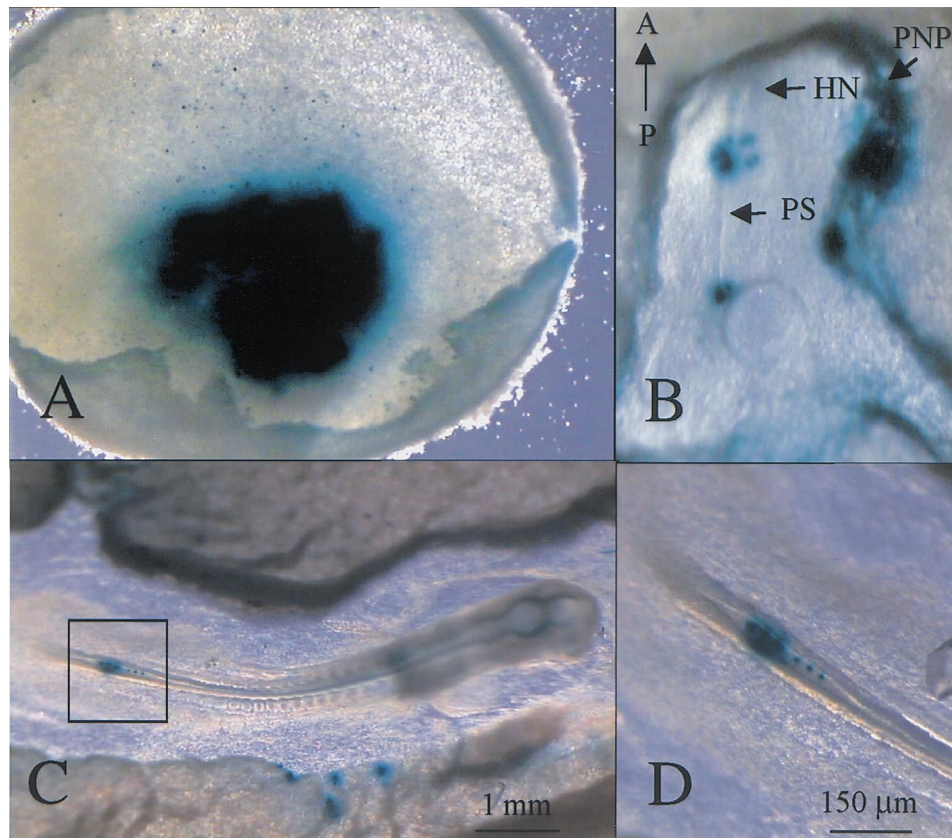


Fig. 2. LacZ activity in chimaeras. 9N2-5 cells were injected into stage X (EG) recipient embryos.  $\beta$ -galactosidase activity was detected in pellucida area of stage XIII embryo (E&G) (A), the primitive streak (PS) and prospective neural plaque (PNP) of stage 5 (H&H) embryo (B) and in the caudal part of stage 13 (H&H) embryo (C). An enlargement of the caudal part of stage 13 (H&H) embryo is shown in panel (D). The embryo shown in panel B was distorted by the chemical treatment and the antero-posterior (P  $\rightarrow$  A) axis is shown on the left.



Sequence analysis of the RACE product revealed 172 bases of a novel sequence we called C1, upstream of the vector splice acceptor site, suggesting a correct use of this site for

maturation of the transcript. To confirm that this novel C1 sequence corresponds to the trapped gene, RT-PCR was performed using RNA from 9N2.5 cells. Reverse transcription was performed by using oligo(dT)15. The primers used for PCR amplification were C1 P4(s) derived from 5'-RACE fragment and antisense LacZ B or LacZ C derived from the gene trap vector. Sizes of the PCR products were, respectively, 331 and 159 pb as expected, confirming that the C1 fragment was transcribed from the host trapped gene (not shown).

The C1 sequence was used as a probe for screening a cDNA library prepared from proliferating non differentiated CES cells. A polyadenylated cDNA of 4.2 kbp was isolated (clone 4a) and called *cENS-1* (for Embryonic Normal Stem cell).

2.4. Sequence of the *cENS-1* mRNA

The sequence of the *cENS-1* mRNA is shown in Fig. 3 (Genbank Accession number: AF327879). It contains an open reading frame, hence referred to as *cENS-1*-ORF encoding a 490 amino acids product of 54 kDa predicted molecular weight. The size of the product was confirmed by in vitro translation of the full-length cDNA (data not shown). Analysis of this protein with the PSORT II computer program revealed the presence of a coiled-coil domain between the residues 80 and 137. This protein is identical to the chick ERNI protein recently described as a marker for neural induction before gastrulation (Streit et al., 2000).

Interestingly the *cENS-1* mRNA contains a direct repeated sequence of 486 bp including the C1 sequence, hence referred to as the C1-repeat, located both at its 5' end and its 3' end immediately upstream of the poly(A)<sup>+</sup> tail.

2.5. Conservation of the *cENS-1* ORF among Gallinaceae

To investigate the level of conservation of the *cENS-1*-ORF sequence among vertebrates, genomic DNAs from various animal species (including chick, quail, duck goose, pigeon, ascidia, drosophila, xenopus, smolt, mouse, rat, monkey and human species) were analyzed on Southern blots with S1, a probe within the censequence. Although we used low stringency conditions a signal was found only with DNAs from birds of the gallinaceous species (not shown). We also compared various chicken strains and observed the same restriction pattern on Southern blot

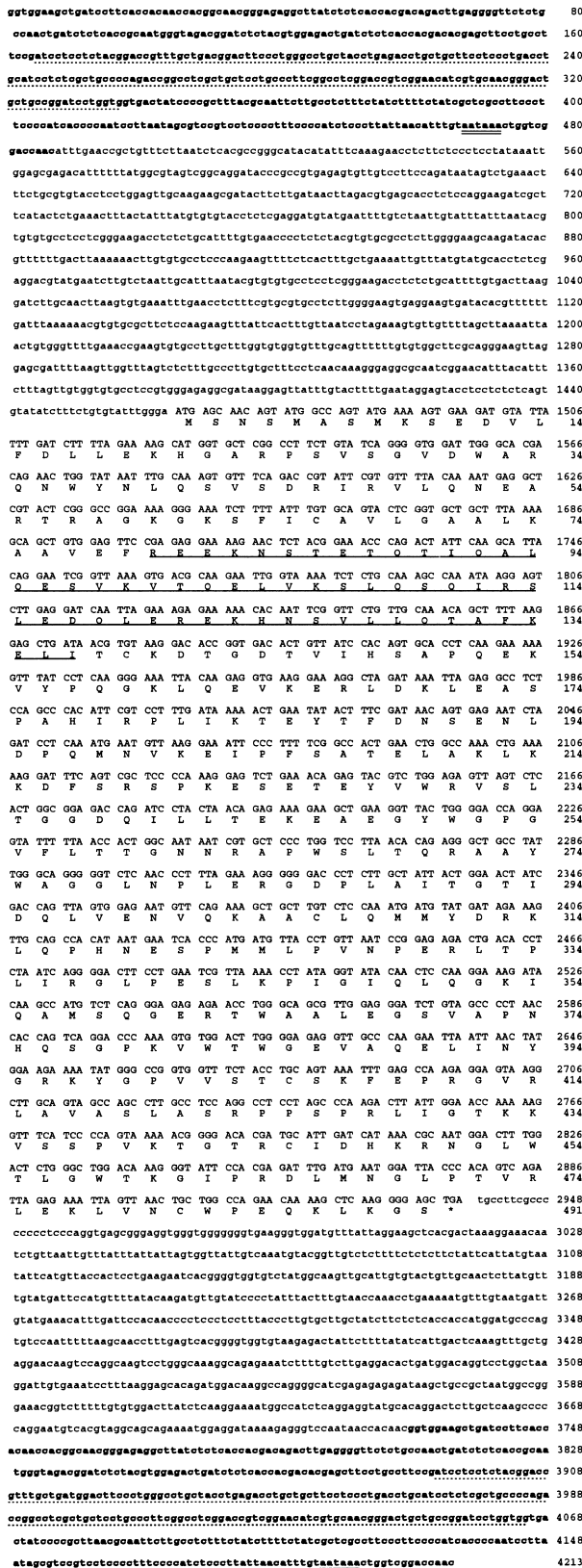


Fig. 3. Nucleotide sequence of the *cENS-1* transcript and corresponding amino acid composition. This sequence corresponds to the nucleotide sequence of the clone 4a (nt-56 to nt-4214) and the additional 5' sequence (nt-1–55) obtained by the 5' RACE method on RNA isolated from CES cells transfected with the promoter 2 reporter. Within the 486 bp direct repeat regions (bold type) corresponding to the R region, the C1 region is assigned with dotted lines and the polyadenylation signal is double underlined. The open reading frame codes for a 490 amino acid protein which contains a coiled-coil region (underlined).

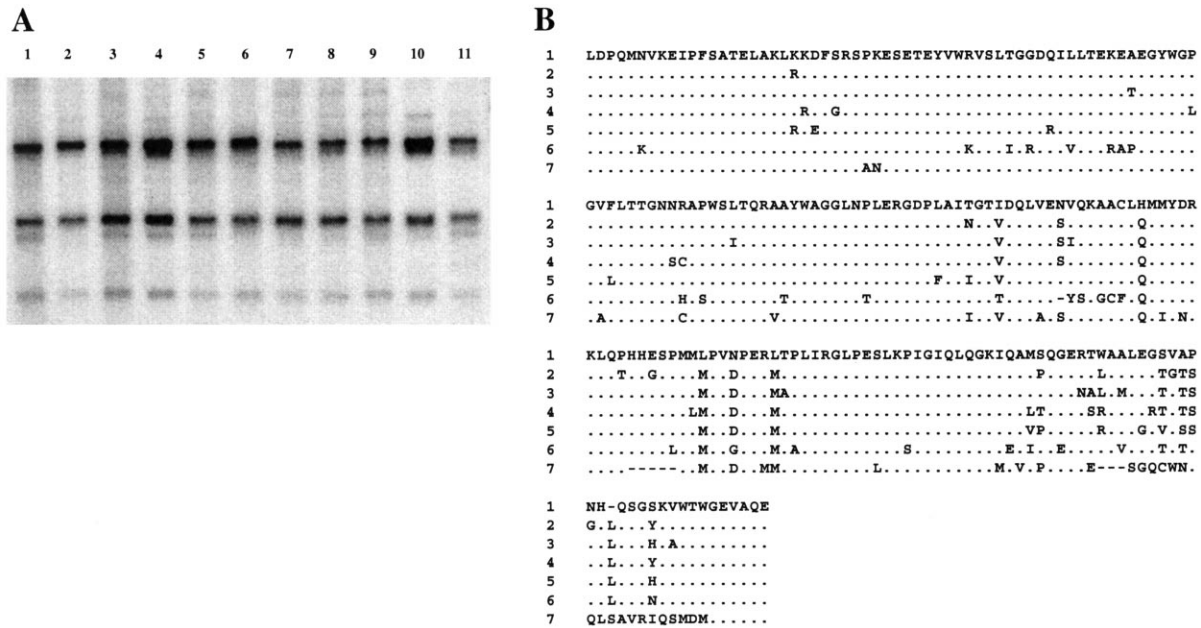


Fig. 4. *ENS-1* is conserved in various chicken lines and in Gallinaceae. (A) Southern blot on BamHI digested genomic DNAs from various chicken lines hybridized with the S1 probe. 1: Fayoumi (Egypt); 2: Geline of Touraine; 3: Picage; 4: Black sheaver; 5: Isabrown; 6: DPF; 7: broiler chick cross Y; 8: broiler chick cross X; 9: golden leghorn; 10: white leghorn; 11: red naked neck. (B) Alignment of amino-acid sequences of partial *ENS-1*-ORF from: 1: chick (*Gallus gallus*), 2: quail (*Coturnix coturnix*), 3: red legged partridge (*Alectoris rufa*), 4: guinea fowl (*Numida meleagris*), 5: gray partridge (*Perdix perdix*), 6: turkey (*Meleagris gallopavo*) and 7: pheasant (*Phasianus colchicus*).

(Fig. 4A). Further attempts to amplify an *cENS-1*-like sequence by PCR was successful only in the Gallinaceae that we tested including quail, pheasant, guinea fowl, partridge and turkey. Sequencing of a 594 bp fragment obtained by PCR with conserved primers revealed in all these species a highly conserved sequence (Fig. 4B).

## 2.6. Expression pattern of the *cENS-1* gene

To verify that the cloned gene is expressed through a cell-specific and developmental-specific pattern as suggested by the LacZ reporter transgene in the 9N2.5 cells, northern blot, RT-PCR and in situ hybridization were performed with *cENS-1* probes.

Parental CES cells and 9N2.5 CES cells RNAs were analyzed with the C1 and S1 probes, respectively (Fig. 5A). The C1 probe revealed three specific transcripts of, respectively, about 10, 4.2 and 2.5 kb (Fig. 5B). The 4.2 kb transcript was also detected in the 48 h chicken embryo but at a much lower level and its level decreased strongly in older embryos and became undetectable between the 4th and 7th day of embryonic development. The S1 probe revealed only the 10 kb and the 4.2 kb species (Fig. 5B). In CES cells induced into differentiation, the expression of the *cENS-1* gene was strongly reduced (Fig. 5C). None of these transcripts was detected in any of tested chicken tissues including brain, heart, bone, liver, gizzard, muscle, skin, kidney and eye (not shown).

These observations strongly suggest that the expression

of *cENS-1* is restricted to CES cells and the earliest stages of embryonic development.

The spatial pattern of the *cENS-1* gene expression was examined during embryogenesis by whole-mount in situ hybridization. Furthermore, to study the expression pattern in more detail, serial 10  $\mu$ m paraffin sections were made from whole-mount stained embryos at different stages.

By using the C1 probe, we observed a very restricted expression pattern during chick embryogenesis. *cENS-1* expression was very high at stages X (E–G&K) and mostly in the area pellucida (Fig. 6A). This expression was restricted to the epiblast and was undetectable in the forming hypoblast (Fig. 6B). In stage XIII (E–G&K) embryo, *cENS-1* expression decrease (Fig. 6C). In the stage 2 (H&H) embryo the expression of *cENS-1* increased and was localized mostly around the Koller's sickle and the nascent primitive streak (Fig. 6E). At the gastrulation stage (stage 4H&H), the expression was restricted to the primitive streak (black arrows) and to the prospective neural plate (white arrow) (Fig. 6F). At stage 5, *cENS-1* expression was localized to the border of the prospective neural plate (white arrow on Fig. 6G) and spots are still detected in the regressing primitive streak (black arrows on Fig. 6G). Sections made in the anterior part of the embryo (dotted line in Fig. 6G) show a strong signal in cluster of cells located in the primitive streak (Fig. 6J). At stage 7 (H&H), a signal was observed in the neural plate of future forebrain and in the regressing primitive streak (black arrows) (Fig. 6H). At later stage (stage 12 H&H) a faint signal was detected in the sinus rhomboidalis (enlarged in the right down corner of Fig. 6I)

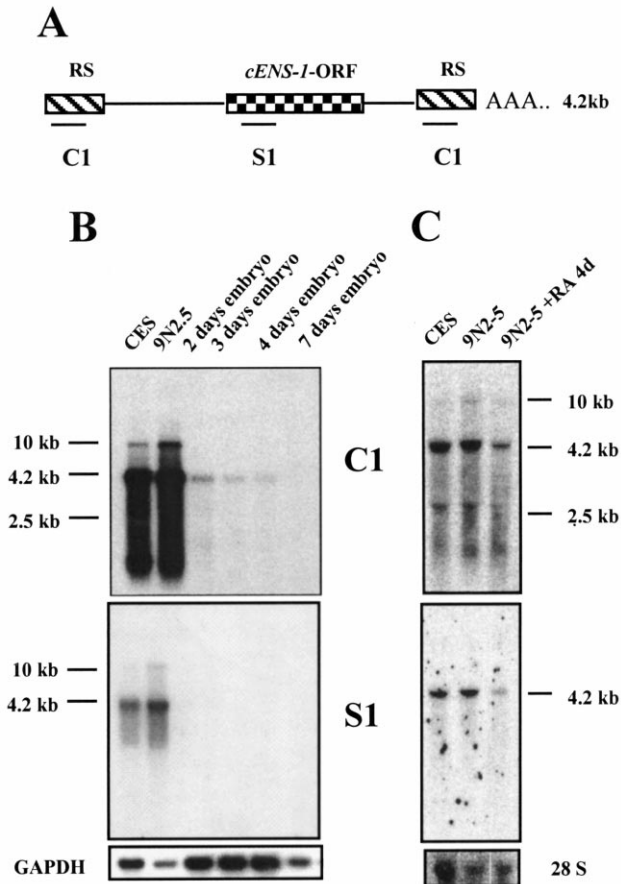


Fig. 5. Expression of *cENS-1* in CES cells and chicken embryos. *cENS-1* transcripts were detected by Northern blot with probes localized in the repeated sequence (RS) of the transcript (probe C1) or in the open reading frame (probe S1) (A). In panel (B) poly(A)<sup>+</sup> RNAs were extracted from CES cells, 9N2-5 cells, 2, 3, 4 and 7-days embryos, and Northern blots were hybridized with C1, S1 or GAPDH probe. In panel (C) total RNAs were extracted from CES cells, 9N2-5 cells and 9N2-5 cells treated for 4 days with retinoic acid (RA). Northern blot were hybridized with the C1 or S1 probes.

and in the closing neural tube (Fig. 6I). Sections made in the anterior part of this embryo (dotted line on Fig. 6I) show a strong signal in the endoderm and in the fore-gut (Fig. 6K). The same patterns at all stages were observed with the S1 probe (not shown).

### 2.7. Structure of the *cENS-1* locus and related loci

A PCR amplification both on genomic DNA and CES cDNA with primers located on each sides of the *cENS-1-ORF* gave fragments with identical sizes. Sequencing of these fragments showed identity between genomic DNA and cDNA (not shown). These data demonstrate that the *cENS-1* locus has no intronic sequences within the *cENS-1-ORF* domain. We have sequenced several *cENS-1*-related cDNA clones isolated with the C1 and S1 probes and observed a significant polymorphism suggesting the occurrence of multiple *cENS-1*-loci in the chicken genome. To

test this hypothesis we performed a southern blot analysis using the restriction enzyme Bgl I, which cuts only once in the *cENS-1-ORF* and EcoRI which does not cut in the *cENS-1-ORF*, and probed with the S2 sequence located downstream of the Bgl I restriction site in the *cENS-1-ORF*. We could detect as many as 11 bands suggesting that the chicken genome contains at least 11 different *cENS-1*-related loci. Highly labeled bands are suggestive of tandem loci (not shown). The C1 probe also identified numerous C1 sequence copies in the chicken genome (not shown).

Others cDNA clones obtained by screening the CES cDNA library and containing a C1 sequence were isolated and analyzed. One of these cDNAs, *cENS-2* (Genbank Accession number: AF327880), was a short version of *cENS-1* containing a truncated form of the *cENS-1-ORF* with a different 3' coding half (Fig. 7A). The occurrence of a *cENS-2* locus devoid of introns was confirmed by PCR on chicken genomic DNA. An other related locus, *cENS-3* (GenBank Accession number: AF329451), with proviral structure was also cloned (see below).

### 2.8. Identification of a *cENS-1*-related proviral structure

As shown in Fig. 7A one clone, *cENS-3*, isolated with the C1 probe contained two adjacent ORFs which, after search in BLAST protein data bases, revealed to encode proteins highly homologous to respectively a mouse ERV-L pro-pol-d-UTPase polyprotein (Benit et al., 1997) and a human ERV3 envelope (Cohen et al., 1985). To check whether an *cENS-1-ORF* like sequence, the pro-pol-d-UTPase, the envelope ORFs and the C1 sequences were collinear in the genomic DNA and were transcribed under a unique transcript, we performed PCRs on chicken genomic DNA and RT-PCRs on CES RNAs using primers specific of each sequence. As shown in Fig. 7B, all the amplifications gave positive results demonstrating the occurrence of proviral-like structures containing both the *cENS-1-ORF* and the proviral sequences.

### 2.9. Characterization of a CES specific promoter

The identification of C1-repeats at both ends of the transcripts containing an *cENS-1-ORF* or an *cENS-1*-related ORF together with the identification of some provirus-like cENS loci strongly suggested that the C1-repeats are part of retrovirus LTR-like structures. In proviruses LTRs are made of three parts, respectively U3, R and U5 (Teich et al., 1984). R and U5 are present in this order at the 5' end of the proviral transcript, whereas U3 and R in this order are present at the 3' end. The U3 region is able to promote transcription, sometimes with a tissue specific control (Seifarth et al., 1998). To test the promoter activity of the *cENS-LTR*-like structure we have cloned parts of this structure upstream of the firefly luciferase reporter gene (Fig. 8A). By analogy with retrovirus LTRs we tentatively delineated the R and U3-R regions in the *cENS-1* transcript. The R region was identified as the repeated sequence at both

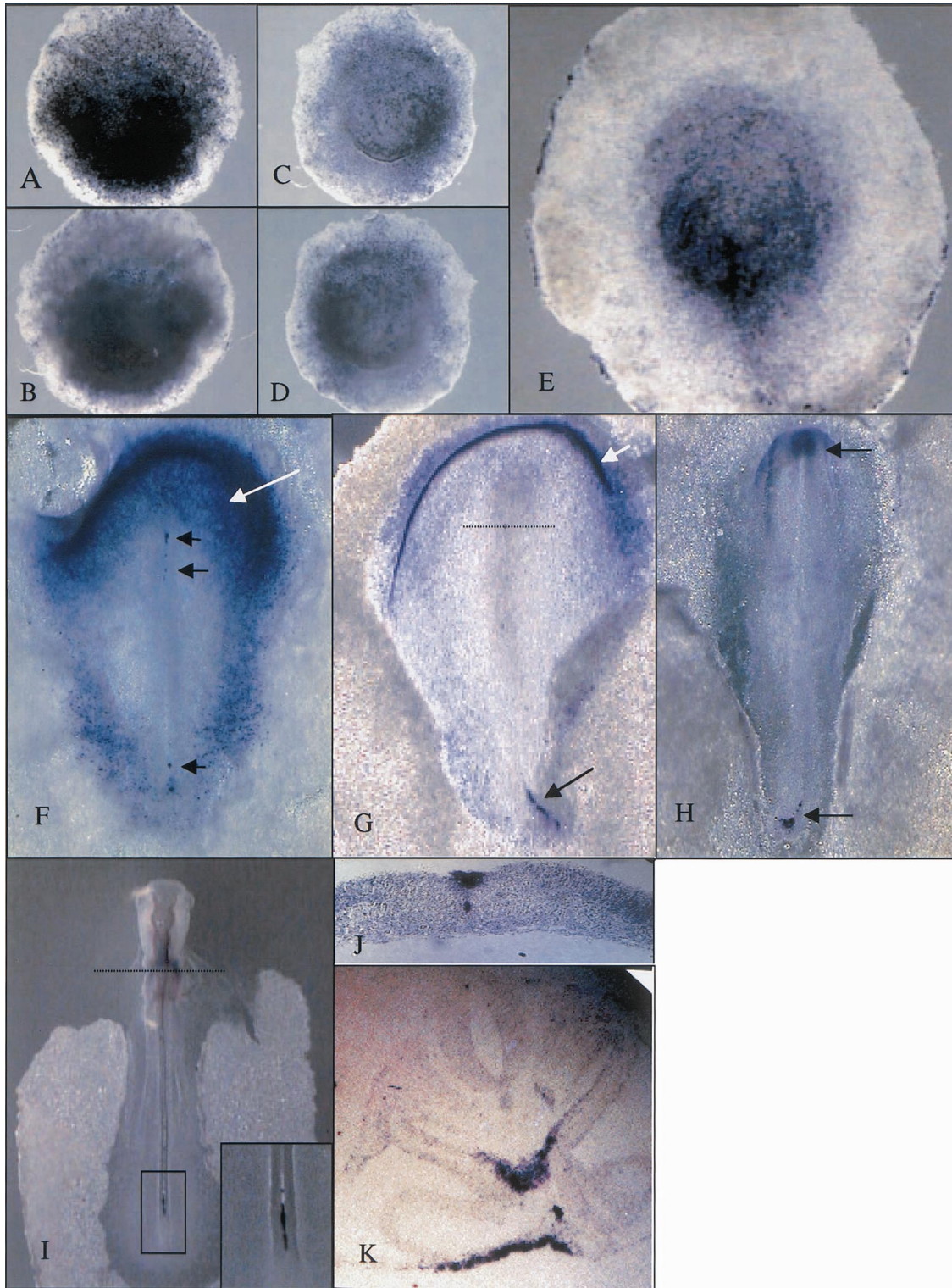


Fig. 6. Expression of *cENS-1* in chicken embryo. Whole-mount in situ hybridization of *cENS-1* mRNA. Embryos were hybridized to a RNA antisense corresponding to the ends of *cENS-1* (C1 probe). (A,B) Stage X (H&H) embryo dorsal side (A) and ventral side (B). (C,D) Stage XIII (H&H) embryo dorsal (C) and ventral side (D). (E) Stage 2. (F,G) Stage 4 and 5 embryos. The white arrow indicates the prospective neural plate (F) and the border of the prospective neural plate (G). The black arrows show signals located in the primitive streak. The dotted line in (G) indicate a 10  $\mu$ m transversal section after paraffin embedding of the same embryo shown in (J). (H) Stage 7 embryo. The black arrow in the anterior part of the embryo indicate the neural plate of the future forebrain and those at the posterior part shows signal in the regressing primitive streak. (I) Stage 12 embryo. The square in the right down corner is an enlargement of the framed region. (K) Ten microgram transversal section after paraffin embedding of stage 12 embryo mid trunk region.

ends of the transcript (bold sequence in Fig. 3). The U3 region was hypothetically located between the 3' end of the *cENS-1*-ORF and the 5' end of R (see Fig. 3). These R and U3-R regions were cloned upstream to the luciferase reporter gene in the two possible orientations corresponding respectively to the promoter 1 and promoter 2 [sense (s) and antisense (as)] vectors (Fig. 8A). These reporter vectors were tested by transfection into CES, QT6, a quail fibroblast cell line (Moscovici et al., 1977), QBr, a quail brain derived cell line (Guilhot et al., 1993) and HeLa cells. A reporter plasmid containing the CMV promoter and a promoter-less plasmid were used as positive and negative controls respectively. Only the promoter-2 construct containing the U3-R structure in the sense orientation showed a significant activ-

ity similar to that of the CMV promoter in the CES only (Fig. 8B). We further tested whether the promoter activity of the U3-R region was specific of the undifferentiated CES cells. The assay was performed in parallel in normal CES cultures and in cultures treated for various time with retinoic acid as a differentiation inducer. As shown in Fig. 8C the promoter activity of U3-R strongly and quickly decreased in differentiating CES cells. In contrast the CMV promoter construct showed a strong activity in differentiating CES cells. To confirm that the decrease activity of the promoter was not due to the retinoic acid treatment, we have induced differentiation of CES cells by transferring into medium lacking growth factors. The promoter activity in these cells was far weaker compared with CES cells whereas

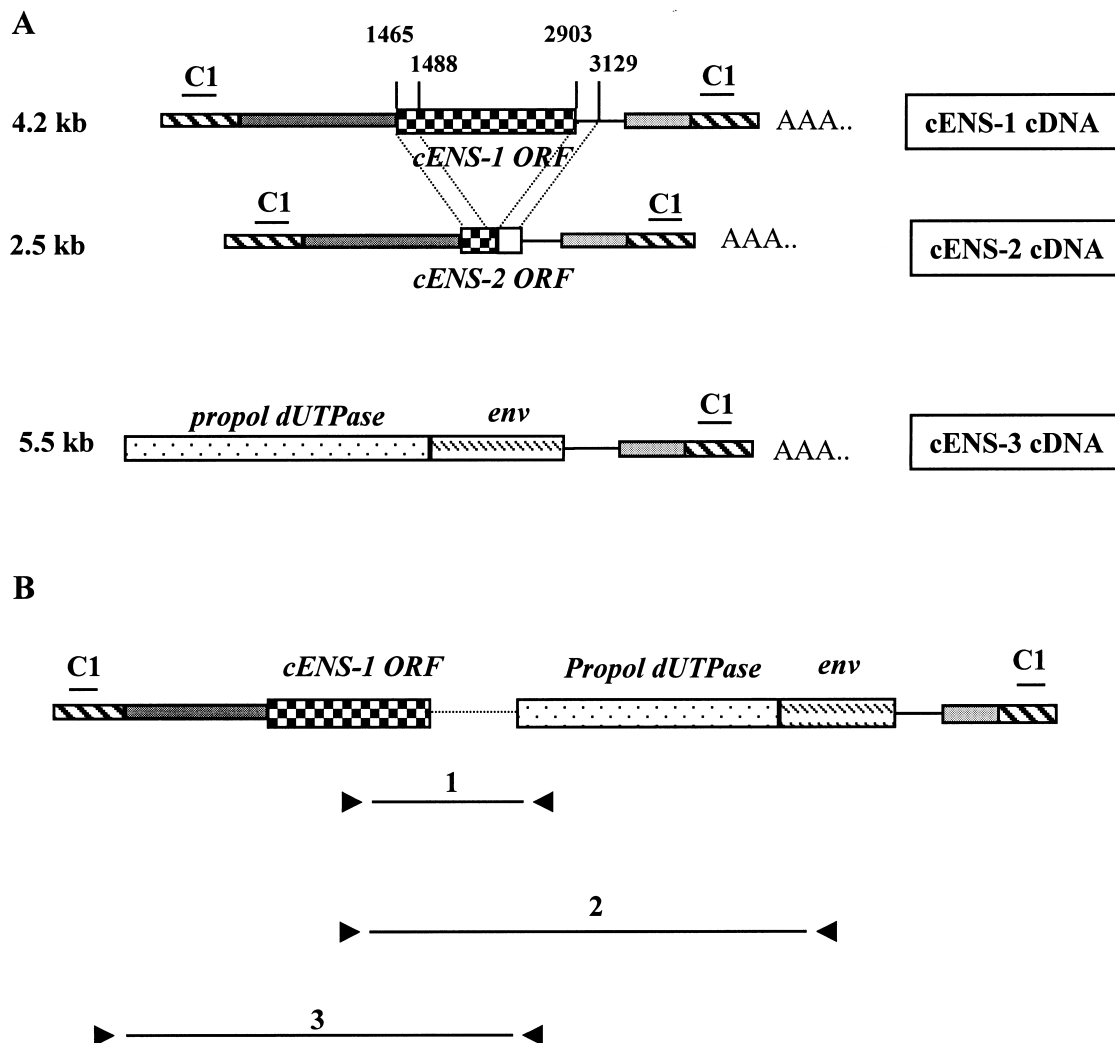


Fig. 7. Structure of some *cENS-1*-related loci. (A) Schematic representation of three cDNA clones containing the C1 sequence. The *cENS2*-ORF is a truncated version of the *cENS1*-ORF: the eight first amino acids are conserved but a 1414 pb deletion generates a frame shift and thereby a smaller protein of 83 amino acids. The *cENS-3* locus shows two open reading frames encoding respectively a propol dUTPase- and an env-related peptides. (B) PCR experiments showing the colinearity between *cENS-1* ORF and *cENS-3* ORFs. Positive amplifications (+) were obtained with the shown primers (black arrow heads) from either CES cDNA or genomic DNA. 1: Amplification positive on genomic DNA and cDNA using cENS1 S1 (5'-CTCACAGATGCGGGCGG-3') and Pol AS1 (5'-GGTAAGGTTTAACACAGGTCA-3') primers. 2: Amplification positive on genomic DNA using cENS1 S1 and env AS1 (5'-TGCCATCCTTTCCTGGGA-3') primers. 3: Amplification positive on genomic DNA using C1S1 (5'-TTCCGATCCTCTCTACG-3') and pol AS1 primers.

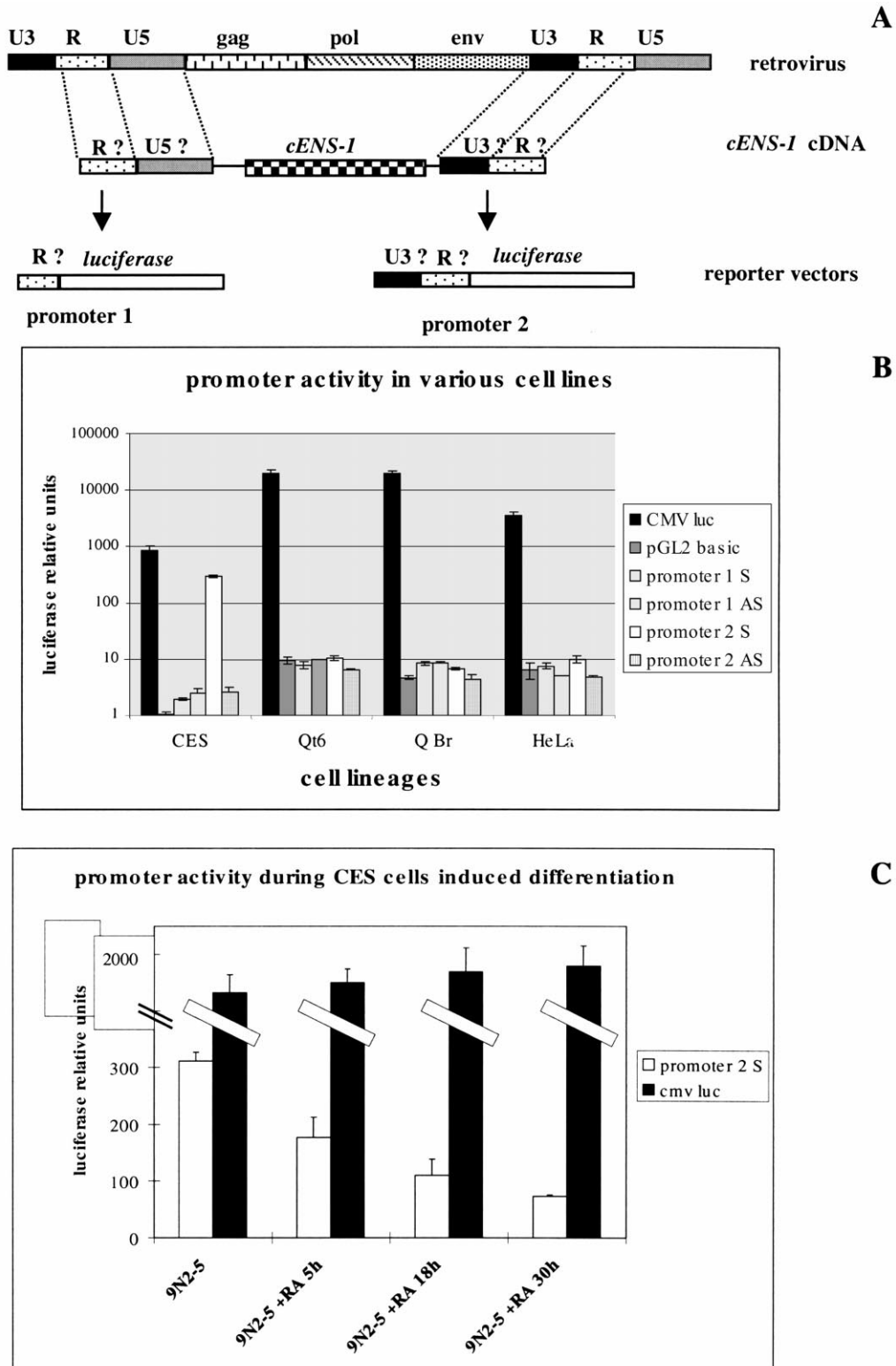


Fig. 8. Test of a CES cells specific promoter. (A) The upper scheme depicts the general structure of a provirus. The putative R and U3-R regions of the *cENS-1* locus were inserted upstream of the Firefly luciferase reporter gene in the sense (S) or antisense (AS) orientation to make respectively the promoter-1 and promoter-2 vectors. (B) Each vector was cotransfected in various cell lines including CES cells with an internal control (pRL CMV) and luciferase activity was measured 2 days later. (C) The transcriptional activity of the promoter-2 sense (S) was measured in 9N2.5 cells after various times following treatment with retinoic acid. The activity of a reporter driven by the CMV promoter is shown in parallel.

the activity of the CMV promoter was maintained confirming that the *cENS-1* promoter activity was specific to undifferentiated CES cells.

In order to further characterize this promoter we have identified the putative transcription initiation site (corresponding to the R region starting point) using a 5'RACE technique on RNA isolated from CES cells transfected with the promoter 2 reporter (not shown). We could then localize the U3-R junction at position 3726–3727 (Fig. 3). The 3' end of R being precisely defined by the presence of a poly(A)<sup>+</sup> tail (Fig. 3), the deduced size of the R region is 486 nt. A TATA-like sequence (5'-GATAAAA-3') (Huang et al., 2000) was found to be located 27 nt upstream of the beginning of R (position 3701 on Fig. 3).

Therefore we identified a functional promoter and a putative U3-R structure in the *cENS-1* locus. Furthermore we showed that the transcriptional activity of this promoter is highly specific of the undifferentiated chicken ES cells.

### 3. Discussion

The aim of this work was to identify genes specifically expressed in chicken embryonic stem cells and thereby in early blastoderm epiblast. We used a gene trap strategy as developed by Friedrich and Soriano (1991). This strategy appeared quite reliable as the pattern of expression of the inserted reporter gene and that of the identified host cellular *cENS-1* gene appeared identical. Moreover as expected from the relationship between CES and early blastoderm epiblast, the identified genes show an expression restricted to the earliest steps of embryonic development and to the multipotent epiblast cells.

The loci identified, named *cENS*, belong to a numerous family counting at least 11 members in the chicken genome. Several observations strongly suggest that some *cENS* loci were introduced into the genome of Gallinacea via retrovirus infection. First they contain direct repeats at each extremity of the transcribed region. Second their coding sequences are devoid of intron. Third some of these loci like the locus *cENS-3*, contain coding sequences highly similar with those of env and pol retrovirus genes. It is not yet known whether the pol- and env-like sequences of *cENS-3* would be efficient in making infectious particles in avian species. These loci might be similar to some endogenous viruses identified in the chicken genome (Dunwiddie et al., 1986). This retrovirus structure might represent remnants of an ancestral retrovirus which infected first members of the gallinaceous family. In agreement with phylogenetic studies of galliform birds based on morphological traits (Johnsgard, 1986) or molecular markers (Kimball et al., 1999) the detection of *ENS-1* in turkey and guinea fowl suggest that the *ENS-1*-ORF is conserved among Gallinacea. The lack of *ENS-1* gene in the anseriform birds tested (pekin and barbarie duck and goose) suggests that the gene

appeared after the radiation of anseriform and galliform birds.

Several of the *cENS* loci contain a coding region for a protein which was recently identified as ERNI, a protein encoded by early response genes induced by signals from the Hensen's node in the chicken embryo (Streit et al., 2000). From the description of the *ERNI* gene, its presence in multiple copies in the chicken genome and its pattern of expression (Streit et al., 2000), it is likely that the *ERNI* gene and the *cENS* loci refer to same genome structures. The *cENS-2* locus does not encode a full length ERNI and the locus *cENS-3* encodes additional products to ERNI. Therefore we propose to keep the terminology of *cENS* for identifying the loci and ERNI to identify the product isolated by Streit et al. The *cENS-1* locus is then likely to be identical to the *cERNI* gene identified by Streit et al. Until all *cENS* loci have been cloned we cannot anticipate they all encode the ERNI product. Streit et al. (2000) proposed that ERNI is an early marker of neural induction for two reasons. First its expression is restricted to the prospective neural plate after gastrulation. Second its expression is induced at the site of ectopic grafting of a Hensen's node in the chicken extra-embryonic epiblast. Our data clearly show that *cERNI/cENS-1* gene has a biphasic kinetics of expression with a high expression level in the epiblast prior to gastrulation (consistent with similar expression in CES cells) and a second peak of expression during neural plate induction. In these conditions the *cERNI/cENS-1* gene might not be considered strictly as marker of neural induction. Another question is on the putative functions and roles of the ERNI protein. In our hands, we were unable to detect the *ERNI* sequence outside of the Gallinacea whereas Streit et al. (2000) could find some conserved sequence in the zebrafish. We might imagine that only very limited sequences or structural motives of the protein are conserved among vertebrates. These motives could be the hallmark of specific biochemical functions. Its high sequence conservation among gallinaceous species whereas it might have been transposed in the genome through highly mutagenic processes like retroviral infection or retrotransposition is a strong argument in favor of an important function during embryogenesis. One might imagine that the function of the ERNI protein relies mostly on its specific coiled-coil structure which might have been conserved through evolution. One approach to the function of the ERNI products will consist in impairing its expression and/or function in CES cells.

In this work we have identified a promoter region for the *cENS* loci. As expected from their provirus-like structure, the transcription promoter was found in the regions that are highly suggestive of retrovirus LTRs. Interestingly the activity of this promoter appears highly specific of undifferentiated CES. The *cENS* promoter provides then a helpful model to identify transcription factors which might specifically control the multipotency and self-renewal of stem cells and cells within the early epiblast. So far very few transcription factors specific of ES cells have been identified. Rex-1

(Hosler et al., 1989) coding for a zinc finger protein and Oct-3/4 (Okazawa et al., 1991) have been described to have a decreasing expression after induction of ES cell differentiation. The POU factor Oct-3/4 is known to play determinant roles in the maintenance of the stem cell phenotype (Niwa et al., 2000) and POU factors may be involved in the control of the *cENS-1* gene expression since some response elements of these gene family are present in the sequence of the *cENS* promoter. However other factors seem to be involved in the maintenance of the multipotent phenotype as shown by the recent discovery using a gene trap strategy of the *taube nuss* gene which is implicated in the control of apoptosis in mouse ES cells and early embryo (Voss et al., 2000).

In conclusion, the gene trap strategy we used in chicken embryonic stem cells was quite fruitful to identify genes specifically expressed in these cells and early embryonic epiblast. The identification of these genes will enable to decipher the mechanisms of transcriptional control of gene expression in relation with the maintenance of the multipotent phenotype. Chicken embryonic stem cells appear then as a good model for the study of early steps of chicken embryonic development.

## 4. Materials and methods

### 4.1. CES cells culture, grafting in vivo and transfection

Isolation and cultivation of CES cells were described previously (Pain et al., 1996). In vitro differentiation of CES cells was induced on gelatin coated dishes in medium supplemented with retinoic acid ( $5 \times 10^{-6}$  M) without growth factors or by the formation of embryoid bodies (Pain et al., 1996). Grafting of CES cells into X-irradiated stage X chicken embryos was described previously (Pain et al., 1996). Embryonic stages were defined according to Eyal Giladi (EG) (Eyal-Giladi and Kochav, 1976) and Hamburger and Hamilton (1951) (H&H).

The ROSA  $\beta$ -geo vector was obtained from P. Soriano (Hutchinson cancer Center, Seattle, WA) and described in Friedrich and Soriano (1991). The supercoiled plasmid was linearized by the *Dra*I enzyme.

CES cell cultures were dissociated when their growth density reached semi confluence on 100 mm plates corresponding roughly to  $5\text{--}7 \times 10^6$  cells per plate. After dissociation with pronase (0.025% in Tris Glucose buffer), the cells were washed, counted and resuspended in Glasgow medium at a number of  $5 \times 10^6$  cells in 0.8 ml. Ten micrograms of linear ROSA  $\beta$ -geo DNA were added and the mixture allowed to rest for 10 min at 4°C. The cell suspension was electroporated twice at 280 V, 500  $\mu$ F in a 1 mm cuvette using a BioRad electroporator. Cells were then kept for 10 min at 4°C. Fresh complete medium was added and the cells were directly plated on fresh feeder cells. Thirty-six hours after the pulse, G418 was added at a final concentration of 250  $\mu$ g/ml.

The selection medium was changed daily during the first 4 days and every 2 days thereafter. Fresh feeder cells were added on the 5th days of selection. Clones were identifiable 6–8 days after beginning of the selection procedure. They were picked after 8–10 days, amplified and frozen in liquid nitrogen.

### 4.2. $\beta$ -galactosidase staining

Cells and embryos were fixed at 4°C in a fixative solution containing 1% formaldehyde, 0.2% glutaraldehyde, and 0.02% Nonidet P-40 in PBS for 30 min for the cells and 30–120 min for the embryos depending on their developmental stage and then washed twice in PBS at room temperature. They were then incubated at 37°C in PBS containing 1 mg/ml 5-bromo-4-chloro-3-indolyl b-D-galactopyranoside (X-gal), 5 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 5 mM K<sub>4</sub>Fe(CN)<sub>6</sub>, and 2 mM MgCl<sub>2</sub> and 0.02% Nonidet P-40. The time of incubation was 1 h for the cells and 1–24 h for the embryos depending on their stage.

### 4.3. RNA extraction and RT-PCR analysis

Total RNAs were isolated using a modified method of Chomczynski and Sacchi (1987). Northern blot analyses were carried out according to the standard glyoxal-DMSO procedure. Poly(A)<sup>+</sup> RNA was isolated from total RNA with Dynabeads Oligo(dT)25 (Dynal).

### 4.4. DNA and RNA probes

The probes C1, S1 and S2 were prepared by PCR respectively with the following primers: C1 P4(s), 5'-AGACCGGCCTCACTGCTC-3' and C1 3'(as1), 5'-AGGACG-GACGCTATTAAGGATT-3'; 2c S1, 5'-AACCCAGAC-TATCAAGC-3' and 2cAS1, 5'-TGGCCAGTTCAGT-GGCTG-3'; 2c S2, 5'-GAGAGGTTGCCCAA- GAAT-3' and 2c AS2, 5'-GAGCTTCCTAATAAACATCC-3'. The C1 and S1 probes used for the in situ hybridization were bidirectionally cloned in the BlueScript vector (Stratagene) for in vitro transcription by T3 or T7 polymerase.

### 4.5. DNA and PCR analysis

Cells, embryos and embryonic tissues were lysed in a lysis buffer (150 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5), 0.5% SDS, 100  $\mu$ g/ml proteinase K) during 12 h at 50°C. DNA was extracted with phenol/chloroform and precipitated with isopropanol. Southern blot analysis of genomic DNA was performed using standard techniques. For the Zooblot, membranes were hybridized overnight at different temperatures from 45 to 65°C and washed twice either with 2 $\times$  SSC, 0.1% SDS or 0.2 $\times$  SSC, 0.1% SDS for 30 min at 45°C.

PCR amplifications of *cENS-1* coding sequence in Gallinaceae were performed using the *cENS-1* S1, 5'-GATCTAGATCCTCAAATGAAT-3' and *cENS-1* AS1, 5'-TCTTGGGCAACCTCCCC-3' primers. The PCR



conditions used were the following: 5 min denaturation at 94°C followed by 10 cycles of 95°C for 30 s, 62°C for 30 s and 72°C for 1 min; followed by 10 cycles of 95°C for 30 s, 57°C for 30 s and 72°C for 1 min, followed by 20 cycles of 95°C for 30 s, 52°C for 30 s and 72°C for 1 min ending with a 5 min extension at 72°C.

#### 4.6. Isolation of trapped gene sequences by 5'RACE

Cloning of the fusion LacZ transcript from the 9N2.5 cells was performed with the Marathon cDNA amplification kit (Clontech) according to the manufacturer's instructions. As suggested by the manufacturer the following primers were used for PCR and nested PCR amplification with Expand™ Long Template PCR System (Boehringer Mannheim): LacZ A, 5'-CCGTGCATCCACCACATACAGG-3', LacZ B, 5'-GGGATCCGCCATGTACAGA-3' and LacZ C, 5'-GTGCCGGAAACCAGGCAAA-3'. PCR products were digested respectively with *NotI* and *HindIII* or *NotI* and *BamHI*, ligated into the pBluescript vector, and transformed into DH5 $\alpha$  bacterial host. Isolated plasmids were sequenced in a PE 373 XL sequencer using the Big Dye™ RR mix.

#### 4.7. Construction of a CES cDNA library and screening

Total RNAs extracted from undifferentiated CES cells were used to construct a cDNA library in the  $\lambda$ ZAP-II EcoRI/CIAP-treated Vector (Stratagene). The library was screened using the PstI/SacII C1 fragment as a <sup>32</sup>P-labelled probe.

#### 4.8. In situ hybridization of whole-mount embryos

Whole-mount in situ hybridization was carried out as described by Wilkinson (1993) and revised by Streit et al. (1997). For histological sections, fixed embryos were dehydrated in methanol during 10 min then in isopropanol during 5 min, embedded in paraffin, and sectioned at 10  $\mu$ m on a Leitz microtome.

#### 4.9. Promoter analysis using luciferase reporter gene

Promoter activity was assessed with the Dual Luciferase Reporter Assay System (Promega, USA). All the luciferase reporter vectors were purchased from Promega. The putative promoter region of the *cENS-1* locus was amplified by PCR using LTR 1S, 5'-CCACGCGTACGACAGACTTGAGGGGT-3', LTR 3S, 5'-ACGCGTGTGGAT-GTTT-ATTAGGAAGC-3' and LTR AS, 5'-CCACGCGTGT-TGGTCCGACCAGTTTA-3' primers and cloned by TA-cloning in pGEM-T easy vector (Promega). The amplified fragment was then inserted into the MluI site of pGL2 basic vector.

CES, QT6, QBr and HeLa cells were seeded in 12-well plates at  $6 \times 10^4$  cells/well. After 24 h, cells were transiently cotransfected using Lipofectamine reagent (Gibco) with 300 ng of vector and 30 ng of the Renilla Luciferase Reporter

Gene pRL-CMV for internal control. After 5 h of incubation in culture medium containing only 2% fetal bovine serum the cells were washed twice with phosphate buffer saline (PBS) and fresh normal growth medium was added to each well. On the following day, cells were washed twice with PBS and lysed with passive lysis buffer, then Firefly and Renilla luciferase luminescence assays were successively performed as described by the manufacturer. Firefly reporter gene values were normalized to the activity of the Renilla luciferase, used as internal control.

#### Acknowledgements

We thank François Mallet, Vincent Laudet and Thierry Heidman for helpful suggestions, Laurent Duret for computer analysis, and Abdel Bensmail, Christelle Prudhomme and Christelle Merouche for technical assistance. This work was supported by grants from the ARC and EEC (no. BIO4-97-2106).

#### References

- Aladjem, M.I., Spike, B.T., Rodewald, L.W., Hope, T.J., Klemm, M., Jaenisch, R., Wahl, G.M., 1998. ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Curr. Biol.* 8, 145–155.
- Benit, L., De Parseval, N., Casella, J.F., Callebaut, I., Cordonnier, A., Heidmann, T., 1997. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. *J. Virol.* 71, 5652–5657.
- Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156–159.
- Cohen, M., Powers, M., O'Connell, C., Kato, N., 1985. The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology* 147, 449–458.
- Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J., Smith, A., 1998. Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev. Biol.* 203, 149–162.
- Doetschman, T., Williams, P., Maeda, N., 1988. Establishment of hamster blastocyst-derived embryonic stem (ES) cells. *Dev. Biol.* 127, 224–227.
- Dunwiddie, C.T., Resnick, R., Boyce-Jacino, M., Alegre, J.N., Faras, A.J., 1986. Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. *J. Virol.* 59, 669–675.
- Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156.
- Eyal-Giladi, H., Kochav, S., 1976. From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev. Biol.* 49, 321–337.
- Friedrich, G., Soriano, P., 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* 5, 1513–1523.
- Guilhot, C., Benchaïbi, M., Flechon, J.E., Samarut, J., 1993. The 12S adenoviral E1A protein immortalizes avian cells and interacts with the avian RB product. *Oncogene* 8, 619–624.
- Hamburger, V., Hamilton, H., 1951. A series of normal stages in the development of the chick embryo. *J. Morphol.* 88, 49–92.

- Hosler, B.A., LaRosa, G.J., Grippo, J.F., Gudas, L.J., 1989. Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. *Mol. Cell Biol.* 9, 5623–5629.
- Huang, K., Serria, M.S., Nakabayashi, H., Nishi, S., Sakai, M., 2000. Molecular cloning and functional characterization of the mouse *mafB* gene. *Gene* 242, 419–426 (Jan 25).
- Johnsgard, P.A., 1986. *The Pheasants of the World*, Oxford University Press, Oxford.
- Kimball, R.T., Braun, E.L., Zwartjes, P.W., Crowe, T.M., Ligon, J.D., 1999. A molecular phylogeny of the pheasants and partridges suggests that these lineages are not monophyletic. *Mol. Phylogenet. Evol.* 11, 38–54.
- Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* 78, 7634–7638.
- Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., Yokota, T., 1999. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* 18, 4261–4269.
- Matsui, Y., Zsebo, K., Hogan, B.L., 1992. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 70, 841–847.
- Moscovici, C., Moscovici, M.G., Jimenez, H., Lai, M.M., Hayman, M.J., Vogt, P.K., 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. *Cell* 11, 95–103.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., Smith, A., 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379–391.
- Niwa, H., Burdon, T., Chambers, I., Smith, A., 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* 12, 2048–2060.
- Niwa, H., Miyazaki, J., Smith, A.G., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells [see comments]. *Nat. Genet.* 24, 372–376.
- Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M., Hamada, H., 1991. The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *EMBO J.* 10, 2997–3005.
- Pain, B., Clark, M.E., Shen, M., Nakazawa, H., Sakurai, M., Samarut, J., Etches, R.J., 1996. Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. *Development* 122, 2339–2348.
- Savatier, P., Huang, S., Szekely, L., Wiman, K.G., Samarut, J., 1994. Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* 9, 809–818.
- Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B., Samarut, J., 1996. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* 12, 309–322.
- Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R., Leib-Mosch, C., 1998. Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J. Virol.* 72, 8384–8391.
- Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., Gearhart, J.D., 1998. Derivation of pluripotent stem cells from cultured human primordial germ cells [published erratum appears in *Proc Natl Acad Sci USA* 1999 Feb 2;96(3):1162]. *Proc. Natl. Acad. Sci. USA* 95, 13726–13731.
- Smith, A.G., Hooper, M.L., 1987. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev. Biol.* 121, 1–9.
- Streit, A., Stern, C.D., 1999. Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mech. Dev.* 82, 51–66.
- Streit, A., Sockanathan, S., Perez, L., Rex, M., Scotting, P.J., Sharpe, P.T., Lovell-Badge, R., Stern, C.D., 1997. Preventing the loss of competence for neural induction: HGF/SF, L5 and Sox-2. *Development* 124, 1191–1202.
- Streit, A., Berliner, A.J., Papanayotou, C., Sirulnik, A., Stern, C.D., 2000. Initiation of neural induction by FGF signalling before gastrulation. *Nature* 406, 74–78 (Jul 6).
- Sukoyan, M.A., Vatolin, S.Y., Golubitsa, A.N., Zhelezova, A.I., Semenova, L.A., Serov, O.L., 1993. Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: comparisons of their pluripotencies. *Mol. Reprod. Dev.* 36, 148–158.
- Sun, L., Bradford, C.S., Ghosh, C., Collodi, P., Barnes, D.W., 1995. ES-like cell cultures derived from early zebrafish embryos. *Mol. Mar. Biol. Biotechnol.* 4, 193–199.
- Teich, N., Varmus, H., Coffin, J., 1984. In: Weiss, R. (Ed.). *RNA Tumor Viruses*, Vol. 1. Cold Spring Harbor laboratory, New York.
- Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., Hearn, J.P., 1995. Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. USA* 92, 7844–7848.
- Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic stem cell lines derived from human blastocysts [see comments] [published erratum appears in *Science* 1998 Dec 4;282(5395):1827]. *Science* 282, 1145–1147.
- Voss, A.K., Thomas, T., Petrou, P., Anastassiadis, K., Scholer, H., Gruss, P., 2000. Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* 127, 5449–5461.
- Wheeler, M.B., 1994. Development and validation of swine embryonic stem cells: a review. *Reprod. Fertil. Dev.* 6, 563–568.
- Wilkinson, D.G., 1993. In: Rickwood, D., Hames, B.D. (Eds.). *Essential Developmental Biology*, Oxford University Press, Oxford.

## 5°) BREVET

Ces travaux ont permis le dépôt d'un brevet dénommé : Cellules ES modifiées et gènes spécifiques de cellules ES. Il protège notamment l'utilisation des gènes ENS comme gènes marqueurs des cellules ES de poulet et l'utilisation des loci et du promoteur ENS pour l'expression ES spécifique de gènes d'intérêt. L'intégralité du texte du brevet est disponible sur le site de l'Institut National de la Propriété Industrielle : <http://www.inpi.fr> .

①⑨ RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

①① N° de publication : **2 808 803**

(à n'utiliser que pour les  
commandes de reproduction)

②① N° d'enregistrement national : **00 06029**

⑤① Int Cl<sup>7</sup> : **C 12 N 15/12**, C 07 K 14/465, 16/18, C 12 N 15/85, 5/  
06, 5/10, C 12 Q 1/68, G 01 N 33/53, 33/02, A 61 K 48/00, 38/  
17, 39/395

①②

## DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 11.05.00.

③③ Priorité :

④③ Date de mise à la disposition du public de la  
demande : 16.11.01 Bulletin 01/46.

⑤⑥ Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule*

⑥① Références à d'autres documents nationaux  
apparentés :

⑦① Demandeur(s) : INSTITUT NATIONAL DE LA  
RECHERCHE AGRONOMIQUE INRA — FR, CENTRE  
NATIONAL DE LA RECHERCHE SCIENTIFIQUE  
CNRS — FR et ECOLE NORMALE SUPERIEURE DE  
LYON — FR.

⑦② Inventeur(s) : ACLOQUE HERVE, BIROT ANNE  
MARIE, RISSON VALERIE, PAIN BERTRAND et  
SAMARUT JACQUES.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : REGIMBEAU.

⑤④ CELLULES ES MODIFIEES ET GENE SPECIFIQUE DE CELLULES ES.

⑤⑦ La présente invention concerne des cellules ES  
d'oiseau modifiées, exprimant spécifiquement un gène exo-  
gène lorsqu'elles possèdent un caractère pluripotent. L'in-  
vention concerne également un acide nucléique et un  
polypeptide exprimés spécifiquement dans les cellules  
d'oiseau pluripotentes, ainsi que des méthodes de détection  
du caractère pluripotent de cellules mettant en oeuvre ces  
acide nucléique et polypeptide.

FR 2 808 803 - A1



## 6°) DISCUSSION ET COMMENTAIRES SUR LES RESULTATS PRESENTES

### a) Approche par *gene-trap*

Cette publication est la première à montrer qu'il est possible de modifier génétiquement des cellules ES de poulet et de conserver leur capacité à coloniser l'embryon, au moins partiellement. Cette approche nous a également permis d'identifier une famille de gènes exprimés spécifiquement dans les cellules ES et dans l'embryon précoce de poulet.

Plusieurs résultats obtenus par cette approche *gene-trap* sont cependant troublants. Comme on l'a vu précédemment, l'idée initiale était d'obtenir des clones exprimant de façon ubiquitaire un gène rapporteur. Or tous les clones obtenus avec le vecteur ROSA  $\beta$ -geo présentaient le même profil d'expression : une expression forte dans les cellules souches non différenciées et une extinction de cette expression dans les cellules induites à se différencier. D'autre part, l'analyse par 5'RACE des différents clones obtenus suggérait qu'ils étaient tous intégrés dans des loci identiques. Il est envisageable qu'une contamination des différents clones ait pu avoir lieu, soit au cours de la culture, soit lors de la manipulation de leurs ARN ou ADN. Il est cependant troublant qu'aucun autre type de clones n'ait été obtenu. Nous pensons que le vecteur ROSA  $\beta$ -geo, qui possède des LTR rétrovirales, a pu s'intégrer de façon ciblée dans ces loci, phénomène qui pourrait être amplifié par le nombre de loci ENS. D'autres types de vecteurs de *gene-trap*, qui ne possèdent pas de LTRs rétrovirales, pourraient être utilisés par la suite pour éviter ce genre de problème.

Une autre difficulté à laquelle nous avons été confrontés réside dans notre impossibilité à identifier le site d'insertion du vecteur de *gene-trap* dans l'ADN génomique des cellules 9N2-5. Nous n'avons jamais réussi à amplifier sur l'ADN génomique la séquence entre le transgène et la séquence ENS endogène. Il se pourrait donc qu'au cours de l'épissage du transcrit primaire, les deux séquences aient été rapprochées, alors qu'elles se situent à plusieurs kilobases dans l'ADN génomique. Ce point nous a particulièrement gêné car nous n'avons jamais pu montrer que l'expression du transgène était dirigée par le même promoteur que celui que nous avons analysé par la suite.

## b) Identification d'une nouvelle famille de gènes

Par le criblage d'une banque d'ADNc de cellules ES de poulet, réalisée par Ryota Kunita, nous avons identifié plusieurs ADNc comportant la même séquence que celle présente en amont du transgène  $\beta$ -geo, sur le transcrit issu du locus piégé par gene trap. Ce travail, initié par Valérie Risson, a été particulièrement surprenant car l'ensemble de nos résultats a abouti à l'identification de nombreux gènes, résumés dans la **Figure 31**. En plus de ceux présentés dans l'article, nous avons obtenu des gènes dont nous n'avons jamais pu valider la correspondance avec de véritables loci et qui semblent provenir d'artéfacts produits par la banque. Ce fut le cas du gène *TRAP150* (Thyroid receptor associated protein) ou de celui de la *cofiline*, un gène du cytosquelette, dont la séquence était homologue à 100% avec celle de la souris (les cellules ES de poulet utilisées pour la réalisation de la banque sont cultivées sur un feeder de cellules murines). Nous n'avons d'ailleurs pas pu valider ces résultats par Northern Blot ou hybridation *in situ*. Il est tout de même étrange de constater qu'une majorité des cadres ouverts de lecture débute au même endroit, comme s'il s'agissait d'un site d'intégration privilégié, que ce soit pour les gènes *TRAP150*, *cofiline* ou *ENS-1*... L'impression générale donnée par ces résultats est que, de façon artéfactuelle ou non, il existe un site d'insertion privilégié dans cette structure, qui a permis la formation du gène ENS-1. Il reste maintenant à déterminer si *ENS-1* existe dans le génome avec une structure « classique », composée d'une alternance d'introns et d'exons et seulement sous une forme biallélique.

## c) Profil d'expression des gènes ENS

Au moment de la rédaction de cet article, il existait peu de données disponibles sur le transcriptome du poulet. A partir de nos données acquises par Northern Blot et hybridation *in situ*, nous avons décrit ce gène comme exprimé spécifiquement dans les cellules ES de poulet et l'embryon précoce. L'accessibilité récente de banques d'EST produites à partir de différents tissus embryonnaires et adultes, nous a permis de constater que ce profil d'expression est certainement plus large que celui décrit initialement. D'autre part l'analyse des chimères produites avec les cellules 9N2-5 (clone issu du gene trap) montre que le gène LacZ est exprimé également dans de nombreux tissus mais seulement dans quelques cellules. Ces données suggèrent donc que les gènes ENS pourraient être exprimés spécifiquement dans certains types cellulaires d'organes ou de tissus embryonnaires et adultes. Il serait intéressant

de vérifier si les gènes *ENS* peuvent aussi être un marqueur des cellules souches adultes présentes chez l'oiseau. Afin de déterminer plus précisément les tissus dans lesquels sont exprimés les protéines ENS-1, nous avons produit un anticorps polyclonal dirigé contre un peptide de la protéine ENS-1. Si l'anticorps semble reconnaître dans les cellules ES une protéine de la taille attendue (par rapport à celle d'ENS-1 traduite *in vitro*), il reconnaît également une protéine de même taille dans toutes les cellules mammifères testées. Afin de savoir si cet anticorps était spécifique chez le poulet, nous avons réalisé par Western Blot son profil d'expression au cours du développement embryonnaire. Comme attendu, le niveau d'expression de la protéine décroît très rapidement dans l'embryon après la gastrulation. Nous avons donc ensuite observé par immunocytochimie la localisation de la protéine dans les cellules ES (**Figure 32**). Celle-ci semble se localiser majoritairement dans le cytoplasme.

#### d) Conservation des gènes *ENS*

Notre étude nous a conduit à supposer que les gènes *ENS* étaient conservés uniquement chez les galliformes. En effet, si ces gènes sont facilement détectables dans l'ADN génomique des espèces de galliformes testés, ils sont indétectables par southern blot et PCR sur l'ADN génomique des ansériformes, le genre le plus apparenté aux galliformes, mais aussi chez le pigeon et tous les autres vertébrés testés. Streit et *al.* (2000) avaient publié que le gène *ERNI (ENS-1)* était également conservé chez le diamant mandarin, un passereau d'Australie, mais ce résultat semble être dû à une contamination, la séquence publiée étant identique à celle du poulet alors que celle de la caille, phylogénétiquement plus proche du poulet que le diamant mandarin, est moins conservée... En discutant avec Vincent Laudet, dont l'équipe est spécialisée dans la phylogénèse des récepteurs nucléaires, j'avais eu envie d'utiliser le gène *ENS-1* pour établir un nouvel arbre phylogénétique des galliformes basé sur un gène nucléaire et non mitochondrial. Le polymorphisme du gène *ENS-1* semblait en effet intéressant pour réaliser ce genre d'étude. Le clonage et le séquençage de ce gène dans de nombreuses espèces de galliformes (poulet, faisan, perdrix rouge et grise, caille des blés et du japon, pintade et dinde) a sensiblement bloqué mes velléités évolutionnistes. En effet, pour avoir des résultats cohérents, il fallait cloner toutes les copies du gène *ENS-1* présentes dans le génome de chaque espèce, de nouvelles insertions ayant pu se produire successivement après la première intégration de ce gène dans l'ancêtre des galliformes. Ce travail étant relativement fastidieux (il existe au moins dix copies du gène *ENS-1* chez le poulet) j'ai renoncé à le mener à terme. La connaissance du génome du poulet permettra d'obtenir la séquence de toutes les

copies du gène *ENS-1* présentes chez le poulet et peut être, alors, de choisir un locus particulier pour mener à bien cette étude.

#### e) Origine rétrovirale des gènes ENS

Les séquences rétrovirales que nous avons identifiées ne présentent pas d'homologue connu chez le poulet. La comparaison des séquences avec celles disponibles sur la base de données GenBank a permis de rattacher la séquence du gène *pol* à celle des rétrovirus du type C et la séquence *env* à celle des rétrovirus du type B. Si ce rétrovirus endogène existe en intégralité dans le génome du poulet, il s'agirait alors d'une nouvelle famille de rétrovirus endogène du poulet, ne s'apparentant pas aux rétrovirus endogènes de type ALV ou EAV. Il est également troublant de constater que la séquence codante pour *ENS-1* se trouve en amont du gène *pol* comme s'il avait été substitué au gène *gag* dans la structure rétroviral. Cependant la protéine ENS-1 ne présente aucune homologie avec celle des protéines gag connues.

Nous n'avons pas focalisé nos recherches sur la caractérisation complète des loci ENS. Il aurait en effet fallu pour cela cribler à nouveau la banque d'ADNc avec des sondes localisées sur les gènes *pol* ou *env* pour obtenir le génome complet du virus comprenant le gène *gag*. Le séquençage du génome du poulet, dont un premier brouillon devrait être disponible courant septembre 2003, permettra de savoir s'il existe un rétrovirus endogène complet apparenté à *ENS-3* et combien de loci différents gouvernent l'expression des gènes *ENS*.

Afin de tester la fonctionnalité d'un rétrovirus endogène complet apparenté à *cENS-3* dans les cellules ES de poulet nous avons évalué, en collaboration avec Philippe Mangeot, alors étudiant en thèse dans le laboratoire de rétrovirologie humaine INSERM U412, l'activité de transcription inverse d'un surnageant de cellules ES de poulet. Celle-ci ne diffère cependant guère de celle du surnageant de cellules COS non infectées. Il n'y aurait donc, dans les cellules ES de poulet, qu'une activité basale de transcription inverse, beaucoup plus faible que dans des cellules où les transcriptases inverses de rétrovirus seraient fortement exprimées et fonctionnelles. Nous n'avons pas, par contre, recherché de particules virales dans le surnageant des cultures de cellules ES.

Une deuxième hypothèse intéressante consiste à proposer que les gènes *cENS-1* et *cENS-2* dériveraient d'un événement de rétrotransposition. Les pseudogènes dérivant d'un événement de transposition et habituellement décrits présentent une structure génomique sans introns, sont encadrés par de courtes séquences répétées (pour revue Mighell et al. 2000) et



codent parfois pour des protéines fonctionnelles (Lyko et al. 2000, Locascio et al. 2002). Les gènes *cENS-1* et *cENS-2* semblent correspondre à cette définition. Il devrait donc exister, dans ce cas, dans le génome du poulet, un locus « conventionnel » pour ces deux gènes, précédant l'événement de rétrotransposition.

Ces deux gènes ont-ils été produits par l'intermédiaire d'un rétrotransposon ou d'un rétrovirus ? La littérature différencie les rétrotransposons à LTRs des rétrovirus endogènes par l'absence de la séquence codante pour *env*. Nos résultats, qui montrent l'existence du gène *env* en colinéarité avec *pol* et encadré par des LTRs, suggèreraient donc plutôt que cet événement de rétrotransposition serait dû à un rétrovirus (endogène ou non) et non à un rétrotransposon (Carlton et al. 1995). Des travaux publiés en 1998 par Lum et Linial montrent qu'il est possible d'observer des rétrotranspositions d'ARN cellulaire dans une lignée de cellules de caille SE21Q1b infectées par le virus RSV. Les auteurs observent également que ce mécanisme nécessite la présence des glycoprotéines d'enveloppe et d'un signal rétroviral d'encapsulation car l'intégration du pseudogène nécessite une rétroinfection.

Si l'origine des gènes ENS est réellement due à un rétrotransposon ou à un rétrovirus, il serait alors particulièrement intéressant de reconstituer un rétrovirus ou un rétrotransposon fonctionnel, comme cela fut réalisé pour le rétrotransposon *Sleeping Beauty* (Ivics et al. 1997). Il serait alors possible de produire un outil intéressant, à la fois pour l'expression de gènes d'intérêt dans les cellules ES mais aussi pour la création de lignées ES mutantes.

#### f) Identification d'une région promoteur

Nous avons montré que la région potentielle U3 des gènes *ENS* possédait bien une activité promoteur, au contraire des régions décrites comme R ou U5. Par la technique du 5'RACE nous avons également identifié un site d'initiation de la transcription, proche d'une boîte TATAAA.

Nous avons également montré que cette activité était particulièrement forte dans les cellules ES de poulet indifférenciées.

Je voudrais insister tout de même sur la difficulté de comparer des activités d'un même promoteur dans différents types cellulaires. En effet plusieurs problèmes peuvent se présenter :

- Le taux de transfection transitoire peut différer fortement d'un type cellulaire à l'autre. C'est le cas avec les cellules ES de poulet où seulement entre 5 et 10% des cellules sont transfectées.

- Le promoteur du gène rapporteur utilisé comme contrôle interne peut avoir une activité variable selon les types cellulaires. La normalisation à partir de promoteurs différents peut changer l'amplitude de l'activité observée.
- Le promoteur du contrôle interne peut réagir aux traitements appliqués aux cellules transfectées. Nous avons ainsi observé que l'acide rétinoïque augmente l'activité du promoteur CMV. Il ne faut donc pas utiliser ce promoteur pour normaliser une activité lors de l'induction de la différenciation des cellules ES. Nous avons donc par la suite vérifié nos résultats en utilisant un gène de normalisation sous le contrôle du promoteur TK (thymidine kinase) dont l'activité ne varie pas lors du traitement par l'acide rétinoïque.
- Le promoteur rapporteur peut entrer en compétition avec le promoteur testé pour un ou plusieurs facteurs de transcription.

L'ensemble de ces observations confirme que la comparaison d'une activité d'un promoteur dans différents types cellulaires est délicate par une approche transitoire. Elle permet tout de même de mettre en évidence des différences si celles-ci sont importantes.

## **II UN FACTEUR DE TRANSCRIPTION DE LA FAMILLE CP2 EST NECESSAIRE A L'ACTIVATION DU PROMOTEUR *ENS***

### 1°) INTRODUCTION AU TRAVAIL EXPERIMENTAL

Le but de ce travail était d'identifier les régulateurs conférant au promoteur *ENS* son activité tissu-spécifique. Dans cette optique, je me suis notamment inspiré des travaux réalisés par l'équipe de Gérard Karsenty à Houston, qui avait montré que la tissu-spécificité du promoteur de l'ostéocalcine dans les ostéoblastes était dépendante de deux éléments de réponse présents dans la partie proximale du promoteur (Ducy et Karsenty 1995). L'analogie de l'un des éléments de réponse avec celui de facteurs de transcription connus leur a permis d'identifier un facteur de transcription spécifique de l'ostéoblaste, *Osf2*, mais appartenant à une famille connue de facteurs de transcription, *Cbfa* (Ducy et *al.* 1997). Parallèlement, ils ont purifié le deuxième facteurs de transcription par chromatographie d'affinité (Schinke et Karsenty 1999). Dans le même ordre d'idée, mon approche était donc de caractériser tout

d'abord des éléments de réponse présents sur le promoteur *ENS* puis d'identifier les facteurs de transcription régulant l'expression des gènes *ENS*.

## 2°) PRESENTATION DU TRAVAIL EFFECTUE

### a) Mise en évidence d'une région activatrice dans le promoteur *ENS*

Pour étudier le promoteur *ENS*, nous avons délimité une séquence incluant la région potentielle U3 allant du résidu -738 au résidu +83 par rapport au site d'initiation de la transcription décrit initialement (**Figure 33**). Ce promoteur a été placé en amont du gène rapporteur codant pour la luciférase.

En réalisant des délétions successives du promoteur à partir de son extrémité 5', nous avons mis en évidence que, dans les cellules ES de poulet, l'activité du promoteur variait peu en supprimant les résidus -738 à -308 mais qu'une région activatrice très forte du promoteur se situait entre les résidus -179 et -308. En effet l'activité du promoteur est multipliée par un facteur 20 entre ces deux constructions (**Figure 34A**). Par contre, dans des cellules ES différenciées ou des cellules QT6 (fibroblastes de caille), cette activation demeurait faible, (**Figure 34B**) suggérant que cette région était beaucoup moins active dans ces cellules. Le contrôle de la différenciation étant délicat et les populations cellulaires différentes d'une expérimentation à l'autre, nous avons choisi de travailler par la suite uniquement avec les cellules QT6, puisque l'activité transcriptionnelle du promoteur *ENS* est identique dans ces cellules et dans les cellules ES de poulet différenciées.

Afin de tester si la région comprise entre les résidus -179 et -308 comportait réellement un activateur de la transcription, nous avons placé une ou trois copies de cette région en amont du promoteur basal SV40 ou du promoteur *ENS* -179/+83. Après transfection dans des cellules ES de poulet ou des cellules QT6 et quantification de l'activité luciférase, nous avons montré que cette région augmentait l'activité du promoteur basal SV40 dans les deux lignées cellulaires testées (**Figure 35A**). Lorsqu'elle était placée en amont du promoteur *ENS* -179/+83, l'activité de cette région était détectable dans les cellules ES de poulet et dans les cellules QT6 (**Figure 35B**), avec cependant une activation plus faible dans les cellules QT6, en comparaison de celle observée dans les cellules CES. Ces données mettent en évidence deux caractéristiques du promoteur *ENS*. Premièrement, la région située entre les résidus -179/-308 comporte un activateur de la transcription, aussi bien actif dans les cellules CES que les cellules QT6 lorsqu'il est placé en amont du promoteur basal SV40.

Deuxièmement, l'activité de cette région est restreinte aux cellules ES de poulet dans le contexte particulier du promoteur ENS -179/+83.

Plusieurs hypothèses peuvent être avancées pour expliquer ce résultat. Soit il existe dans le promoteur ENS -179/+83 un site de fixation pour un coactivateur qui augmente fortement le niveau basal d'activité de la région -308/-179. Cependant l'activité de la région -308/-179 demeure élevée dans le contexte du promoteur basal SV40. Soit il existe dans cette région un site de fixation pour un facteur qui bloque l'activité de la région activatrice dans les cellules QT6. Nous verrons plus loin que plusieurs arguments sont en faveur de cette deuxième hypothèse. Finalement ces résultats montrent que la région -308/-179 est importante pour l'activité du promoteur ENS mais semble insuffisante, seule, à en assurer l'activité tissu-spécifique.

Afin d'identifier les facteurs qui pouvaient se fixer sur cette région activatrice, nous avons réalisé une série d'expériences de retard de migration sur gel (EMSA) avec des oligonucléotides chevauchants, couvrant la région -308/-179 (**Figure 36A**). Cette analyse nous a conduit à délimiter deux petits domaines d'une vingtaine de paire de bases sur lesquels se fixaient des facteurs nucléaires présents dans les cellules ES mais aussi dans les cellules QT6 (**Figure 36B**).

Pour identifier les nucléotides plus particulièrement impliqués dans l'interaction ADN-protéine, nous avons utilisé la technique de l'interférence de méthylation. Le principe de cette technique est expliquée sur la **Figure 37**. Afin de méthyler les sondes A et B, nous avons utilisé un agent méthylant appelé DMS (diméthylsulfate). Ce réactif ne méthyle que les résidus guanines et ne permet donc de détecter que les résidus de type guanine importants pour l'interaction. Nous avons ainsi identifié plusieurs résidus délimitant deux éléments de réponse, appelés A (-240/-232) et B (-294/-280) dont les séquences respectives sont les suivantes : TGATGGACA et CAAGTCCAGGCAAGT (**Figure 38**).

b) La région A n'est pas capable de trans-activer le promoteur ENS

Nous avons tout d'abord vérifié que l'élément de réponse défini par la technique de l'interférence de méthylation était bien celui observé dans l'interaction ADN-protéine. Pour cela, nous avons réalisé des gels retard où la sonde A marquée entrait en compétition avec des sondes mutées pour les résidus localisés dans la région A. Ces expériences ont confirmé les données obtenues précédemment par la technique de l'interférence de méthylation.

Nous avons ensuite étudié la fonctionnalité de cet élément A pour l'activation du promoteur ENS. Dans cette optique, nous avons placé 1 ou 3 copies de la région A en amont du promoteur basal SV40 ou du promoteur ENS -179/+83. Après transfection de ces constructions dans les cellules ES de poulet ou dans les cellules QT6, l'activité enhancer a été estimée proportionnellement à l'activité des vecteurs seuls. Nos résultats suggèrent que l'élément A ne possède aucune fonction enhancer.

### c) Caractérisation de l'élément B

La détermination des résidus nécessaires à la formation du complexe ADN-protéine sur les sondes comportant l'élément B, a été réalisée par des expériences de retard sur gel. La sonde comportant l'élément B a été incubée en présence d'un excès de différents oligonucléotides mutés dans la région B (**Figure 39A**). Nous avons ainsi montré que les guanines, localisées sur les brin sens et antisens, identifiées par interférence de méthylation, étaient chacune nécessaires à l'interaction (**Figure 39B**). La séquence de l'élément B est composée d'une répétition directe 5'-CAAGT CCAGG CAAGT-3'. Nos résultats montrent que les résidus cytosines ou guanines de chaque répétition sont indispensables à l'interaction. Afin de déterminer si les autres résidus présents dans chaque partie de la répétition étaient importants pour l'interaction ADN-protéine, nous avons réalisé des expériences de retard sur gel avec des sondes mutées pour chaque résidu (**Figure 39C**). Nos résultats montrent que, finalement, seuls les résidus cytosines et guanines de cette répétition sont importants pour l'interaction. La mutation du résidu tyrosine affecte également l'interaction mais sans l'abolir complètement (**Figure 39C** ligne 5). La répétition est séparée par une séquence de 5 nucléotides (5'-CCAGG-3'). Nous avons testé l'influence de cet espacement pour la formation du complexe ADN-protéine. Nous avons donc utilisé des sondes auxquelles des bases, situées entre les éléments de la répétition, avaient été ajoutées ou enlevées. Les résultats des expériences de retard sur gel obtenus avec ces sondes sont présentés sur la **Figure 40**. Une délétion de 3 nucléotides ne perturbe que partiellement la formation du complexe ADN-protéine (**Fig.40** ligne 6). Au contraire, l'addition de 3 nucléotides affecte fortement la formation du complexe (**Fig.40** ligne 10) tandis que celui-ci devient quasiment indétectable avec l'addition de 6 nucléotides (**Fig.40** ligne 14).

L'élément de réponse B optimal peut donc être résumé par la séquence 5'-CAAGTNNNNNCAAGT où les caractères gras représentent les résidus nécessaires à la formation du complexe ADN-protéine

d) L'élément B est nécessaire à l'activation du promoteur ENS

Pour tester la fonctionnalité de l'élément, B nous avons placé 1 ou 3 copies de cet élément, muté ou non, en amont du promoteur basal SV40 ou du promoteur ENS -179/+83. Les différentes constructions ont ensuite été transfectées dans les cellules ES de poulet ou les cellules QT6 (**Figure 41**). Il ressort de ces expériences que l'élément B possède une activité enhancer aussi bien dans les cellules ES de poulet que dans les cellules QT6 lorsqu'il est placé en amont du promoteur basal SV40 (**Fig.41A**). En amont du promoteur ENS -179/+83 l'activité de l'élément B est également mise en évidence dans les cellules ES de poulet (**Fig.41B**) alors que dans les cellules QT6, l'activité globale demeure très faible.

D'autre part la destruction de l'élément B par mutagenèse dirigée dans le promoteur ENS -456/+83 en diminue fortement l'activité dans les cellules ES de poulet (**Figure 42**). L'ensemble de ces résultats confirme que l'élément B est nécessaire et important pour la trans-activation du promoteur ENS dans les cellules ES de poulet.

e) L'élément B fixe un facteur de transcription appartenant à la famille CP2

L'analyse de l'élément B montre que sa séquence est similaire à celle décrite précédemment pour fixer le facteur de transcription CP2 (Lim et *al.* 1993). Cet élément de réponse, dont le consensus est 5'-CNRG-N<sub>6</sub>-CNRG-3', est lui aussi composé d'une répétition directe de deux éléments, chacun nécessaire pour la fixation du facteur CP2.

Afin de tester si le complexe protéique fixé sur l'élément B comportait la protéine CP2, nous avons réalisé une expérience de retard sur gel en présence d'un antisérum de lapin dirigé contre la protéine CP2 humaine (Yoon et *al.* 1994). Lorsque l'antisérum était ajouté au mélange réactionnel comportant les extraits nucléaires puis la sonde marquée, la fixation du complexe sur l'élément B était complètement abolie (**Figure 43**). Au contraire, l'ajout d'un sérum préimmun de lapin ne perturbait pas la fixation du complexe. D'autre part, afin de vérifier la spécificité de l'antisérum CP2, nous avons également montré que l'ajout de l'antisérum CP2 ne perturbait pas la formation du complexe ADN-protéine se fixant sur l'élément A. L'ensemble de ces résultats suggère fortement que le complexe protéique se fixant sur l'élément B comporte un facteur appartenant à la famille de facteurs de transcription CP2.

f) Clonage et profil d'expression du facteur CP2 dans les cellules ES de poulet

La protéine CP2 de poulet a été décrite précédemment comme un régulateur du promoteur de l' $\alpha$ A-crystalline du poulet (Murata et *al.* 1998). En s'appuyant sur les séquences publiées dans cet article, malheureusement indisponibles sur différentes bases de données (dont Genbank), nous avons amplifié par PCR puis cloné le gène CP2 de poulet à partir d'ARN totaux extraits de cellules ES de poulet. La séquence codante est très similaire aux séquences humaine (Yoon et *al.* 1994) et murine (Lim et *al.* 1992) et identique à la séquence de poulet précédemment décrite (Murata et *al.* 1998), avec cependant une petite délétion de 10 acides aminés (de la position 323 à la position 332), située dans un domaine d'interaction avec d'autres protéines (Shirra et *al.* 1998). Nous avons par la suite quantifié par PCR en temps réel le niveau d'expression du gène *CP2* dans différents types cellulaires, en le normalisant par rapport au niveau d'expression du gène *GAPDH*. Les résultats montrent que le niveau d'expression du gène CP2 est semblable entre les cellules ES de poulet, induites ou non à se différencier, et les cellules QT6. La protéine CP2 est également exprimée dans ces cellules, comme le montre des immunoblots réalisés avec l'antisérum CP2 ou un anticorps monoclonal commercial (BD Biosciences Pharmingen). L'ensemble de ces résultats s'accorde avec ceux de l'étude fonctionnelle de la région B, qui était active dans ces différents types cellulaires. Ils confirment également qu'un facteur de transcription de la famille CP2 est nécessaire pour l'expression des gènes *cENS*.

Finalement, si le facteur de transcription CP2 est nécessaire à l'activité du promoteur ENS dans les cellules ES de poulet, son activité semble réprimée dans le contexte de ce promoteur dans d'autres types cellulaires, que ce soit les cellules ES différenciées ou les cellules QT6. Nos résultats suggèrent également qu'une (ou plusieurs) région localisée dans le promoteur ENS -179/+83 serait impliquée dans la régulation transcriptionnelle spécifique des gènes *ENS* dans les cellules ES de poulet indifférenciées.

3°) MANUSCRIPT

Cf annexes

#### 4°) DISCUSSIONS ET COMMENTAIRES SUR LES RESULTATS PRESENTES

##### a) Identification d'une région activatrice

L'analyse du promoteur ENS par des délétions successives en 5' nous a conduit à identifier une région importante pour l'activation du promoteur, située entre les positions -313 et -180. Les premiers résultats présentés sur la figure 34A pouvait suggérer qu'un activateur, présent spécifiquement dans les cellules ES de poulet, se fixait sur la région -308/-179. L'analyse fonctionnelle de cette région a cependant montré qu'elle était aussi bien active dans les cellules ES de poulet que dans les cellules QT6, lorsqu'elle était placée en amont d'un promoteur ubiquitaire. Par contre, en amont du promoteur ENS -179/+83, l'activité de cette région était freinée dans les cellules QT6 où le promoteur ENS est également très faiblement actif. Ces résultats suggèrent que l'activité de la région -313/-180 est régulée par la région -179/+83. Plusieurs hypothèses peuvent être avancées pour expliquer l'activité enhancer de cette région :

- La région -313/-180 est constitutivement active dans de nombreux types cellulaires quand elle est placée en amont d'un promoteur ubiquitaire comme le promoteur basal SV40
- Il est possible que l'activité de la région -313/-180 soit bloquée dans les cellules QT6 car le promoteur basal ENS n'est pas fonctionnel ou peu fonctionnel dans ces cellules. Plusieurs travaux récents montrent en effet l'importance du promoteur basal dans le recrutement des facteurs de transcription assurant le contrôle de la tissu-spécificité (Smale 2001, Conkright et *al.* 2003).
- Il existe, dans le promoteur ENS -179/+83, des sites de fixation pour des facteurs réprimant la transcription dans les cellules QT6. Ces répresseurs pourraient recruter des partenaires inhibant l'activité transcriptionnelle, comme les histones déacétylases.

##### b) L'élément B est nécessaire à l'activation du promoteur ENS

L'élément B a été caractérisé par des expériences de retard sur gel et d'interférence de méthylation. Il se compose d'une répétition directe séparée par une séquence de 5 nucléotides. La séquence optimale décrite ici est 5'- CAAGTNNNNNCAAGT-3' où les caractères gras représentent les résidus indispensables pour l'interaction. L'intégrité de chaque élément de la répétition est indispensable à la formation du complexe ADN-protéine. Cette caractéristique



suggère que l'élément B pourrait être fixé par des homo- ou des hétérodimères protéiques, comme c'est le cas de nombreux éléments de réponse composés de répétitions directes fixant notamment le récepteur aux rétinoïdes X, RXR (Sacchetti et *al.* 2002) ou le facteur HNF4 (Stroup et Chiang 2000). D'autre part, l'élément B, d'une longueur de 10 bases (si l'on exclue la dernière thymidine) correspond à l'intervalle entre deux tours d'hélice d'ADN. Le rapprochement des éléments de la répétition réduit l'interaction mais ne l'abolit pas complètement, comme son éloignement. Celui-ci affecte soit l'accessibilité de l'un de ces éléments par des facteurs nucléaires, soit la formation du complexe protéique car l'éloignement des monomères ne permet plus leur dimérisation.

La fonctionnalité de l'élément B a été testée dans les cellules CES et dans les cellules QT6. Dans ces deux types cellulaires, l'élément B est capable d'activer le promoteur basal SV40. Par contre, lorsqu'il est placé en amont du promoteur ENS -179/+83, une augmentation de l'activité du promoteur n'est décelable que dans les cellules CES, les activités détectées étant trop faibles pour être significatives dans les cellules QT6. Ces résultats s'accordent avec ceux obtenus pour la région -313/-180 et les mêmes arguments peuvent être avancés pour en expliquer les particularités.

- c) Un facteur de transcription appartenant à la famille de facteurs de transcription CP2 fixe l'élément B

La forte similarité entre la séquence de l'élément B et celle précédemment décrite pour le facteur de transcription CP2 nous a amené à vérifier si ce facteur appartenait au complexe protéique se fixant sur l'élément B. Pour cela, nous avons demandé au laboratoire de Robert Roeder à New York de nous envoyer un antisérum dirigé contre la partie N-terminale de la protéine humaine LBP-1c (identique à CP2) (Yoon et *al.* 1994). La très forte conservation de ce gène chez l'homme, la souris et le poulet rendait possible le fait que l'anticorps reconnaisse les protéines des différentes espèces animales, ce qui est en effet le cas (données non présentées). L'incubation de l'anticorps en présence d'extraits nucléaires abolit complètement l'interaction du complexe protéique avec la sonde comportant l'élément B, démontrant ainsi la présence de CP2 dans ce complexe. On aurait pu s'attendre à un retard de migration du complexe reconnu par l'anticorps mais la disparition du complexe a également été décrite par Yoon et *al.* (1994), qui indiquent que l'incubation des extraits nucléaires avec l'antisérum CP2, avant l'ajout de la sonde, abolit l'interaction alors que l'ajout de la sonde, avant l'antisérum CP2, s'accompagne d'une modification de la migration du complexe ADN-

protéine. Il est probable que certains des anticorps présents dans l'antisérum CP2 reconnaissent le domaine de fixation à l'ADN de la protéine CP2 et empêche ainsi la formation du complexe ADN-protéine. En effet Shirra et *al.* (1998) ont montré qu'une grande partie de la région N-terminale de la protéine CP2 était nécessaire pour sa fixation sur l'ADN (cf **Figure 28**).

CP2 a été décrit comme un facteur de transcription ubiquitaire activant ou réprimant un nombre important de promoteurs viraux et cellulaires (cf partie D). CP2 a également été décrit comme s'associant avec des partenaires lui conférant des fonctions particulières suivant le contexte cellulaire dans lequel il était exprimé. Ainsi, il a été montré que CP2 s'associait avec la protéine NF-E4 pour activer l'expression de la globine embryonnaire  $\gamma$  (Zhou et *al.* 2000). C'est l'association des deux protéines qui confère au promoteur sa spécificité spatiale et temporelle. Il a été également montré que CP2 pouvait s'associer avec le facteur YY1 pour réprimer l'activité des LTRs du virus HIV-1 (Romério et *al.* 1997) en recrutant des histones déacétylases comme HDAC-1 (Coull et *al.* 2000).

Comme on le voit ici, l'activité du facteur de transcription ubiquitaire CP2 peut être modulée en fonction des partenaires qui lui sont associés. Il en est peut-être de même pour le promoteur *ENS*. En effet, si l'élément B fixant CP2 est aussi actif dans les cellules CES et les cellules QT6 lorsqu'il est placé en amont d'un promoteur ubiquitaire, il semble réprimé dans les cellules QT6 en amont du promoteur *ENS* -179/+83. On peut avancer l'hypothèse qu'il existerait dans les cellules QT6 un partenaire de CP2 capable de réprimer l'activité du promoteur *ENS* dans ces cellules. L'activité de ce partenaire serait effective sur le promoteur *ENS* -179/+83 car celui-ci posséderait des sites de liaison pour ce facteur, comme le suggère la différence d'activité constatée entre le promoteur basal SV40 et le promoteur *ENS* dans les cellules QT6.

---

## PERSPECTIVES ET CONCLUSION GENERALE

---

### I MISE EN EVIDENCE D'UNE REGION REPRESSEUR SPECIFIQUE DES CELLULES DIFFERENCIEES

Comme on l'a vu précédemment, la tissu-spécificité du promoteur ENS n'est pas assurée par sa région activatrice mais semble due, au contraire, à la région localisée entre les positions -179 et +83. D'autres expériences que nous avons réalisées confirment cette observation. Si l'on enlève la région localisée entre les positions -238 et -32 dans ce promoteur complet ENS, on observe une augmentation de l'activité du promoteur ENS d'environ 10 fois dans les cellules QT6, alors qu'il n'y a aucune différence dans les cellules CES. Si l'on enlève la région -238/+83 au promoteur ENS on observe une augmentation forte dans les cellules QT6 (environ 10 fois) et légère dans les cellules CES (environ 3 fois) (**Figure 44**). Ce dernier résultat montre qu'il existe probablement un deuxième promoteur localisé dans la région -456/-238 puisque l'on observe une activité transcriptionnelle alors que le premier promoteur est absent. Ces résultats s'accordent aussi pour proposer qu'il existe dans la région -179/+83 un régulateur de l'activité du promoteur ENS dans les cellules différenciées. Nous sommes actuellement en train d'analyser cette région par mutagenèse et gels retards pour identifier les séquences impliquées dans cette répression.

Une deuxième approche, qu'il serait nécessaire de réaliser pour analyser ce mécanisme de répression, serait d'évaluer le niveau de méthylation de l'ADN et d'acétylation des histones sur les promoteurs ENS endogènes des cellules ES indifférenciées et différenciées. Corréler ces mesures avec la présence de répresseurs capables de recruter des histones déacétylases par exemple, permettrait d'apporter une cohérence forte à notre analyse.

### II ACTIVITE DU PROMOTEUR *ENS* DANS L'EMBRYON DE POULET

Si le promoteur ENS est particulièrement actif dans les cellules ES de poulet, du moins transitoirement, il est indispensable de valider son activité à long terme dans les cellules CES, mais aussi *in vivo* dans l'embryon de poulet. Pour cela j'ai réalisé plusieurs

constructions comportant la séquence codante pour la GFP ou Lac Z sous le contrôle du promoteur ENS, et un gène de sélection (néomycine ou hygromycine) sous le contrôle du promoteur SV40. J'avais réalisé auparavant des vecteurs bicistroniques mais les résultats avaient été très décevants car je n'avais obtenu que très peu de clones cellulaires après transfection et sélection de ces constructions dans les cellules CES, que ce soit avec la construction contrôle comportant le promoteur CMV ou celle avec le promoteur ENS. C'est pourquoi je suis revenu à un vecteur classique comportant deux promoteurs. L'obtention de clones de cellules CES exprimant de façon stable un gène rapporteur sous le contrôle du promoteur ENS nous permettra ensuite de réaliser des chimères d'injection dans l'embryon de poulet et d'observer ainsi le profil d'activité du promoteur ENS *in vivo*.

### **III ACTIVITE DU PROMOTEUR *ENS* DANS D'AUTRES ESPECES ANIMALES**

Il serait également particulièrement intéressant de savoir si le promoteur ENS est actif dans d'autres cellules animales et, si oui, avec la même tissu-spécificité. Les résultats de tests en transfection transitoire dans les cellules ES de souris suggèrent que le promoteur ENS serait également actif dans ces cellules. Dans cette optique, nous sommes actuellement en train de produire des lignées de souris transgéniques pour l'expression d'un gène rapporteur (GFP ou LacZ) placé sous le contrôle du promoteur ENS. Nous sommes aussi en train d'essayer de sélectionner des clones de cellules ES de souris exprimant un gène rapporteur sous le contrôle du promoteur *ENS*. Ces cellules pourront alors être injectées dans un embryon receveur pour établir ensuite des lignées de souris transgéniques. D'autre part, nous collaborons avec le laboratoire de Daniel Boujard, à Rennes, pour suivre l'activité du promoteur ENS dans l'embryon de Xénope.

### **IV CONCLUSIONS**

Au cours de ces quatre années passées au Laboratoire de Biologie Moléculaire et Cellulaire dans l'équipe de Jacques Samarut, j'ai consacré l'essentiel de mon temps à la caractérisation de la famille de gènes *ENS*, en produisant des résultats intéressants même si de nombreuses questions n'ont malheureusement pas encore trouvé de réponses.

Ce travail a permis d'identifier une nouvelle famille de gènes, dont la structure génomique est originale et dont le profil d'expression est d'un intérêt tout particulier puisqu'il est restreint aux cellules ES de poulet et aux premiers stades du développement embryonnaire du poulet. La découverte d'une région promoteur, en amont des séquences codantes pour ces gènes, et spécifiquement active dans les cellules ES de poulet a sans doute été une étape clé dans le déroulement de ce travail de thèse. L'essentiel des travaux expérimentaux a donc été par la suite axé sur les mécanismes régulant l'activité de ce promoteur. Nous avons ainsi identifié une région activatrice fixant un facteur de transcription apparenté à CP2 et nécessaire à l'activité du promoteur ENS. Nous avons également mis en évidence une région répresseur, active dans les cellules différenciées, dont la caractérisation n'est pas encore achevée. Cette région répresseur est probablement importante pour conférer au promoteur sa tissu-spécificité.

J'espère également que la connaissance prochaine de la séquence du génome du poulet apportera quelques réponses aux questions que je me suis posées au cours de ce travail. En voici quelques unes : le promoteur décrit ici est-il en amont de tous les gènes ENS ? Existe-t-il un rétrovirus endogène ENS ? Combien de loci contenant les séquences répétées caractéristiques des gènes *ENS* sont présents dans le génome du poulet ? Les ADNc que nous avons classés dans la catégorie artéfact, existent-ils dans le génome ? Si oui, il faudrait alors réanalyser avec beaucoup plus de circonspection leurs expressions dans les cellules ES de poulet.

De nombreuses autres questions s'enchaînent naturellement en écrivant ces lignes. Quelle est la fonction du gène *cENS-1/cERNI* ? Comment relier l'activation *in vivo* du gène *cENS-1/cERNI* par le nœud de Hensen ou le FGF-8 aux résultats exposés ici ?...

Il est heureux de constater que ce sujet reste ouvert et digne d'intérêt. J'espère que notre équipe apportera des réponses à quelques-unes de ces interrogations dans les années à venir.

## RÉFÉRENCES

- **Acloque, H., Risson, V., Birot, A. M., Kunita, R., Pain, B. and Samarut, J.** (2001). Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo. *Mech Dev* **103**, 79-91.
- **Allioli, N.** (1995). Approche de transfert de gènes, chez le poulet, par manipulation in vitro des cellules germinales primordiales. Utilisation de vecteurs rétroviraux dérivés des virus ALSV., (ed. Lyon: Université Claude Bernard Lyon I.
- **Aristote.** Histoire des animaux. Paris: Edition Gallimard.
- **Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. and Lonai, P.** (1998). Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc Natl Acad Sci U S A* **95**, 5082-7.
- **Astrin, S. M., Buss, E. G. and Haywards, W. S.** (1979). Endogenous viral genes are non-essential in the chicken. *Nature* **282**, 339-41.
- **Astrin, S. M., Crittenden, L. B. and Buss, E. G.** (1980). Ev 2, a genetic locus containing structural genes for endogenous virus, codes for Rous-associated virus type 0 produced by line 72 chickens. *J Virol* **33**, 250-5.
- **Aubert, J., Dunstan, H., Chambers, I. and Smith, A.** (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* **20**, 1240-5.
- **Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R.** (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-40.
- **Baltimore, D.** (1992). Viral RNA-dependent DNA polymerase. 1970. *Biotechnology* **24**, 3-5.
- **Barbot, W., Dupressoir, A., Lazar, V. and Heidmann, T.** (2002). Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res* **30**, 2365-73.
- **Barnea, E. and Bergman, Y.** (2000). Synergy of SF1 and RAR in activation of Oct-3/4 promoter. *J Biol Chem* **275**, 6608-19.
- **Barnhart, K. M., Kim, C. G., Banerji, S. S. and Sheffery, M.** (1988). Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine alpha-globin gene. *Mol Cell Biol* **8**, 3215-26.
- **Basu, A., Dong, B., Krainer, A. R. and Howe, C. C.** (1997). The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. *Mol Cell Biol* **17**, 677-86.
- **Beck, F., Erler, T., Russell, A. and James, R.** (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn* **204**, 219-27.
- **Bellorini, M., Dantonel, J. C., Yoon, J. B., Roeder, R. G., Tora, L. and Mantovani, R.** (1996). The major histocompatibility complex class II Ea

promoter requires TFIID binding to an initiator sequence. *Mol Cell Biol* **16**, 503-12.

- **Ben-Shushan, E., Thompson, J. R., Gudas, L. J. and Bergman, Y.** (1998). Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Mol Cell Biol* **18**, 1866-78.
- **Bing, Z., Reddy, S. A., Ren, Y., Qin, J. and Liao, W. S.** (1999). Purification and characterization of the serum amyloid A3 enhancer factor. *J Biol Chem* **274**, 24649-56.
- **Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M. A., Jun, Z., Walter, J. and Jaenisch, R.** (2002). Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol Cell Biol* **22**, 2124-35.
- **Boeuf, H., Merienne, K., Jacquot, S., Duval, D., Zeniou, M., Hauss, C., Reinhardt, B., Huss-Garcia, Y., Dierich, A., Frank, D. A. et al.** (2001). The ribosomal S6 kinases, cAMP-responsive element-binding, and STAT3 proteins are regulated by different leukemia inhibitory factor signaling pathways in mouse embryonic stem cells. *J Biol Chem* **276**, 46204-11.
- **Boiani, M., Eckardt, S., Scholer, H. R. and McLaughlin, K. J.** (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev* **16**, 1209-19.
- **Bosselman, R. A., Hsu, R. Y., Boggs, T., Hu, S., Bruszewski, J., Ou, S., Kozar, L., Martin, F., Green, C., Jacobsen, F. et al.** (1989). Germline transmission of exogenous genes in the chicken. *Science* **243**, 533-5.
- **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**, 2073-90.
- **Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E.** (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-6.
- **Brinster, R. L.** (1965). Studies on the development of mouse embryos in vitro. IV. Interaction of energy sources. *J Reprod Fertil* **10**, 227-40.
- **Brown, J. P., Wei, W. and Sedivy, J. M.** (1997). Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. *Science* **277**, 831-4.
- **Bruni, P., Minopoli, G., Brancaccio, T., Napolitano, M., Faraonio, R., Zambrano, N., Hansen, U. and Russo, T.** (2002). Fe65, a ligand of the Alzheimer's beta-amyloid precursor protein, blocks cell cycle progression by down-regulating thymidylate synthase expression. *J Biol Chem* **277**, 35481-8.
- **Burdon, T., Smith, A. and Savatier, P.** (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432-8.
- **Burdon, T., Stracey, C., Chambers, I., Nichols, J. and Smith, A.** (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol* **210**, 30-43.
- **Carlton, M. B., Colledge, W. H. and Evans, M. J.** (1995). Generation of a pseudogene during retroviral infection. *Mamm Genome* **6**, 90-5.

- **Carsience, R. S., Clark, M. E., Verrinder Gibbins, A. M. and Etches, R. J.** (1993). Germline chimeric chickens from dispersed donor blastodermal cells and compromised recipient embryos. *Development* **117**, 669-75.
- **Casolaro, V., Keane-Myers, A. M., Swendeman, S. L., Steindler, C., Zhong, F., Sheffery, M., Georas, S. N. and Ono, S. J.** (2000). Identification and characterization of a critical CP2-binding element in the human interleukin-4 promoter. *J Biol Chem* **275**, 36605-11.
- **Challita, P. M. and Kohn, D. B.** (1994). Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. *Proc Natl Acad Sci U S A* **91**, 2567-71.
- **Challita, P. M., Skelton, D., el-Khoueiry, A., Yu, X. J., Weinberg, K. and Kohn, D. B.** (1995). Multiple modifications in cis elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. *J Virol* **69**, 748-55.
- **Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A.** (2003). Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55.
- **Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F. and Beck, F.** (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**, 84-7.
- **Chen, X., Oh, S. W., Zheng, Z., Chen, H. W., Shin, H. H. and Hou, S. X.** (2003). Cyclin D-Cdk4 and cyclin E-Cdk2 regulate the Jak/STAT signal transduction pathway in *Drosophila*. *Dev Cell* **4**, 179-90.
- **Cheng, A. M., Saxton, T. M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C. G., Cardiff, R. D., Cross, J. C., Muller, W. J. et al.** (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* **95**, 793-803.
- **Cherry, S. R., Biniszkiwicz, D., van Parijs, L., Baltimore, D. and Jaenisch, R.** (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. *Mol Cell Biol* **20**, 7419-26.
- **Christy, R. J. and Huang, R. C.** (1988). Functional analysis of the long terminal repeats of intracisternal A-particle genes: sequences within the U3 region determine both the efficiency and direction of promoter activity. *Mol Cell Biol* **8**, 1093-102.
- **Clarke, D. L., Johansson, C. B., Wilbertz, J., Veress, B., Nilsson, E., Karlstrom, H., Lendahl, U. and Frisen, J.** (2000). Generalized potential of adult neural stem cells. *Science* **288**, 1660-3.
- **Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P. N. and Lovell-Badge, R.** (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* **122**, 509-20.
- **Conkright, M. D., Guzman, E., Flechner, L., Su, A. I., Hogenesch, J. B. and Montminy, M.** (2003). Genome-Wide Analysis of CREB Target Genes Reveals A Core Promoter Requirement for cAMP Responsiveness. *Mol Cell* **11**, 1101-8.
- **Coull, J. J., He, G., Melander, C., Rucker, V. C., Dervan, P. B. and Margolis, D. M.** (2002). Targeted derepression of the human immunodeficiency virus type 1 long terminal repeat by pyrrole-imidazole polyamides. *J Virol* **76**, 12349-54.



- **Coull, J. J., Romerio, F., Sun, J. M., Volker, J. L., Galvin, K. M., Davie, J. R., Shi, Y., Hansen, U. and Margolis, D. M.** (2000). The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J Virol* **74**, 6790-9.
- **Cristofari, G.** (2002). Mécanismes moléculaires de la rétrotransposition des éléments Ty de *Saccharomyces cerevisiae*., (ed. Lyon: ENS Lyon).
- **Crittenden, L. B. and Salter, D. W.** (1990). Expression of retroviral genes in transgenic chickens. *J Reprod Fertil Suppl* **41**, 163-71.
- **Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J. et al.** (1998). Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev Biol* **203**, 149-62.
- **Dannenbergh, J. H., van Rossum, A., Schuijff, L. and te Riele, H.** (2000). Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev* **14**, 3051-64.
- **Dodge, J. E., Ramsahoye, B. H., Wo, Z. G., Okano, M. and Li, E.** (2002). De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289**, 41-8.
- **Dooley, T. P., Miranda, M., Jones, N. C. and DePamphilis, M. L.** (1989). Transactivation of the adenovirus E1a promoter in the absence of adenovirus E1A protein is restricted to mouse oocytes and preimplantation embryos. *Development* **107**, 945-56.
- **Driesch, H.** (1891). Entwicklungsmechanische Studien I. Der Wert der beiden ersten Furchungszellen in der Echinodermenentwicklung. Experimentelle Erzeugung von Teil- und Doppelbildungen. *Zeitschrift für wissenschaftliche Zoologie* **53**, 160-183.
- **Ducy, P. and Karsenty, G.** (1995). Two distinct osteoblast-specific cis-acting elements control expression of a mouse osteocalcin gene. *Mol Cell Biol* **15**, 1858-69.
- **Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. and Karsenty, G.** (1997). *Osf2/Cbfa1*: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747-54.
- **Dunwiddie, C. and Faras, A. J.** (1985). Presence of retrovirus reverse transcriptase-related gene sequences in avian cells lacking endogenous avian leukosis viruses. *Proc Natl Acad Sci U S A* **82**, 5097-101.
- **Dunwiddie, C. T., Resnick, R., Boyce-Jacino, M., Alegre, J. N. and Faras, A. J.** (1986). Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. *J Virol* **59**, 669-75.
- **Dupressoir, A. and Heidmann, T.** (1996). Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. *Mol Cell Biol* **16**, 4495-503.
- **Duval, D., Reinhardt, B., Keding, C. and Boeuf, H.** (2000). Role of suppressors of cytokine signaling (Socs) in leukemia inhibitory factor (LIF) - dependent embryonic stem cell survival. *Faseb J* **14**, 1577-84.
- **Ellermann, V. and Bang, O.** (1908). Experimentelle Leukämie bei Hühnern. *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. Orig.* **46**, 595-609.

- **Ernst, M., Oates, A. and Dunn, A. R.** (1996). Gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways. *J Biol Chem* **271**, 30136-43.
- **Evans, M. J.** (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J Embryol Exp Morphol* **28**, 163-76.
- **Evans, M. J. and Kaufman, M. H.** (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.
- **Eyal-Giladi, H., Ginsburg, M. and Farbarov, A.** (1981). Avian primordial germ cells are of epiblastic origin. *J Embryol Exp Morphol* **65**, 139-47.
- **Eyal-Giladi, H. and Kochav, S.** (1976). From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev Biol* **49**, 321-37.
- **Ezashi, T., Ghosh, D. and Roberts, R. M.** (2001). Repression of Ets-2-induced transactivation of the tau interferon promoter by Oct-4. *Mol Cell Biol* **21**, 7883-91.
- **F.H., W. J. D. C.** (1953). Molecular Structure of nucleic acids. A structure for desoxyribose nucleic acid. *Nature* **171**, 737-738.
- **Falzon, M. and Kuff, E. L.** (1988). Multiple protein-binding sites in an intracisternal A particle long terminal repeat. *J Virol* **62**, 4070-7.
- **Falzon, M. and Kuff, E. L.** (1989). Isolation and characterization of a protein fraction that binds to enhancer core sequences in intracisternal A-particle long terminal repeats. *J Biol Chem* **264**, 21915-22.
- **Falzon, M. and Kuff, E. L.** (1991). Binding of the transcription factor EBP-80 mediates the methylation response of an intracisternal A-particle long terminal repeat promoter. *Mol Cell Biol* **11**, 117-25.
- **Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. and Goldfarb, M.** (1995). Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246-9.
- **Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D. and Olson, E. N.** (1998). Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nat Genet* **18**, 266-70.
- **Flamant, F.** (2001). De l'oeuf à la poule. Paris: Belin.
- **Friedrich, G. and Soriano, P.** (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**, 1513-23.
- **Fuhrmann, G., Chung, A. C., Jackson, K. J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H. R. and Cooney, A. J.** (2001). Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Dev Cell* **1**, 377-87.
- **Fuhrmann, G., Sylvester, I. and Scholer, H. R.** (1999). Repression of Oct-4 during embryonic cell differentiation correlates with the appearance of TRIF, a transiently induced DNA-binding factor. *Cell Mol Biol (Noisy-le-grand)* **45**, 717-24.
- **Gael, C.**
- **Gardner, R. L.** (1968). Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature* **220**, 596-7.
- **Gearing, D. P., Thut, C. J., VandeBos, T., Gimpel, S. D., Delaney, P. B., King, J., Price, V., Cosman, D. and Beckmann, M. P.** (1991). Leukemia

inhibitory factor receptor is structurally related to the IL-6 signal transducer, gp130. *Embo J* **10**, 2839-48.

- **Gidekel, S. and Bergman, Y.** (2002). A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element. *J Biol Chem* **277**, 34521-30.
- **Grez, M., Akgun, E., Hilberg, F. and Ostertag, W.** (1990). Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proc Natl Acad Sci U S A* **87**, 9202-6.
- **Gross, L.** (1951). Spontaneous leukemia developing in C3H mice following inoculation in infancy, with Ak-leukemic extracts, or AK-embryos. *Proc. Soc. Exp. Biol. Med.* **76**, 27-32.
- **Hahnel, A. C. and Eddy, E. M.** (1983). Syngeneic antiserum to Nulli SCC1 embryonal carcinoma cells recognizing surface antigens of embryonic cells. *J Reprod Immunol* **5**, 371-82.
- **Hahnel, A. C. and Eddy, E. M.** (1987). The distribution of two cell surface determinants of mouse embryonal carcinoma and early embryonic cells. *J Reprod Immunol* **10**, 89-110.
- **Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J. and Surani, M. A.** (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23.
- **Hamburger, V. and Hamilton, H. L.** (1992). A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* **195**, 231-72.
- **Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S. and Labosky, P. A.** (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* **16**, 2650-61.
- **He, G. and Margolis, D. M.** (2002). Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. *Mol Cell Biol* **22**, 2965-73.
- **Heidmann, O. and Heidmann, T.** (1991). Retrotransposition of a mouse IAP sequence tagged with an indicator gene. *Cell* **64**, 159-70.
- **Hibi, M. and Hirano, T.** (2000). Gab-family adapter molecules in signal transduction of cytokine and growth factor receptors, and T and B cell antigen receptors. *Leuk Lymphoma* **37**, 299-307.
- **Hong, Y., Winkler, C. and Schartl, M.** (1998). Production of medakafish chimeras from a stable embryonic stem cell line. *Proc Natl Acad Sci U S A* **95**, 3679-84.
- **Huang, H. C., Sundseth, R. and Hansen, U.** (1990). Transcription factor LSF binds two variant bipartite sites within the SV40 late promoter. *Genes Dev* **4**, 287-98.
- **Humpherys, D., Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W. M., 3rd, Binischkiewicz, D., Yanagimachi, R. and Jaenisch, R.** (2001). Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95-7.
- **Hynes, R. O.** (1996). Targeted mutations in cell adhesion genes: what have we learned from them? *Dev Biol* **180**, 402-12.
- **Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A. and Lemischka, I. R.** (2002). A stem cell molecular signature. *Science* **298**, 601-4.

- **Ivics, Z., Hackett, P. B., Plasterk, R. H. and Izsvak, Z.** (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501-10.
- **Jaenisch, R., Fan, H. and Croker, B.** (1975). Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. *Proc Natl Acad Sci U S A* **72**, 4008-12.
- **Jahner, D., Stuhlmann, H., Stewart, C. L., Harbers, K., Lohler, J., Simon, I. and Jaenisch, R.** (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623-8.
- **Jane, S. M., Nienhuis, A. W. and Cunningham, J. M.** (1995). Hemoglobin switching in man and chicken is mediated by a heteromeric complex between the ubiquitous transcription factor CP2 and a developmentally specific protein. *Embo J* **14**, 97-105.
- **Jones-Villeneuve, E. M., McBurney, M. W., Rogers, K. A. and Kalnins, V. I.** (1982). Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol* **94**, 253-62.
- **Kagami, H., Tagami, T., Matsubara, Y., Harumi, T., Hanada, H., Maruyama, K., Sakurai, M., Kuwana, T. and Naito, M.** (1997). The developmental origin of primordial germ cells and the transmission of the donor-derived gametes in mixed-sex germline chimeras to the offspring in the chicken. *Mol Reprod Dev* **48**, 501-10.
- **Kahan, B. W. and Ephrussi, B.** (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J Natl Cancer Inst* **44**, 1015-36.
- **Kang, J. S., Lee, H. B., Rhee, S. G., Park, K. and Yoo, O. J.** (1997). The 5'-upstream region of the rat phospholipase C-beta 3 gene contains two critical Sp1 sites and an HIV Inr-like element. *Gene* **197**, 19-28.
- **Karagenc, L., Cinnamon, Y., Ginsburg, M. and Petite, J. N.** (1996). Origin of primordial germ cells in the prestreak chick embryo. *Dev Genet* **19**, 290-301.
- **Kim, C. H., Heath, C., Bertuch, A. and Hansen, U.** (1987). Specific stimulation of simian virus 40 late transcription in vitro by a cellular factor binding the simian virus 40 21-base-pair repeat promoter element. *Proc Natl Acad Sci U S A* **84**, 6025-9.
- **Kraft, H. J., Mosselman, S., Smits, H. A., Hohenstein, P., Piek, E., Chen, Q., Artzt, K. and van Zoelen, E. J.** (1996). Oct-4 regulates alternative platelet-derived growth factor alpha receptor gene promoter in human embryonal carcinoma cells. *J Biol Chem* **271**, 12873-8.
- **Kuff, E. L. and Lueders, K. K.** (1988). The intracisternal A-particle gene family: structure and functional aspects. *Adv Cancer Res* **51**, 183-276.
- **Kunath, T., Strumpf, D., Rossant, J. and Tanaka, S.** (2001). Trophoblast stem cells. In *Stem Cell Biology*, (ed. D. R. Marshak R. L. Gardner and D. Gottlieb), pp. 267-287. New York: Cold Spring Harbor Laboratory Press.
- **LaBarge, M. A. and Blau, H. M.** (2002). Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. *Cell* **111**, 589-601.
- **Labosky, P. A. and Kaestner, K. H.** (1998). The winged helix transcription factor Hfh2 is expressed in neural crest and spinal cord during mouse development. *Mech Dev* **76**, 185-90.

- **Lamb, B. T., Satyamoorthy, K., Solter, D., Basu, A., Xu, M. Q., Weinmann, R. and Howe, C. C.** (1992). A DNA element that regulates expression of an endogenous retrovirus during F9 cell differentiation is E1A dependent. *Mol Cell Biol* **12**, 4824-33.
- **Le Douarin, N.** (2000). Des chimères, des clones et des gènes. Paris: Editions Odile Jacob.
- **Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R. and Li, E.** (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195-205.
- **Li, E., Bestor, T. H. and Jaenisch, R.** (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26.
- **Li, M., Sendtner, M. and Smith, A.** (1995). Essential function of LIF receptor in motor neurons. *Nature* **378**, 724-7.
- **Lim, L. C., Fang, L., Swendeman, S. L. and Sheffery, M.** (1993). Characterization of the molecularly cloned murine alpha-globin transcription factor CP2. *J Biol Chem* **268**, 18008-17.
- **Lim, L. C., Swendeman, S. L. and Sheffery, M.** (1992). Molecular cloning of the alpha-globin transcription factor CP2. *Mol Cell Biol* **12**, 828-35.
- **Linial, M. L. and Miller, A. D.** (1990). Retroviral RNA packaging: sequence requirements and implications. *Curr Top Microbiol Immunol* **157**, 125-52.
- **Liu, L. and Roberts, R. M.** (1996). Silencing of the gene for the beta subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. *J Biol Chem* **271**, 16683-9.
- **Locascio, A., Vega, S., de Frutos, C. A., Manzanares, M. and Nieto, M. A.** (2002). Biological potential of a functional human SNAIL retrogene. *J Biol Chem* **277**, 38803-9.
- **Lu, C. C., Brennan, J. and Robertson, E. J.** (2001). From fertilization to gastrulation: axis formation in the mouse embryo. *Curr Opin Genet Dev* **11**, 384-92.
- **Luedeking-Zimmer, E., DeKosky, S. T., Nebes, R. and Kamboh, M. I.** (2003). Association of the 3' UTR transcription factor LBP-1c/CP2/LSF polymorphism with late-onset Alzheimer's disease. *Am J Med Genet* **117B**, 114-7.
- **Lueders, K. K., Fewell, J. W., Morozov, V. E. and Kuff, E. L.** (1993). Selective expression of intracisternal A-particle genes in established mouse plasmacytomas. *Mol Cell Biol* **13**, 7439-46.
- **Lueders, K. K. and Kuff, E. L.** (1977). Sequences associated with intracisternal A particles are reiterated in the mouse genome. *Cell* **12**, 963-72.
- **Lum, R. and Linial, M. L.** (1998). Retrotransposition of nonviral RNAs in an avian packaging cell line. *J Virol* **72**, 4057-64.
- **Lutticken, C., Wegenka, U. M., Yuan, J., Buschmann, J., Schindler, C., Ziemiecki, A., Harpur, A. G., Wilks, A. F., Yasukawa, K., Taga, T. et al.** (1994). Association of transcription factor APRF and protein kinase Jak1 with the interleukin-6 signal transducer gp130. *Science* **263**, 89-92.
- **Lyko, F., Whittaker, A. J., Orr-Weaver, T. L. and Jaenisch, R.** (2000). The putative Drosophila methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. *Mech Dev* **95**, 215-7.

- **Markert, C. L. and Petters, R. M.** (1978). Manufactured hexaparental mice show that adults are derived from three embryonic cells. *Science* **202**, 56-8.
- **Martin, G. R.** (1980). Teratocarcinomas and mammalian embryogenesis. *Science* **209**, 768-76.
- **Martin, G. R.** (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8.
- **Martin, G. R. and Evans, M. J.** (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* **72**, 1441-5.
- **Marzullo, G.** (1970). Production of chick chimaeras. *Nature* **225**, 72-3.
- **Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T. and Yokota, T.** (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo J* **18**, 4261-9.
- **Matsui, Y., Zsebo, K. and Hogan, B. L.** (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-7.
- **Matsuo, I., Kuratani, S., Kimura, C., Takeda, N. and Aizawa, S.** (1995). Mouse Otx2 functions in the formation and patterning of rostral head. *Genes Dev* **9**, 2646-58.
- **Mayr, E.** (1982). The Growth of biological thought. Diversity, Evolution and Inheritance. New Haven: Harvard University Press.
- **Meno, C., Shimono, A., Saijoh, Y., Yashiro, K., Mochida, K., Ohishi, S., Noji, S., Kondoh, H. and Hamada, H.** (1998). lefty-1 is required for left-right determination as a regulator of lefty-2 and nodal. *Cell* **94**, 287-97.
- **Metcalf, D.** (2003). The unsolved enigmas of leukemia inhibitory factor. *Stem Cells* **21**, 5-14.
- **Mighell, A. J., Smith, N. R., Robinson, P. A. and Markham, A. F.** (2000). Vertebrate pseudogenes. *FEBS Lett* **468**, 109-14.
- **Minopoli, G., de Candia, P., Bonetti, A., Faraonio, R., Zambrano, N. and Russo, T.** (2001). The beta-amyloid precursor protein functions as a cytosolic anchoring site that prevents Fe65 nuclear translocation. *J Biol Chem* **276**, 6545-50.
- **Mintz, B.** (1962). Formation of genotypically mosaic mouse embryos. *American Zoologist* **2**, 432.
- **Mintz, B. and Illmensee, K.** (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc Natl Acad Sci U S A* **72**, 3585-9.
- **Minucci, S., Botquin, V., Yeom, Y. I., Dey, A., Sylvester, I., Zand, D. J., Ohbo, K., Ozato, K. and Scholer, H. R.** (1996). Retinoic acid-mediated down-regulation of Oct3/4 coincides with the loss of promoter occupancy in vivo. *Embo J* **15**, 888-99.
- **Mitrani, E., Coffin, J., Boedtker, H. and Doty, P.** (1987). Rous sarcoma virus is integrated but not expressed in chicken early embryonic cells. *Proc Natl Acad Sci U S A* **84**, 2781-4.
- **Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S.** (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**, 631-42.

- **Moore, F. L., Jaruzelska, J., Fox, M. S., Urano, J., Firpo, M. T., Turek, P. J., Dorfman, D. M. and Pera, R. A.** (2003). Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc Natl Acad Sci U S A* **100**, 538-43.
- **Murata, T., Nitta, M. and Yasuda, K.** (1998). Transcription factor CP2 is essential for lens-specific expression of the chicken alphaA-crystallin gene. *Genes Cells* **3**, 443-57.
- **Narducci, M. G., Fiorenza, M. T., Kang, S. M., Bevilacqua, A., Di Giacomo, M., Remotti, D., Picchio, M. C., Fidanza, V., Cooper, M. D., Croce, C. M. et al.** (2002). TCL1 participates in early embryonic development and is overexpressed in human seminomas. *Proc Natl Acad Sci U S A* **99**, 11712-7.
- **Nichols, J., Chambers, I., Taga, T. and Smith, A.** (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**, 2333-9.
- **Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A.** (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91.
- **Nikiforov, M. A. and Gudkov, A. V.** (1994). ART-CH: a VL30 in chickens? *J Virol* **68**, 846-53.
- **Nishimoto, M., Fukushima, A., Okuda, A. and Muramatsu, M.** (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* **19**, 5453-65.
- **Niwa, H.** (2001). Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct Funct* **26**, 137-48.
- **Niwa, H., Burdon, T., Chambers, I. and Smith, A.** (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-60.
- **Niwa, H., Masui, S., Chambers, I., Smith, A. G. and Miyazaki, J.** (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol Cell Biol* **22**, 1526-36.
- **Niwa, H., Miyazaki, J. and Smith, A. G.** (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-6.
- **Oka, M., Tagoku, K., Russell, T. L., Nakano, Y., Hamazaki, T., Meyer, E. M., Yokota, T. and Terada, N.** (2002). CD9 is associated with leukemia inhibitory factor-mediated maintenance of embryonic stem cells. *Mol Biol Cell* **13**, 1274-81.
- **Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H.** (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-72.
- **Okano, M., Bell, D. W., Haber, D. A. and Li, E.** (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57.
- **Okano, M., Xie, S. and Li, E.** (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* **26**, 2536-40.

- **Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Boon, K., Keaveney, M., Stunnenberg, H. G. et al.** (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *Embo J* **17**, 2019-32.
- **Osborne, C. S., Pasceri, P., Singal, R., Sukonnik, T., Ginder, G. D. and Ellis, J.** (1999). Amelioration of retroviral vector silencing in locus control region beta-globin-transgenic mice and transduced F9 embryonic cells. *J Virol* **73**, 5490-6.
- **Pain, B., Clark, M. E., Shen, M., Nakazawa, H., Sakurai, M., Samarut, J. and Etches, R. J.** (1996). Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. *Development* **122**, 2339-48.
- **Palmieri, S. L., Peter, W., Hess, H. and Scholer, H. R.** (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* **166**, 259-67.
- **Pannell, D. and Ellis, J.** (2001). Silencing of gene expression: implications for design of retrovirus vectors. *Rev Med Virol* **11**, 205-17.
- **Pannell, D., Osborne, C. S., Yao, S., Sukonnik, T., Pasceri, P., Karaiskakis, A., Okano, M., Li, E., Lipshitz, H. D. and Ellis, J.** (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *Embo J* **19**, 5884-94.
- **Papaioannou, V. E., McBurney, M. W., Gardner, R. L. and Evans, M. J.** (1975). Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70-73.
- **Park, T. S. and Han, J. Y.** (2000). Derivation and characterization of pluripotent embryonic germ cells in chicken. *Mol Reprod Dev* **56**, 475-82.
- **Petitte, J. N., Clark, M. E., Liu, G., Verrinder Gibbins, A. M. and Etches, R. J.** (1990). Production of somatic and germline chimeras in the chicken by transfer of early blastodermal cells. *Development* **108**, 185-9.
- **Powell, C. M., Rudge, T. L., Zhu, Q., Johnson, L. F. and Hansen, U.** (2000). Inhibition of the mammalian transcription factor LSF induces S-phase-dependent apoptosis by downregulating thymidylate synthase expression. *Embo J* **19**, 4665-75.
- **Poznanski, A. A. and Calarco, P. G.** (1991). The expression of intracisternal A particle genes in the preimplantation mouse embryo. *Dev Biol* **143**, 271-81.
- **Pu, W. T., Wickman, K. and Clapham, D. E.** (2000). ICln is essential for cellular and early embryonic viability. *J Biol Chem* **275**, 12363-6.
- **Qu, C. K. and Feng, G. S.** (1998). Shp-2 has a positive regulatory role in ES cell differentiation and proliferation. *Oncogene* **17**, 433-9.
- **Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. and Melton, D. A.** (2002). "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597-600.
- **Ramamurthy, L., Barbour, V., Tuckfield, A., Clouston, D. R., Topham, D., Cunningham, J. M. and Jane, S. M.** (2001). Targeted disruption of the CP2 gene, a member of the NTF family of transcription factors. *J Biol Chem* **276**, 7836-42.



- **Rappolee, D. A., Basilico, C., Patel, Y. and Werb, Z.** (1994). Expression and function of FGF-4 in peri-implantation development in mouse embryos. *Development* **120**, 2259-69.
- **Reddy, S. T., Stoker, A. W. and Bissell, M. J.** (1991). Expression of Rous sarcoma virus-derived retroviral vectors in the avian blastoderm: potential as stable genetic markers. *Proc Natl Acad Sci U S A* **88**, 10505-9.
- **Reik, W., Dean, W. and Walter, J.** (2001). Epigenetic reprogramming in mammalian development. *Science* **293**, 1089-93.
- **Resnick, J. L., Bixler, L. S., Cheng, L. and Donovan, P. J.** (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550-1.
- **Riley, P., Anson-Cartwright, L. and Cross, J. C.** (1998). The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet* **18**, 271-5.
- **Rodda, S., Sharma, S., Scherer, M., Chapman, G. and Rathjen, P.** (2001). CRTR-1, a developmentally regulated transcriptional repressor related to the CP2 family of transcription factors. *J Biol Chem* **276**, 3324-32.
- **Rodda, S., Sharma, S., Scherer, M., Chapman, G. and Rathjen, P.** (2001). CRTR-1, a developmentally regulated transcriptional repressor related to the CP2 family of transcription factors. *J Biol Chem* **276**, 3324-32.
- **Roeser, T., Stein, S. and Kessel, M.** (1999). Nuclear beta-catenin and the development of bilateral symmetry in normal and LiCl-exposed chick embryos. *Development* **126**, 2955-65.
- **Rogers, M. B., Hosler, B. A. and Gudas, L. J.** (1991). Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. *Development* **113**, 815-24.
- **Romerio, F., Gabriel, M. N. and Margolis, D. M.** (1997). Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. *J Virol* **71**, 9375-82.
- **Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. and Staudt, L. M.** (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-92.
- **Rossant, J. and Cross, J. C.** (2001). Placental development: lessons from mouse mutants. *Nat Rev Genet* **2**, 538-48.
- **Rous, P.** (1911). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.* **13**, 397-411.
- **Sacchetti, P., Dwornik, H., Formstecher, P., Rachez, C. and Lefebvre, P.** (2002). Requirements for heterodimerization between the orphan nuclear receptor Nurr1 and retinoid X receptors. *J Biol Chem* **277**, 35088-96.
- **Sage, J., Mulligan, G. J., Attardi, L. D., Miller, A., Chen, S., Williams, B., Theodorou, E. and Jacks, T.** (2000). Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev* **14**, 3037-50.
- **Saijoh, Y., Fujii, H., Meno, C., Sato, M., Hirota, Y., Nagamatsu, S., Ikeda, M. and Hamada, H.** (1996). Identification of putative downstream genes of Oct-3, a pluripotent cell-specific transcription factor. *Genes Cells* **1**, 239-52.
- **Salter, D. W., Smith, E. J., Hughes, S. H., Wright, S. E., Fadly, A. M., Witter, R. L. and Crittenden, L. B.** (1986). Gene insertion into the chicken germ line by retroviruses. *Poult Sci* **65**, 1445-58.

- **Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M. and Nakano, T.** (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech Dev* **113**, 91-4.
- **Satyamoorthy, K., Park, K., Atchison, M. L. and Howe, C. C.** (1993). The intracisternal A-particle upstream element interacts with transcription factor YY1 to activate transcription: pleiotropic effects of YY1 on distinct DNA promoter elements. *Mol Cell Biol* **13**, 6621-8.
- **Savatier, P., Huang, S., Szekely, L., Wiman, K. G. and Samarut, J.** (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9**, 809-18.
- **Savatier, P., Lapillonne, H., van Grunsvan, L. A., Rudkin, B. B. and Samarut, J.** (1996). Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309-22.
- **Savatier, P., Morgenstern, J. and Beddington, R. S.** (1990). Permissiveness to murine leukemia, virus expression during preimplantation and early postimplantation mouse development. *Development* **109**, 655-65.
- **Schinke, T. and Karsenty, G.** (1999). Characterization of Osf1, an osteoblast-specific transcription factor binding to a critical cis-acting element in the mouse Osteocalcin promoters. *J Biol Chem* **274**, 30182-9.
- **Scholer, H. R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P.** (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-9.
- **Schoorlemmer, J., van Puijenbroek, A., van Den Eijnden, M., Jonk, L., Pals, C. and Kruijer, W.** (1994). Characterization of a negative retinoic acid response element in the murine Oct4 promoter. *Mol Cell Biol* **14**, 1122-36.
- **Seale, P., Asakura, A. and Rudnicki, M. A.** (2001). The potential of muscle stem cells. *Dev Cell* **1**, 333-42.
- **Seidel, F.** (1952). Die Entwicklungspotenzen einer isolierten blastomere der Zweizellenstadiums in Saugetier Ei. *Naturwissenschaften* **39**.
- **Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R. and Leib-Mosch, C.** (1998). Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J Virol* **72**, 8384-91.
- **Sherr, C. J. and Roberts, J. M.** (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13**, 1501-12.
- **Shirra, M. K. and Hansen, U.** (1998). LSF and NTF-1 share a conserved DNA recognition motif yet require different oligomerization states to form a stable protein-DNA complex. *J Biol Chem* **273**, 19260-8.
- **Shirra, M. K., Zhu, Q., Huang, H. C., Pallas, D. and Hansen, U.** (1994). One exon of the human LSF gene includes conserved regions involved in novel DNA-binding and dimerization motifs. *Mol Cell Biol* **14**, 5076-87.
- **Smale, S. T.** (2001). Core promoters: active contributors to combinatorial gene regulation. *Genes Dev* **15**, 2503-8.
- **Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D.** (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90.

- **Smith, T. A. and Hooper, M. L.** (1983). Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp Cell Res* **145**, 458-62.
- **Solis, C., Aizencang, G. I., Astrin, K. H., Bishop, D. F. and Desnick, R. J.** (2001). Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. *J Clin Invest* **107**, 753-62.
- **Soodeen-Karamath, S. and Gibbins, A. M.** (2001). Apparent absence of oct 3/4 from the chicken genome. *Mol Reprod Dev* **58**, 137-48.
- **Soriano, P., Friedrich, G. and Lawinger, P.** (1991). Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J Virol* **65**, 2314-9.
- **Spemann, H.** (1938). Embryonic Development and induction. New Haven: Yale University Press.
- **Stahl, N., Boulton, T. G., Farruggella, T., Ip, N. Y., Davis, S., Witthuhn, B. A., Quelle, F. W., Silvennoinen, O., Barbieri, G., Pellegrini, S. et al.** (1994). Association and activation of Jak-Tyk kinases by CNTF-LIF-OSM-IL-6 beta receptor components. *Science* **263**, 92-5.
- **Stahl, N., Farruggella, T. J., Boulton, T. G., Zhong, Z., Darnell, J. E., Jr. and Yancopoulos, G. D.** (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349-53.
- **Stevens, L. C.** (1964). Experimental production of testicular teratomas in mice. *Proc Natl Acad Sci U S A* **52**, 654-661.
- **Stevens, L. C. and C.C., L.** (1954). Spontaneous testicular teratomas in an inbred strain of mice. *Proc Natl Acad Sci U S A* **40**, 1080-1087.
- **Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Kontgen, F. and Abbondanzo, S. J.** (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-9.
- **Stewart, C. L., Stuhlmann, H., Jahner, D. and Jaenisch, R.** (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proc Natl Acad Sci U S A* **79**, 4098-102.
- **Stewart, T. A. and Mintz, B.** (1981). Successive generations of mice produced from an established culture line of euploid teratocarcinoma cells. *Proc Natl Acad Sci U S A* **78**, 6314-8.
- **Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. and Stern, C. D.** (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**, 74-8.
- **Strickland, S., Smith, K. K. and Marotti, K. R.** (1980). Hormonal induction of differentiation in teratocarcinoma stem cells: generation of parietal endoderm by retinoic acid and dibutyryl cAMP. *Cell* **21**, 347-55.
- **Stroup, D. and Chiang, J. Y.** (2000). HNF4 and COUP-TFII interact to modulate transcription of the cholesterol 7alpha-hydroxylase gene (CYP7A1). *J Lipid Res* **41**, 1-11.
- **Suemori, H., Hashimoto, S. and Nakatsuji, N.** (1988). Presence of the adenovirus E1A-like activity in preimplantation stage mouse embryos. *Mol Cell Biol* **8**, 3553-5.
- **Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D. A., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Klemsz, M. et al.** (1996).

Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J Biol Chem* **271**, 23126-33.

- **Swendeman, S. L., Spielholz, C., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. and Sheffery, M.** (1994). Characterization of the genomic structure, chromosomal location, promoter, and development expression of the alpha-globin transcription factor CP2. *J Biol Chem* **269**, 11663-71.
- **Symes, A., Stahl, N., Reeves, S. A., Farruggella, T., Servidei, T., Gearan, T., Yancopoulos, G. and Fink, J. S.** (1997). The protein tyrosine phosphatase SHP-2 negatively regulates ciliary neurotrophic factor induction of gene expression. *Curr Biol* **7**, 697-700.
- **Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T. and Akira, S.** (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc Natl Acad Sci U S A* **94**, 3801-4.
- **Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. and Rossant, J.** (1998). Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072-5.
- **Tarkowski, A. K.** (1961). Mouse chimeras developed from fused eggs. *Nature* **190**, 857-860.
- **Tarkowski, A. K.** (1963). Studies on mouse chimeras developed from eggs fused in vitro. *National Cancer Institute Monography* **11**, 51-71.
- **Thoraval, P., Afanassieff, M., Cosset, F. L., Lasserre, F., Verdier, G., Coudert, F. and Dambrine, G.** (1995). Germline transmission of exogenous genes in chickens using helper-free ecotropic avian leukosis virus-based vectors. *Transgenic Res* **4**, 369-77.
- **Tokuzawa, Y., Kaiho, E., Maruyama, M., Takahashi, K., Mitsui, K., Maeda, M., Niwa, H. and Yamanaka, S.** (2003). Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol Cell Biol* **23**, 2699-708.
- **Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M. and Okuda, A.** (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* **30**, 3202-13.
- **Tuckfield, A., Clouston, D. R., Wilanowski, T. M., Zhao, L. L., Cunningham, J. M. and Jane, S. M.** (2002). Binding of the RING polycomb proteins to specific target genes in complex with the grainyhead-like family of developmental transcription factors. *Mol Cell Biol* **22**, 1936-46.
- **Urven, L. E., Erickson, C. A., Abbott, U. K. and McCarrey, J. R.** (1988). Analysis of germ line development in the chick embryo using an anti-mouse EC cell antibody. *Development* **103**, 299-304.
- **Vogt, P. K.** (1997). Historical introduction to the general properties of retroviruses. In *Retroviruses*, (ed. J. M. Coffin S. H. Hughes and H. E. Varmus), pp. 1-26: Cold Spring Harbor Laboratory Press.
- **Volker, J. L., Rameh, L. E., Zhu, Q., DeCaprio, J. and Hansen, U.** (1997). Mitogenic stimulation of resting T cells causes rapid phosphorylation of the transcription factor LSF and increased DNA-binding activity. *Genes Dev* **11**, 1435-46.
- **Voss, A. K., Thomas, T., Petrou, P., Anastassiadis, K., Scholer, H. and Gruss, P.** (2000). Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* **127**, 5449-61.

- **Walsh, C. P., Chaillet, J. R. and Bestor, T. H.** (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-7.
- **Watson, J. D. and Crick, F. H.** (1953). Molecular Structure of nucleic acids. A structure for desoxyribose nucleic acid. *Nature* **171**, 737-738.
- **Wilanowski, T., Tuckfield, A., Cerruti, L., O'Connell, S., Saint, R., Parekh, V., Tao, J., Cunningham, J. M. and Jane, S. M.** (2002). A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. *Mech Dev* **114**, 37-50.
- **Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M.** (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7.
- **Wilson, S. I., Graziano, E., Harland, R., Jessell, T. M. and Edlund, T.** (2000). An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo. *Curr Biol* **10**, 421-9.
- **Wilson, S. I., Rydstrom, A., Trimborn, T., Willert, K., Nusse, R., Jessell, T. M. and Edlund, T.** (2001). The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature* **411**, 325-30.
- **Wu, F. K., Garcia, J. A., Harrich, D. and Gaynor, R. B.** (1988). Purification of the human immunodeficiency virus type 1 enhancer and TAR binding proteins EBP-1 and UBP-1. *Embo J* **7**, 2117-30.
- **Xie, X., Chan, R. J. and Yoder, M. C.** (2002). Thrombopoietin acts synergistically with LIF to maintain an undifferentiated state of embryonic stem cells homozygous for a Shp-2 deletion mutation. *FEBS Lett* **529**, 361-4.
- **Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. R.** (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94.
- **Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. and Smith, A.** (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-6.
- **Yoder, J. A., Walsh, C. P. and Bestor, T. H.** (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**, 335-40.
- **Yoon, J. B., Li, G. and Roeder, R. G.** (1994). Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. *Mol Cell Biol* **14**, 1776-85.
- **Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T. and Kishimoto, T.** (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech Dev* **45**, 163-71.
- **Yoshida, K., Taga, T., Saito, M., Suematsu, S., Kumanogoh, A., Tanaka, T., Fujiwara, H., Hirata, M., Yamagami, T., Nakahata, T. et al.** (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. *Proc Natl Acad Sci U S A* **93**, 407-11.
- **Yuan, H., Corbi, N., Basilico, C. and Dailey, L.** (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-45.

- **Zagris, N. and Podimatas, T.** (1994). 5-Azacytidine changes gene expression and causes developmental arrest of early chick embryo. *Int J Dev Biol* **38**, 741-4.
- **Zappone, M. V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A. L., Lovell-Badge, R., Ottolenghi, S. et al.** (2000). Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367-82.
- **Zhang, Y., Niu, Z., Cohen, A. J. and Adams, S. L.** (1999). The internal chondrocyte-specific promoter of the chick type III collagen gene is activated by AP1 and is repressed in fibroblasts by a complex containing an LBP1-related protein. *Nucleic Acids Res* **27**, 4090-9.
- **Zhong, Z., Wen, Z. and Darnell, J. E., Jr.** (1994). Stat3 and Stat4: members of the family of signal transducers and activators of transcription. *Proc Natl Acad Sci U S A* **91**, 4806-10.
- **Zhou, W., Clouston, D. R., Wang, X., Cerruti, L., Cunningham, J. M. and Jane, S. M.** (2000). Induction of human fetal globin gene expression by a novel erythroid factor, NF-E4. *Mol Cell Biol* **20**, 7662-72.

# ANNEXES

**A member of the CP2 transcription factor family controls gene expression in chicken embryonic stem cells**

Hervé Acloque, Anne Marie Birot, Anne Mey, Bertrand Pain, and Jacques Samarut\*

Laboratoire de Biologie Moléculaire et Cellulaire de l'Ecole Normale Supérieure de Lyon,  
CNRS UMR5665, INRA UMR , 46 Allée d'Italie 69364 Lyon Cedex 07, France

\* To whom correspondence should be addressed: Tel:+33-4-7272-8536; Fax: +33-4-7272-8080  
*email address:* Jacques.Samarut@ens-lyon.fr

**Key Words:** ES cells, CP2, ERNI, ENS, promoter regulation, early embryo, chick

**Running Title:** cERNI/ENS promoter regulation



## **Summary:**

*cENS-1/cERNI* genes have been shown to be expressed very early during chicken embryonic development and as well as in pluripotent chicken embryonic stem (CES) cells. We have identified previously a promoter region, which is specifically active in CES cells compared to differentiated cells. In order to understand the molecular mechanisms, which confer this specific activity to the *cENS-1/cERNI* promoter, we analysed the *cis*-acting elements of this promoter in CES and differentiated cells. Deletion analyses of the *cENS-1* promoter show that a region located between -313 and -180 is necessary to activate the *cENS-1* promoter in CES cells. Two different binding sites for nuclear factors expressed in CES cells, called A and B, have been identified in this 134bp promoter region. The A element, TGATGGACA located between -240 and -232 binds a non identified factor that doesn't exhibit any enhancer activity. The B element, 5'-CAAGT CCAGG CAAGT-3', which is located between -294 and -280, is composed of a direct repeat, and increases the activity of a SV40 basal promoter in QT6 and CES cells. Mutation of the B region in the whole *cENS-1* promoter strongly decreases the promoter activity in CES cells, suggesting that this region is essential for activating the promoter. The B response element is similar to those previously described for the transcription factor CP2. Supershift experiments using an antiserum directed against human CP2 show that a protein complex containing a CP2 like transcription factor is bound to this B response element. We have also demonstrated that a region located between position -179/-32 strongly decreases the *cENS-1* promoter activity in differentiated cells, suggesting that this region could confer its ES cell-specificity to the *cENS-1* promoter. All these results identify a nuclear factor belonging to the CP2 transcription factor family that is crucial for the activation of the *cENS-1/cERNI* promoter. Transcriptional activation by CP2 is specifically downregulated in differentiated cells by a region located in the -179/-32 ENS promoter.

## **Introduction:**

Chicken embryonic stem cells (CES) derive from the epiblast of chicken early embryos (stage X according to (1)). These cells are pluripotent, can be grown for long term *in vitro* and can contribute to many embryonic tissues including germ line when they are grafted into an host embryo (2). They provide then a very good model to investigate molecular mechanisms underlying the control of pluripotency and differentiation commitment. In addition to the mouse ES cell model, it could help to define more precisely the pathways conserved and involved in the maintenance of pluripotency during embryonic development.

In the mouse embryo, few regulators of ES cell pluripotency have been characterized so far. It has been shown that *Oct-3/4* is a key regulator of ES cells pluripotency and is involved in the control of other genes like *FoxD3* or *FGF4* (see (3) for a review) . *Sox2*, *FoxD3*, *FGF4* and *FGFR2* have also been shown to be involve in the regulation of embryonic cell pluripotency *in vivo* and *in vitro* (4-7). The *in vivo* knock-out of these genes leads to a slight decrease of the number of cells originated from the epiblast and to a rapid embryonic lethality. It has been impossible *in vitro* to derive ES cells from embryos, carrying an inactivation of these genes.

In order to identify new sets of critical genes for ES cell biology we used a strategy based on gene trap, in chicken embryonic stem (CES) cells. We have identified a gene family which is specifically expressed in chicken embryonic stem cells and early embryo (8). The *Embryonic Normal Stem cell (ENS)* genes are strongly expressed in CES cells, but their expression decreases rapidly after the induction of CES cells differentiation by retinoic acid, or during embryoid bodies formation. In the embryo the expression of *ENS* genes is at a high level in the epiblast prior to gastrulation, and is then restricted to the prospective neural plate, and later to the border of the neural plate. Up to now three genes of this family have been characterized (8): 1) *cENS-1* which contains an open reading frame identical to *cERNI* (Early

Response to Neural Induction) (9), inserted between two repeated sequences; 2) *cENS-2* which is a truncated form of *cENS-1*; 3) *cENS-3* which possesses two open reading frames corresponding to *pol* and *env*-retrovirus related gene. The coding region for *ENS-1* does not show any known homology with retroviral gene and does not seem to be conserved among other birds than galliforms. However the occurrence of LTR-like repeated structures at both ends of *ENS* genes suggested that these genes might derive from an ancestral retrovirus or transposon structure (8). In the putative U3 region of the LTR-like structures, we have previously isolated a promoter region, and shown that its activity is much stronger in chicken embryonic stem cells than in differentiated cells (8). This result was surprising because it has been shown that retroviral LTRs are usually rapidly silenced in mouse ES cells and early mouse embryo, correlated with a high methylation of these sequences (10),(11). This silencing is probably due to a three step mechanism involving 1) *trans* specific factors, 2) recruitment of chromatin remodelling factors and 3) DNA methyltransferases (for a review see (12)). However, such a mechanism has not yet been observed in chicken blastula compared to mouse early embryo. It has been demonstrated that early chicken blastodermal cells (from which chicken ES cells are derived) are permissive for infection by Rous Sarcoma Virus (RSV) and readily expressed RSV genes (13). These data suggest that CES cells could be more permissive to retroviral gene expression (like *ENS* genes) than mouse ES cells.

It is also possible that *ENS* LTRs are protected from gene silencing by specific molecular regulators or by a positive position effect. The restricted expression pattern of *ENS* genes in embryonic stem cells and early embryo is of particular interest, and led us to study the mechanisms underlying *ENS* genes expression. Based on a systematic analysis of the regulatory elements of *cENS* genes, we report here that two cis-acting elements, belonging to the first 310 bp of *cENS-1/cERNI* promoter, are responsible for its activity in chicken ES cells, and that a region located between position -179/-32 strongly represses the *cENS-1*

promoter activity in differentiated cells. This suggests that this last region could confer its tissue-specificity to the ENS promoter. These results as a whole reveal the existence of a nuclear factor, which belongs to the family of the CP2 transcription factor. This factor is crucial in activating the *cENS-1/cERNI* promoter but its activity is made specific to CES cells by a specific region of *cENS-1* promoter.

## **Material and methods**

### **cell culture and DNA transfection**

The 9N2-5 chicken ES cell line was cultivated and transfected as previously described (8) with few modifications. Briefly CES cells were cultivated in Dulbecco's modified Eagle's medium (DMEM)/F12 (Invitrogen) –5% fetal bovine serum (FBS; PAN) supplemented with IGF1 1ng/ml(Sigma), mSCF 0.5 ng/ml (R&D), hIL-6 1ng/ml (Peprotek), hIL6-sR 1ng/ml (peprotek),  $\beta$ -mercaptoethanol  $10^{-4}$ M, sodium pyruvate and non essential amino acids on an irradiated feeder of STO cells. QT6 cells (14) were cultivated in HamF12 medium (Invitrogen)-2% FBS complemented with 10% TPB.

The day before transfection, the cells were plated in 12-wells dishes at a density of  $5 \times 10^4$  cells per well for the 9N2-5 cells or  $1 \times 10^5$  per well for QT6 cells. Cells were cotransfected using Lipofectamine reagent (Invitrogen) with 300 ng of DNA construct and 30 ng of the Renilla Luciferase Reporter construct, pRL-TK, to normalize for transfection efficiency. After 5 hours of incubation in diluted culture medium containing only 1% FBS, the cells were washed twice with phosphate-buffered saline (PBS) and regular medium was added to each well. On the following day, cells were washed twice with PBS and lysed with passive lysis buffer (Promega), then Firefly and Renilla luciferase luminescence assays were successively performed using Dual Luciferase Assay (Promega) as described by the manufacturer. Firefly reporter gene values were normalised to the activity of the Renilla luciferase, used as internal control.

### **DNA constructs**

All the luciferase reporter vectors were purchased from Promega.

ENS reporter constructs were produced using PCR amplification with multiple primers that introduced a 5'-*MluI* restriction site at positions -738, -456, -308, -179 and -32 respectively paired with a single 3'-antisense primer that generated a 3'-*MluI* restriction site at position +83. Primer sequences are available in Table 1. PCR products and pGL2-basic vector were digested with *MluI*, purified and then ligated to produce respectively p738-luc, p456-luc, p308-luc, p179-luc and p32-luc.

Reporter constructs and the pGL2 basic negative control vectors were used for transient transfection studies.

Monomers or trimers of the -313/-180 region were constructed as follows: the -313/-180 region was amplified by PCR with *BamHI* -313 S and *BglIII* -180 AS primers (Table1). After digestion with *BamHI/BglIII*, PCR products were ligated in the BamHI site of pGEM-7Z. *BglIII* restriction enzyme was then added to the ligation mix to eliminate multimers with various orientations. Multimers were then excised from pGEM-7Z by *NsiI/KpnI* and ligated into p179-luc or pGL2-promoter (Promega). Multimer copy numbers and identities were confirmed by DNA sequencing.

Multimers (one or three copies) of the wild-type or mutated A and B response element (see Table 1) were cloned in the *SacI* restriction site of the p179-luc construct or of pGL2-promoter. Insert identities were confirmed by DNA sequencing.

Targeted mutagenesis on p456-luc for the B response elements has been performed with the Quick Change Mutagenesis kit (Stratagene) using the primers: B mutagenesis S and B mutagenesis AS.

p456del-238/-32-luc was constructed as follow : -456/-238 region was amplified by PCR using *KpnI* -456 S and *SacI* -238 AS primers. After digestion with *KpnI/SacI*, PCR products were ligated in the *KpnI/SacI* sites of p32-luc to give p456del-238/-32-luc construct.

### **Nuclear extracts**

Nuclear extracts were prepared with the Active Motif nuclear extract kit (Active Motif) and stored at  $-80^{\circ}\text{C}$  until further use. Protein concentrations were estimated using the Biorad Protein Assay Kit I.

### **DNA binding assays**

Probes used for DNA-binding assay were labeled by using T4 polynucleotide kinase and  $[\gamma\text{-}^{32}\text{P}]$  ATP. The labeled fragments were purified by acrylamide gel electrophoresis. Approximately  $2 \times 10^4$  cpm of the probe were added to a 30- $\mu\text{l}$  reaction mixture in the presence or absence of 10 $\mu\text{g}$  of nuclear extract in a buffer containing 20 mM HEPES (pH 7.9), 20 % glycerol, 100mM KCl, 2mM MgCl<sub>2</sub>, 0.2mM EDTA, 0.5mM dithiothreitol and 1 $\mu\text{g}$  of poly(dI-dC)-poly(dI-dC). For competition experiments, a 100-fold molar excess of unlabeled double-stranded oligonucleotide was incubated for 10 min with nuclear extract prior to the addition of labeled double-stranded probe. After incubation on ice for 30 minutes, the DNA-protein complexes and unbound probes were separated by electrophoresis on a 4% glycerol and 5% polyacrylamide gels in 0.5X TBE at 160 V for 120 minutes. The gels were dried and autoradiographed. EMSA supershift assays were performed with 1  $\mu\text{l}$  of antiserum to CP2 (Yoon et al. 1994) or 1  $\mu\text{l}$  of rabbit preimmune serum, added to the reaction mix prior to the addition of the labeled probe.

### **Methylation interference assays**

Probes were labeled as previously described (15). For sense or antisense oligoprobes, only the sense or antisense oligonucleotide was labeled. For the guanine methylation procedure,  $1 \times 10^6$  cpm of the probe was methylated using 1  $\mu\text{l}$  of dimethyl sulfate (DMS) (Merck) for 3 minutes at room temperature in 200  $\mu\text{l}$  of 50 mM sodium cacodylate (pH 8.0)-1mM EDTA. The

reaction was stopped by the addition of 50  $\mu$ l of stop buffer (5 M sodium acetate [pH 7.0], 1M  $\beta$ -mercaptoethanol). The DNA was then precipitated with ethanol, resuspended in 30  $\mu$ l of water. The electrophoretic mobility shift assays were carried out as described above, but with  $2 \times 10^5$  cpm of modified  $^{32}\text{P}$ -labeled probe and 30 $\mu$ g of crude nuclear extracts. After autoradiography, the DNA was electro-eluted from the bands corresponding to the retarded probe and the free unbound probe. After precipitation with ethanol, the amounts of radioactivity of the bound and free probe were adjusted and the DNA was cleaved in 100  $\mu$ l of 1M piperidine for 30 minutes at 90°C. The cleaved products were separated by electrophoresis through a denaturing 18% polyacrylamide gel and were autoradiographed.

### **Real time PCR**

Total mRNA were extracted from chicken ES cells by using Rneasy kits (Qiagen). Real-time RT-PCR was performed by using the QuantiTect SYBR Green PCR Kit (Qiagen). The obtention of the single denaturation peak upon heating and of a single band after agarose gel electrophoresis (data not shown) confirmed that the real time RT-PCR reaction was specific. The PCR conditions used were the following: 15mn denaturation at 95°C followed by 50 cycles of 95°C for 15s denaturation step, 58°C for 15s and 72°C for 15s; followed by a melting curve from 60°C to 95°C.

The primers used for real time RT-PCR are:

QPCR GAPDH S: 5'-TGGGTGTCAACCATGAGAAA-3'

QPCR GAPDH AS: 5'-CATCCACCGTCTTCTGTGTG-3'

QPCR CP2 S: 5'-CTATCTGGAGGAGCTGACGG-3'

QPCR CP2 AS 5'-TCTGAATCATCTCGTCGCTGA-3'



### **In situ hybridisation of whole-mount embryos**

Whole-mount in situ hybridisation was carried out as described by Wilkinson (16) and revised by Streit et al. (17). The probe used for the experiment was amplified by PCR with the following primers: CP2 S 5'-GAATTCATGGCCTGGGCGCTGAA-3', CP2 QPCR AS 5'-CCCTTGGTTGAGGTAGGTGA-3' and covers the first 280 bp of the CP2 coding sequence.

## **Results:**

### **A –308/+83 cENS/cERNI promoter fragment is necessary for promoter activation in chicken ES (CES) cells**

In a previous study, we have identified a promoter region in the putative U3 region of *ENS* genes, which was specifically active in CES cells (8). The promoter sequence is shown in figure 1. Analysis of the promoter sequence using MatInspector Release professional 6.1 (18) revealed several putative *cis*-acting elements for factors implicated in regulating constitutive gene expression (like AP1). It also revealed three well conserved binding elements for GATA-1 in the proximal region of the promoter. However the analysis did not identify any candidates known to be involved in regulating ES cells specific genes. In order to localize functional *cis*-acting element(s) responsible for tissue-specific *cENS* promoter activity in CES cells and in early embryo, some luciferase reporter constructs were designed and generated for use in transient transfection assays. We first performed DNA transfection experiments by using 5' deletion mutants of the *cENSI* promoter fused to the luciferase reporter gene (cENS-luc). Our largest construct contained the first 738 bp of the promoter. We chose this fragment because it includes the putative U3 region of *ENS* genes (8). Successive deletions of cENS-luc from –738 to –308 did not reduce the expression level of the reporter gene. In contrast, a further deletion to position –179 led to a 20-fold decrease, indicating that critical *cis*-acting elements are located between –308 and –179 (Fig.2A). Compared to CES cells, none of these constructs showed a strong transcriptional activity either in CES cells induced to differentiate with retinoic acid for 48 hours or in QT6 cells, a quail immortalized fibroblast line (14) (Fig.2B). QT6 cells were used in the following experiments as a model for differentiated cells.

## **A -313/-180 region is a key region for promoter activation in chicken ES cells**

To address the possible enhancer function of the DNA sequences, that are located between -313 and -180, we cloned this 134bp region in one or three copies upstream of a SV40 basal promoter, which drives the expression of the luciferase reporter gene (SV40 basal promoter in pGL2-promoter). These constructs were transfected in chicken ES cells or in QT6 cells. DNA transfection assays with these constructs showed first a 10-fold increased activity in QT6 cells and a 4-fold increased activity in CES cells with one copy of the -313/-180 region, and second, a 18-fold and 13-fold increased activity, respectively in QT6 and CES cells with three copies of the -313/-180 region as compared to the basal SV40 promoter construct (Fig.3A). Transfection assays with the -313/-180 region in opposite orientation exhibited the same tendency. To confirm these results, we placed one or three copies of the -313/-180 region upstream of the -179/+83 *cENS-I* promoter region (p179-luc construct). This latter region contains a TATA-like box as a known *cis*-acting element, and has a weak transcriptional activity in CES cells. Transfection assays in CES and QT6 cells showed respectively a 4-fold and 2-fold increase of the promoter activity with one copy, reaching 12-fold and 4-fold with 3 copies of the -313/-180 region (Fig.3B). In QT6 cells, the activation level decreased strongly (from 17-fold to 4-fold with three copies of the -313/-180 region and from 10-fold to 3-fold with one copy), when the -313/-80 region was placed upstream of the -179/+83 *cENS-I* promoter, compared with the basal SV40 promoter (compare Fig.3A and 3B).

These results as a whole indicate that the -313/-180 region acts as a transcriptional enhancer for the expression of a reporter gene as well in undifferentiated CES cells as in QT6 cells. Moreover, when located upstream of the -179/+83 region, this -313/-180 region is specifically active in undifferentiated CES cells.

## **Two distinct protein-DNA interacting regions are located in the –313/-180 region**

To investigate potential DNA-protein binding sites located in the –313/-180 region and involved in *cENS-1/ERN1* promoter activity, we performed gel retardation experiments using chicken ES cells and QT6 cells nuclear extracts. Seven overlapping oligonucleotides were designed to cover the –310/-180 region (Fig.4A). The putative AP-1 site identified *in silico* and located in probe I was apparently not bound *in vitro* by nuclear proteins from CES cells (Fig.4B, lane 2). In contrast the oligonucleotides II, V and VI were bound by CES nuclear factors (Fig.4B, lanes 5, 14 and 17). These three complexes were competed for by 100-fold molar excess of the respective unlabeled oligonucleotides (Fig.4B lanes 6, 15 and 18) confirming the binding specificity. The DNA-protein complexes observed with probes V and VI migrated to the same positions suggesting that they are identical and located on a DNA sequence situated between positions –251 and –230 of the ENS-1 promoter, in the overlapping region of the V and VI oligonucleotides. Complexes migrating to the same positions, as in the case of CES cells, were observed when using QT6 nuclear extracts with probes II, V and VI (data not shown). These results are consistent with our previous data showing that the –313/-180 region was active in both cell lines when placed upstream of a SV40 basal promoter. The overlapping sequence of the oligonucleotides V and VI was called the A region and the sequence located on oligonucleotide II was designated as the B region.

### **Fine mapping of the A and B response elements.**

To define more accurately the nucleotides involved in DNA-protein interaction on A and B regions, we used double stranded oligonucleotides chemically modified with dimethyl

sulfate (methylation interference analysis). On both the coding and the non-coding strands of A region, strong interferences were detected on guanine residues which allowed us to delimit a binding region between positions –240 and –232 in the sequence 5'-TGATGGACA-3' (Fig.5A).

The same experiment was performed on both the coding and the non-coding strands of the B region, where strong interferences were detected on some guanine residues between positions –296 and –280 in the sequence 5'-AACAAAGTCCAGGCAAGT-3' (Fig.5A). This response element contains a direct repeat 5'-CAAGT CCAGG CAAGT- 3', which is important for DNA-protein interaction, as suggested by the strong interferences on the same guanine residues on each CAAGT box (Fig.5A). The respective places of this two response elements in the *cENS-1* promoter is shown in figure 1 and figure 5B.

**The A region (-240/-232) binds nuclear factors present in chicken ES cells but is devoid of enhancer activity.**

To confirm that the residues identified by methylation interference analysis were essential for DNA-protein interactions, we performed band shift experiments with a labeled probe containing the A region, in competition with wild-type or mutated oligonucleotides for the A region(Fig.6). The radioactively labeled probe, incubated with nuclear extracts from CES cells bound two prominent protein complexes (Fig.6 lane 2) that were also revealed in QT6 cells (not shown). A 100-fold molar excess of wild-type A oligonucleotide totally abolished the interactions between the labeled probe and nuclear factors (Fig.6, lane 3). On the other hand, a 100-fold molar excess of A oligonucleotides with mutations localised either in 5' or in the 3' part of the sequence did not compete with the labeled probe (Fig.6, lane 4 and 5). Gel retardation assays with probes containing mutated A region did not exhibit any

DNA-protein complex (not shown). These results confirmed that the residue identified by methylation interference analysis in the A region were essential for DNA-protein interaction. We expected that various mutations in the A sequence would affect preferentially one of the two complexes previously described. However our results suggested that these two complexes probably interact with the same DNA binding element or with elements very close to each other.

To investigate the role of the A binding site in *cENS-1* promoter activity we placed multimers of the A sequence upstream of the p179-luc construct or the SV40 basal promoter. The different constructs generated were transfected in CES or in QT6 cells and the enhancer activity was estimated with respect to the activity of empty vectors. None of these constructs revealed any enhanced activity as compared to empty vectors (not shown). These data then suggest that the A element does not have any enhancer function in the *cENS-1/cERN1* promoter region.

### **Characterization of the B region (-294/-280)**

After having identified the residues by the methylation interference method, we confirmed that they were essential for DNA-protein interaction through gel shift experiments with an oligonucleotide labeled probe covering the B region (Fig. 7A); the results are shown in figure 7B. A unique DNA-protein complex was detected with the labeled probe incubated with CES nuclear extract (Fig. 7B lane 2). A competition with 10 or 100-fold molar excess of wild type B oligonucleotide strongly decreased the signal corresponding to the DNA-protein complex, confirming the specificity of this interaction (Fig. 7B lanes 3 and 4). A competition with a 100-fold molar excess of oligonucleotides mutated respectively on the C<sub>1</sub>, G<sub>4</sub>, C<sub>11</sub> and G<sub>14</sub> residues (Fig. 7A), which are important for DNA-protein interaction as proven by methylation interference analysis, did not decrease the signal (Fig. 7B lane 5 to 8). These

results confirmed that each of these four nucleotides are essential for the interaction with nuclear binding proteins. It demonstrates also that each part of the direct repeat is necessary for DNA-protein complex formation. To determine if each of the nucleotides of the B sequence participates to this interaction, mutated labeled probes covering the first part of the repeat were used in gel shift retardation assays (Fig. 7A and 7C). Results showed that the interaction with nuclear proteins was abrogated only when the C<sub>1</sub> or the G<sub>4</sub> residues were mutated (Fig.7C lanes 1 and 4). A probe, in which the T<sub>5</sub> residue is mutated, exhibited a weaker signal compared to the wild-type probe (Fig.7C lane 5), suggesting that this nucleotide improves the stabilization of the DNA-protein complex. Same results were obtained with probes mutated in the second part of the direct repeat (not shown). The two parts of this direct repeat are separated by a linker of five nucleotides (CCAGG). In order to check the function of this spacing sequence we tested probes containing spacers of different lengths. Band shift assays performed with these oligonucleotides are shown in figure 7D. Deletion or addition of three nucleotides to the linker decreased the interaction between DNA and nuclear factors (Fig.7D lanes 6 and 10), while the addition of 6 bases totally abolished the interaction (Fig.7D lanes 14). The optimal B response element is then summarized by the sequence **CAAGTNNNNNCAAGT** where bold characters indicate bases essential for the DNA-protein complex formation.

### **The B region is essential for promoter activation in chicken ES cells.**

To test the enhancer activity of the B response element, multimers of this region were placed upstream of the minimal *cENS-1* promoter (region -179/+83) or of the SV40 basal promoter. The different constructs were transfected in parallel into chicken ES cells and QT6 cells. One copy of the B region in sense or antisense orientation increased the activity of the basal SV40 promoter from 5-7 fold in CES cells to 10 fold in QT6 cells. When the *cENS-1* minimal

promoter was used a 13-fold increased in CES cells was observed (Fig.8A and 8B). Three copies of the B region in sense or antisense orientation also strongly increased the activity of the basal SV40 promoter from 10-fold in CES cells to 20-fold in QT6 cells and that of the *cENS-1* minimal promoter to 10-fold in CES cells (Fig.8A and 8B). In comparison, the mutated B response element was inactive in the two cell lines. These data strongly suggest that the B region is a key activator of the *cENS-1* promoter. These results were also confirmed by targeted mutagenesis of the B response element in the p456-luc construct. Compared to the unmutated -456/+83 *cENS-1* promoter, the promoter with a mutated B element exhibited a strong decrease of its activity (Fig.9). All these data support the fact that the B element plays an essential role in the *cENS-1* promoter activation in chicken ES cells.

### **The B response element is bound by a transcription factor belonging to CP2 family**

Analysis of the B response element revealed that it is similar to the CP2 consensus binding sequence CNRG-N<sub>6</sub>-CNRG (Lim et al. 1993). The CP2 response element, like the B response element, is composed of a direct repeat essential for protein binding. In order to test if the B response element is bound in vitro by a CP2 like protein, we performed supershift experiments using a previously described rabbit antiserum, which recognizes human CP2 (19). Incubation of CES cells nuclear extract with anti-CP2 serum prior to the addition of the B labeled DNA probe totally abolished the formation of the specific DNA-protein complex (Fig.10, lane 5), while the addition of a preimmune rabbit serum did not interfere with the DNA-protein complex (Fig.10 lane 6). Yoon et al. (19) also reported that addition of anti-CP2 serum to nuclear extract prior to addition of DNA probe totally abolishes formation of DNA-protein complex. This could be explained by the fact that the antiserum probably recognizes the DNA binding site or the oligomerization domain of the CP2 protein.



On the contrary, a CP2 antiserum or a rabbit preimmune serum incubated with nuclear extracts of CES cells did not interfere with the DNA-protein complex formed with the A response element (Fig.10B lanes 2 and 3). This indicates that the anti-CP2 serum specifically interacts with the complex bound to the B region by blocking either the DNA binding site or the oligomerisation domain of the CP2 protein, or both of them. These results show in the end that the B response element binds a protein belonging to the CP2 transcription factor family.

**CP2 response element activity is repressed in QT6 cells by a region located between positions -179/-32**

We have previously shown that the -313/-180 region is less active in QT6 cells than in CES cells when it is placed upstream of the *cENS-1* minimal promoter -179/+83 (Fig.3B). More precisely, when 3 copies of the CP2 response element (B region) were placed upstream of the *cENS-1* -179/+83 minimal promoter, their activation level was 6 times weaker in QT6 cells than in the construct with the SV40 basal promoter (Fig.11A). On the contrary, no difference was observed between these two constructs in CES cells (Fig.11A). These data suggest that a region located in the *cENS-1* -179/+83 promoter could repress the enhancer activity of the CP2 response element (B region) in QT6 cells, while it does not have any effect in CES cells. In order to check this assumption, we have generated a construct in which the -238/-32 region is deleted from the -456/+83 *cENS-1* promoter. After transfection in CES and QT6 cells, we analysed the promoter activity in both cell lines with respect to the activity of the -456/+83 *cENS-1* promoter. While both promoters exhibit the same activity in CES cells, a 10-fold increased of the luciferase activity was observed in QT6 cells when the -238/-32 region has been deleted (Fig.11B). Taken together, these data suggest that if the CP2 binding element is important for *cENS-1* promoter activation, its activity is repressed in QT6 cells by

a region located between positions -238 and -32. Our previous results using the *cENS-1* -179/+83 promoter also suggest that this repressive region could be located more precisely between positions -179/-32.

### **Expression pattern of chicken *CP2* in chicken ES cells and early embryo**

Chicken *CP2* has been previously shown to regulate the  $\alpha$ A-crystallin promoter (20). Based on the *cCP2* cDNA sequence published in that reported work (but not available on Genbank) we amplified the chicken *CP2* by PCR from RNAs extracted from chicken ES cells indicating that *cCP2* is expressed in CES cells. Analysis of the *cCP2* coding sequence (Genbank Accession number: AY298725) revealed an amino acid sequence, which is very similar to those of human (19) and mouse (21) and is also identical to the complete sequence of chicken CP2 (20), but with a small deletion of 10 amino acid (from position 323 to position 332). By mean of Real Time PCR we observed that *cCP2* transcript is expressed at the same level in undifferentiated and differentiated chicken ES cells and in QT6 cells (data not shown). The protein expression was also detected in both cells by immunoblot using either LBP-1c antiserum or a commercial anti CP2 antibody (anti LSF from BD Biosciences Pharmingen)(not shown). These results are consistent with the fact that the B region is active in undifferentiated and differentiated cells when placed upstream of the SV40 basal promoter. The spatial pattern of the *cCP2* gene expression was examined during embryogenesis by whole-mount in situ hybridisation. By using a probe located in the coding sequence of *cCP2*, we observed a particular expression pattern during early chick embryogenesis. *cCP2* is expressed in the epiblaste of embryos at the pre-primitive streak stage (fig.12 A-B). In embryos at the primitive streak stage, it is also expressed in the extending primitive streak and in the prospective neural plate (9) (fig.12C-F). At stage 7 and 8, the *cCP2* expression was

mostly localized in the neural folds and the somites of the embryo (Fig.12 G-I) but also in the regressing primitive streak. Later at stage12, *cCP2* was ubiquitously expressed in the whole embryo (Fig.12 J).

## **Discussion**

The *cENS/cERNI* gene is expressed specifically in chicken ES cells, in the epiblast of chicken blastula as well as in the prospective neural plate of chick gastrula (8,9). Even if the functions of *cENS-1/cERNI* remain to day elusive either for the biology of chicken ES cells or for the neural induction of embryonic ectoderm, its specific expression pattern is of particular interest to understand the mechanisms which lead to both the formation of the neural plate and the maintenance of cell pluripotency. These genes are so far the only genes identified as having an expression specific to undifferentiated chicken embryonic stem cells (8). They can be then considered as endogenous reporter genes to identify transcription mechanisms and signalling pathways that might account for determining the pluripotency state of embryonic stem cells. Few mechanisms, which regulate ES-specific genes expression, have been so far identified. *Oct-3/4* is one of the genes, whose specific expression in pluripotent cells (22-24) and implication for transcriptional regulation of ES-specific genes (3) are the best understood in the mouse. Analysing the transcriptional regulation of *cENS-1/cERNI* will then provide new insights on ES-specific transcription regulatory mechanisms but also on neural induction. Therefore, the primary aims of our studies were to identify first specific regulatory elements located on the ENS promoter and then *trans*-activating factors that might be involved in the cell-specific activity of this promoter. Functional dissection of the *cENS-1* promoter through testing in pluripotent ES cells versus differentiated cells, either differentiated cells derived from ES cells or fibroblast cells, lead us to identify two main signalling ways regulating the *cENS-1* promoter. One way mediated by the CP2 transcription factor works as a transcriptional activating pathway. Another still not fully identified pathway , induces a selective repression of *cENS-1* promoter in differentiated cells.

## **Identification of CP2 transcription factor as essential for *cENS-1* promoter activation.**

Molecular and functional assays clearly led to identify a consensus CP2 responsive element (5'-CAAGTCCAGGCAAGT-3') in the *cENS-1* promoter. Specific mutation of this response element strongly abrogates the function of the promoter in transient transfection assay in CES cells. Moreover, bandshift assay clearly demonstrated the presence in CES cells of a nuclear factor very closely related to CP2. Cloning of the chicken *CP2* product in CES cells confirmed the identity and the expression of *cCP2* in these cells. *CP2* (also known as *LSF* or *LBP-1c*) belongs to a transcription factor family whose founding member is the drosophila *grainyhead* (also known as *NTF-1* or *Elf-1*) gene, an important factor in developmental patterning in the fly (25-27). The phylogenetic tree of this transcription factor family is composed of two distinct branches (28). The first contains *grainyhead* and its homologue *mammalian grainyhead (MGR)*, *brother of MGR (BOM)* and *sister of MGR (SOM)*. The second contains *dCP2* and the mammalian *CP2*, *LBP1-a (NF2d9)* for the mouse homologue) and *LBP-9 (CRTR-1)* in the mouse) genes. *CP2* controls a wide range of gene expressions in human, mouse or chicken cells (19-21,29-34). RT-PCR amplification of *cCP2* and immunoblots using two distinct anti-CP2 antibodies reveals that it is expressed in QT6 and CES cells (not shown), in addition to published data indicating that *cCP2* is expressed ubiquitously in embryonic tissues from stage 13 embryos (20). Our present data show that *cCP2* is differently expressed in cell layers and tissues of early chicken embryo. *cCP2* expression is stronger in the primitive streak and the hensen node and is absent in the non neural ectoderme of stage 3 and 4 embryos. This is consistent with a previous study which demonstrates that a graft of the hensen node induced *cENS-1/cERNI* expression in the area opaca of chicken embryo (9). This also suggests that *cCP2* is essential for *cENS-1/cERNI*

expression in vivo because its lack in the non neural ectoderme is correlated with the absence of *cENS-1/cERNI* in this tissue.

As expression of the *cENS-1* promoter in transient transfection assays strongly depends on the integrity of the CP2-response element, we might consider that CP2 is the major transcription activator of *cENS-1* gene. However we can not exclude less contribution of other transcription factors whose binding site are identified in the promoter sequence, in the regulation of the endogenous *cENS-1* promoter in the chromatin. However, the observation that major expression of *cCP2* in the early embryo (before and around gastrulation) strongly correlates with that of *cENS-1* (8), is highly supportive of a major role attributable to *cCP2* in the regulation of the expression of *cENS-1* in early chicken embryogenesis.

**Activation of *cENS-1* promoter by CP2 is downregulated in differentiated cells.** Although *cCP2* is ubiquitously expressed, some studies have shown that it is also involved in the regulation of tissue-specific gene transcription. For example it has been shown that CP2 is involved in the spatial and temporal control of embryonic  $\gamma$ -globin expression during haemoglobin switching. The SSE (Stage Selector Element) region of the  $\gamma$ -globin promoter is bound to a protein complex called SSP (Stage Selector Protein) which contains CP2 (35). In this complex, CP2 interacts with NF-E4, a gene specifically expressed in foetal liver, cord and bone marrow cells to activate  $\gamma$ -globin expression (36). Some related CP2 proteins are also expressed specifically in some tissues, like CRTR-1, which has been described to be expressed specifically in mouse ES cells (37).

In the present study we clearly show that the ES cell specific activation of *cENS-1* promoter by *cCP2* is controlled by a sequence downstream to the *CP2* responsive element and close to the TATA box. This regulatory sequence specifically abrogates activation by CP2 in differentiated cells and not in pluripotent ES cells. As this downstream sequence is not

necessary for the function of CP2 in ES, the most likely explanation is that it binds negative regulatory factors that are absent in ES cells but that should appear upon commitment of ES cells into differentiated progenies.

This repressive sequence contains many putative binding sites for several transcription factors. Functional and molecular dissection should allow to identify the repressor that inactivate the response to CP2 in differentiated cells but not in ES cells. As the *cENS-1* promoter is also functional in mouse ES cells (not shown), we might anticipate that this repressor is well conserved among vertebrates.

Even if many promoters are driven in differentiated cells by specific enhancer (38,39), it has also been shown that tissue-specific expression could be controlled by the repression of the promoter activity in other tissues. Oct-3/4 expression has been correlated in primordial germ cells with the absence of Germ Cell Nuclear Factor that binds Oct-3/4 promoter and represses its activity *in vitro* (22,40). We presently do not know whether this putative repressor acts in partnership with CP2 or is a global repressor of transcription. Repressors have been shown to be required for regulation of cell-specific gene expression by recruited histone deacetylase or methyltransferase (41,42). Further analysis to quantify variations on histone tail acetylation or methylation in the *cENS-1* promoter during CES cell differentiation will bring new insights on such regulation.

In conclusion we identified in this study an enhancer region essential for *cENS-1/cERN1* promoter activity. This enhancer is bound *in vitro* by a CP2 related protein present in CES nuclear extracts. Its activity is also down regulated in differentiated cells by a repressor region localised in the proximal ENS promoter. These results then demonstrate that during differentiation of ES cells, gene expression program might be controlled by transcription repressive signals. Understanding these signalling pathway should help in understanding molecular bases of pluripotency.

## Figure Legends:

Fig.1. **Structure of *cENS-1* promoter.** The sequence from -738 to +83 relative to the transcription start site (TSS) is presented. The TSS is marked with right-angled arrows (↗) and the putative TATA box in bold. This sequence was analysed by Matinspector professional release 6.1. Transcription factors with 1.00 core similarity and more than 0.95 matrix similarities are highlighted in grey and labeled above the sequence. The A and B response elements identified in this study are underlined.

Fig. 2. **Functional analysis of deletion mutants of the *cENS-1* promoter.** *A*, deletions were generated as described in Materials and Methods; a schematic representation of each reporter construct is shown on the left. These constructs were transfected into chicken ES cells. Firefly luciferase activity was normalized for transfection efficiency using the renilla luciferase activity. Values are relative to the activity obtained with the largest promoter fragment (p738-luc) for three independent experiments, each performed in triplicate. *B*, deletions mutants of *cENS-1* promoter were transfected into CES cells, CES cells induced to differentiate with retinoic acid for 48h and into QT6 cells. Firefly luciferase activity was normalized for transfection efficiency using renilla luciferase activity. Values are relative to the activity obtained with the p738-luc in CES cells.

Fig. 3. **Functional analysis of the -313/-180 region.** *A*, one or three copies of the -313/-180 region fused to the pGL2 promoter construct were assayed by transfection experiments in QT6 and chicken ES cells. Values are relative to the activity obtained in each cell line with the empty pGL2 promoter construct. *B*, One or three copies of the -313/-180 region fused to the p179-luc construct were assayed by transfection into CES cells. Values are relative to the activity obtained with p179-luc construct. Data represent ratios of firefly luciferase versus renilla luciferase activities and values are means of at least three independent transfection experiments. The orientation of the -313/-180 is shown by the sense of the arrows.

Fig. 4. **The -313/-180 region contains two binding sites for chicken ES cells nuclear factors.** *A*, map of the 7 probes used to cover the -313/-180 region. Each probe is designated with a roman number. *B*, EMSAs were performed with each double stranded radiolabeled oligonucleotide probe and nuclear extracts prepared from CES cells as described in Materials and Methods. The positions of DNA-protein complexes throughout are indicated by arrows. Lanes 1,4,7,10,13,16,19: labeled probes alone; lanes 2,5,8,11,14,17,20 : labeled probes with nuclear extracts from CES cells; lanes 3,6,9,12,15,18,21: competition experiments with 100-fold molar excess of wild-type unlabeled probes.

Fig. 5. **Fine mapping of A and B binding sites by methylation interference.** *A*, methylation interference assays were performed with methylated -251/-223 (A region) or -297/-277 (B region) probes radiolabeled on the 5'-end of the sense or antisense strands. The probes were incubated with chicken ES cells nuclear extracts. The free (F) and retarded (R) probes were then sequenced and compared with the G+A sequence obtained with the non methylated probes. The G residues implicated were identified by a disappearance or a decrease of the signal. *B*, summary of the interactions. The guanine nucleotides identified by methylation interference assays are presented in bold on a double strand DNA fragment.



Fig. 6. **The A region binds two protein complexes.** *A*, labeled double-stranded wild-type probe alone (lane 1) was incubated with CES cells nuclear extract in absence of specific competitors (lane 2). Black arrows indicate the two DNA-protein complexes. Competitions experiments were performed with a 100-fold molar excess of either the unlabeled wild-type A oligonucleotide (lane 3), the first mutant A oligonucleotide (lane 4) or the second mutant A oligonucleotide (lane 5). *B*, sequences of the oligonucleotides used. Mutated residues are designated by an asterisk on oligonucleotide sequences and the DNA binding sequence is underlined.

Fig. 7. **Binding of chicken nuclear proteins to the B region.** *A*, sequences of the oligonucleotides used. DNA binding sequence is underlined and mutations are marked with an asterisk and bold characters. Base pairs of the DNA binding sequence are numbered. *B*, EMSAs were performed with the B wild-type labeled probe alone (lane 1) or with nuclear extracts prepared from CES cells (lane 2). The position of DNA-protein complex is indicated by an arrow. Competitions experiments were performed with a 10-fold molar excess (lane 3) or 100-fold molar excess (lane 4) of the unlabeled wild-type B oligonucleotide (lane 3) or with 100-fold molar excess of the B mutated C<sub>1</sub>, G<sub>4</sub>, C<sub>11</sub> and G<sub>14</sub> unlabeled oligonucleotide (respectively lane 5, 6, 7 and 8). *C*, B mutated C<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, G<sub>4</sub>, T<sub>5</sub>, C<sub>6</sub>, C<sub>7</sub> oligonucleotides were labeled and used for EMSAs with CES cells nuclear extracts (respectively lanes 1,2,3,4,5,6,7). *D*, EMSAs were performed with each double-stranded radiolabeled probe alone (lanes 1,5,9,13) or with CES cells nuclear extracts (lanes 2,6,10,14). The position of the DNA-protein complex is indicated by an arrow. Competition experiments were performed with a 100-fold molar excess of either the mutated B unlabeled oligonucleotide (lanes 3,7,11,15) or the B wild-type unlabeled oligonucleotide (lanes 4,8,12,16).

Fig. 8. **Functional analysis of the B region.** *A*, multimerized wild-type or mutated B region fused to the pGL2-promoter construct were assayed by transfection into QT6 and CES cells. Values are relative to the activity obtained with the pGL2-promoter construct. *B*, multimerized wild-type or mutated B region fused to the p179-luc construct were assayed by transfection into chicken ES cells. Values are relative to the activity obtained with the p179-luc construct. Data represent ratios of firefly luciferase versus renilla luciferase activities, and values are means of three independent experiments.

Fig. 9. **Functional analysis of the B region by targeted mutagenesis.** The B region was disrupted by targeted mutagenesis on the p456-luc construct to give p456 Bmut-luc. These two constructs were transfected into CES cells and promoters activity assayed. Values are relative to the activity obtained with the p456-luc.

Fig. 10: **The B region binds a factor recognized by a rabbit anti-CP2 serum.** *A*, labeled wild-type B oligonucleotide (lane 1) was used as a probe with CES cells nuclear extracts (lane 2). Complex is indicated by a black arrow. Competition with 100-fold molar excess of unlabeled mutated or wild-type B oligonucleotides is shown on lanes 3 and 4. Anti-CP2 and rabbit preimmune sera (1 $\mu$ l) were incubated with nuclear extract before the addition of B labeled probe (respectively lanes 5 and 6). *B*, Labeled wild-type A oligonucleotide was incubated with chicken ES cells nuclear extracts (lane 1). Complex is indicated by a black arrow. Anti-CP2 and rabbit preimmune sera were incubated with nuclear extract before the addition of A labeled probe (lanes 2 and 3).

Fig. 11: **Functional identification of a repressive sequence in the *cENS-1* promoter.** *A*, Three copies of the B response element were placed upstream of the -179/+83 (p179-luc)

or the SV40 promoter (pGL2-promoter) in luciferase expressing vectors and were transfected into CES and QT6 cells. Values are relative to the activity obtained with the p179-luc or the pGL2-promoter constructs. *B*, Deletion between position -238/-32 was generated in the promoter region of p456-luc vector. The mutant and wild-type constructs were transfected into CES and QT6 cells. Firefly luciferase activity was normalized for transfection efficiency using the renilla luciferase activity. Values are relative to the activity obtained with the p456-luc construct in CES cells for three independent experiments, each performed in triplicate.

**Fig.12: Expression of CP2 in chicken embryo.** Whole-mount in situ hybridisation of *CP2* mRNA. Embryos were hybridised to a RNA antisense located in the coding sequence of *CP2*. *A*, Stage X embryo (EG) dorsal side, a.o.: area opaca, m.z.: marginal zone. *B*, Stage XIII embryo (EG) dorsal side, k.s.: koller's sickle. *C*, Stage 3 embryo dorsal side, p.s.: primitive streak, p.n.p.: prospective neural plate. *D*, Stage 3+ embryo dorsal side. *E*, stage 4+ embryo, n.n.e.: non neural ectoderm, n.p.: neural plate. *F*, Stage 5 embryo, ch.: chord. *G*, Stage 7+ embryo, n.f.: neural fold, s.: somite. *H*, Stage 8 embryo., *I*, Stage 9 embryo. *J*, Stage 12 embryo.

## References

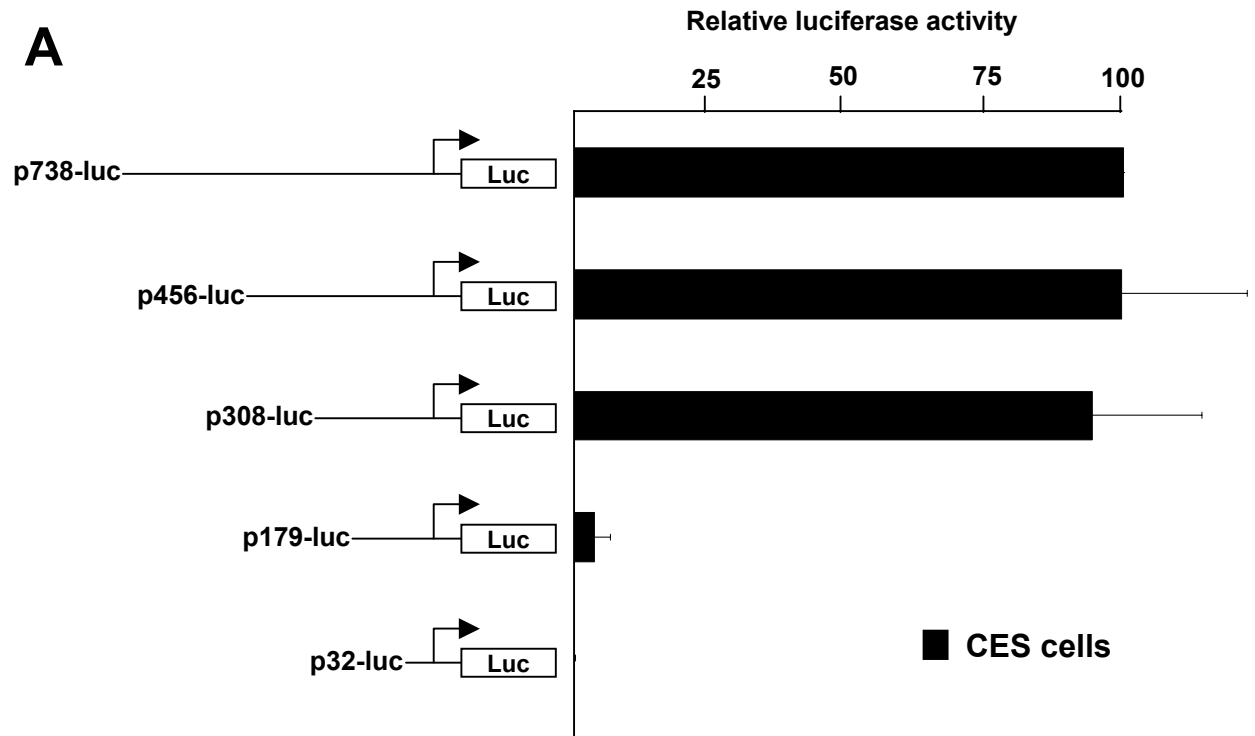
1. Eyal-Giladi, H., and Kochav, S. (1976) *Dev Biol* **49**, 321-337
2. Pain, B., Clark, M. E., Shen, M., Nakazawa, H., Sakurai, M., Samarut, J., and Etches, R. J. (1996) *Development* **122**, 2339-2348
3. Niwa, H. (2001) *Cell Struct Funct* **26**, 137-148
4. Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) *Genes Dev* **17**, 126-140
5. Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S., and Labosky, P. A. (2002) *Genes Dev* **16**, 2650-2661
6. Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M., and Goldfarb, M. (1995) *Science* **267**, 246-249
7. Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K., and Lonai, P. (1998) *Proc Natl Acad Sci U S A* **95**, 5082-5087
8. Acloque, H., Risson, V., Birot, A. M., Kunita, R., Pain, B., and Samarut, J. (2001) *Mech Dev* **103**, 79-91
9. Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A., and Stern, C. D. (2000) *Nature* **406**, 74-78
10. Jahner, D., Stuhlmann, H., Stewart, C. L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982) *Nature* **298**, 623-628
11. Gautsch, J. W., and Wilson, M. C. (1983) *Nature* **301**, 32-37
12. Pannell, D., and Ellis, J. (2001) *Rev Med Virol* **11**, 205-217
13. Reddy, S. T., Stoker, A. W., and Bissell, M. J. (1991) *Proc Natl Acad Sci U S A* **88**, 10505-10509
14. Moscovici, C., Moscovici, M. G., Jimenez, H., Lai, M. M., Hayman, M. J., and Vogt, P. K. (1977) *Cell* **11**, 95-103
15. Gruffat, H., Manet, E., Rigolet, A., and Sergeant, A. (1990) *Nucleic Acids Res* **18**, 6835-6843
16. Wilkinson, D. G., and Nieto, M. A. (1993) *Methods Enzymol* **225**, 361-373
17. Streit, A., Sockanathan, S., Perez, L., Rex, M., Scotting, P. J., Sharpe, P. T., Lovell-Badge, R., and Stern, C. D. (1997) *Development* **124**, 1191-1202
18. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res* **23**, 4878-4884
19. Yoon, J. B., Li, G., and Roeder, R. G. (1994) *Mol Cell Biol* **14**, 1776-1785
20. Murata, T., Nitta, M., and Yasuda, K. (1998) *Genes Cells* **3**, 443-457
21. Lim, L. C., Swendeman, S. L., and Sheffery, M. (1992) *Mol Cell Biol* **12**, 828-835
22. Fuhrmann, G., Chung, A. C., Jackson, K. J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H. R., and Cooney, A. J. (2001) *Dev Cell* **1**, 377-387
23. Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H. R. (1996) *Development* **122**, 881-894
24. Barnea, E., and Bergman, Y. (2000) *J Biol Chem* **275**, 6608-6619
25. Bray, S. J., and Kafatos, F. C. (1991) *Genes Dev* **5**, 1672-1683
26. Liaw, G. J., Rudolph, K. M., Huang, J. D., Dubnicoff, T., Courey, A. J., and Lengyel, J. A. (1995) *Genes Dev* **9**, 3163-3176
27. Huang, J. D., Dubnicoff, T., Liaw, G. J., Bai, Y., Valentine, S. A., Shirokawa, J. M., Lengyel, J. A., and Courey, A. J. (1995) *Genes Dev* **9**, 3177-3189
28. Wilanowski, T., Tuckfield, A., Cerruti, L., O'Connell, S., Saint, R., Parekh, V., Tao, J., Cunningham, J. M., and Jane, S. M. (2002) *Mech Dev* **114**, 37-50
29. Huang, H. C., Sundseth, R., and Hansen, U. (1990) *Genes Dev* **4**, 287-298
30. Powell, C. M., Rudge, T. L., Zhu, Q., Johnson, L. F., and Hansen, U. (2000) *Embo J* **19**, 4665-4675
31. Bing, Z., Reddy, S. A., Ren, Y., Qin, J., and Liao, W. S. (1999) *J Biol Chem* **274**, 24649-24656
32. Casolaro, V., Keane-Myers, A. M., Swendeman, S. L., Steindler, C., Zhong, F., Sheffery, M., Georas, S. N., and Ono, S. J. (2000) *J Biol Chem* **275**, 36605-36611
33. Zhang, Y., Niu, Z., Cohen, A. J., and Adams, S. L. (1999) *Nucleic Acids Res* **27**, 4090-4099
34. Kang, J. S., Lee, H. B., Rhee, S. G., Park, K., and Yoo, O. J. (1997) *Gene* **197**, 19-28
35. Jane, S. M., Nienhuis, A. W., and Cunningham, J. M. (1995) *Embo J* **14**, 97-105
36. Zhou, W., Clouston, D. R., Wang, X., Cerruti, L., Cunningham, J. M., and Jane, S. M. (2000) *Mol Cell Biol* **20**, 7662-7672
37. Rodda, S., Sharma, S., Scherer, M., Chapman, G., and Rathjen, P. (2001) *J Biol Chem* **276**, 3324-3332
38. Ducy, P., and Karsenty, G. (1995) *Mol Cell Biol* **15**, 1858-1869
39. Carr, B. A., Wan, J., Hines, R. N., and Yost, G. S. (2003) *J Biol Chem* **278**, 15473-15483
40. Lim, L. C., Fang, L., Swendeman, S. L., and Sheffery, M. (1993) *J Biol Chem* **268**, 18008-18017

41. Andres, M. E., Burger, C., Peral-Rubio, M. J., Battaglioli, E., Anderson, M. E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999) *Proc Natl Acad Sci U S A* **96**, 9873-9878
42. You, A., Tong, J. K., Grozinger, C. M., and Schreiber, S. L. (2001) *Proc Natl Acad Sci U S A* **98**, 1454-1458

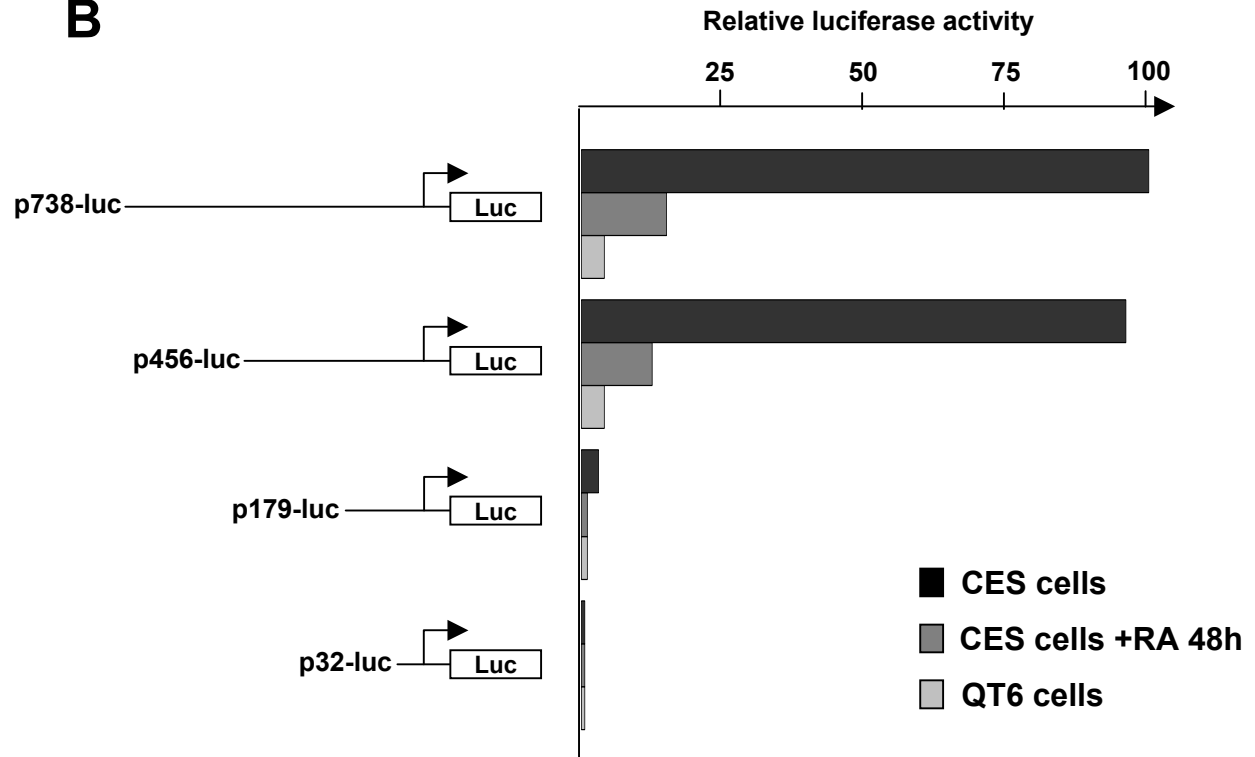


**Figure2**

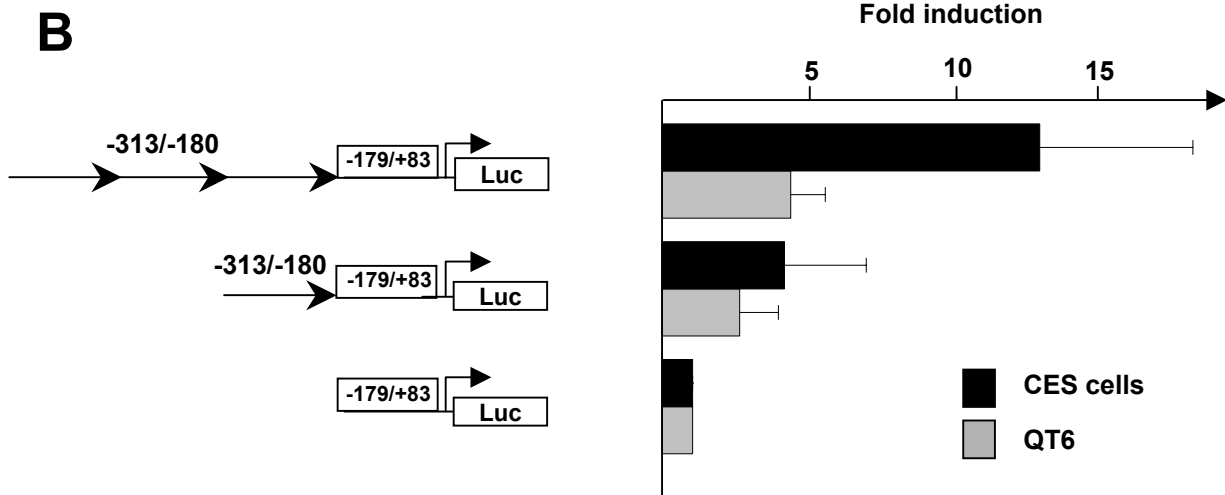
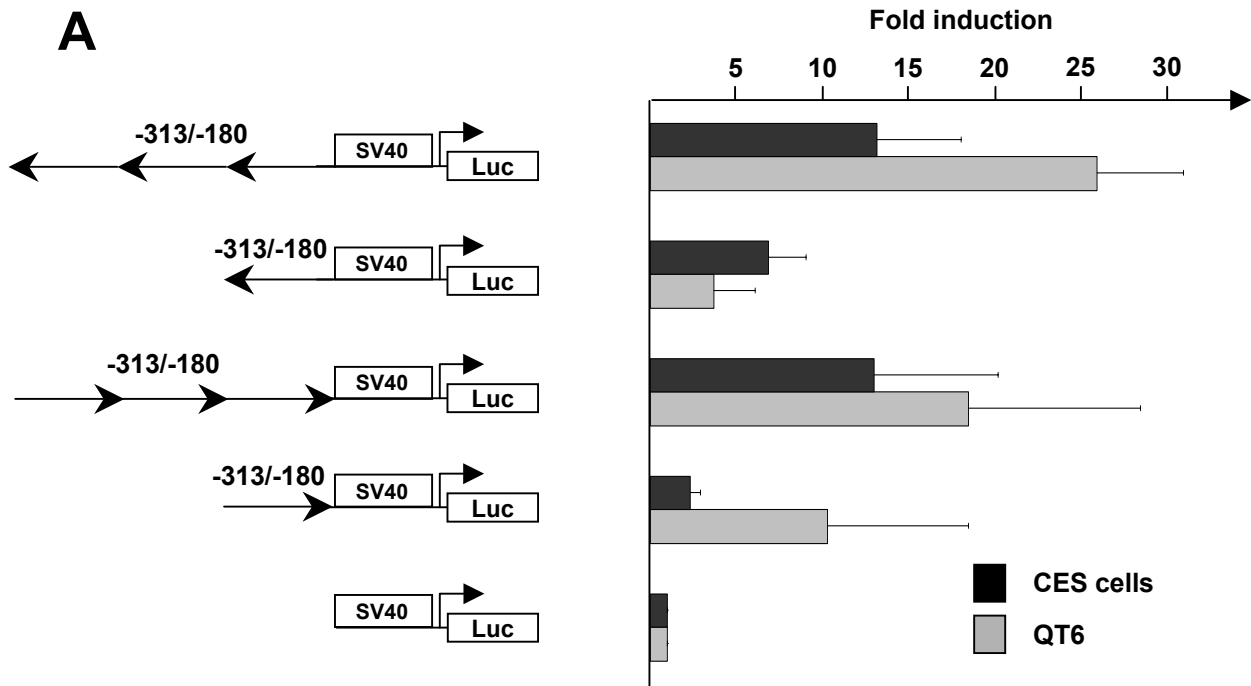
**A**



**B**

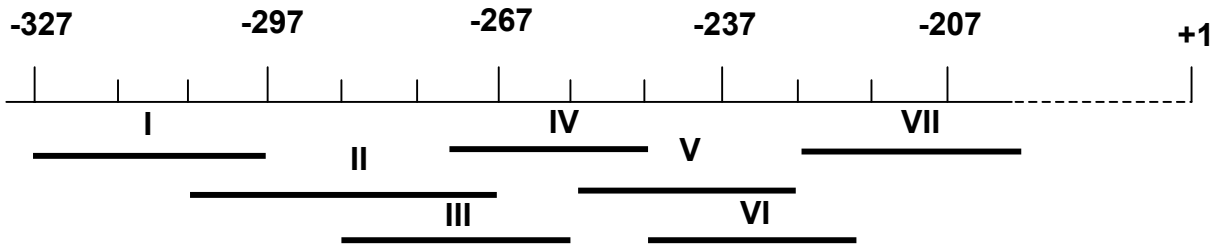


**Figure 3**



**Figure 4**

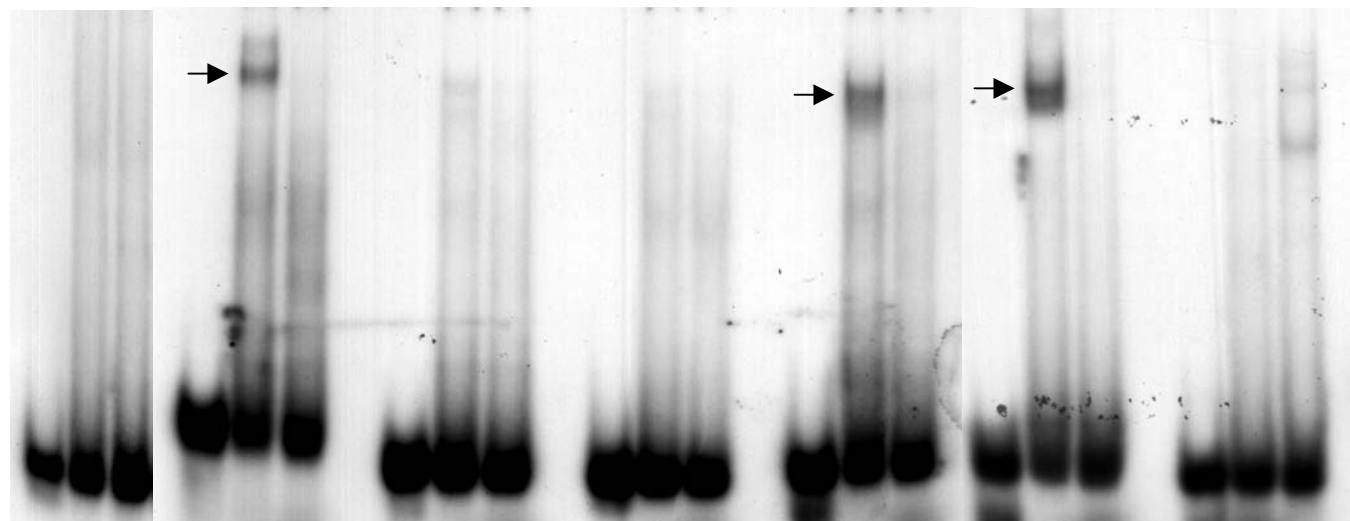
**A**



**B**

I                      II                      III                      IV                      V                      VI                      VII

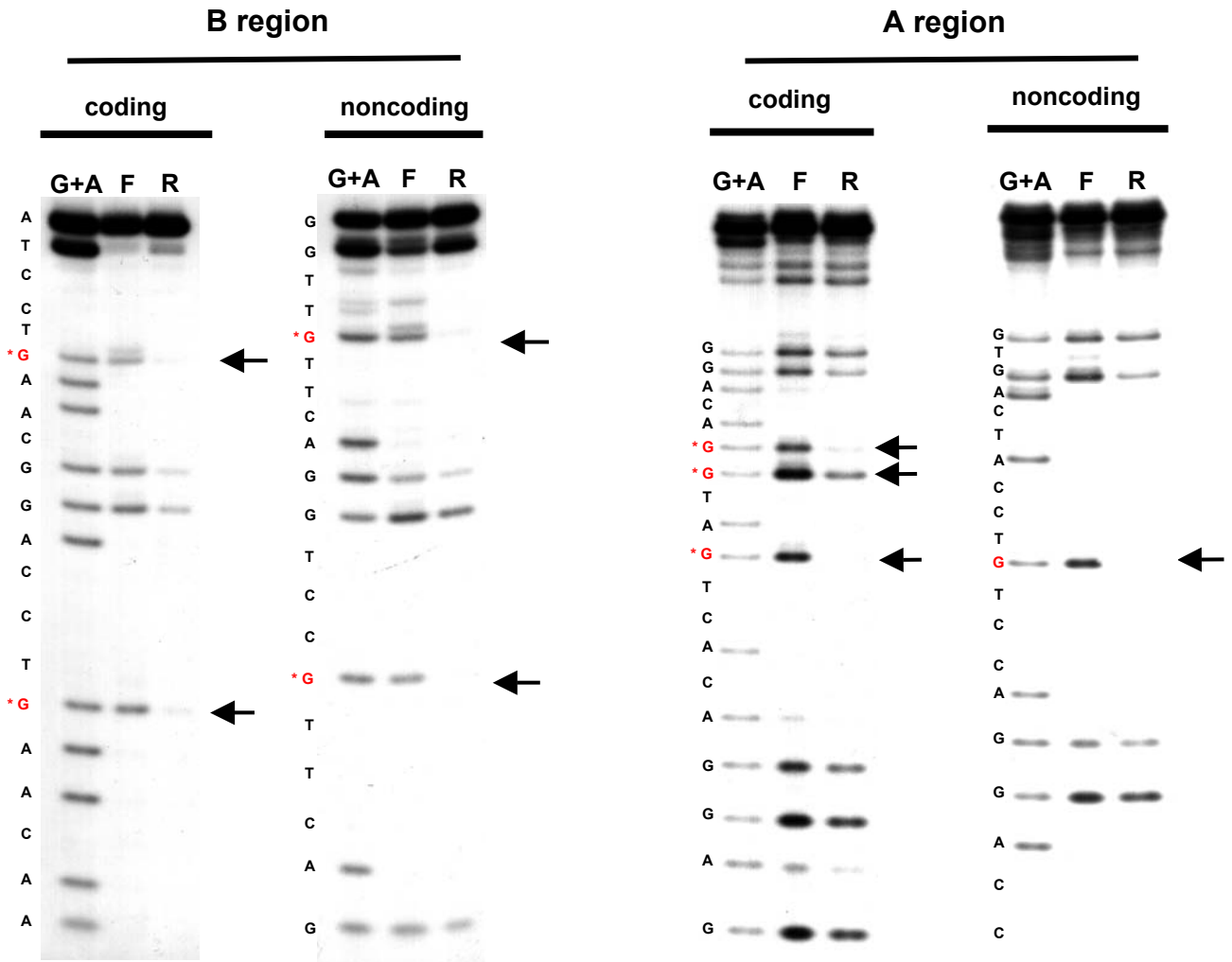
1 2 3                      4 5 6                      7 8 9                      10 11 12                      13 14 15                      16 17 18                      19 20 21



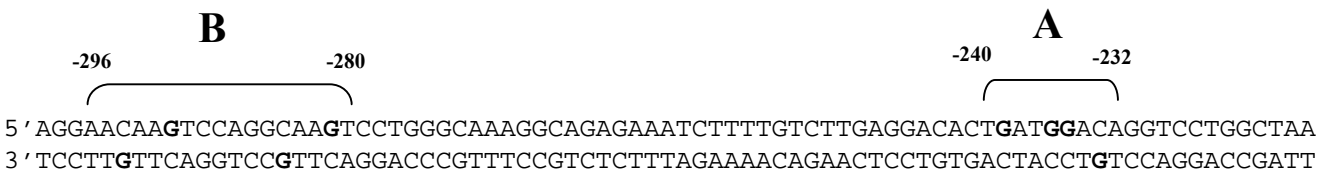


**Figure 5**

**A**



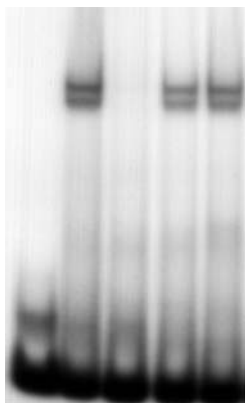
**B**



**Figure 6**

**A**

1 2 3 4 5



**B**

Wild type A : CAC TGATGGACA GGAGCT

first A mutant : CAC **A\*C\*T**\*TGGACA GGAGCT

second A mutant : CAC TGATGGT\***G**\*T\* GGAGCT

**Figure 7**

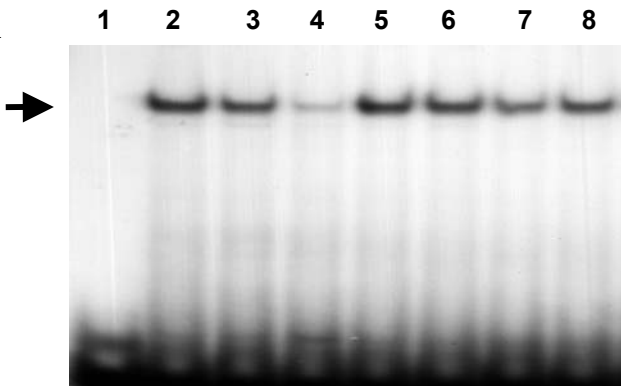
**A**

**B RE probe:** GAA C<sub>1</sub>A<sub>2</sub>A<sub>3</sub>G<sub>4</sub>T<sub>5</sub>C<sub>6</sub>C<sub>7</sub>AGGC<sub>11</sub>AAG<sub>14</sub>T CCTA

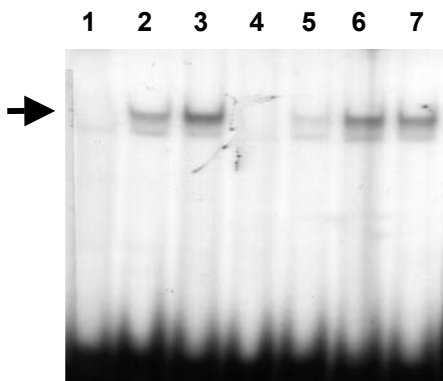
**B RE mutated C<sub>1</sub>:** GAA T\*AAGTCCAGGCAAGT CCTA  
**B RE mutated A<sub>2</sub>:** GAA CG\*AGTCCAGGCAAGT CCTA  
**B RE mutated A<sub>3</sub>:** GAA CAG\*GTCCAGGCAAGT CCTA  
**B RE mutated G<sub>4</sub>:** GAA CAAA\*TCCAGGCAAGT CCTA  
**B RE mutated T<sub>5</sub>:** GAA CAAGC\*CCAGGCAAGT CCTA  
**B RE mutated C<sub>6</sub>:** GAA CAAGTT\*CAGGCAAGT CCTA  
**B RE mutated C<sub>7</sub>:** GAA CAAGTCT\*AGGCAAGT CCTA  
**B RE mutated C<sub>11</sub>:** GAA CAAGTCCAGGT\*AAGT CCTA  
**B RE mutated G<sub>14</sub>:** GAA CAAGTCCAGGCAAA\*T CCTA

**B RE -3bp probe:** GAA CAAGTCCCAAGT CCTA  
**B RE +3bp probe:** GAA CAAGTCCAGGAGGCAAGT CCTA  
**B RE +6bp probe:** GAA CAAGTCCAGGAGGAGGCAAGT CCTA

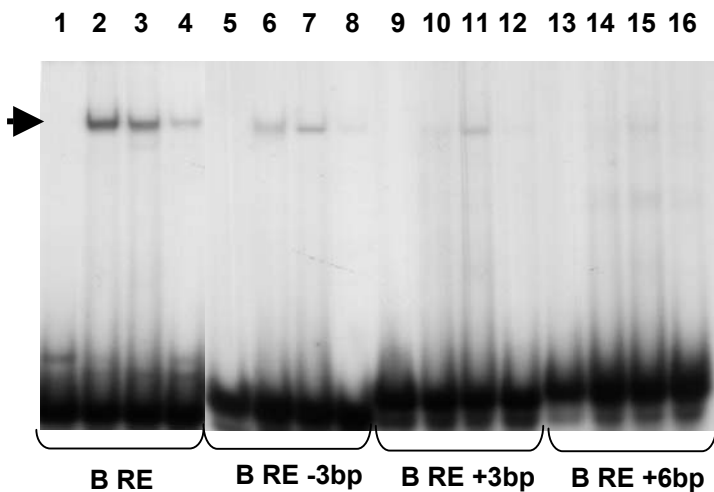
**B**



**C**

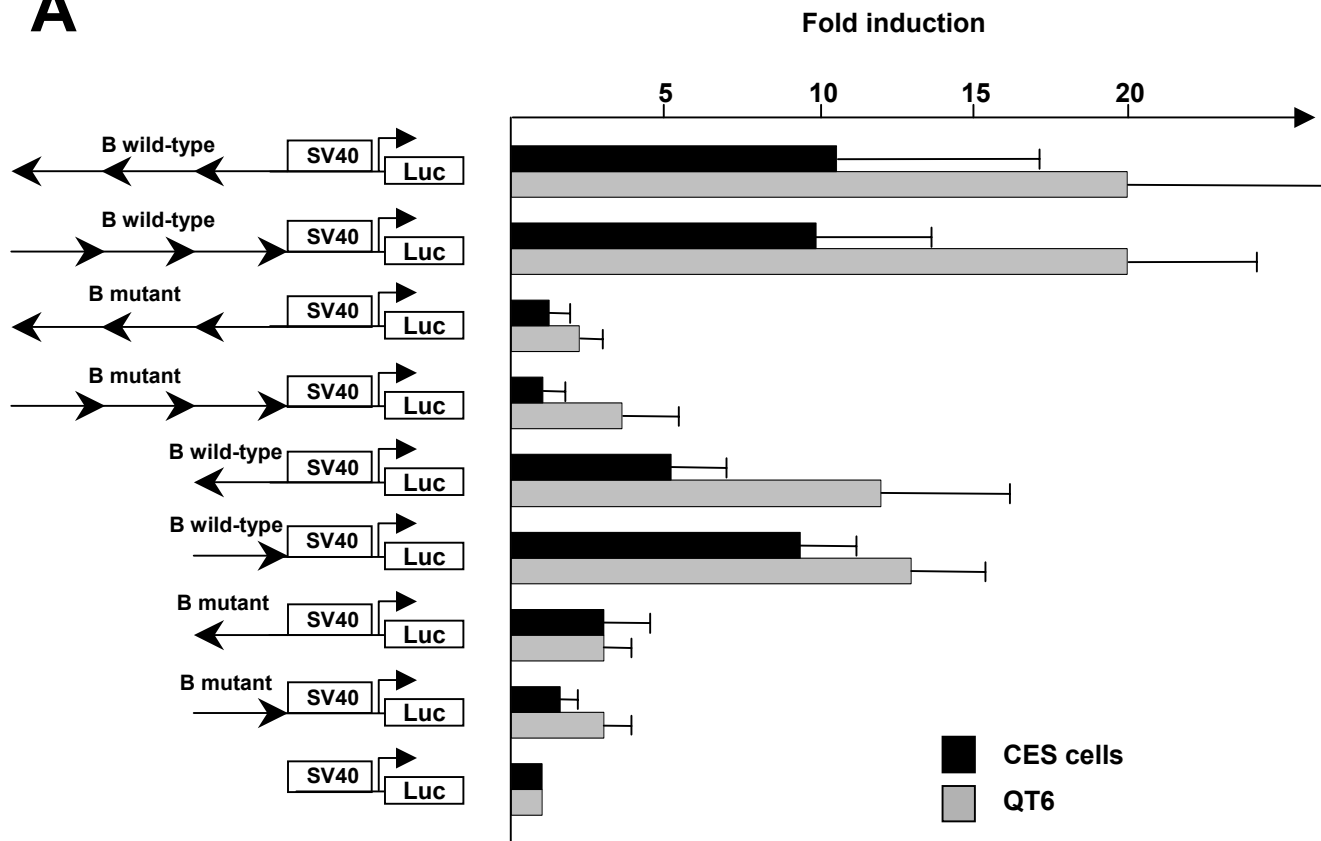


**D**

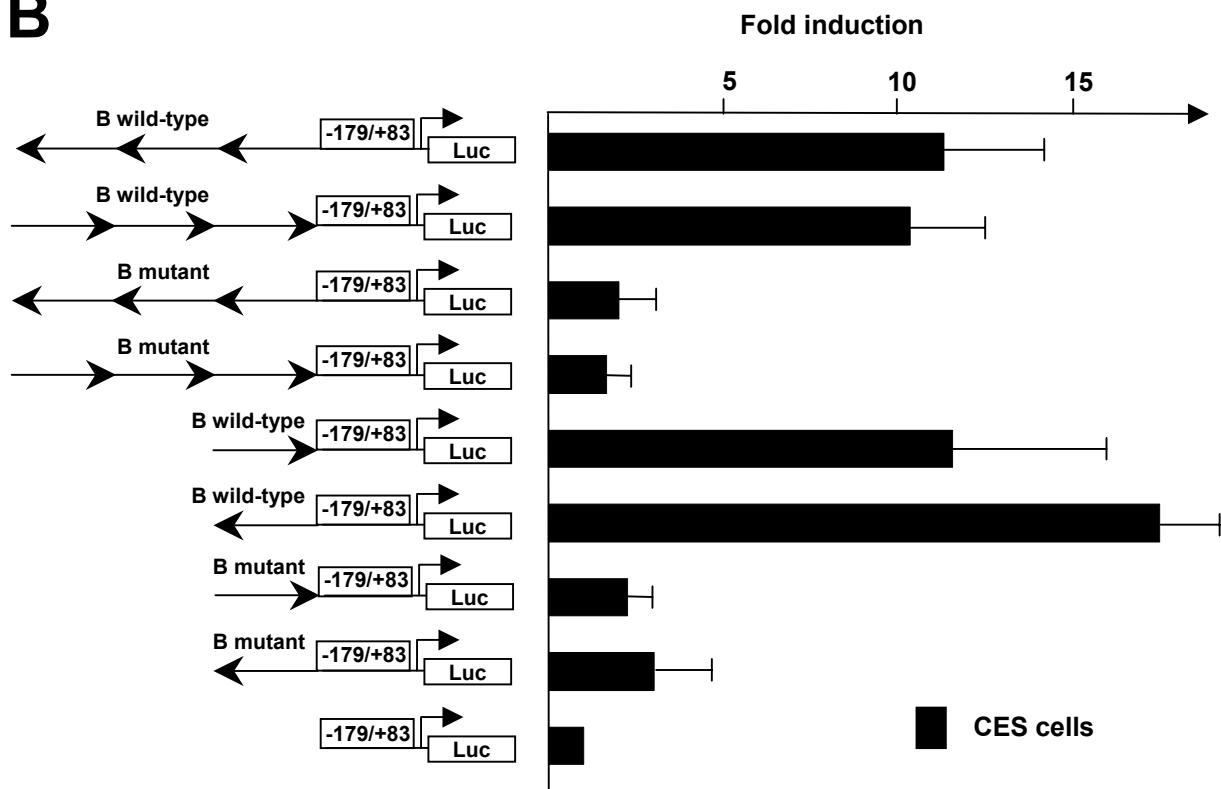


**Figure 8**

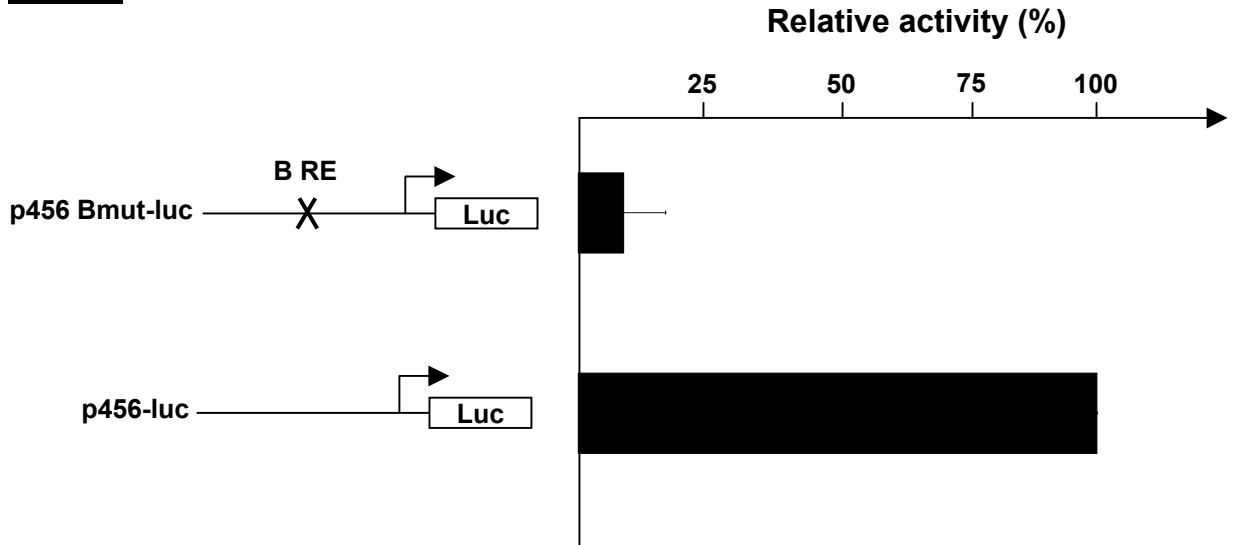
**A**



**B**

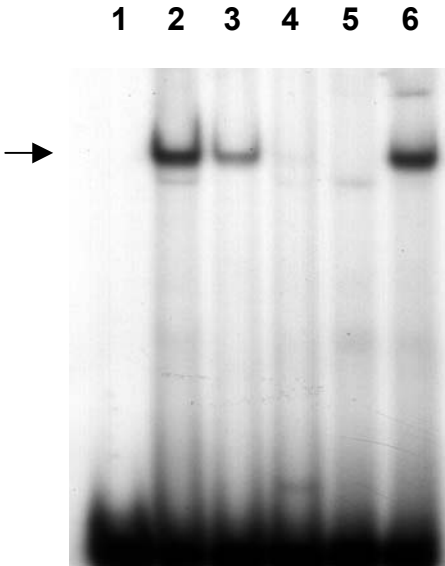


**Figure 9**

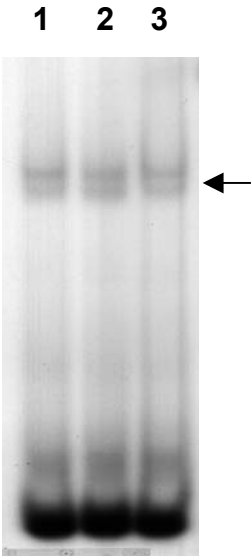


**Figure 10**

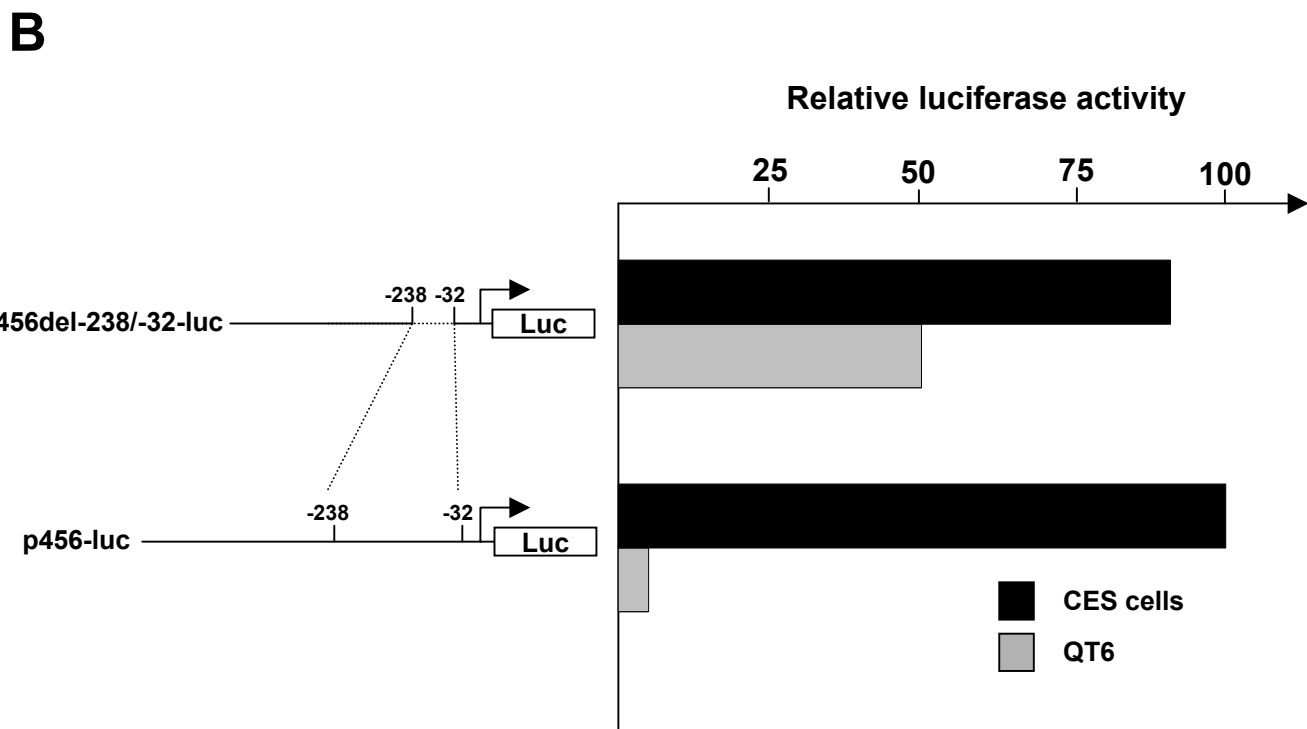
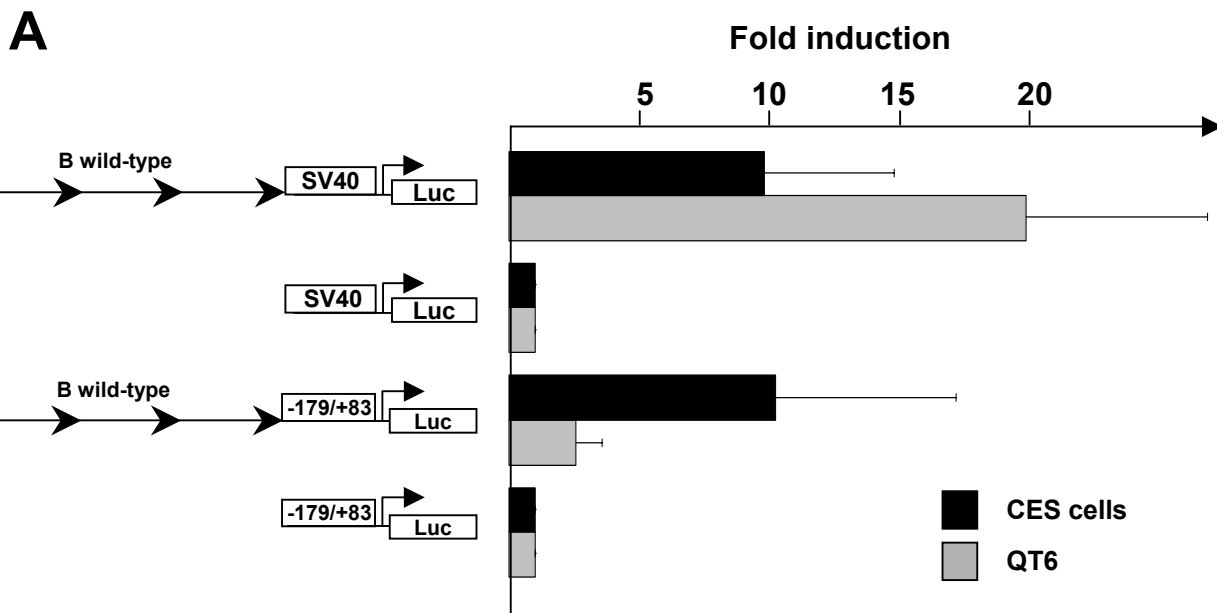
**A**

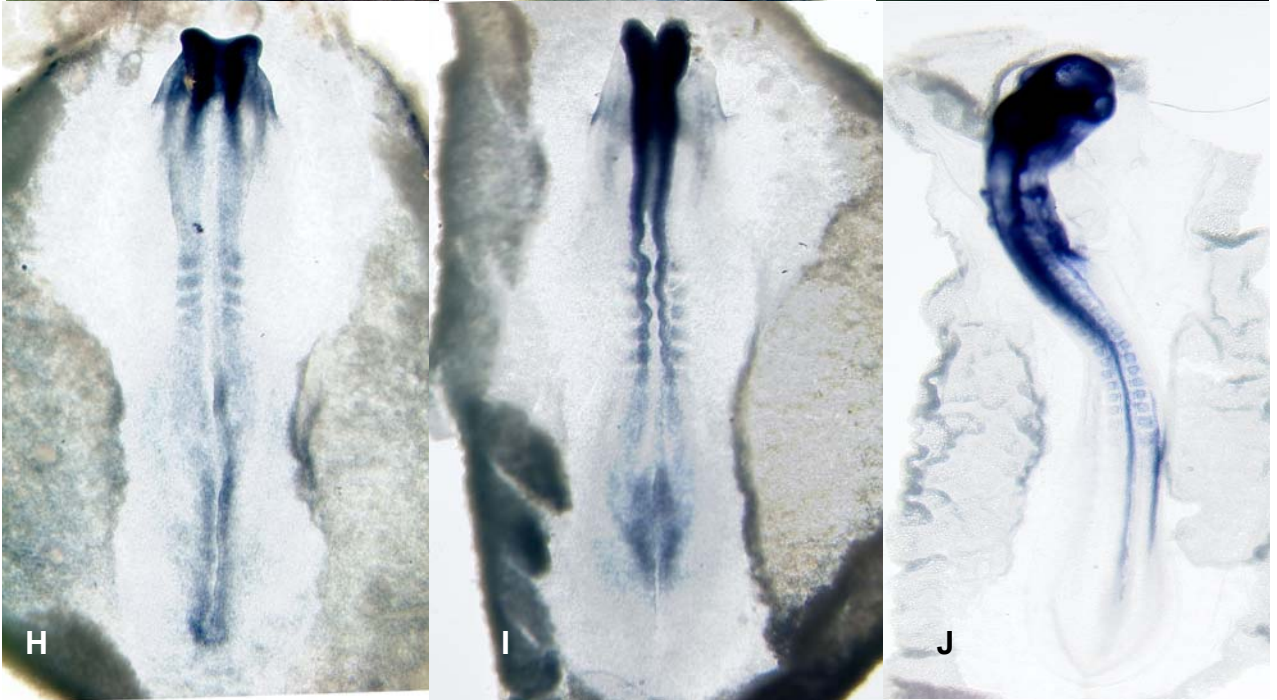
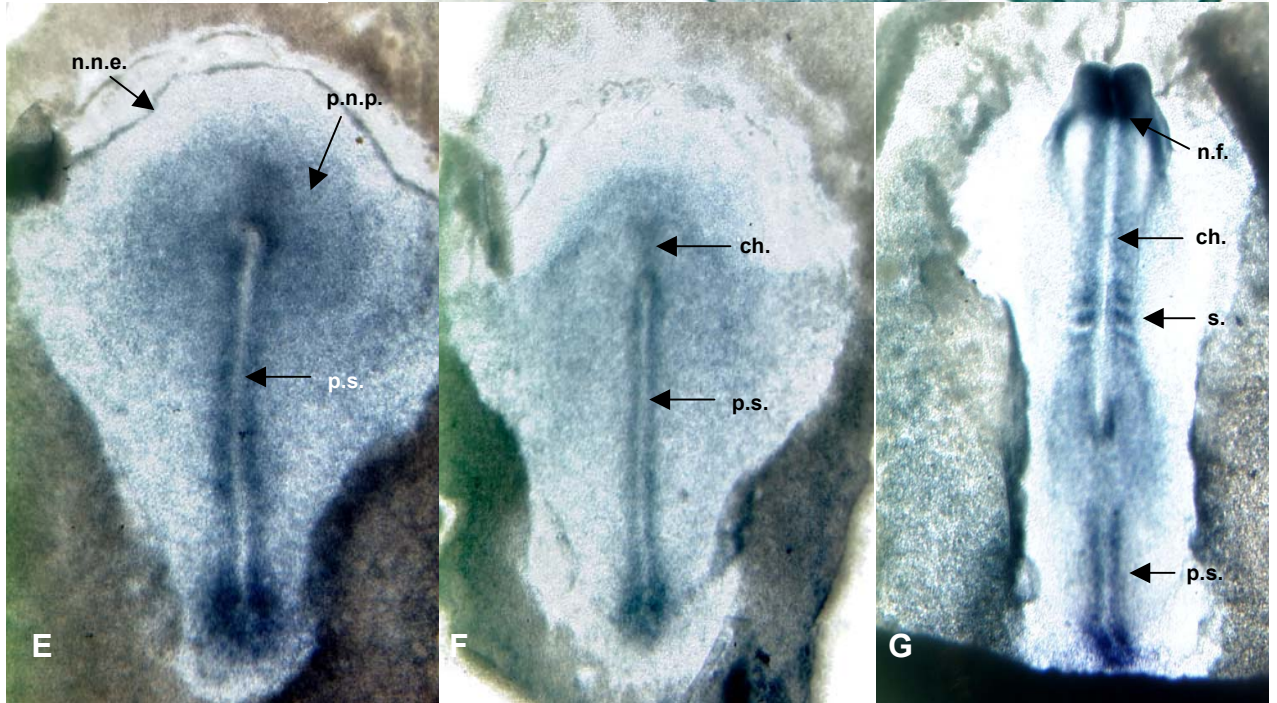
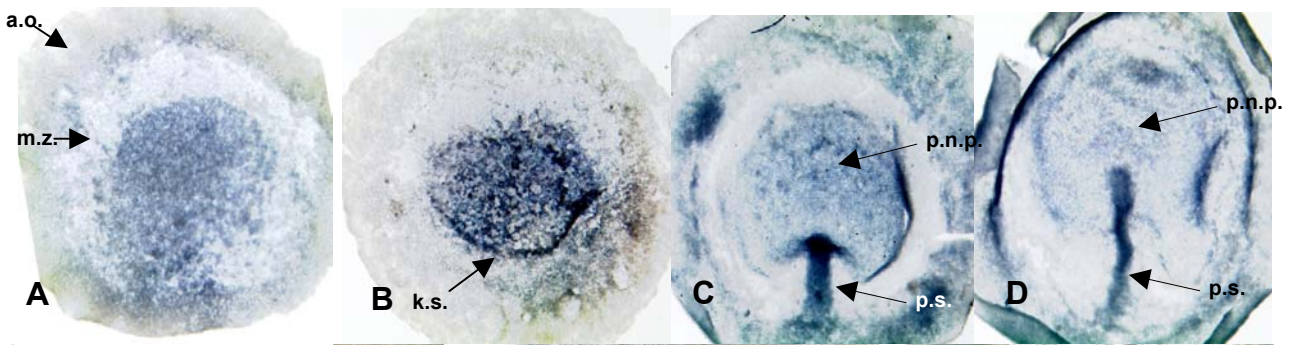


**B**



**Figure 11**







# FIGURES

Aureste, l'œuf n'est à proprement parler que ce qu'on appelle *placenta*, dont l'enfant, après y avoir demeuré un certain temps tout courbé & comme en peloton, brise en s'étendant & en s'allongeant le plus qu'il peut, les membranes qui le couvroient, & posant ses pieds contre le *placenta*, qui reste attaché au fond de la matrice, se pousse ainsi avec la tête hors de sa prison; en quoi il est aidé par la mere, qui agitée par la douleur qu'elle en sent, pousse le fond de la matrice en bas, & donne par consequent d'autant plus d'occasion à cet enfant de

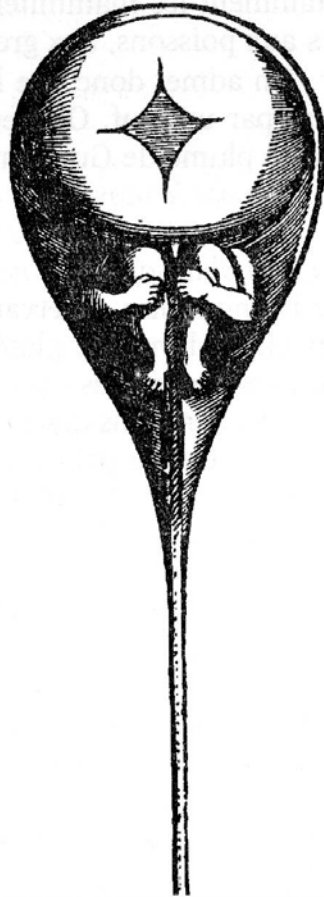


Figure 1 : *Fœtus humain dans un spermatozoïde: l'homunculus* (Nicolas Hartsoeker, 1694)

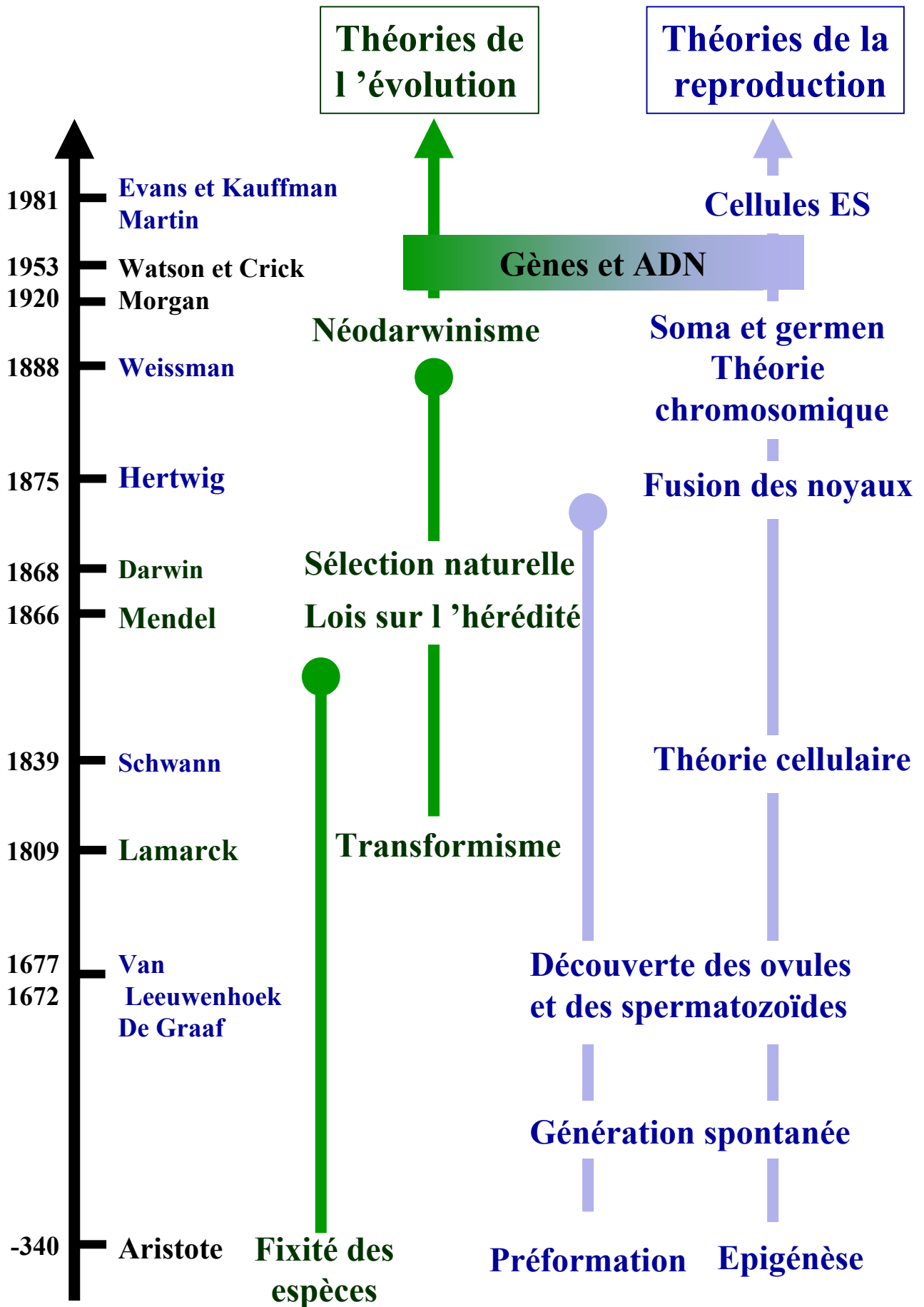
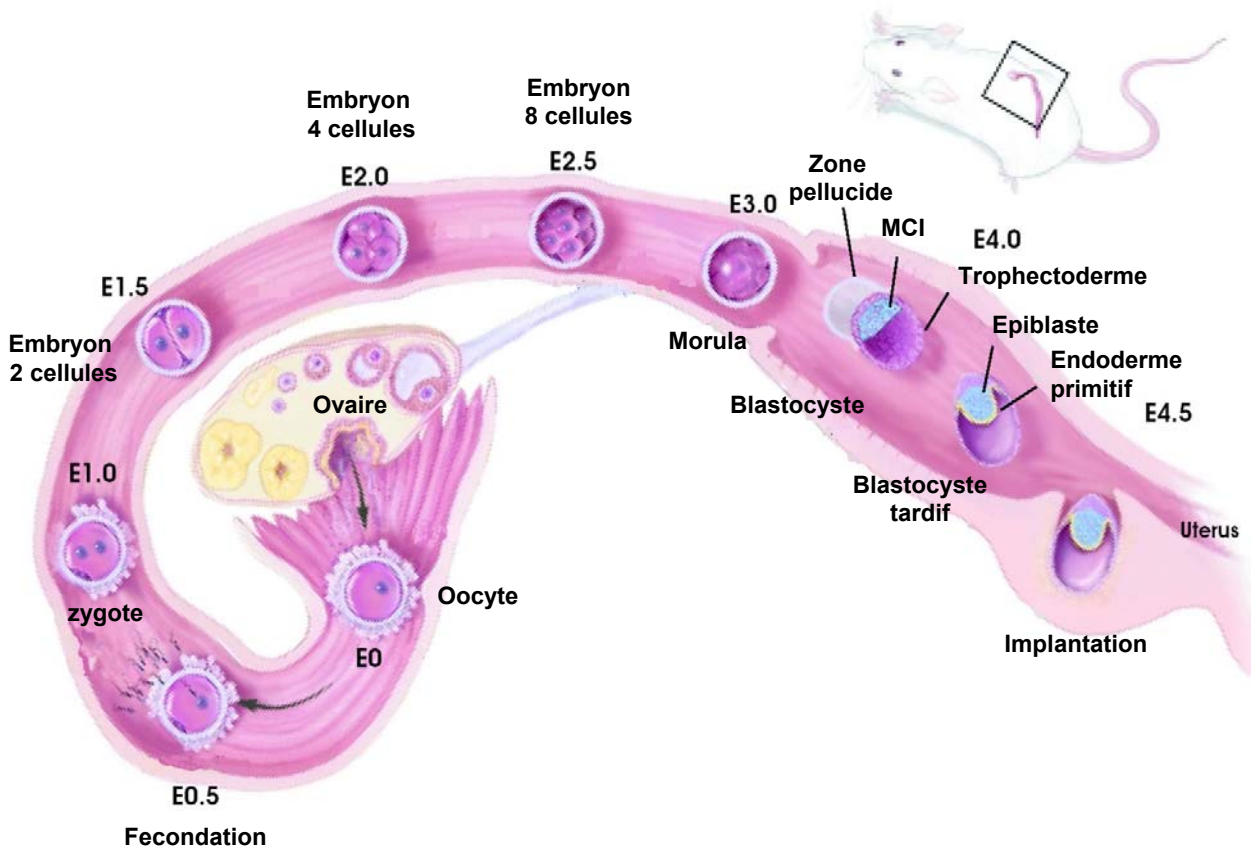
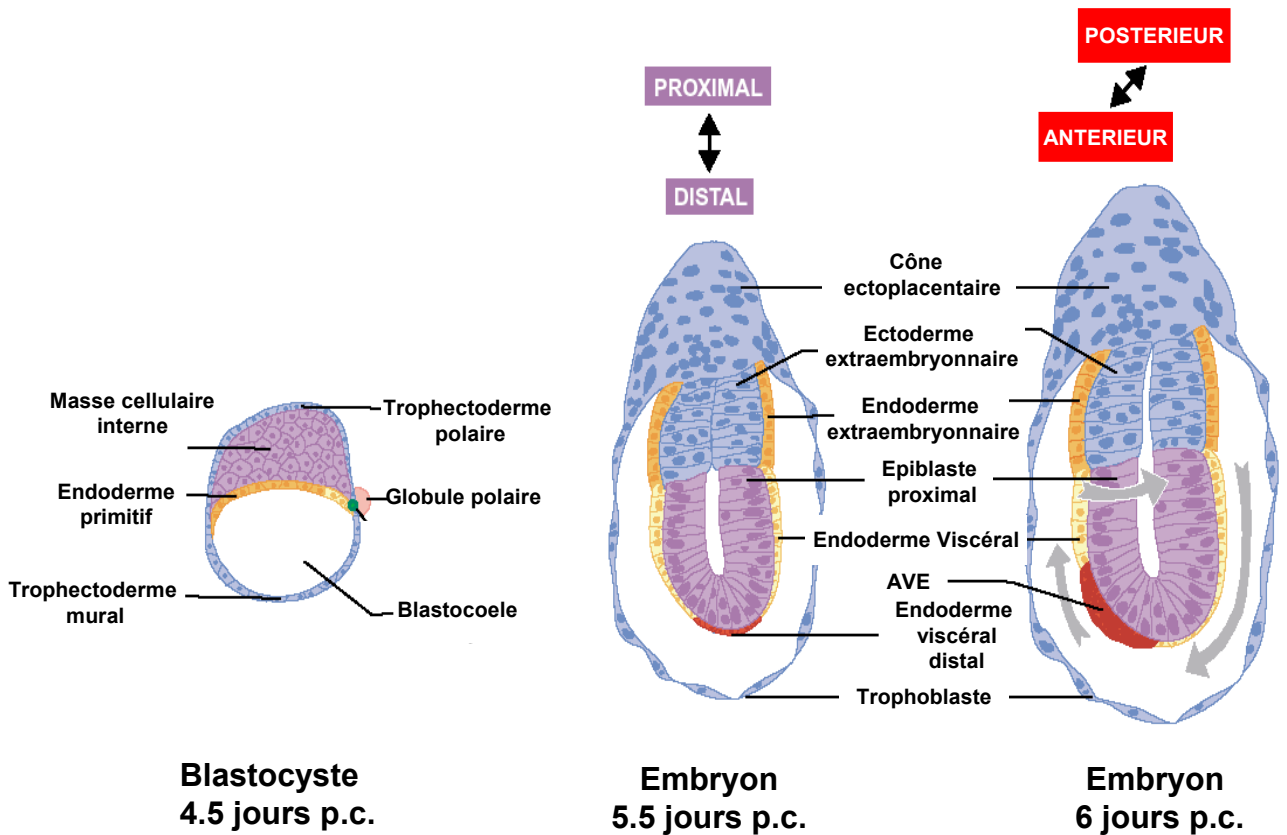


Figure 2: schéma récapitulatif illustrant les grands courants de pensées et quelques découvertes majeures concernant les théories sur l'évolution et sur la reproduction des espèces animales



**Figure 3: Le développement précoce de l'embryon de souris du jour 0 post coïtum (E0) au jour 5 (E5)**

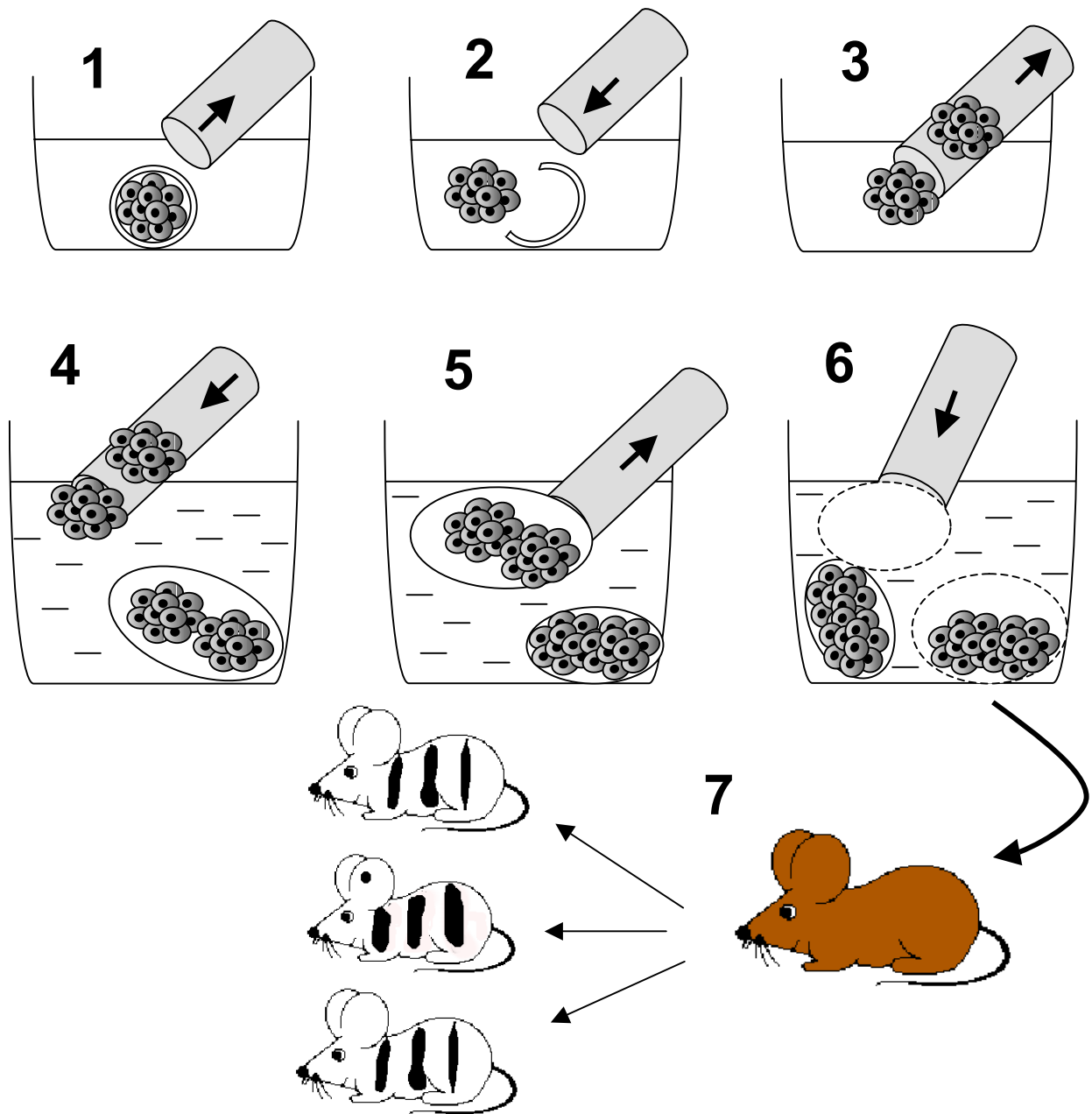
Cette figure récapitule les différentes étapes du développement précoce de la souris et de leurs localisations spatiales dans l'appareil génital femelle, depuis la fécondation jusqu'à l'implantation du blastocyste dans l'utérus (MCI: masse cellulaire interne).



**Figure 4: le développement péri-implantatoire de l'embryon de souris (d'après Lu et al. 2001)**

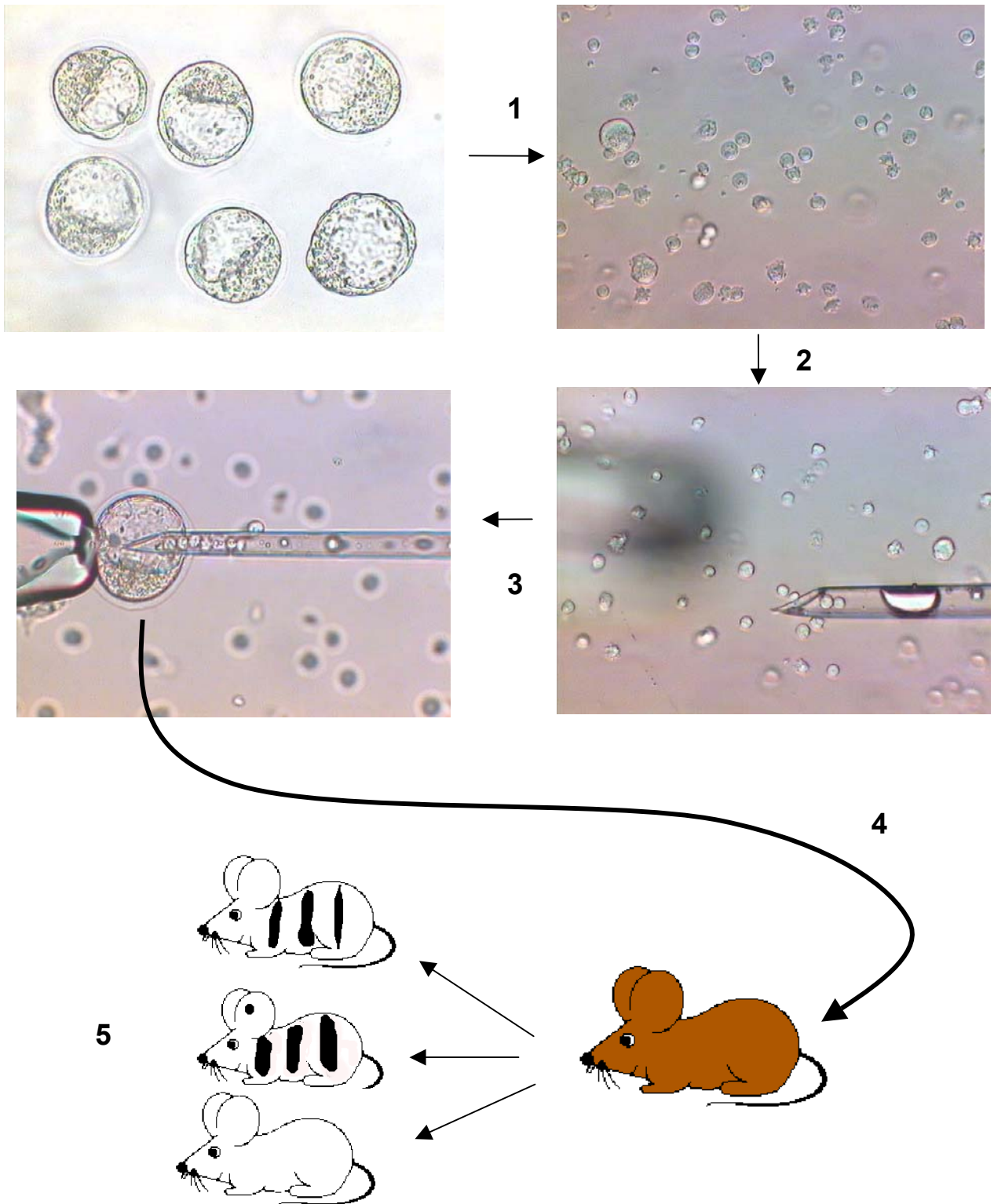
Le blastocyste va s'implanter dans la muqueuse utérine. Il est alors constitué de trois tissus différents: les cellules pluripotentes de la masse cellulaire interne (violet), l'endoderme primitif (en jaune orangé) et les cellules du trophoctoderme (en bleu).

Après l'implantation, les cellules de la masse cellulaire interne vont proliférer pour donner l'épiblaste (en violet), celles de l'endoderme primitif vont donner l'endoderme viscéral et l'endoderme extraembryonnaire. Une partie de l'endoderme viscéral (AVE pour Anterior Visceral Endoderm, en rouge) est important pour la mise en place de l'axe antéro-postérieur. Le trophoctoderme va donner les cellules du cône ectoplacentaire, les cellules de l'ectoderme extraembryonnaire et les cellules géantes du trophoblaste.



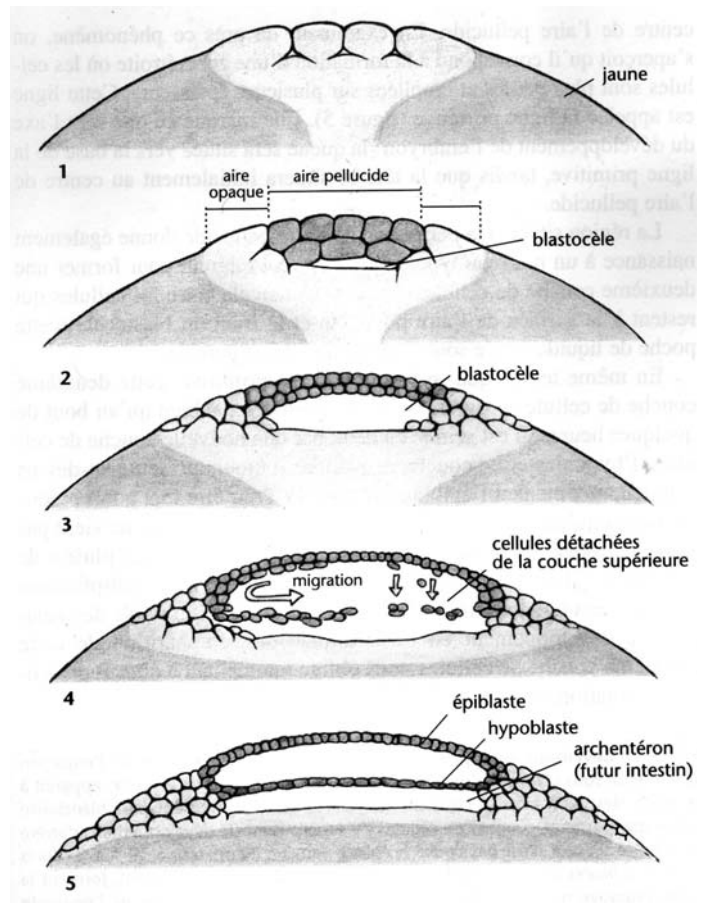
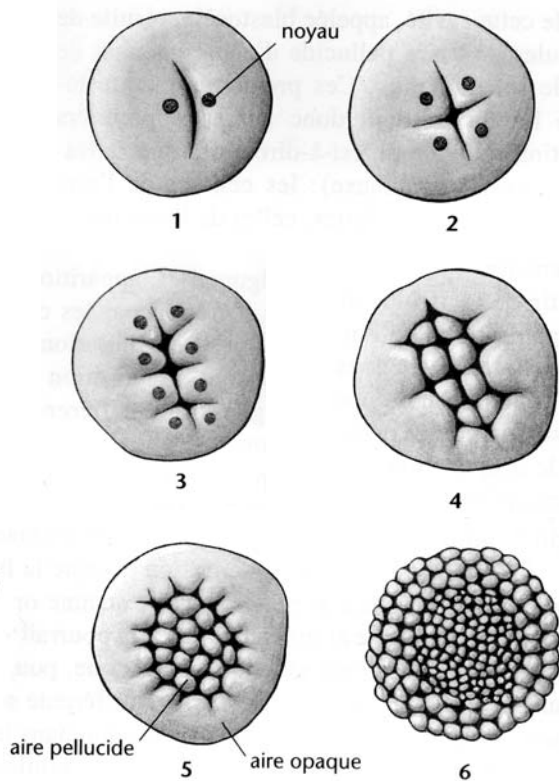
**Figure 5 : les chimères d'agrégation (Tarkowski 1962)**

- 1-2: La zone pellucide des embryons est cassée mécaniquement à l'aide d'une micropipette.
- 3: Les morulas (l'une issue d'une souris au pelage noir, l'autre d'une souris au pelage blanc), dont la zone pellucide a été détruite, sont groupées par deux et prélevées dans une micropipette.
- 4: Les deux morulas prélevées dans un milieu aqueux sont déposées dans un récipient contenant de l'huile minérale.
- 5: Le milieu aqueux est peu à peu aspiré par la micropipette entraînant la mise en contact des deux morulas
- 6: Une fois les deux morulas fusionnées, le milieu de culture est ajouté.
- 7: Transfert des embryons chimères dans une souris porteuse et obtention de souris mosaïques pour la couleur du pelage, blanc et noir.



**Figure 6 : les chimères d'injection (photos D. Aubert)**

- 1: les cellules des blastocystes donneurs (phénotype poil noir) sont dissociées par la trypsine.
- 2: les cellules sont prélevées dans une micropipette.
- 3: les cellules sont réinjectées à l'intérieur d'un blastocyste receveur (phénotype poil blanc).
- 4: le blastocyste receveur est réimplanté dans une souris porteuse.
- 5: certains souriceaux ont un pelage mosaïque noir et blanc. Ce sont des chimères composées de cellules de génotypes différents.



**Figure 7 : les premiers stades du développement de l'embryon de poulet (d'après Flamant 2001)**

**A: vue dorsale des embryons en cours de segmentation**

**1:** stade 2 cellules

**2:** stade 4 cellules

**3:** stades 8 cellules

**4:** stade 16 cellules (stade II (EG) selon Eyal-Giladi et Kochav) environ 5 heures après fécondation

**5 et 6:** stade blastula (respectivement stade VIII (EG) environ 20 heure après fécondation et X (EG) dans un œuf fraîchement pondu): on distingue la formation de deux tissus constituant l'aire pellucide et l'aire opaque.

**B: vue en coupe transversale des embryons**

**1:** stade 8 cellules

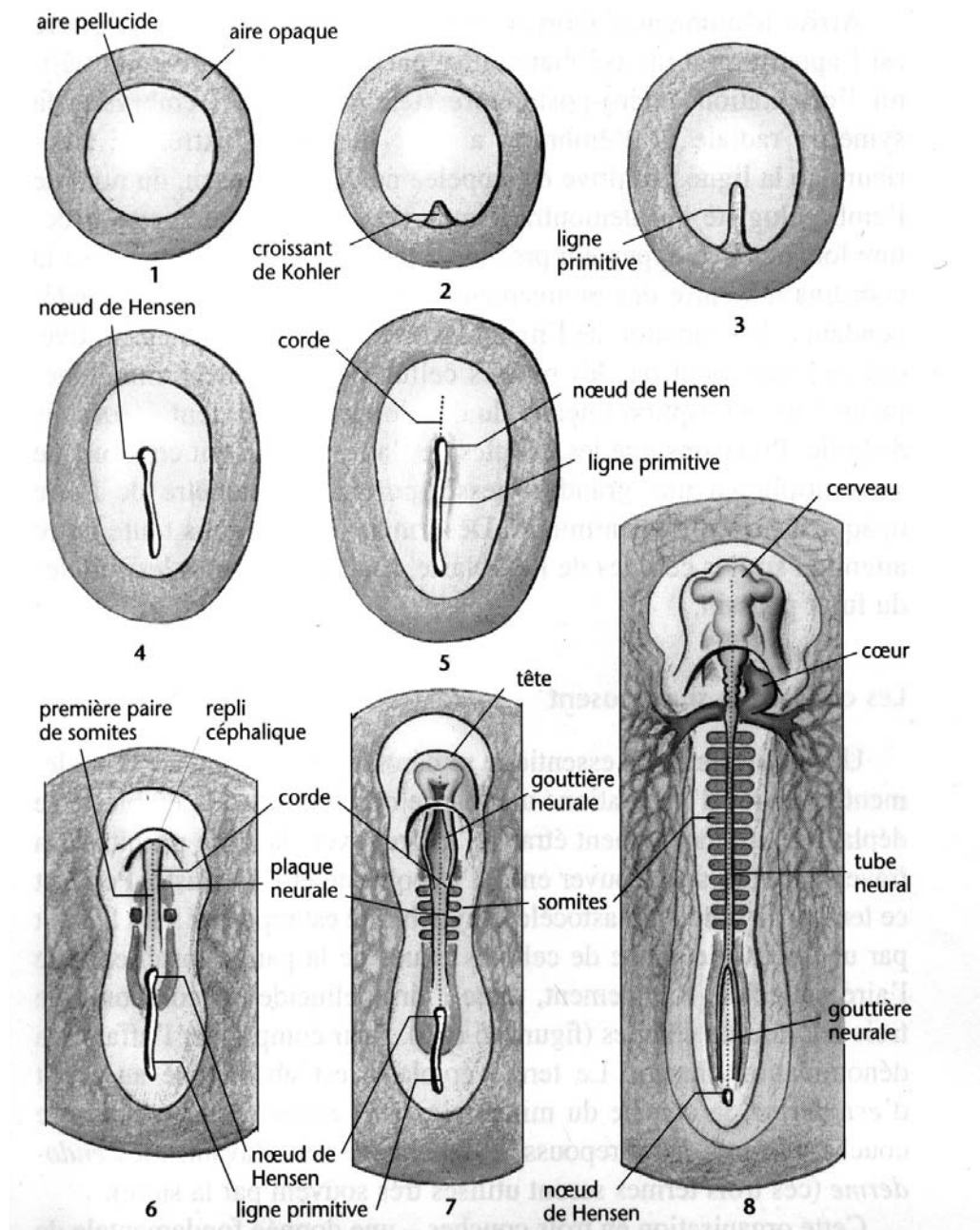
**2:** stade 16-32 cellules: on remarque la formation de l'aire opaque et de l'aire pellucide ainsi que de la cavité appelée blastocoele.

**3:** blastula de poulet au moment de la ponte (stade X (EG)). L'embryon n'est encore formé que d'une couche de cellule, l'épiblaste, séparé du vitellus par une cavité, le blastocoele.

**4-5:** blastulas de poulet (respectivement stade XII et stade XIII soit 3 à 6h après la ponte ).

Formation de l'hypoblaste dans la blastula de poulet par délamination et migration des cellules de l'épiblaste.





**Figure 8: Développement précoce de l'embryon de poulet après la ponte (d'après Flamant 2001)**

**1:** stade X (EG): blastula de poulet au moment de la ponte

**2:** stade XIV (EG) ou stade 2 (Hamburger et Hamilton) environ 6-7h après la ponte: blastula de poulet constitué par l'hypoblaste et de l'épiblaste. Formation de la ligne primitive. Début de la gastrulation.

**3:** stade 3 (HH) environ 12-13h après la ponte: la ligne primitive s'étend.

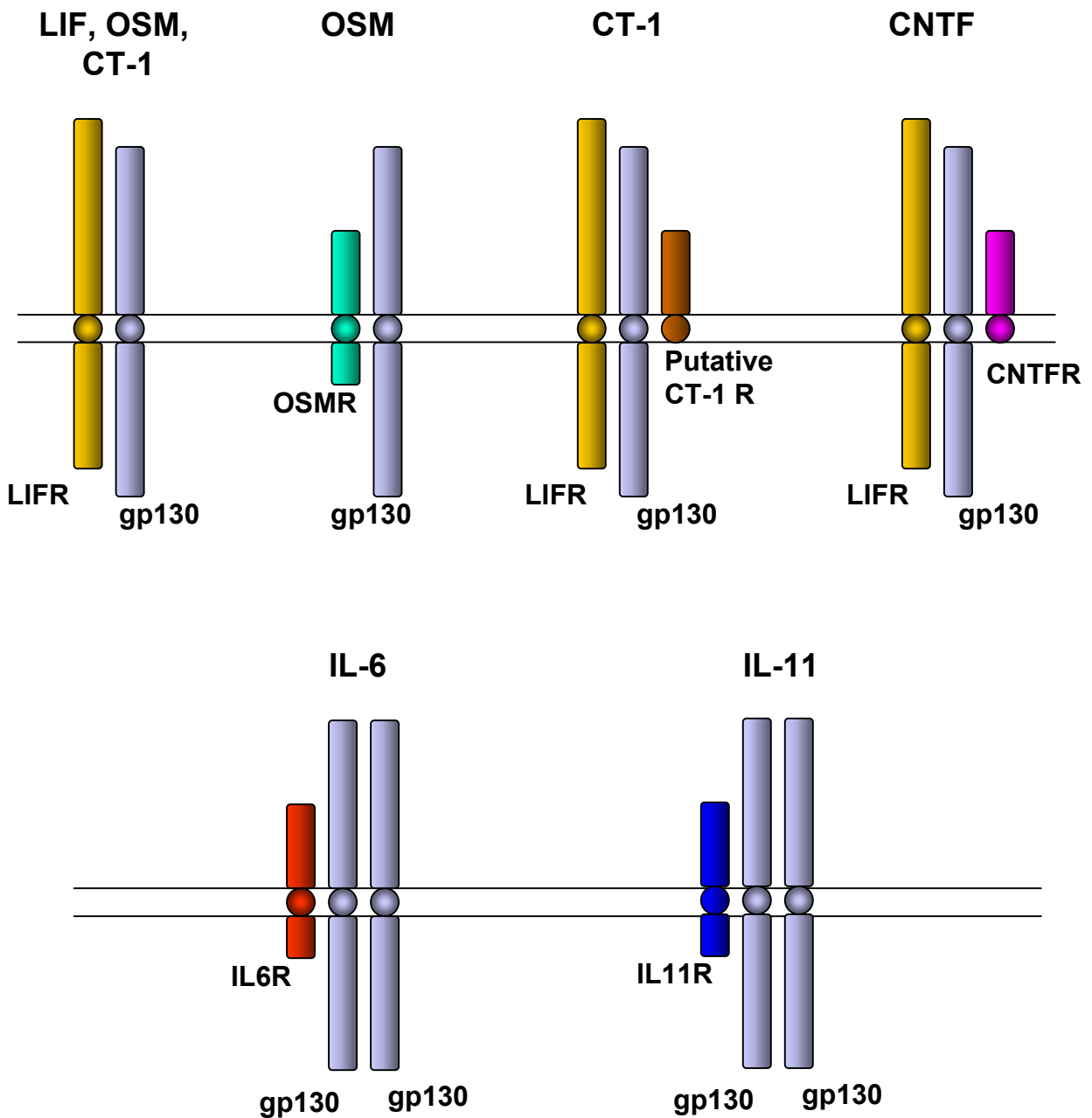
**4:** stade 4 (HH) environ 18-19h après la ponte: la ligne primitive atteint sa taille maximale, le nœud de Hensen est formé.

**5:** stade 5-6 (HH) environ 20-24h après la ponte: régression de la ligne primitive et formation de la corde et de la tête.

**6:** stade 7 (HH) environ 26h après la ponte: formation de la première paire de somites, du tube neural et du repli céphalique.

**7:** stade 9 (HH) environ 30-33h après la ponte: 8 paires de somites.

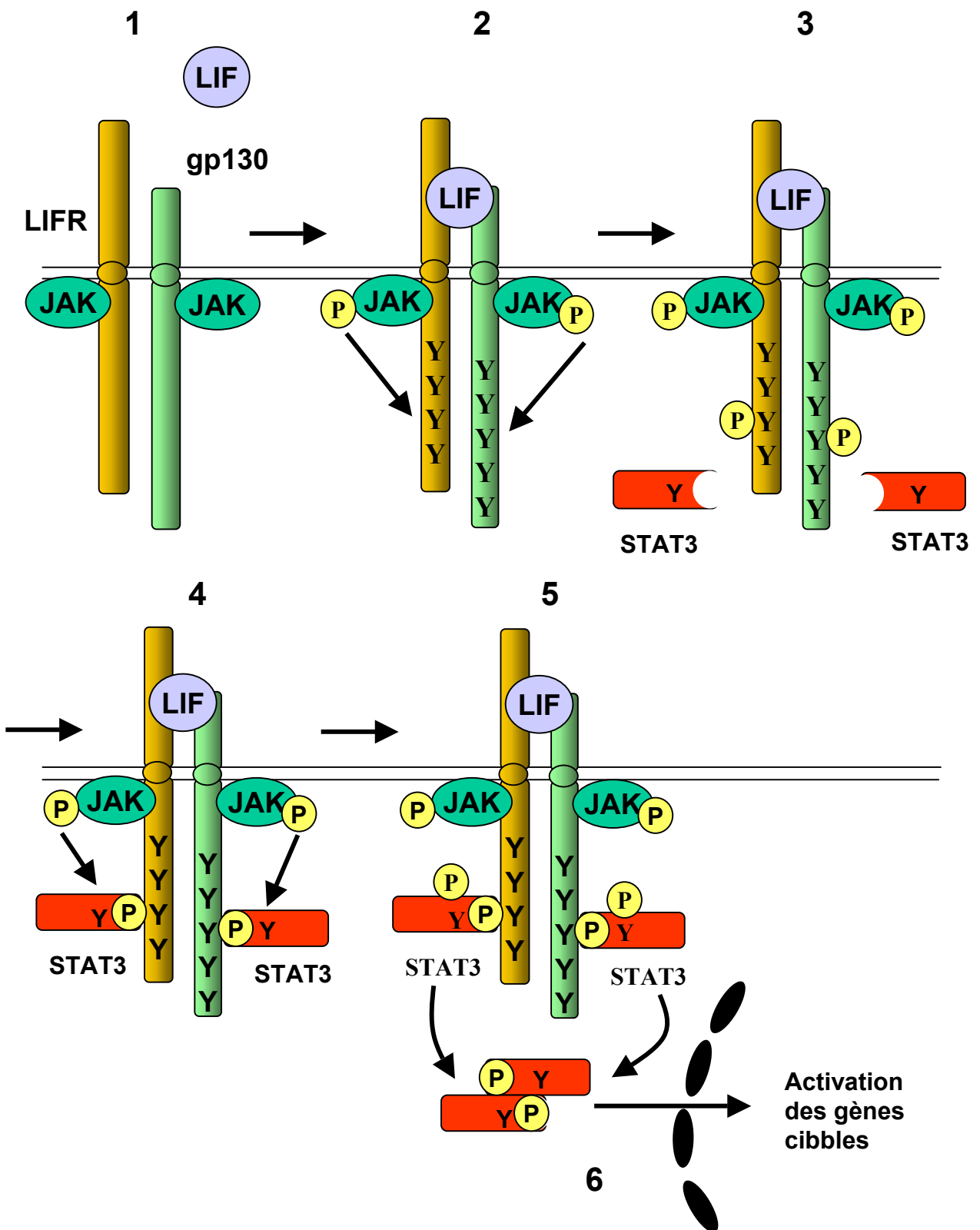
**8:** stade 11 (HH) environ 40-45h après la ponte: 12 paires de somites, le cœur est formé. La ligne primitive a complètement régressé.



**Figure 9 : Les récepteurs des cytokines de la famille du LIF/IL-6 (d'après Metcalf 2003)**

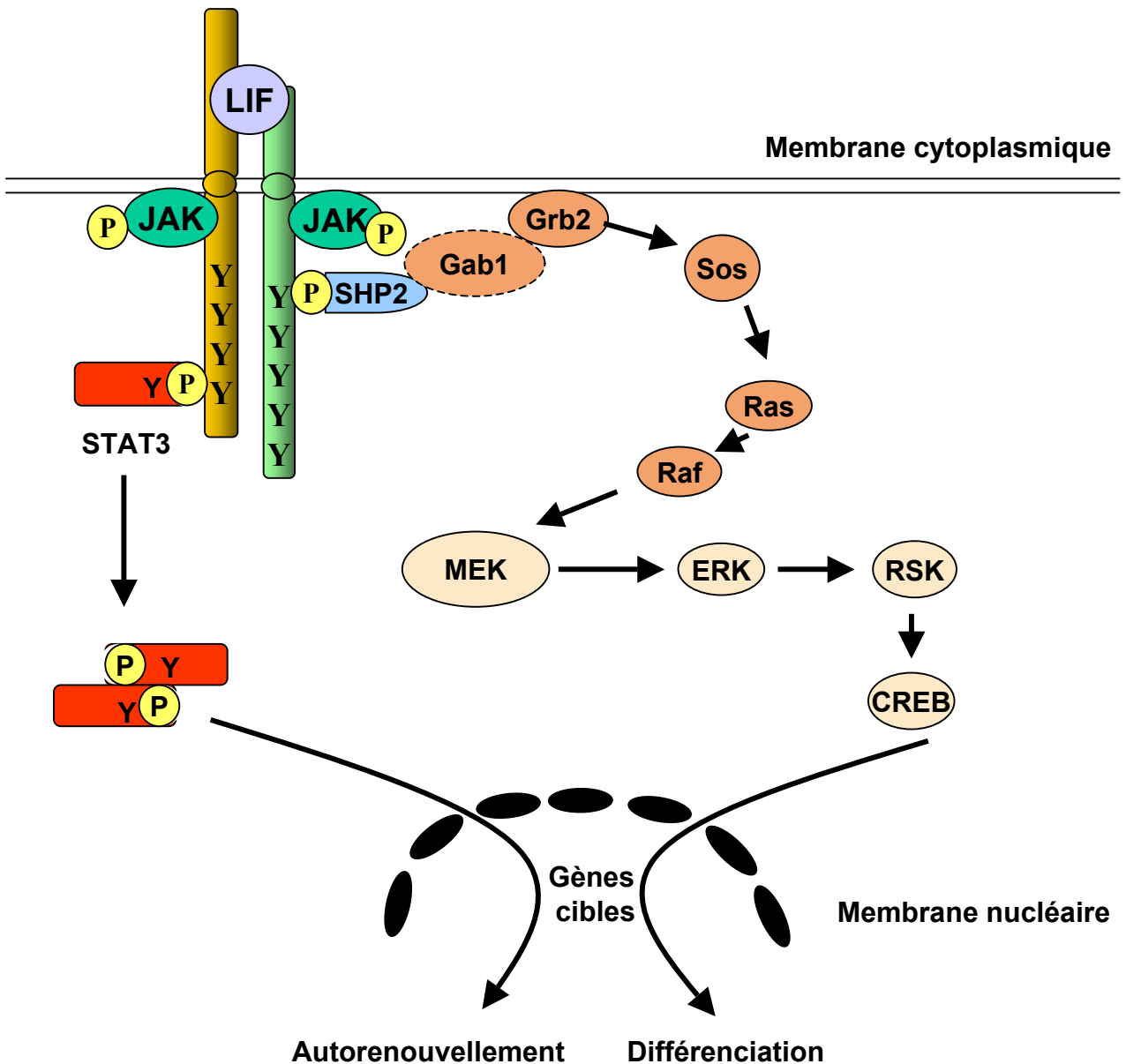
Tous les récepteurs de cette famille de cytokines sont:

- soit des homodimères de gp130 en association avec un troisième récepteur (cas de l'IL-6 et IL-11)
- soit des hétérodimères de gp130 en association avec le récepteur du LIF (jaune orangé) (cas du LIF, de l'OSM et de CT-1) et un troisième récepteur (cas du CNTF et probablement de CT-1) ou en association avec un récepteur spécifique (cas de l'OSM).



**Figure 10 : Activation de STAT3 par le LIF (d'après Sakabedoyan 2002)**

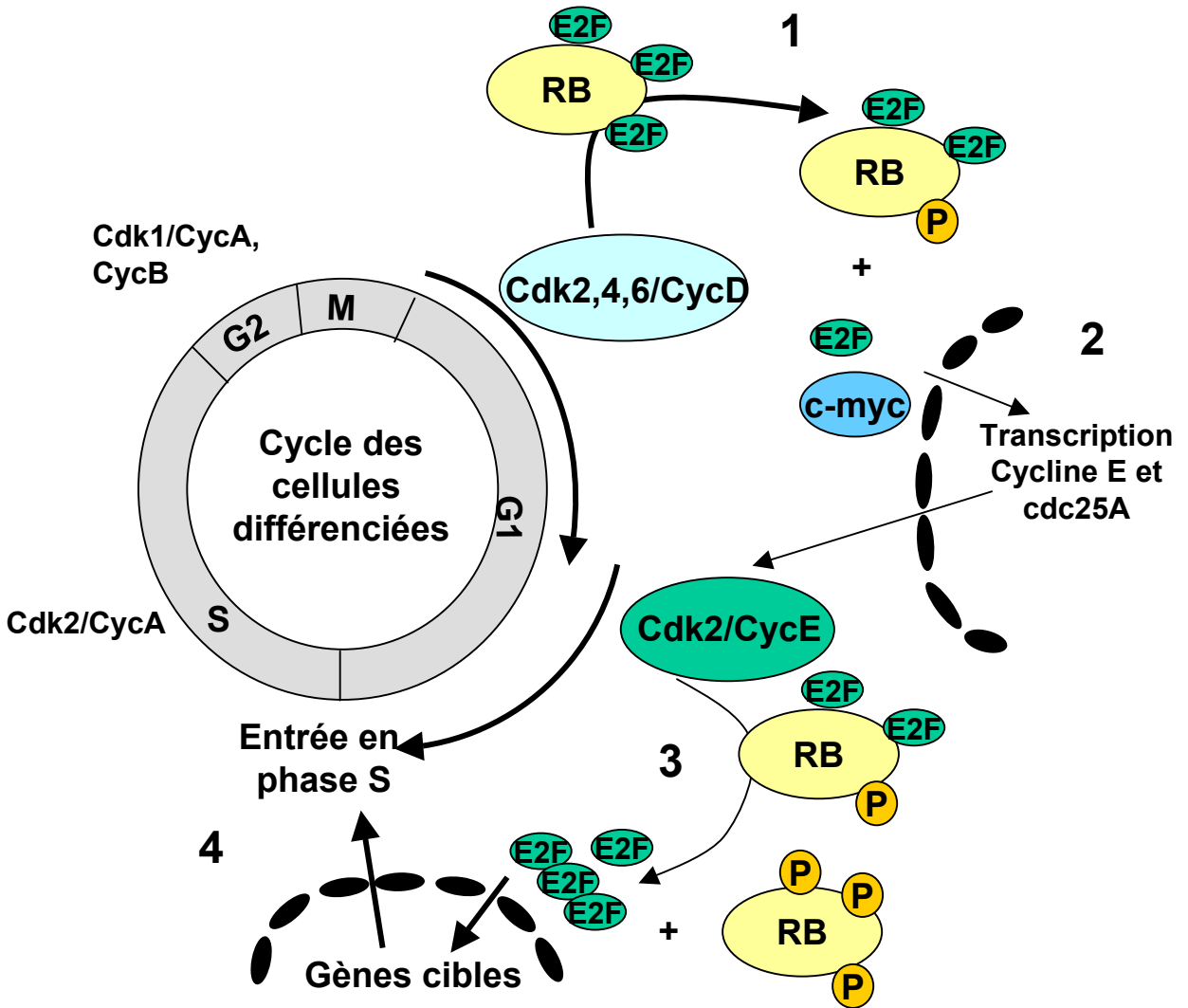
- 1: Liaison du LIF à son récepteur, constitué d'un hétérodimère LIFR et gp130.
- 2: Phosphorylation des JAKs associées au récepteur.
- 3: Phosphorylation par les JAKs des tyrosines du domaine intracytoplasmique du récepteur.
- 4: Fixation de STAT3 par son domaine SH2 sur les tyrosines phosphorylées du récepteur.
- 5: Phosphorylation de STAT3 par les JAKs sur ses résidus tyrosines.
- 6: Dissociation, dimérisation et translocation de STAT3 dans le noyau. Liaison à l'ADN et activation de la transcription des gènes cibles.



**Figure 11 : Equilibre autorenouvellement/différenciation des cellules ES contrôlé par la voie gp130 (d'après Sakabedoyan 2002)**

Suite à l'activation du récepteur par la fixation du ligand, deux signaux contradictoires sont produits:

- La mobilisation de STAT3 active les gènes impliqués dans le maintien de la pluripotence des cellules ES.
- Le recrutement de SHP2 par gp130 entraîne l'activation de MEK (probablement par le biais de gab1). L'activation séquentielle de ERK, RSK et CREB va permettre l'expression de gènes induisant la différenciation des cellules ES.



**Figure 12 : la transition G1/S dans une cellule différenciée**

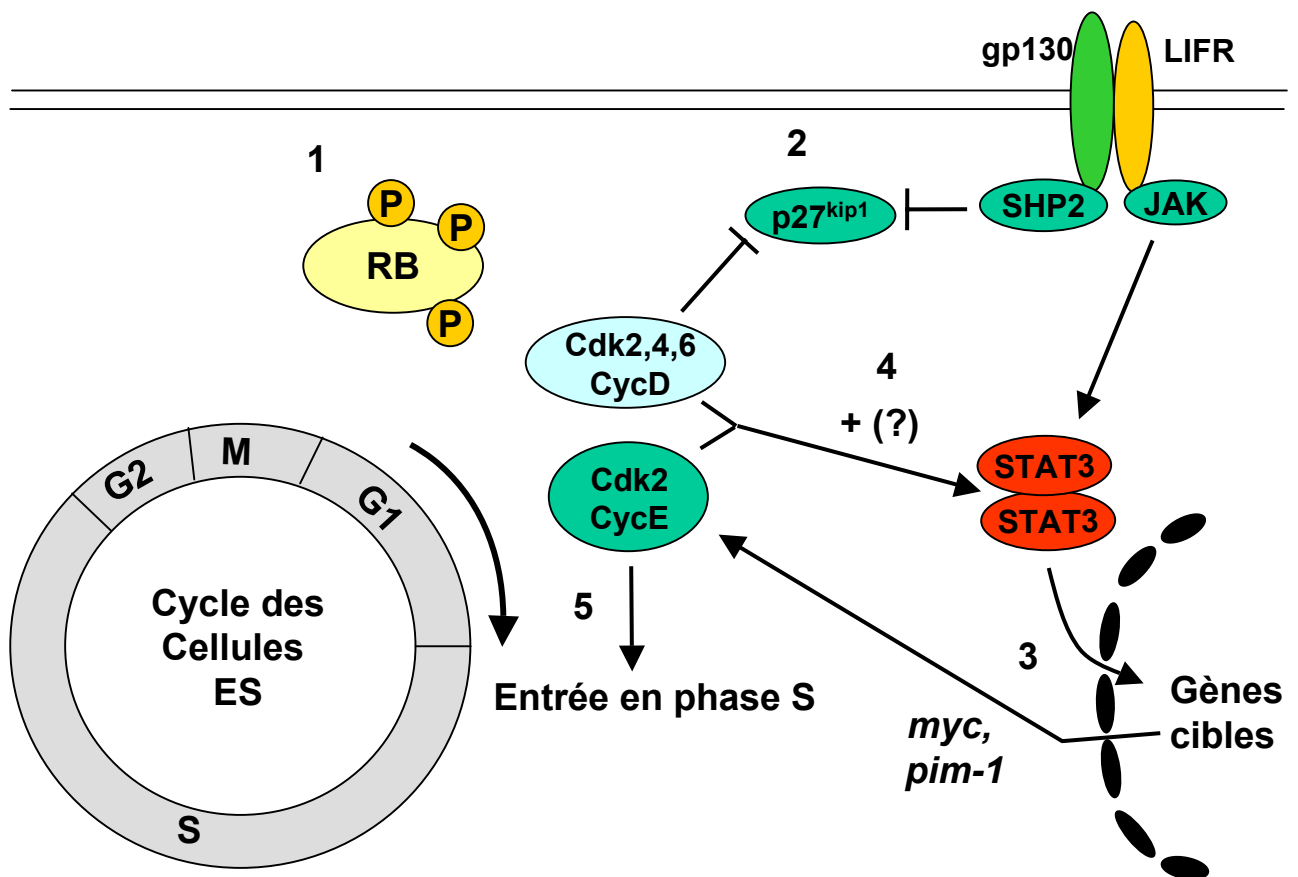
1: Au cours de la phase G1 le complexe CyclineD/Cdk2,4 ou 6 phosphoryle RB.

2: Les facteurs E2F libérés ainsi que c-myc vont activer la transcription des gènes codant la cyclineE et cdc25A.

3: Le complexe cyclineE/Cdk2 est alors activé et phosphoryle RB.

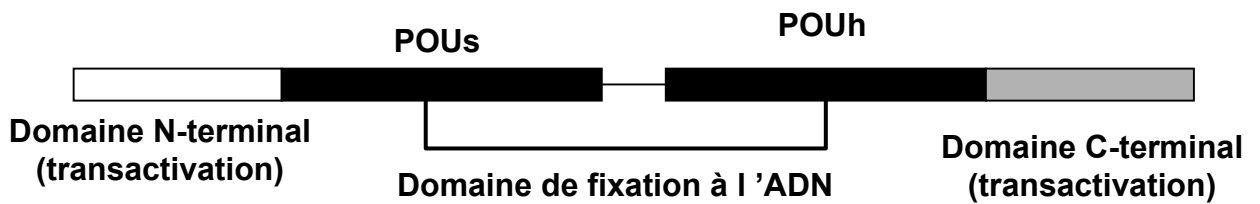
4: La libération d'un grand nombre de facteurs E2F entraîne l'entrée de la cellule en phase S.

N.B.: Pour des raisons de commodité deux noyaux sont représentés sur le schéma mais il s'agit du même noyau.



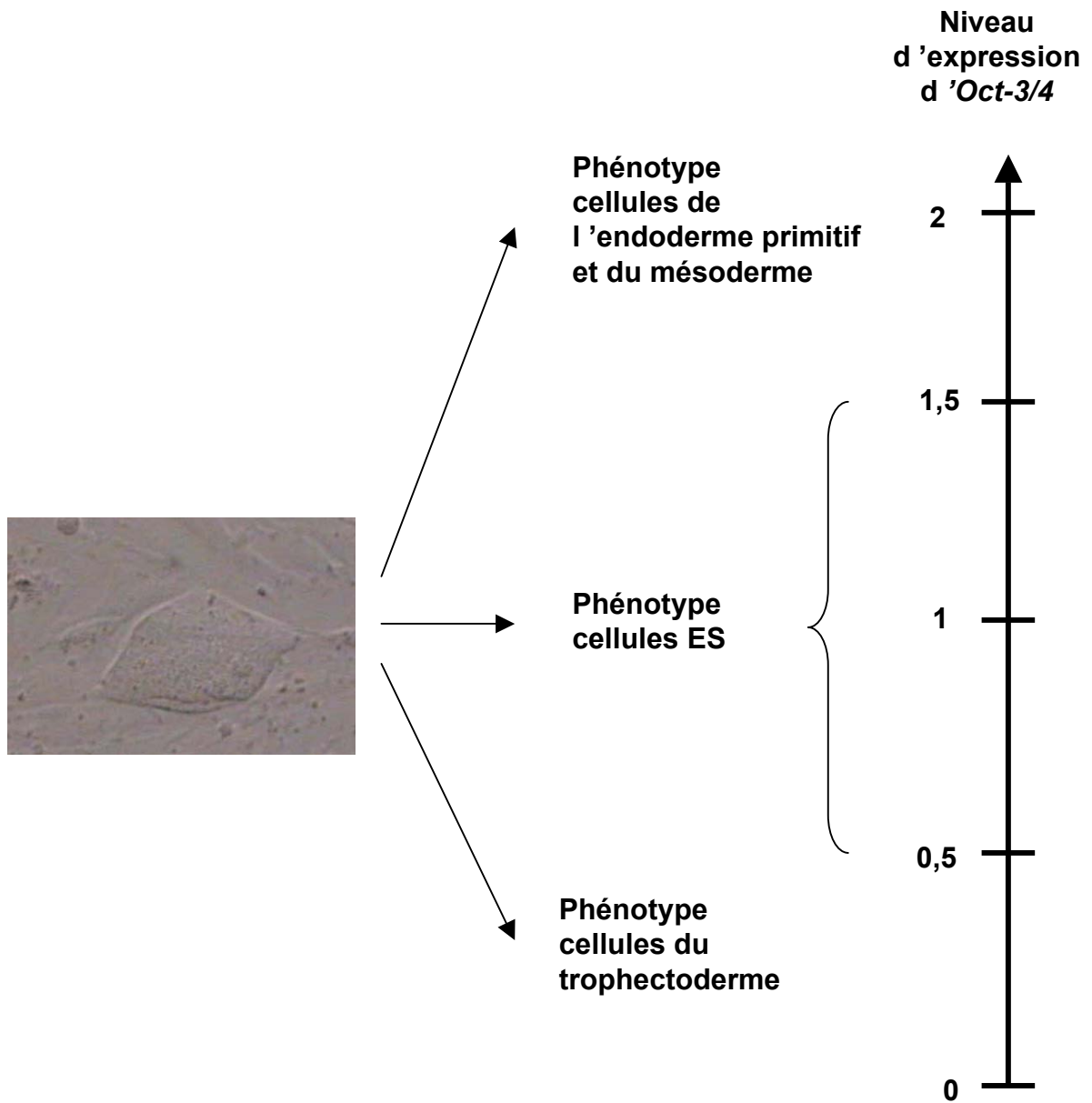
**Figure 13 : la transition G1/S dans une cellule ES: rôle de la voie gp130/STAT3**

- 1: Au cours de la phase G1, la protéine RB est toujours hyperphosphorylée.
- 2: Le complexe CyclineD/CDK2,4,6 réprime l'activité de l'inhibiteur du cycle p27<sup>kip1</sup> de même que SHP2 par l'intermédiaire de Gab1 ou d'une de ses isoformes spécifiques des ES.
- 3: Des gènes cibles de STAT3 comme *myc* ou *pim-1* pourraient activer *in fine* le complexe cyclineE/CDK2.
- 4: Les complexes CyclineD/CDK2,4,6 et CyclineE/CDK2 pourraient activer le facteur de transcription STAT3 entraînant une boucle d'autoactivation.
- 5: Le complexe CyclineE/CDK2 entraîne le passage de la cellule en phase S.



**Figure 14 : structure de la protéine Oct3/4**

La protéine Oct3/4 se compose de deux domaines POU de fixation à l'ADN et de deux domaines de transactivation localisés aux extrémités N et C-terminales.



**Figure 15 : importance du niveau d'expression d'*Oct-3/4* dans la détermination du phénotype des cellules ES (d'après Niwa 2001)**

Le niveau d'expression d'*Oct-3/4* est défini par rapport à son niveau endogène dans les cellules ES en culture. Les cellules ES conservent leur phénotype pluripotent si ce niveau d'expression est compris entre 0.5 et 1.5 fois le niveau endogène normal. Si ce niveau d'expression est inférieur à 0.5 fois le niveau endogène, les cellules acquièrent un phénotype de cellules du trophoctoderme. Si le niveau d'expression est supérieur à 1.5 fois le niveau endogène, les cellules acquièrent un phénotype de cellules de l'endoderme primitif et du mésoderme.



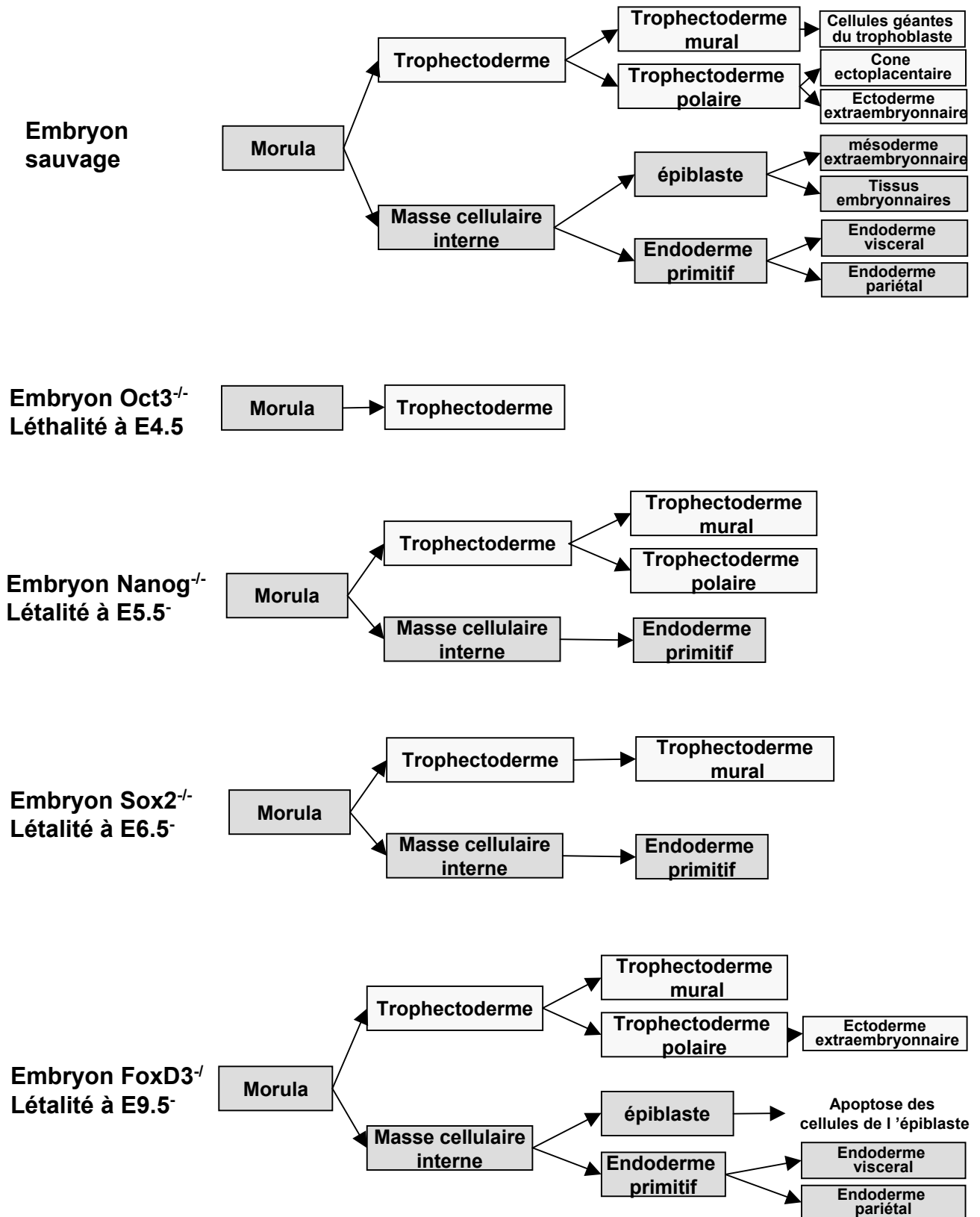
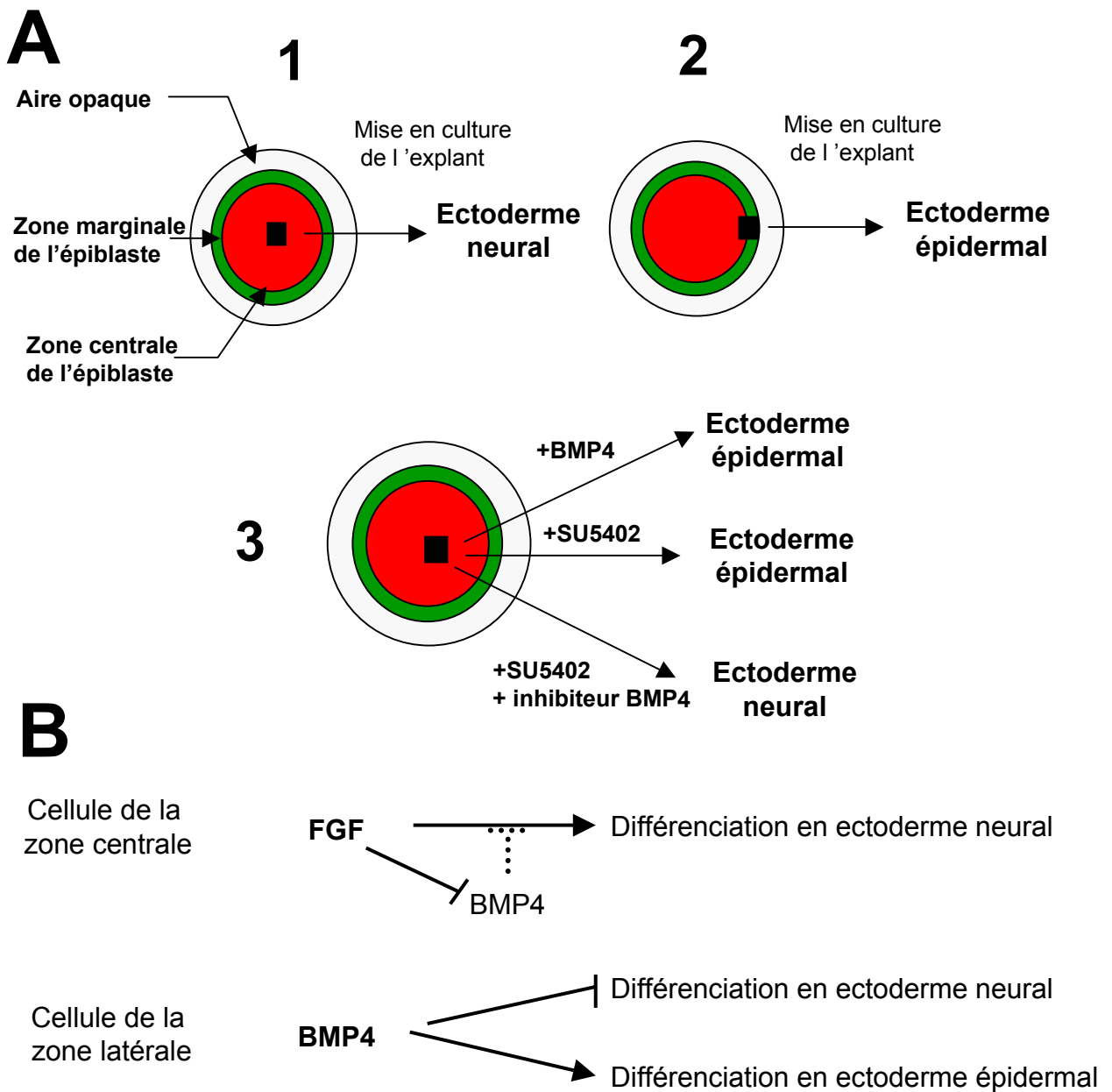


Figure 16: Les premières différenciations embryonnaires et leurs perturbations dans des embryons mutants pour les gènes *Oct3/4*, *FoxD3* et *Sox2*.

Les différents fonds génétiques vont perturber les premières différenciations embryonnaires, soit en induisant l'apoptose des cellules d'un tissu, soit en bloquant certaines voies de différenciation.

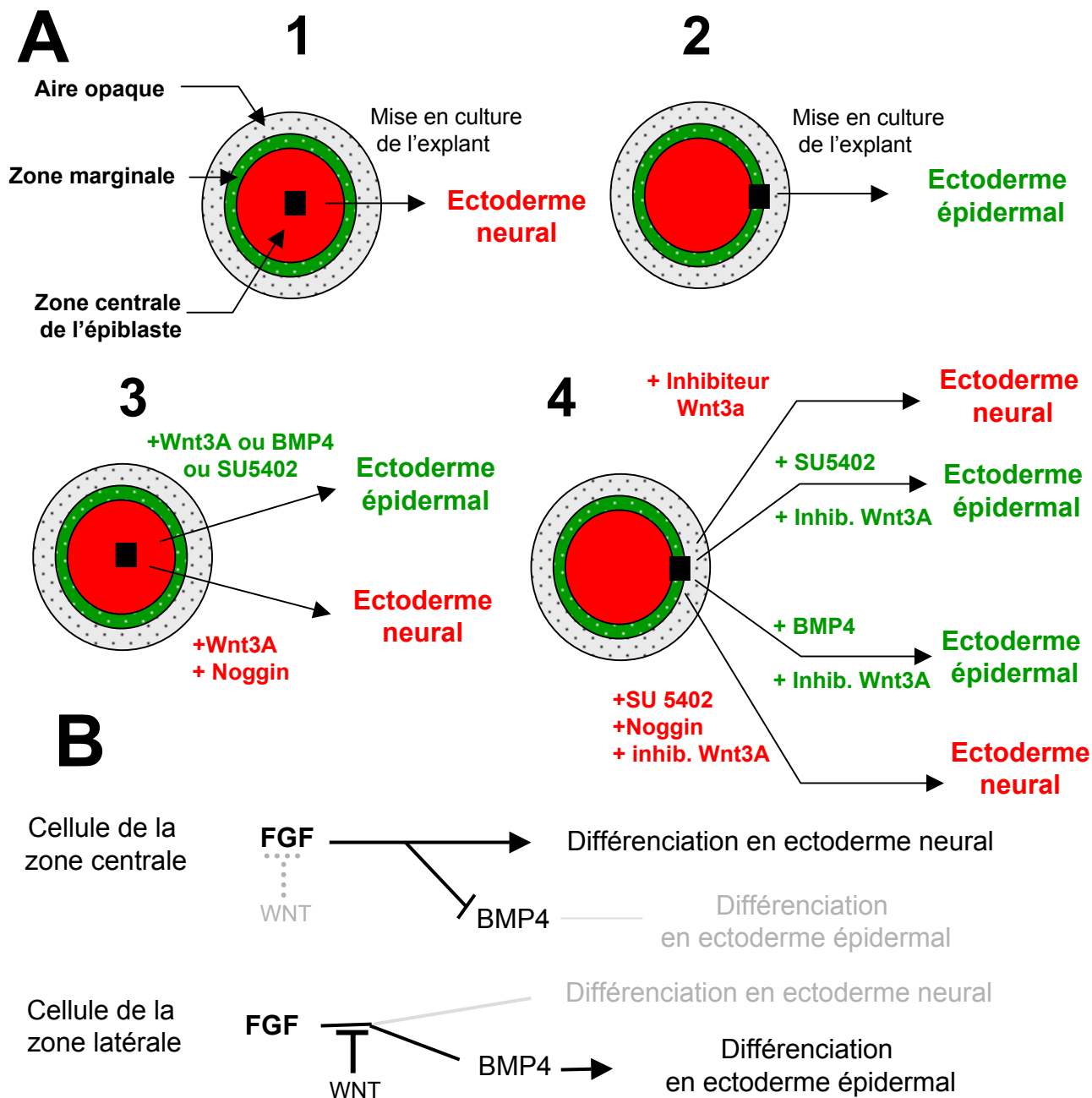


**Figure 17 : Présentation des travaux de Wilson et *al.* 2000 sur les mécanismes de l'induction neural dans l'embryon de poulet au stade XII (EG)**

**A:** Présentation des expériences réalisées: (1) des explants en culture de la zone centrale de l'épiblaste d'embryon de poulet au stade XII (EG) se différencient en cellules de l'ectoderme neural alors que (2) des cellules de la zone marginale se différencient en ectoderme épidermal. L'ajout de BMP4 dans le milieu induit une différenciation des explants de la zone centrale (3) en cellules de l'ectoderme épidermal. L'ajout de l'inhibiteur du récepteur 2 aux FGFs, SU5402 induit aussi une différenciation des explants centraux en cellules de l'ectoderme épidermal. Enfin l'ajout de SU5402 et d'un dominant négatif de BMP4 entraîne une différenciation des explants centraux en cellules de l'ectoderme neural.

**B:** Modélisation des résultats obtenus. Selon les auteurs, les cellules de la zone centrale se différencient par défaut en cellules de l'ectoderme neural. Dans ces cellules, le FGF bloque l'action de BMP4 qui induirait sinon une différenciation de ces cellules en ectoderme épidermal.

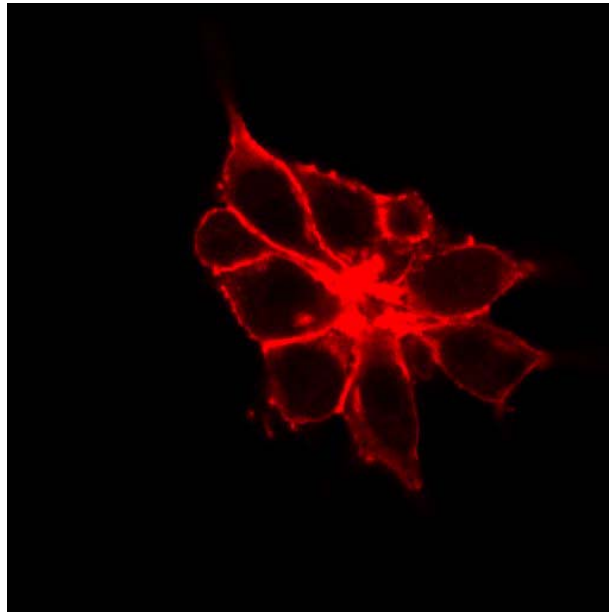
Dans les cellules de la zone latérale, le FGF n'agit pas sur BMP4, celui-ci induit donc une différenciation en cellules de l'ectoderme épidermal.



**Figure 18 : Présentation des travaux de Wilson et al. 2002 sur les mécanismes de l'induction neural dans l'embryon de poulet au stade XII (EG)**

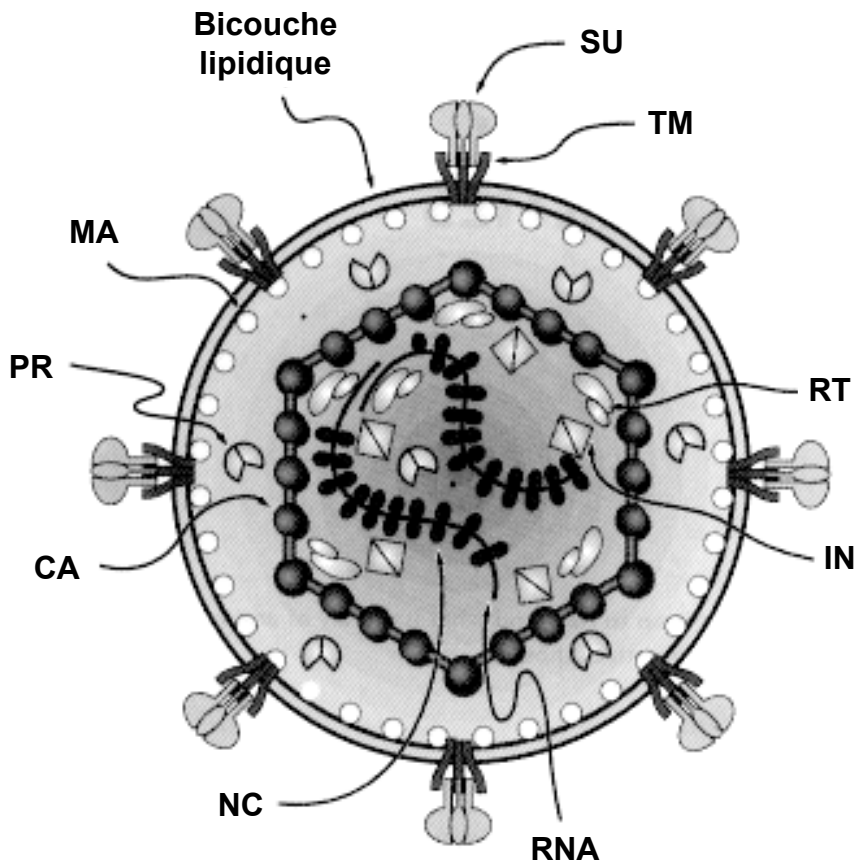
**A :** Les zones ponctuées (aire opaque et zone marginale) correspondent aux cellules avec une localisation nucléaire de la  $\beta$ -caténine. Présentation des expériences réalisées: **(1,2)** rappel des expériences précédentes; **(3)** l'ajout de wnt3A, de BMP4 ou de SU5402 (inhibiteur FGFR2) dans le milieu de culture induit une différenciation des explants centraux en cellules de l'ectoderme épidermal alors que l'ajout de SU5402 et d'un inhibiteur de BMP4 entraîne une différenciation des explants centraux en cellules de l'ectoderme neural; **(4)** l'ajout d'un inhibiteur de wnt3A ou des inhibiteurs du FGFR2 (SU5402), de BMP4 et de Wnt3A dans le milieu de culture induit une différenciation des explants latéraux en cellules de l'ectoderme neural alors que l'ajout de SU5402 et d'un inhibiteur de Wnt3A ou de BMP4 et d'un inhibiteur de Wnt3A dans le milieu de culture induit une différenciation des explants latéraux en cellules de l'ectoderme épidermal.

**B :** Modélisation des résultats obtenus. Selon les auteurs les cellules de la zone centrale et latérale se différencient par défaut en cellules de l'ectoderme neural. Dans ces cellules centrales, le FGF bloque l'action de BMP4 qui induirait sinon une différenciation en cellules de l'ectoderme épidermal. Dans les cellules de la zone latérale, l'action du FGF, bloquée par Wnt, ne réprime plus BMP4, qui induit alors une différenciation de ces cellules en ectoderme épidermal.



**Figure 19: localisation de la  $\beta$ -caténine dans un amas de cellules ES de poulet**

Immunocytochimie réalisée sur des cellules ES de poulet avec un anticorps primaire dirigé contre un peptide conservé de la  $\beta$ -caténine humaine. La localisation de la  $\beta$ -caténine est uniquement membranaire dans ces cellules.

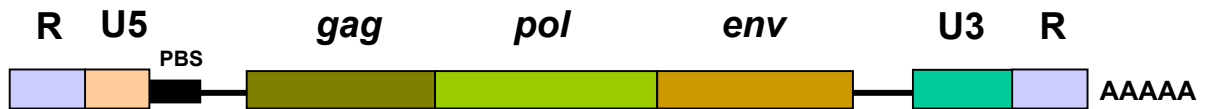


**Figure 20 : structure générale de la particule virale (d'après Vogt 1997)**

Les initiales représentent les différentes protéines du virion:  
 SU: protéine de surface, TM: protéine transmembranaire  
 RT: transcriptase inverse; IN: intégrase; PR: protéase  
 MA: protéine de la matrice; CA: capsid; NC: nucléocapsid

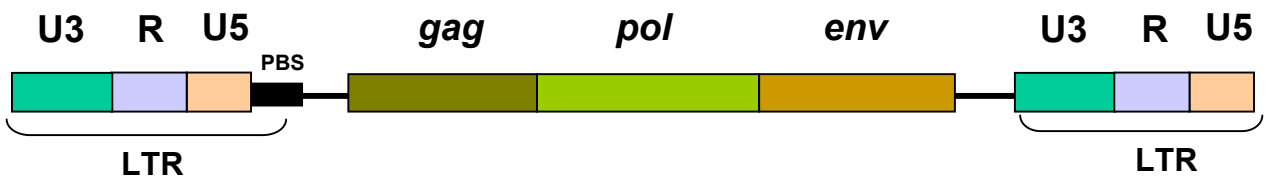
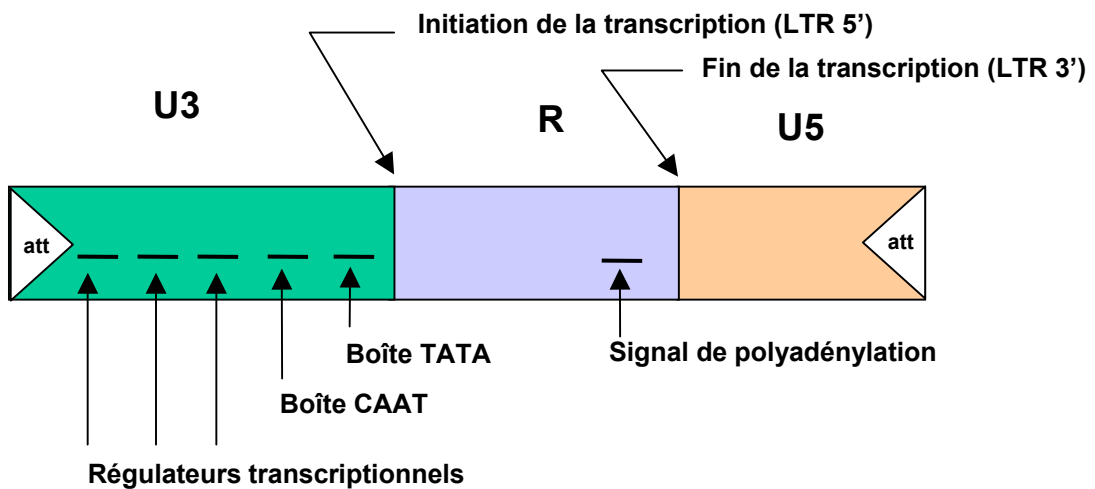
**A**

ARN génomique simple brin



Transcription inverse

ADN proviral double brin

**B****Figure 21: Structure des LTR rétrovirales**

**A:** Après l'étape de transcription inverse, l'ARN génomique simple brin est converti en une molécule d'ADN double brin comportant les LTRs (Long Terminal Repeat) complètes à chaque extrémité, encadrant la séquence PBS (Primer Binding Site) et les gènes viraux.

**B:** Les LTRs sont subdivisées en trois régions U3, R et U5, elles-même bordées par de courtes séquences inversées répétées (att) nécessaires à l'intégration du provirus. Les informations génétiques nécessaires à l'initiation de la transcription (promoteur, boîtes CAAT et TATA) ainsi que de nombreuses séquences régulatrices sont contenues dans la région U3. La région R comporte le signal de polyadénylation.

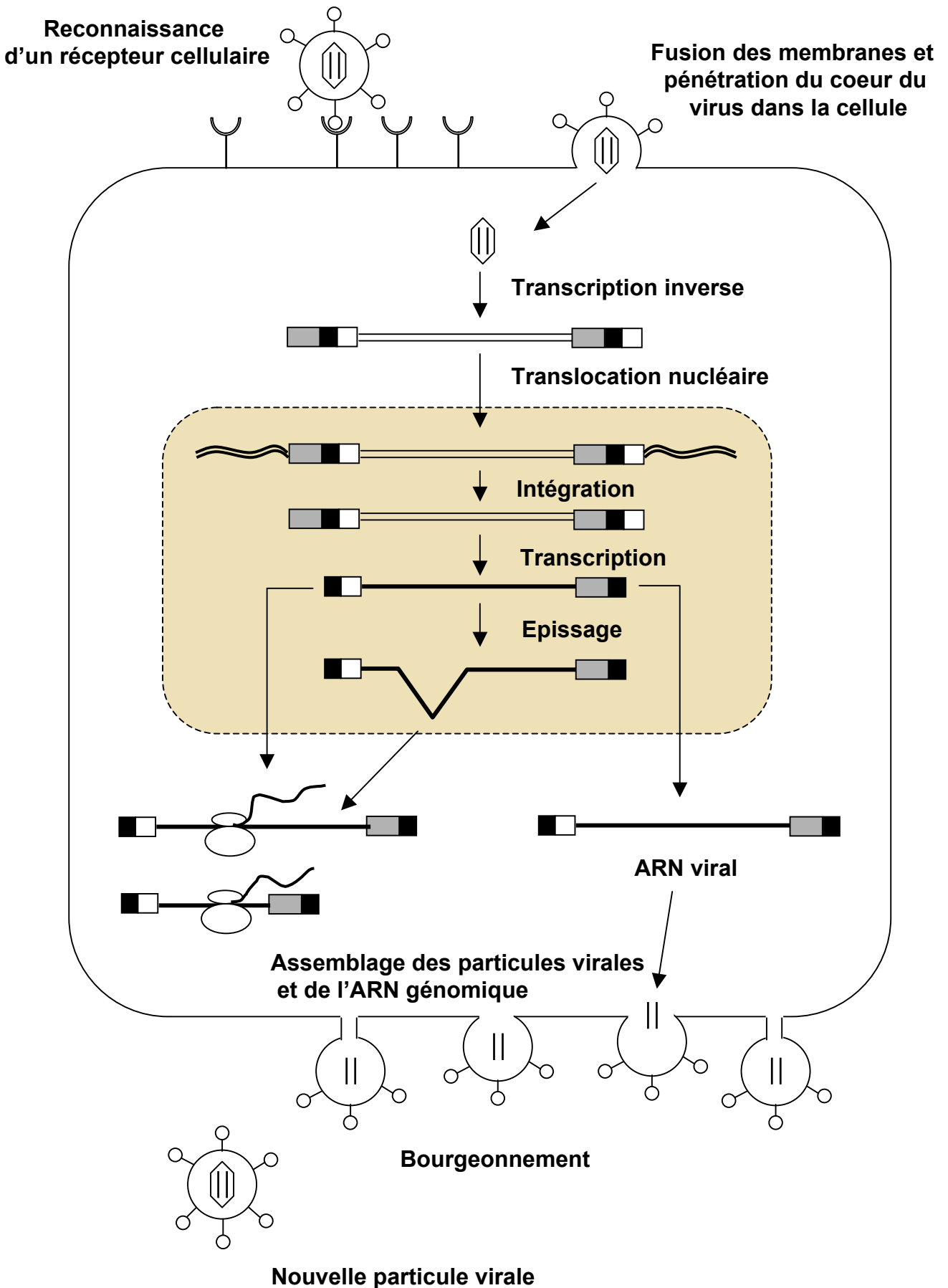
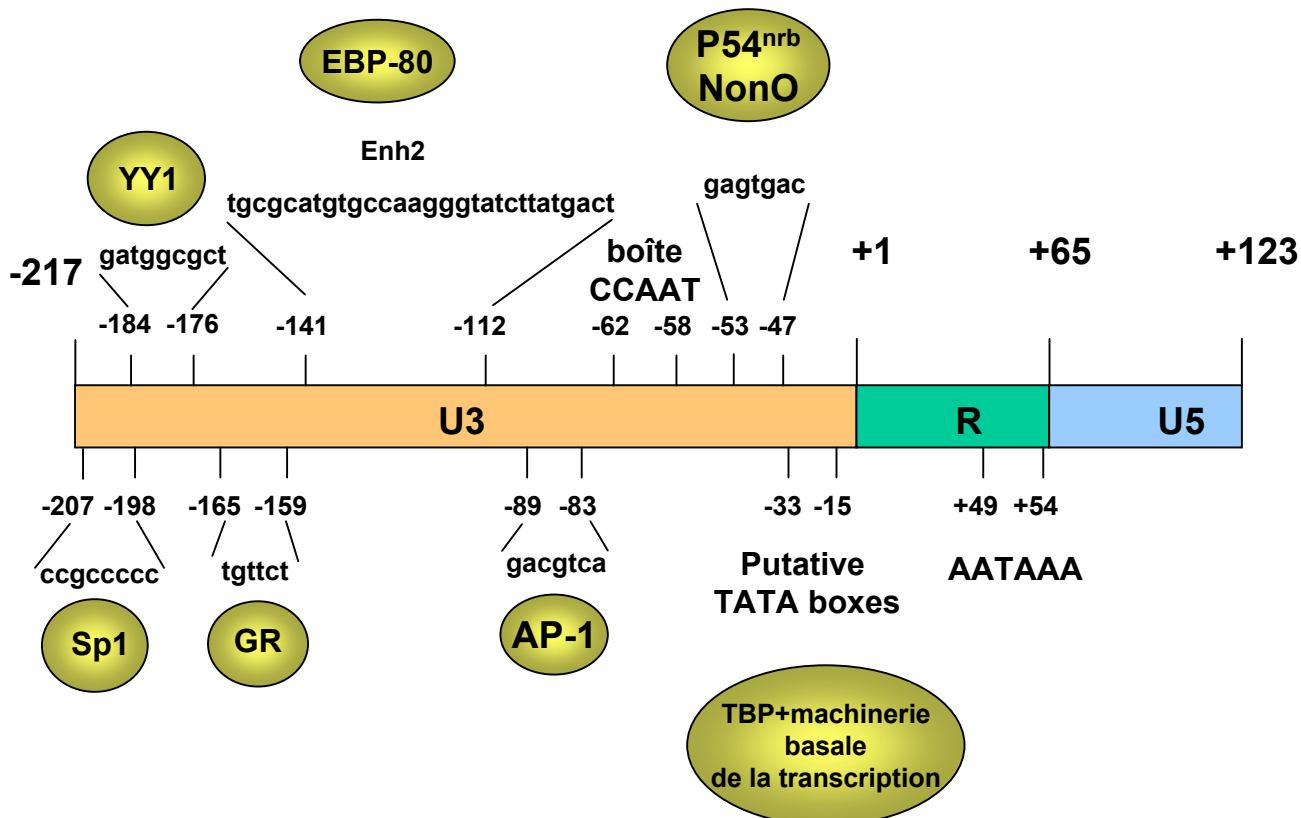


Figure 22 : Présentation du cycle rétroviral (d'après Vogt 1997)

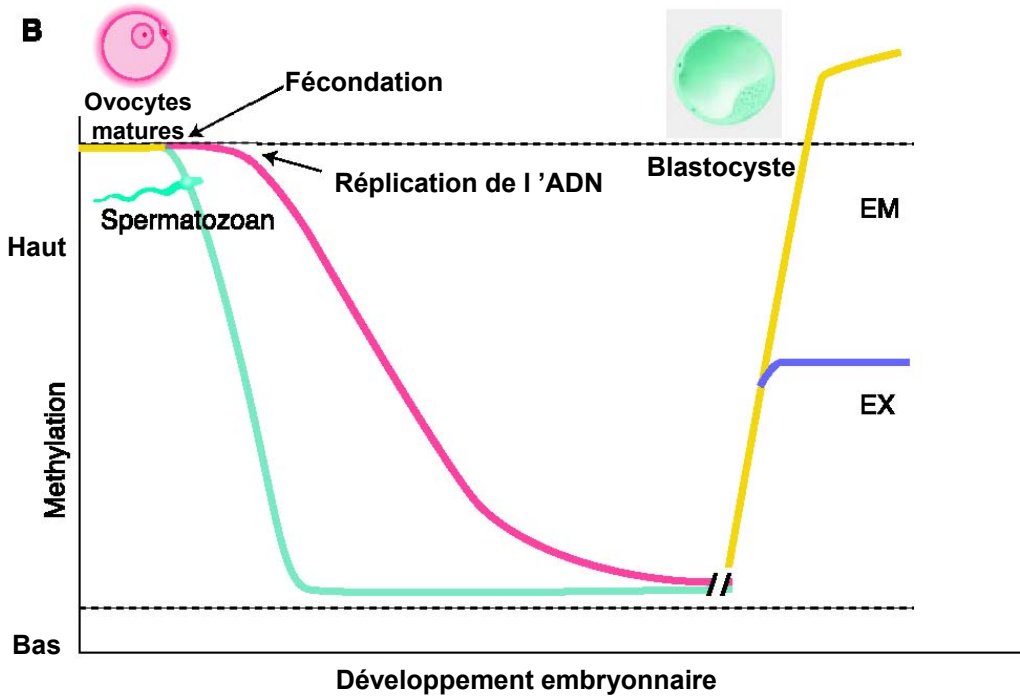
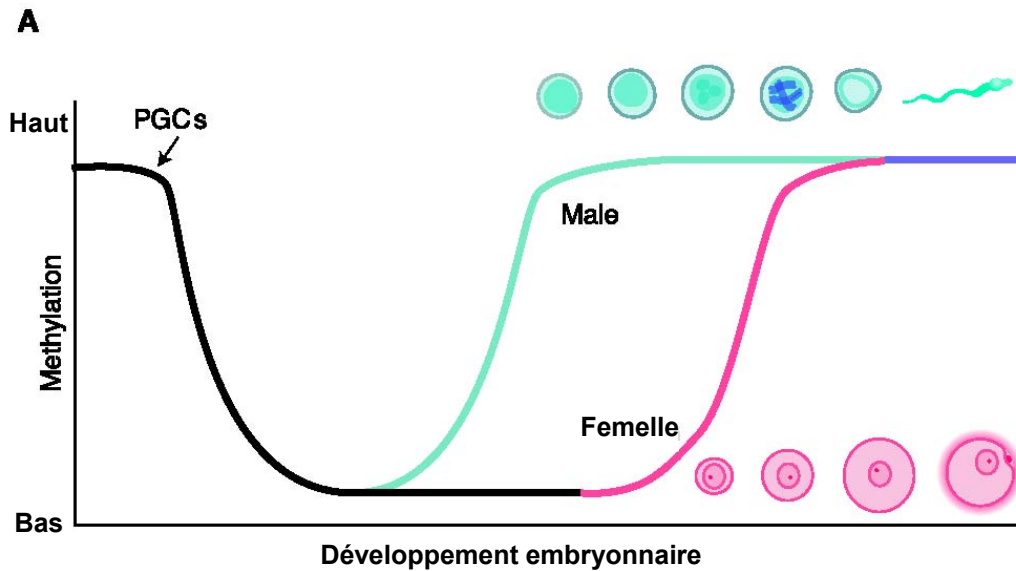
Les différentes étapes du cycle rétroviral sont résumées sur le schéma ci-dessus.



**Figure 23 :** Cartographie des sites de fixation des facteurs de transcription dans la LTR des IAPs

La région U3 de la LTR des IAPs comprend un promoteur basal incluant une boîte TATA et une boîte CAAT, sur lesquelles se fixent la machinerie basale transcriptionnelle. Plusieurs éléments de réponse pour des facteurs de transcription connus sont représentés. La région R comporte un signal de polyadénylation.





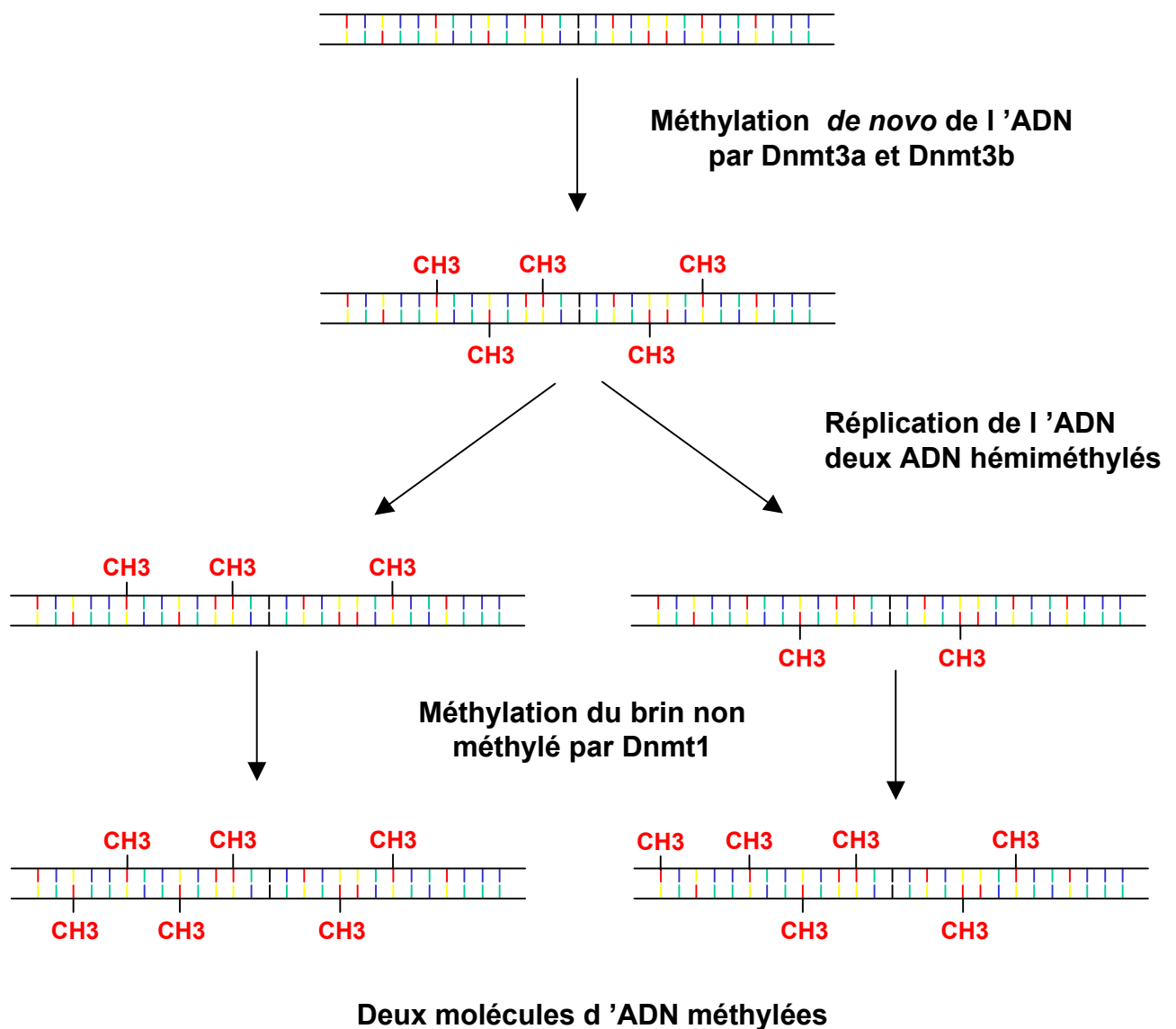
**Figure 24 : Reprogrammation du génome par la méthylation (d'après Reik et al. 2001)**

**A: dans la lignée germinale**

Les cellules germinales primordiales (PGCs) sont déméthylées précocément au cours du développement embryonnaire. La méthylation de leur génome survient dans les préspermatogonies à E16 chez le mâle et après la naissance chez la femelle.

**B: dans l'embryon préimplantatoire**

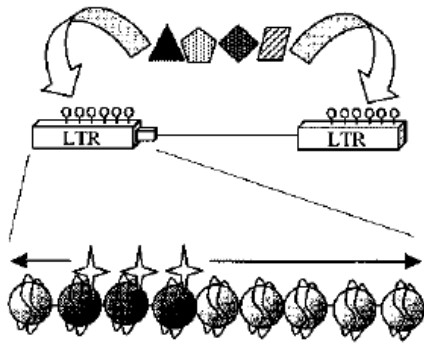
Le génome paternel (en vert) est déméthylé par un mécanisme actif dès la fécondation. Le génome maternel est déméthylé passivement par la réplication successive de l'ADN. Les deux génomes sont ensuite reméthylés pendant la période de l'implantation suivant leur devenir en tissus embryonnaires (EM) ou extraembryonnaires (EX). Certains gènes soumis à l'empreinte ne seront pas déméthylés (pointillé en haut) ou ne seront pas méthylés (pointillé en bas).



**Figure 25 : Méthylation de l'ADN: un mécanisme en deux étapes**

L'ADN non méthylé est reconnu par les méthyltransférases *de novo* Dnmt3a et Dnmt3b. Au cours de la réplication de l'ADN, les deux brins néoformés ne sont pas méthylés. La méthyltransférase Dnmt1 reconnaît l'ADN hémiméthylé et va méthylé les deux brins néoformés. On obtient ainsi une molécule d'ADN méthylé sur les deux brins. Au cours des réplifications suivantes le niveau de méthylation de l'ADN est maintenu par la Dnmt1.

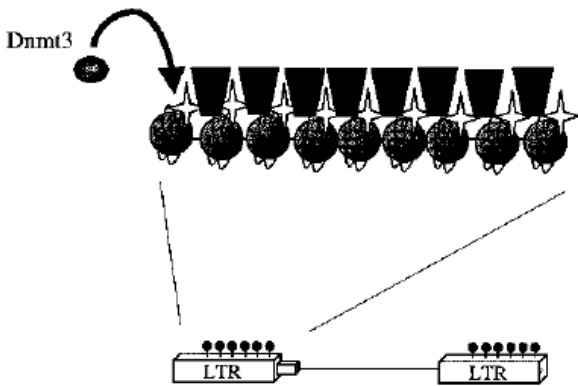
**Des facteurs négatifs agissant en trans se fixent sur l'ADN proviral**  
YY1, ELP...



**Remodelage de la chromatine**  
NuRD, SWI  
HDACs?



**Protéines associées à la chromatine**  
Polycomb groups

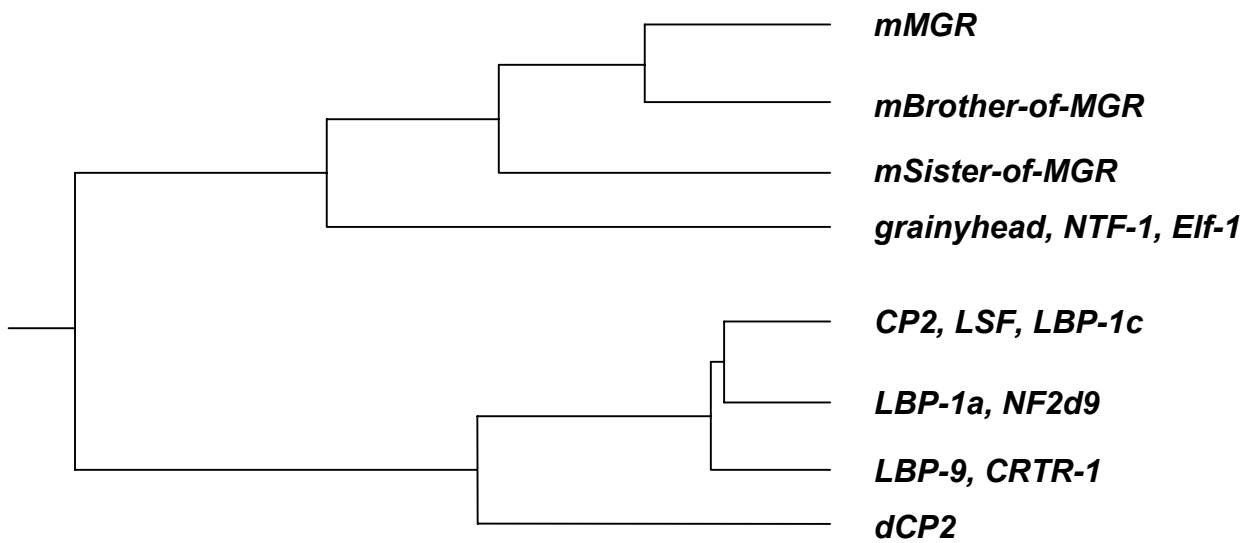


**Méthylation *de novo***  
dnmt3a/dnmt3b

**Provirus silencieux**  
maintenu silencieux par dnmt1  
et les complexes MBD/HDAC

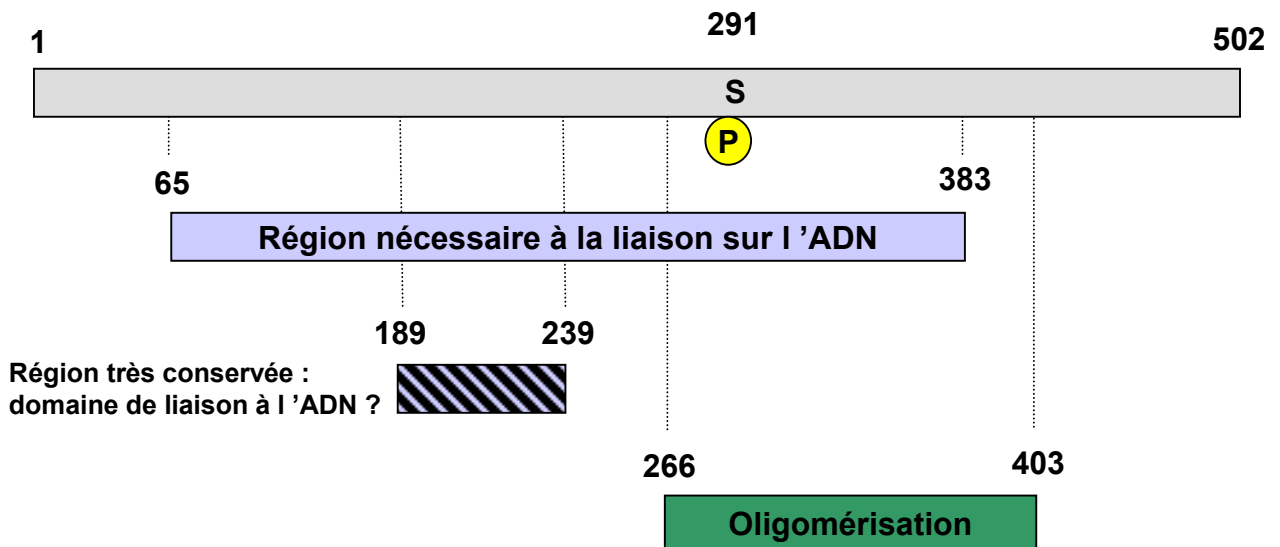
**Figure 26 : Modèle proposé par Pannell et Ellis (2001) pour expliquer l'extinction de l'expression virale dans les cellules souches.**

La chromatine autour du provirus est remodelée (nucléosomes en noir) et des modifications (représentées ici par des étoiles) comme des groupements méthyls ou acétyls sont apportées aux histones. Ces modifications peuvent être reconnues par des protéines comme celles de la famille *Polycomb* pour induire un état répressif de la chromatine. Le recrutement de méthylases *de novo* permet la méthylation des LTRs retrovirales et permet la conservation d'un état silencieux au cours des différenciations cellulaires.



**Figure 27 : schématisation de l'arbre phylogénétique représentant les facteurs de transcription de la famille NTF (d'après Wilanowski et al. 2002)**

On observe deux branches bien distinctes regroupant d'un côté les gènes homologues au facteur *grainyhead* de drosophile et de l'autre ceux homologues au facteur *dCP2* de drosophile.

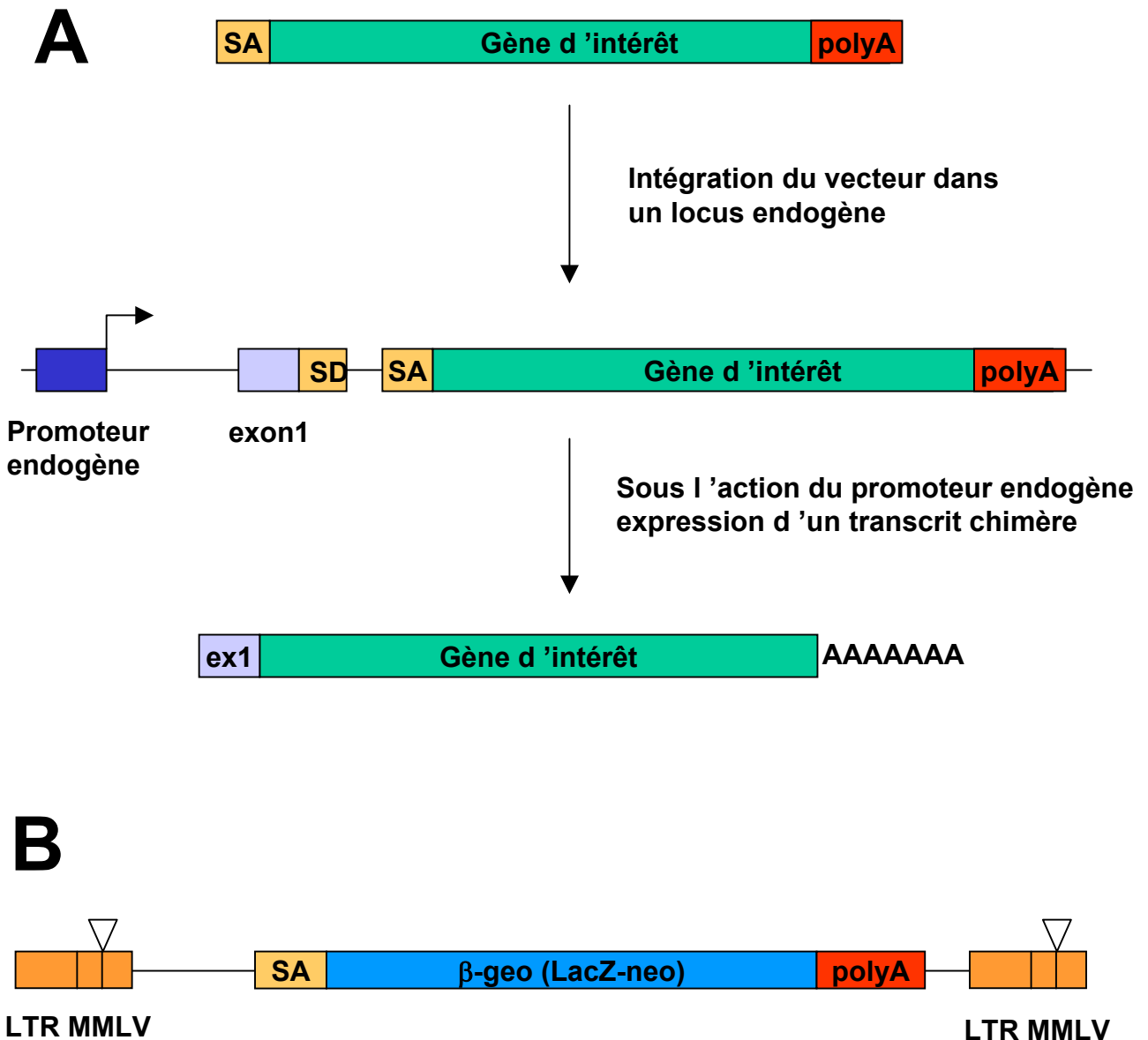


**Figure 28: Présentation des différentes régions fonctionnelles de la protéine CP2**

La région nécessaire à la liaison sur l'ADN (boîte violette) n'est pas le domaine de liaison à l'ADN au sens strict. Il comprend d'ailleurs une grande partie du domaine d'oligomérisation (boîte verte). Il semble d'ailleurs que CP2 se fixe à l'ADN en multimères pour stabiliser cette interaction.

Le domaine de liaison à l'ADN au sens strict pourrait bien être la région très conservée (boîte en violet rayée) entre les différents CP2 murin, humain et aviaires et NF-E1.

La phosphorylation de la sérine 291 est nécessaire pour l'activation du facteur CP2.



**Figure 29 : la stratégie du piégeage de gène (gene trap)**

**A: stratégie générale**

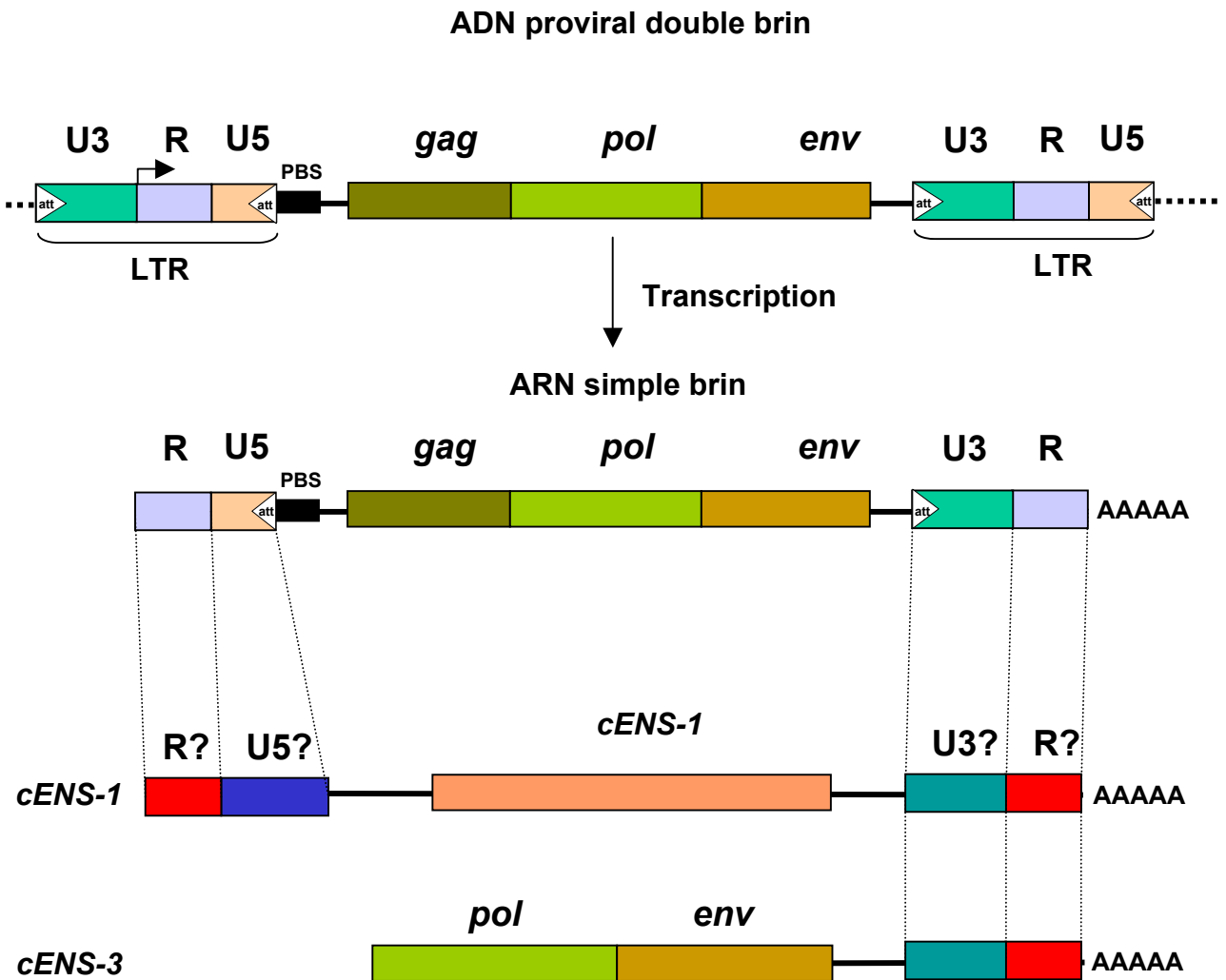
Un vecteur sans promoteur mais comportant un site accepteur d'épissage est introduit dans les cellules par transfection.

L'insertion aléatoire du transgène peut placer celui-ci dans un locus fonctionnel sous le contrôle d'un promoteur endogène.

L'épissage du transcrit primaire fusionne le transgène et les exons du locus endogène.

**B: le vecteur ROSA  $\beta$ -geo**

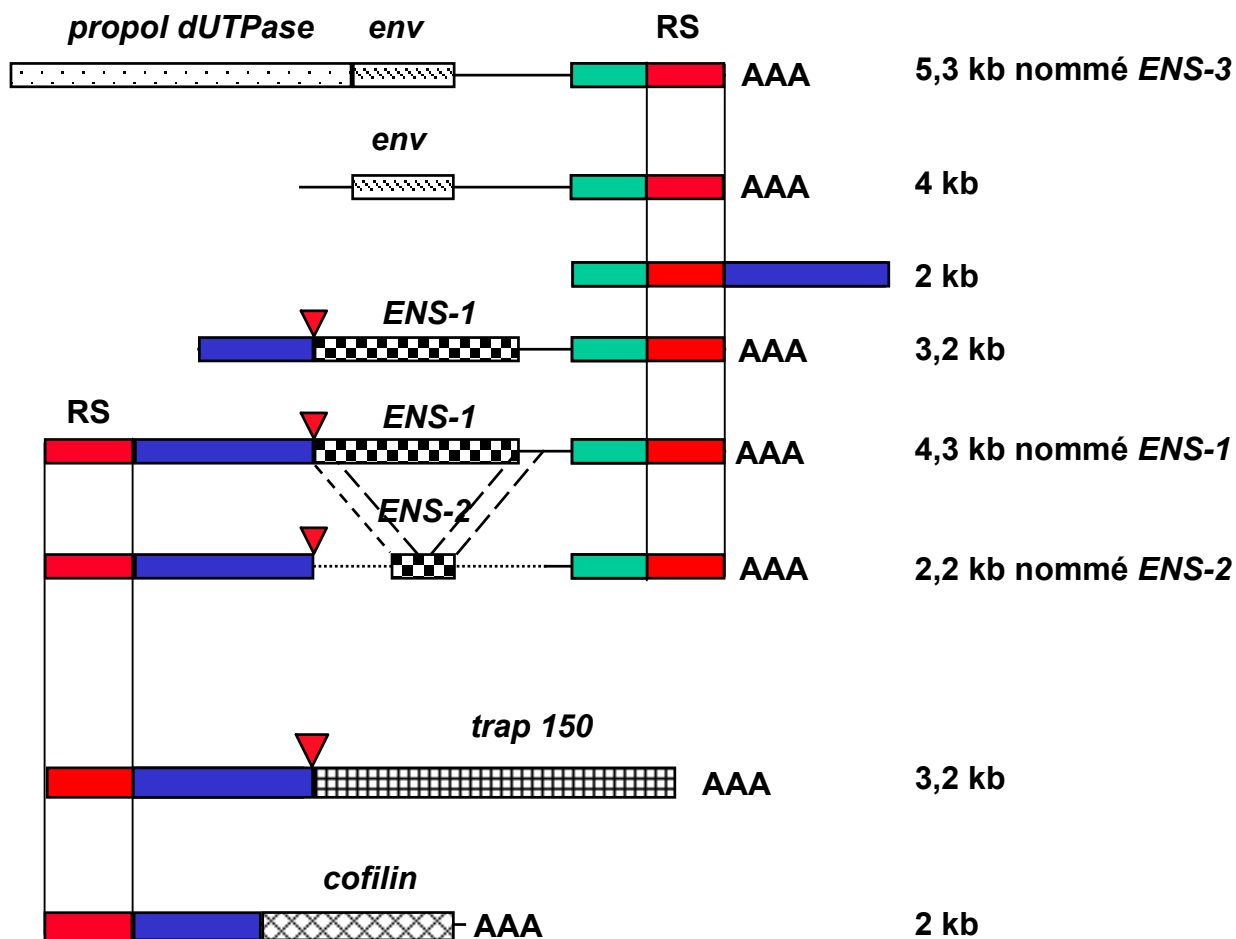
Le vecteur ROSA $\beta$ -geo se compose de deux LTRs mutées (triangle blanc) du virus MMLV encadrant un site accepteur d'épissage et un gène fusion  $\beta$ -geo pour le gène LacZ et celui de résistance à la néomycine. Un signal de polyadénylation est présent à la fin de la séquence codante du gène  $\beta$ -geo.



**Figure 30: Délimitation des régions U5, R et U3 potentielles du transcrit *cENS-1***

Par analogie avec la structure des LTRs rétrovirales, nous avons délimité les régions potentielles correspondant à U3, R et U5 sur le transcrit *cENS-1*.

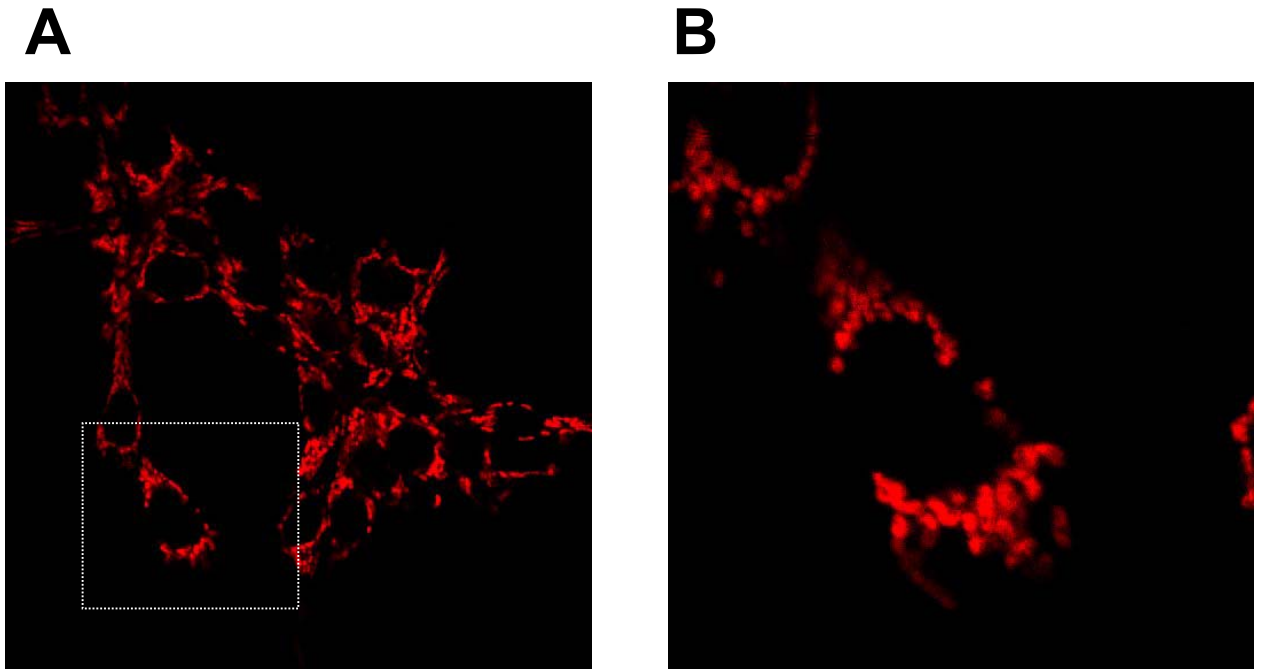
La région R correspond à la région répétée à chaque extrémité du transcrit. Comme nous ne disposons pas des séquences génomiques nous permettant de délimiter avec exactitudes les régions U3 et U5, les extrémités 5' de la région U3 et 3' de la région U5 sont actuellement indéfinies. Nous avons fixé arbitrairement la limite 5' de la région U3 à la région conservée à l'extrémité 3' des différents ADNc ENS.



**Figure 30: Présentation des différents clones obtenus lors du criblage de la banque d'ADNc**

La région répétée (RS) aux extrémités des transcrits est présentée en rouge. Les régions non codantes conservées entre les différents clones sont présentées en vert et en bleu. Elles sont susceptibles de contenir les régions U3 (pour la région verte) et U5 (pour la région bleue). Les clones que nous avons retenus pour la publication sont ceux qui avaient l'air complet et dont le profil d'expression était cohérent avec celui observé pour le gène rapporteur LacZ. La flèche rouge indique un site conservé dans de nombreux clones à partir duquel commence la phase codante des différents gènes.





**Figure 31: localisation intracellulaire de la protéine ENS-1 dans les cellules ES de poulet**

Immunocytochimie réalisée sur des cellules ES de poulet avec un anticorps polyclonal primaire dirigé contre un peptide de la protéine ENS-1.

**A.** Microscopie confocale sur un amas de cellules ES de poulet

**B.** Microscopie confocale sur une cellules ES de poulet

```

          HFH-8                HNF-3b                Dlx-3
-738 GTGGATGTTT ATTAGGAAGC TCACGACTAA AGGAAACAAT CTGTTAATTG TTTATTTATT
          c-Myb
-678 ATTAGTGGTT ATTGTCAAAT GTACGGTTGT CTCTTTTCTC TCTTCTATTC ATTATGTAAT

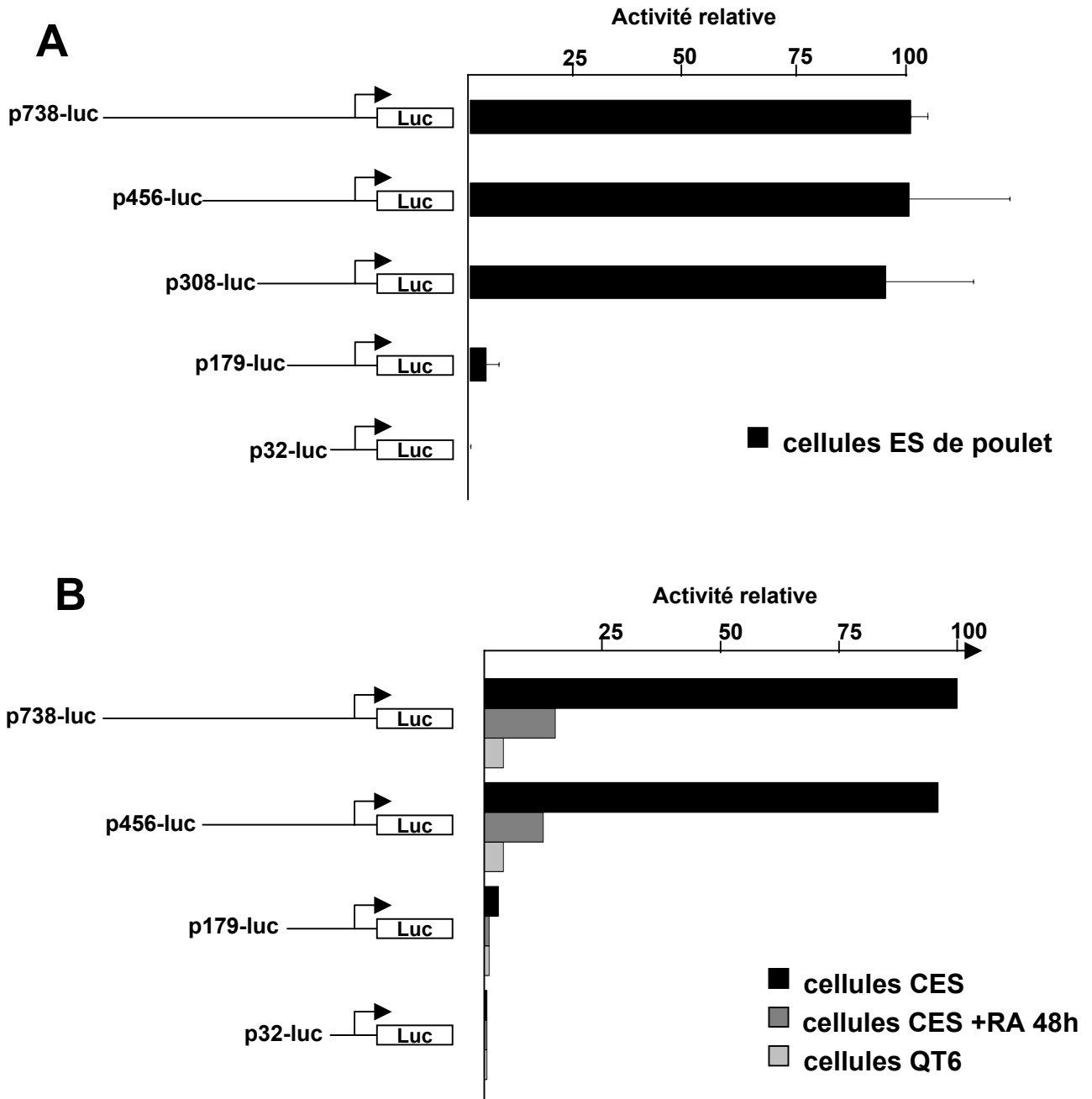
-618 ATTCATGTTA CCACTCCTGA AGAATCACGG GGTGGTGTCT ATGGCAAGTT GCATTGTGTA
          c-Myb
-558 CTGTTGCAAC TCTTATGTTT GTATGATTCC ATGTTTTATA CAAGATGTTG TATCCCCTAT

-498 TTACTIONGTA ACCAAACCTG AAAAAATGTTT GTAATGATTG TATGAAACAT TTGATTCCAC
          GATA-1
-438 AACCCCTCCC TCCTTTACCC TTGTGCTTGC TATCTTCTCT CACCACCATG GATGCCCAGT
          AP-1
-378 GTCCAATTTT TAAGCAACCT TTGAGTCACG GGGTGGTGTA AGAGACTATT CTTTTATATC
          AP-1                B region
-318 ATTGACTCAA AGTTTGCTGA GGAACAAGTC CAGGCAAGTC CTGGGCAAAG GCAGAGAAAT
          A region
-258 CTTTTGTCTT GAGGACACTG ATGGACAGGT CCTGGCTAAG GATTGTGAAA TCCTTTAAGG
          GATA-1
-198 AGCACAGATG GACAAGGCCA GGGGCATCGA GAGAGAGATA AGCTGCCGCT AATGGCCGGG
          GATA-1
-138 AAACGGTCTT TTTGTGTGGA CTTATCTCAA GGAAAATGGC CATCTCAGGA GGTATGCACA
          TEAF                TATA box
-78  GGACTCTTGC TCAAGCCCCC AGGAATGTCA CGTAGGCAGC AGAAAATGGA GGATAAAGA
          └─┬─▶
-18  GGGTCCAATA ACCACAACGG TGGAAGCTGA TCCTTCACCA CAACCACGGC AACGGGAGAG
          GATA-1
+42  GCTTATCTCT CACCACGACA GACTTGAGGG GTTCTCTGCC A

```

**Figure 33 : séquence du promoteur ENS**

La séquence des résidus -738 à +83 par rapport au site d'initiation de la transcription est présentée ici. Le site d'initiation de la transcription est soulignée par une flèche brisée (↵) et la boîte TATA est en caractère gras. L'analyse de cette séquence par le logiciel MatInspector professionnel release 6.1 (1.00 core similarity et 0.95 matrix similarity) a permis d'identifier quelques sites de fixation pour des facteurs de transcription connus (surlignés en gris). Les éléments de réponse A et B caractérisés dans cette étude sont soulignés.

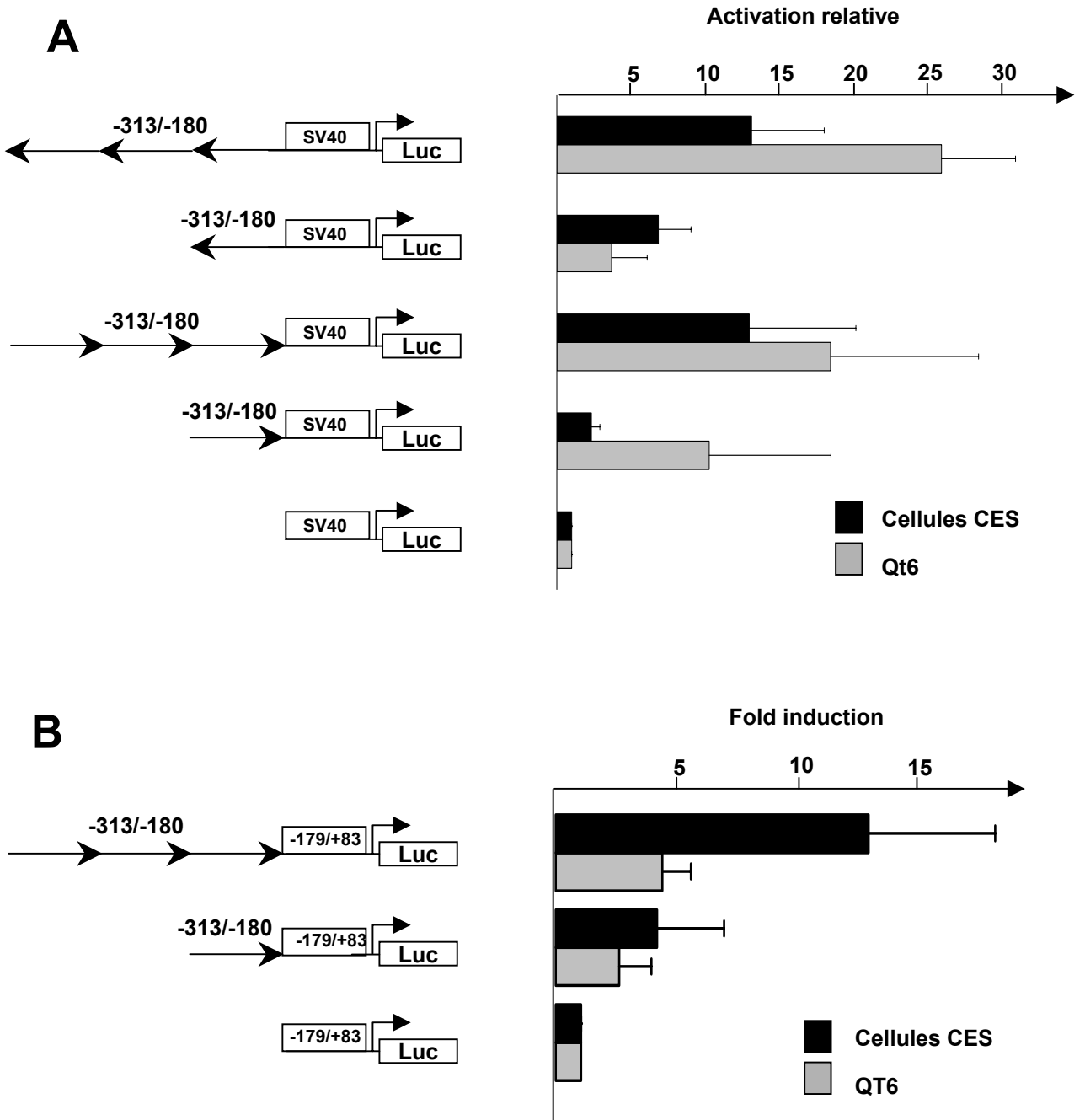


**Figure 34: Analyse fonctionnelle des mutants de délétion du promoteur ENS par transfections transitoires dans les cellules ES de poulet.**

Les différentes délétions ont été produites par PCR. Une représentation schématique de chaque construction est représentée sur la gauche.

**A:** Ces constructions ont été transfectées dans des cellules ES de poulet et l'activité de la luciférase de luciole a été normalisée par le rapporteur interne, luciférase de renilla. Les données sont proportionnelles à celles obtenues avec la construction la plus longue (p738-luc) et ce, pour 3 expériences indépendantes, réalisées en triplicats.

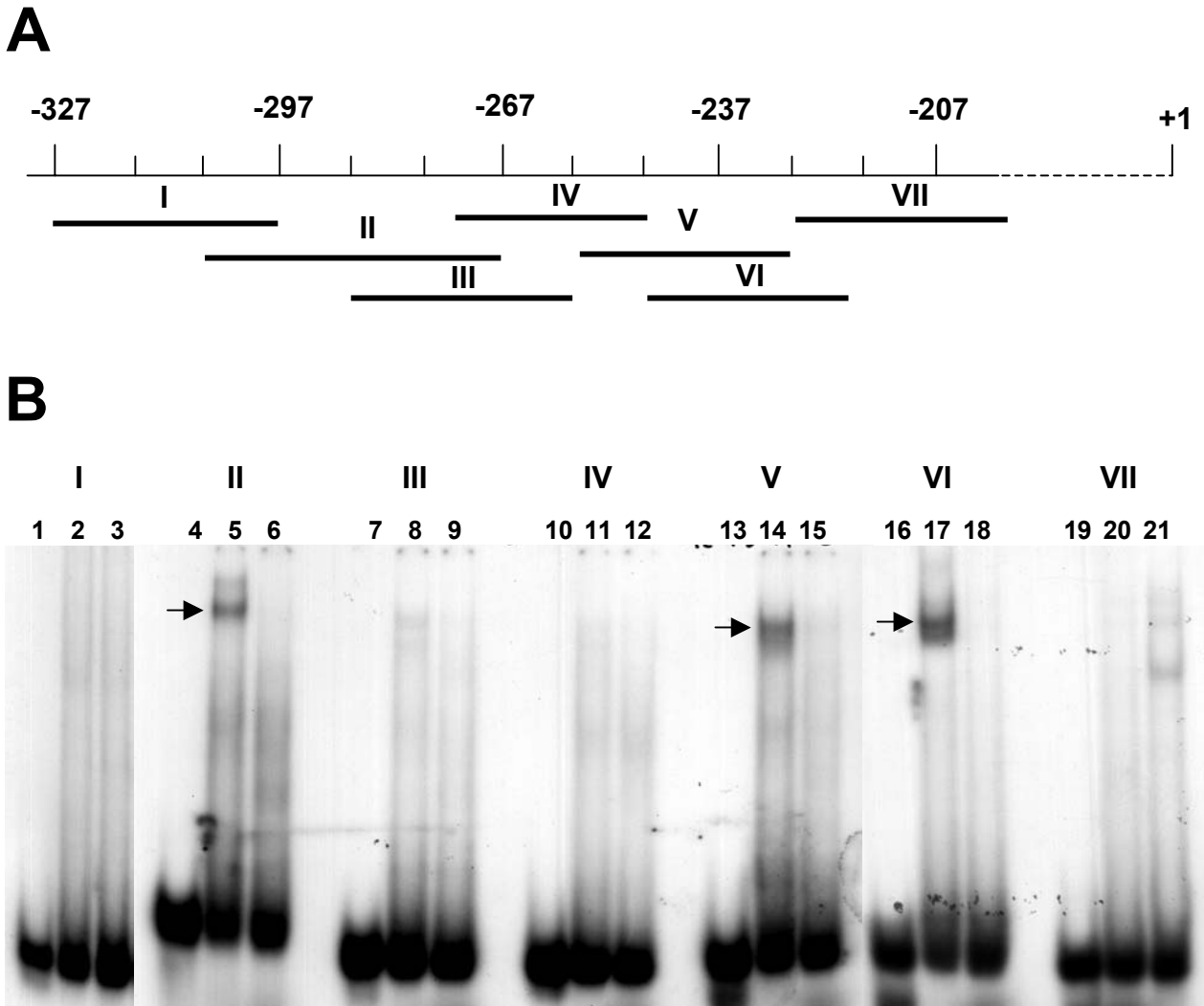
**B:** Ces constructions ont été transfectées dans les cellules CES, CES induites à se différencier par l'acide rétinoïque et QT6. Les valeurs sont proportionnelles à celles obtenues avec la construction p-738-luc dans les cellules CES.



**Figure 35: Analyse fonctionnelle de la région -313/-180 par transfections transitoires.**

**A:** une ou trois copies de la région -313/-180 en amont du promoteur basal SV40 ont été transfectées dans des cellules ES de poulet ou des cellules Qt6. Les flèches indiquent le sens des régions -313/-180. Les valeurs sont proportionnelles à celles obtenues avec le promoteur basal SV40 seul.

**B:** une ou trois copies de la région -313/-180 en amont du promoteur ENS -179/+83 ont été transfectées dans des cellules ES de poulet et des cellules QT6. Les valeurs sont proportionnelles à celles obtenues avec le promoteur ENS -179/+83 dans chaque lignée cellulaire.



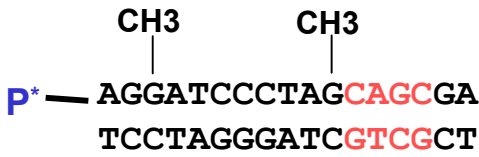
**Figure 36: La région -313/-180 contient deux sites de fixation pour des facteurs nucléaires présents dans les cellules ES de poulet.**

**A:** représentation des 7 oligonucléotides différents utilisés pour couvrir la région -313/-180. Chaque sonde est désignée par un chiffre romain.

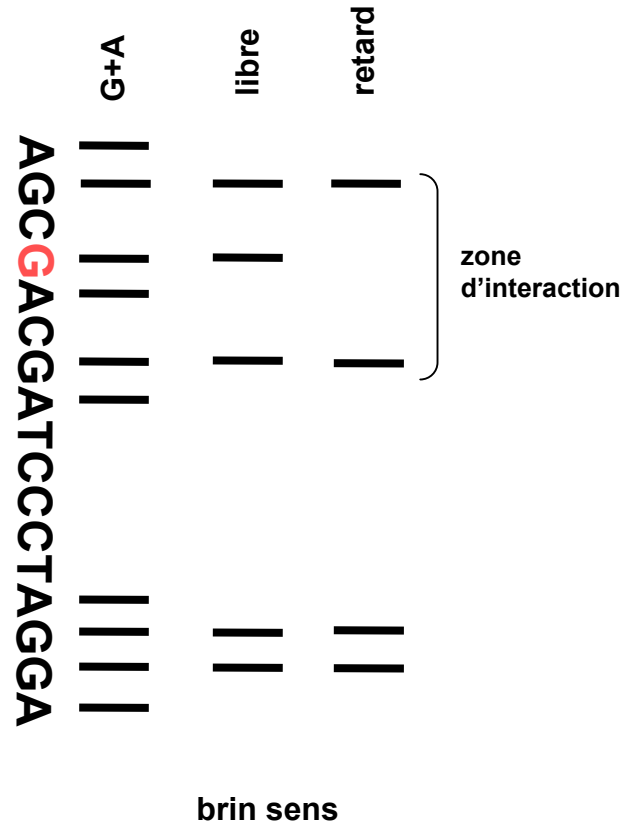
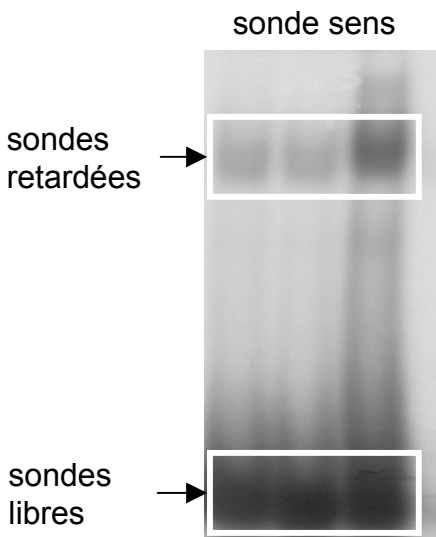
**B:** les expériences de retard sur gel ont été réalisées avec des sondes ADN marquées et des extraits nucléaires de cellules CES. Les positions des complexes ADN-protéines sont indiquées par des flèches. Lignes 1,4,7,10,13,16,19: sonde marquée seule; lignes 2,5,8,11,14,17,20: sonde marquée en présence d'extraits nucléaires; lignes 3,6,9,12,15,18,21: compétition avec la sonde non marquée, en excès molaire de 100-fois.

1<sup>ère</sup> Etape: méthylation partielle de la sonde au DMS

3<sup>ème</sup> Etape: séquençage de l'ADN par traitement à la piperidine et révélation sur gel acrylamide-urée



2<sup>ème</sup> Etape: gel retard avec les sondes méthylées



Les sondes retardées et les sondes libres sont découpées puis l'ADN est électroélué.

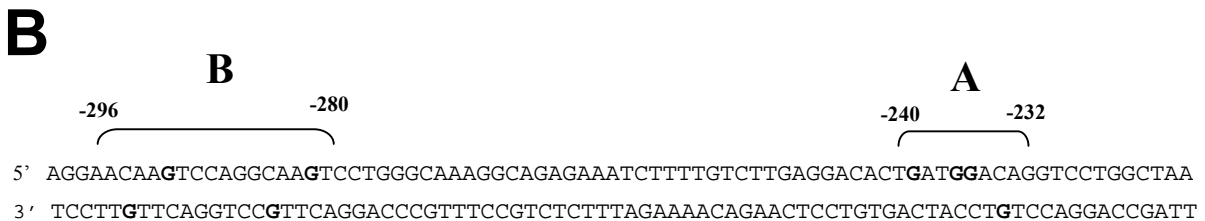
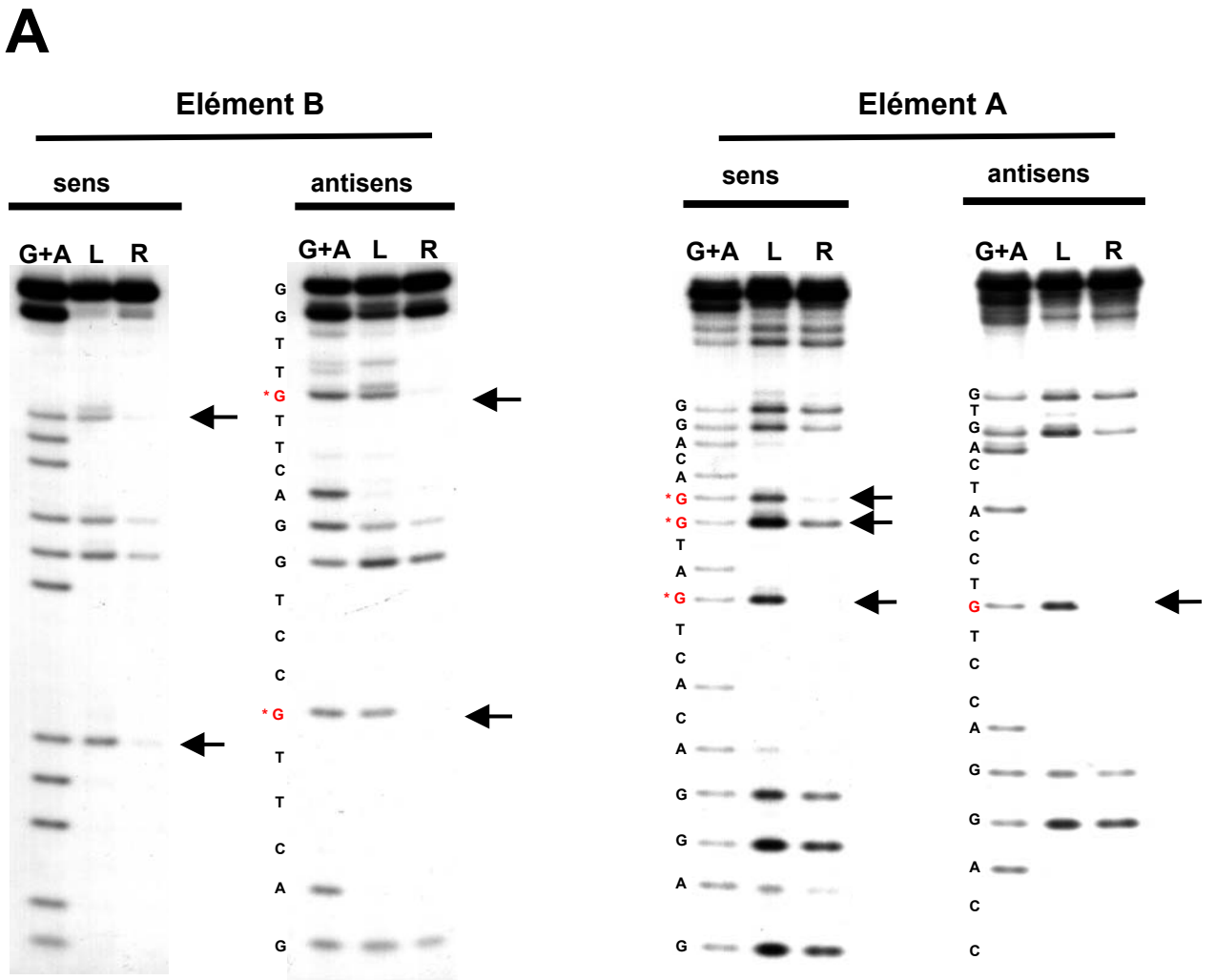
**Figure 37: Principe de l'analyse par interférence de méthylation**

La sonde, marquée au  $\gamma^{32}\text{P}$  sur un brin (sens ou antisens) à son extrémité 5', est méthylée partiellement par un agent méthylant comme le DMS (diméthylsulfate). Celui-ci ne méthyle que les résidus guanines.

La sonde méthylée est ensuite utilisée en gel retard. Si un résidu guanine important pour l'interaction est méthylé alors celui-ci ne pourra pas fixer de complexe protéique, il se retrouvera donc toujours dans la fraction contenant les sondes libres et jamais dans celle contenant les sondes retardées. Après révélation, les sondes retardées et les sondes libres sont ensuite découpées et électroéluées.

Chaque sonde (libre ou retardée) est ensuite traitée à la piperidine, une molécule qui va permettre le clivage de l'ADN au niveau des résidus méthylés.

Les produits de réaction sont normalisés et déposés sur gel acrylamide-urée, en même temps qu'une séquence G+A de la sonde. Après révélation, les bandes correspondant aux résidus guanines important pour l'interaction n'apparaissent pas dans la migration correspondant à la sonde retardée, mais apparaissent dans celle de la sonde libre.



**Figure 38: Caractérisation des éléments de réponse A et B par interférence de méthylation.**

**A:** Les interférences de méthylation ont été réalisées sur les sondes partiellement méthylées - 251/-223 (région A) et -297/-277 (région B) marquées radioactivement en 5' sur le brin sens ou antisens. Les sondes modifiées ont été incubées en présence d'extraits nucléaires de cellules ES de poulet pour la réalisation de gels retards. Les sondes libres (L) ou retardées (R) ont été séquencées et comparées avec la séquence G+A obtenues avec les sondes non méthylées. Les résidus guanines impliqués dans l'interaction ont été identifiés par la perte ou la diminution du signal leur correspondant.

**B:** Les guanines identifiées par interférence de méthylation sont présentées en gras.

# A

Sonde B RE : GAA C<sub>1</sub>A<sub>2</sub>A<sub>3</sub>G<sub>4</sub>T<sub>5</sub>C<sub>6</sub>C<sub>7</sub>AGGC<sub>11</sub>AAG<sub>14</sub>T CCTA

B RE muté en C<sub>1</sub>: GAA T\*AAGTCCAGGCAAGT CCTA

B RE muté en A<sub>2</sub>: GAA CG\*AGTCCAGGCAAGT CCTA

B RE muté en A<sub>3</sub>: GAA CAG\*GTCCAGGCAAGT CCTA

B RE muté en G<sub>4</sub>: GAA CAAA\*TCCAGGCAAGT CCTA

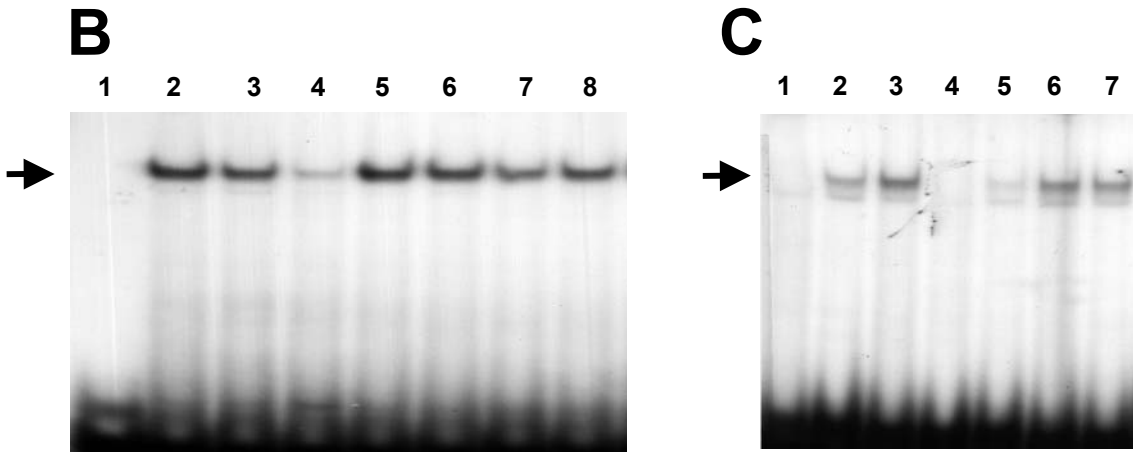
B RE muté en T<sub>5</sub>: GAA CAAGC\*CCAGGCAAGT CCTA

B RE muté en C<sub>6</sub>: GAA CAAGT\*T\*CAGGCAAGT CCTA

B RE muté en C<sub>7</sub>: GAA CAAGTCT\*T\*AGGCAAGT CCTA

B RE muté en C<sub>11</sub>: GAA CAAGTCCAGGT\*AAGT CCTA

B RE muté en G<sub>14</sub>: GAA CAAGTCCAGGCAAA\*T CCTA



**Figure 39: Caractérisation de l'élément de réponse B**

**A:** séquences des oligonucléotides utilisés lors des expériences de retard sur gel.

**B:** Gels retards réalisés avec la sonde seule (ligne1) ou en présence d'extraits nucléaires de cellules CES (ligne 2). La position du complexe ADN-protéine est indiquée par une flèche. Des expériences de compétitions ont été réalisées avec un excès molaire de 10 fois (ligne3) ou 100 fois (ligne 4) d'un oligonucléotide comportant l'élément B, ou d'un excès molaire de 100 fois d'oligonucléotides mutés pour les résidus C<sub>1</sub>, G<sub>4</sub>, C<sub>11</sub> et G<sub>14</sub> de l'élément B (respectivement lignes 5,6,7 et 8).

**C:** Des sondes mutés pour les résidus C<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, G<sub>4</sub>, T<sub>5</sub>, C<sub>6</sub>, C<sub>7</sub> ont été marquées radioactivement et utilisées pour des gels retards avec des extraits nucléaires de cellules CES (respectivement lignes 1,2,3,4,5,6 et 7).



# A

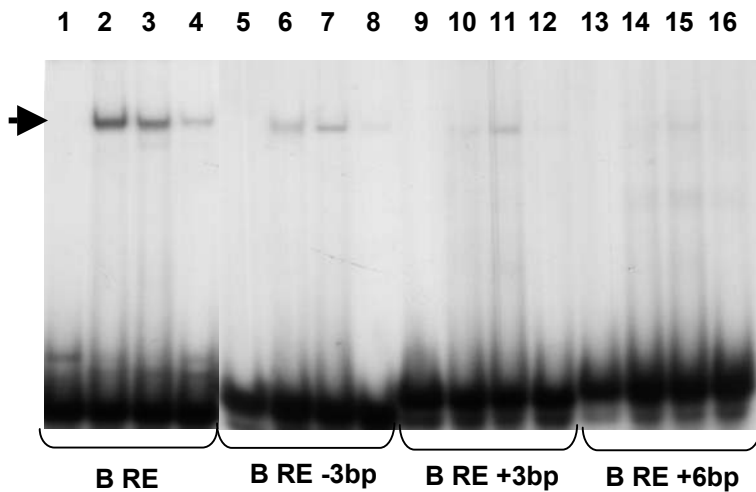
**B RE probe:** GAA CAAGTCCAGGCAAGT CCTA

**B RE -3bp probe:** GAA CAAGTCCAAGT CCTA

**B RE +3bp probe:** GAA CAAGTCCAGGAGGCAAGT CCTA

**B RE +6bp probe:** GAA CAAGTCCAGGAGGAGGCAAGT CCTA

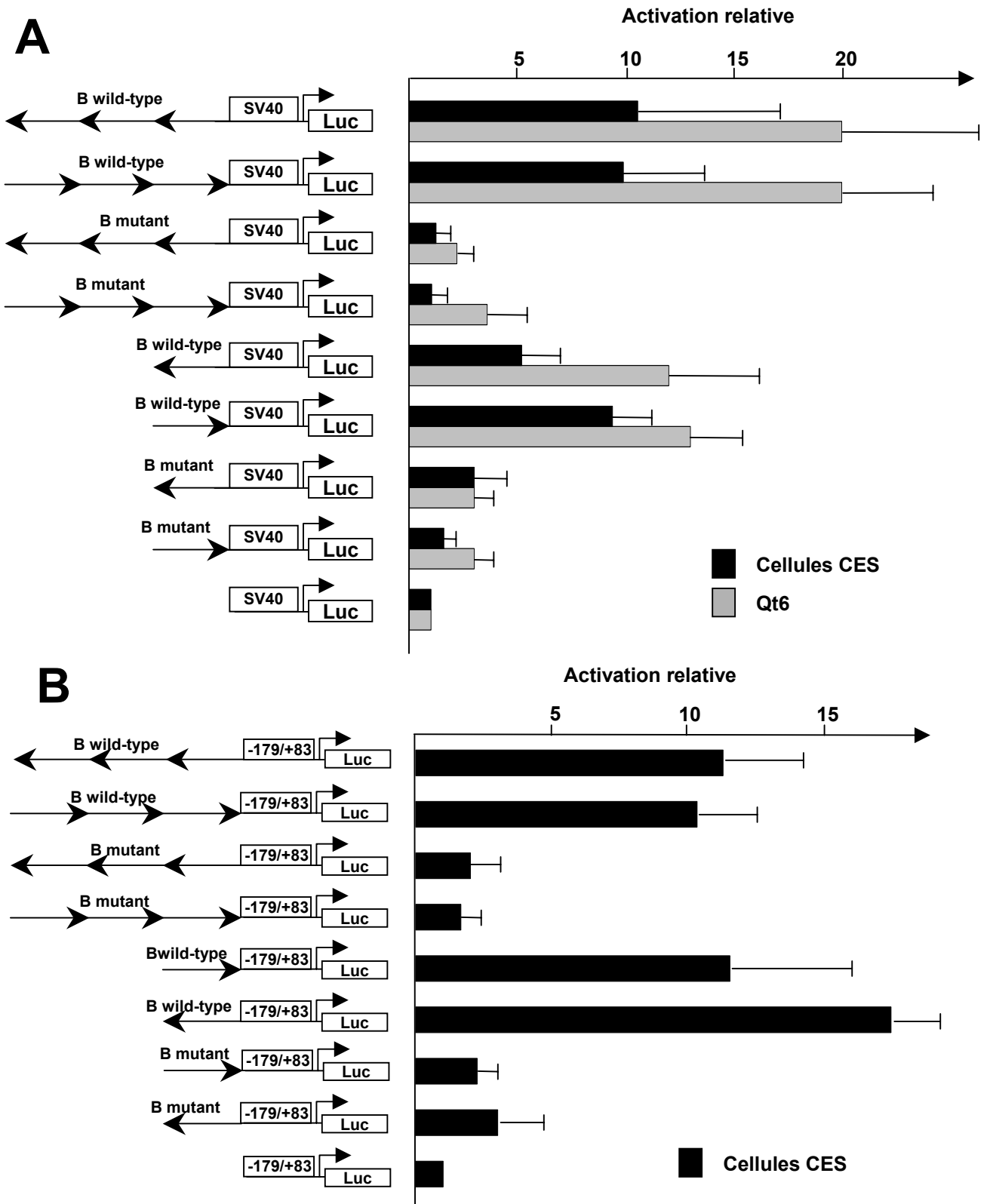
# B



**Figure 40: l'intervalle entre les motifs de la répétition directe de l'élément B influence la fixation du complexe protéique.**

**A:** séquences des oligonucléotides utilisés.

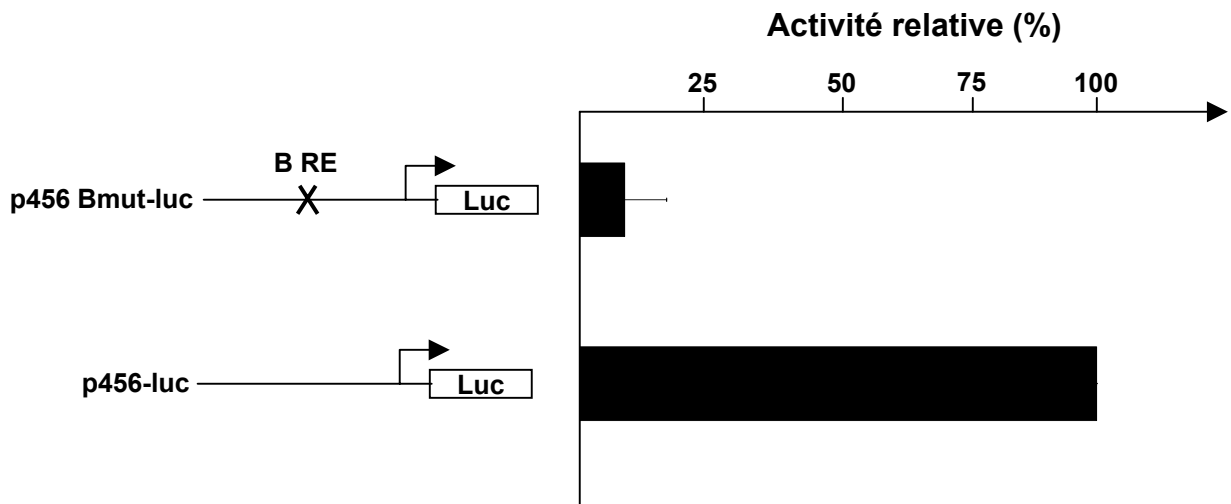
**B:** des expériences de retard sur gel ont été réalisées avec chacune des sondes marquées, soit seules (lignes 1,5,9,13), soit en présence d'extraits nucléaires de cellules CES (ligne 2,6,10 et 14). La position du complexe ADN-protéine est indiquée par une flèche. Les compétitions ont été réalisées avec un excès molaire de 100 fois de l'oligonucléotide comportant un élément B muté (ligne 3,7,11 et 15) ou un élément B (ligne 4,8,12 et 16).



**Figure 41: Analyse fonctionnelle de l'élément B**

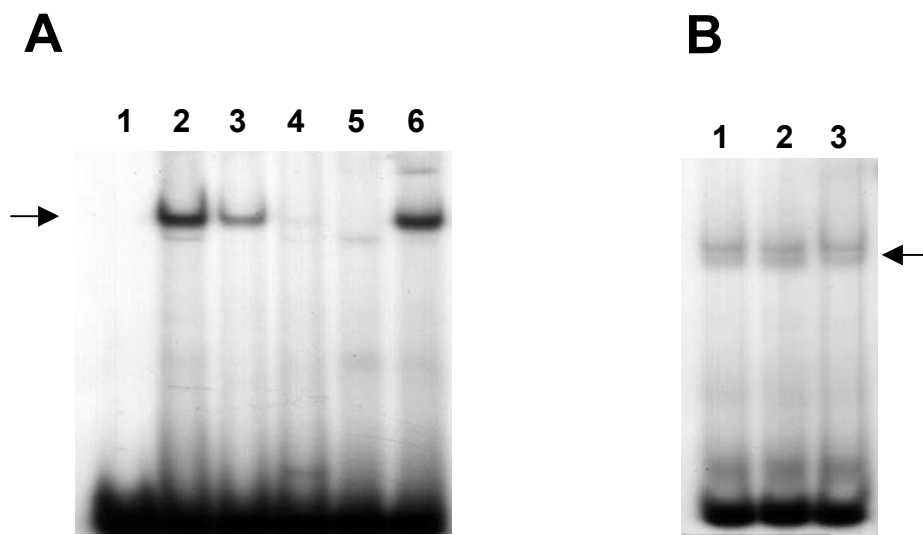
**A:** 1 ou 3 copies de l'élément B, muté ou non, ont été placées en amont du promoteur basal SV40 et transfectées dans les cellules ES de poulet et les cellules QT6. Les valeurs sont proportionnelles à celles obtenues avec le promoteur basal SV40 seul.

**B:** 1 ou 3 copies de l'élément B, muté ou non, ont été placées en amont du promoteur ENS -179/+83 et transfectées dans les cellules ES de poulet. Les valeurs sont proportionnelles à celles obtenues avec le promoteur ENS -179/+83 seul.



**Figure 42: Analyse fonctionnelle de l'élément B par mutagenèse dirigée sur le promoteur ENS**

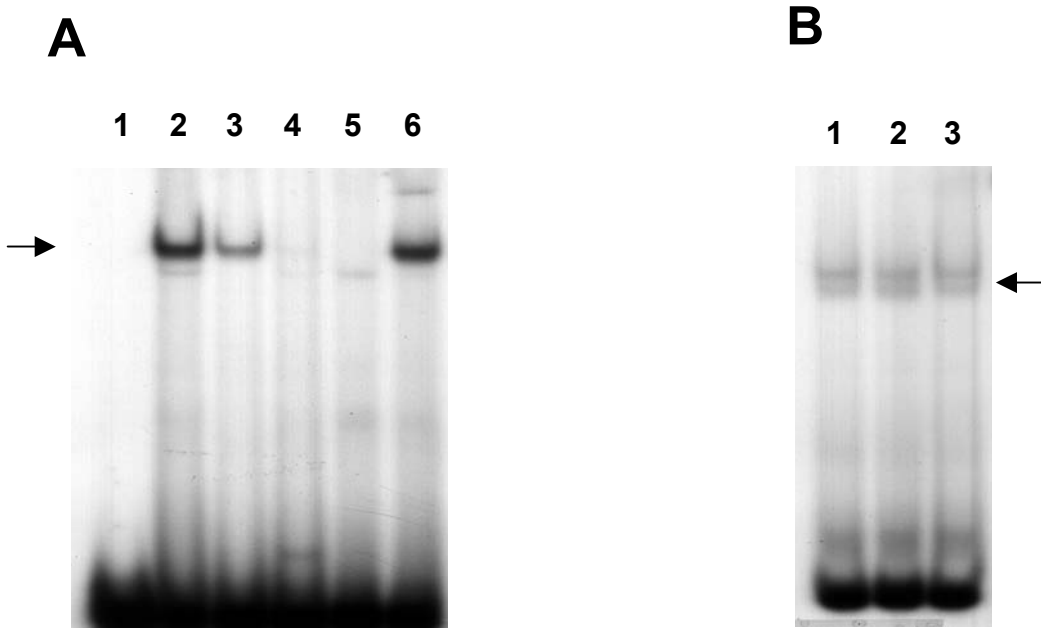
L'élément B a été détruit par mutagenèse dirigée sur le promoteur ENS -456/+83. Les constructions avec ou sans mutations ont été transfectées dans les cellules CES. Les valeurs sont proportionnelles à celles obtenues avec le promoteur ENS -456/+83.



**Figure 43: l'élément B fixe une protéine reconnue par un antisérum CP2.**

**A:** gels retards avec une sonde marquée seule, comportant l'élément B (ligne 1), ou en présence d'extraits nucléaires de cellules CES (ligne 2). Le complexe ADN-protéine est indiquée par une flèche noire. Des compétitions ont été réalisées avec un excès molaire de 100 fois d'oligonucléotides mutés pour l'élément B (ligne 3) ou non muté (ligne 4). L'antisérum CP2 et le sérum préimmun de lapin (1µl) ont été incubés avec des extraits nucléaires de cellules CES avant l'addition de la sonde B marquée (respectivement ligne 5 et 6).

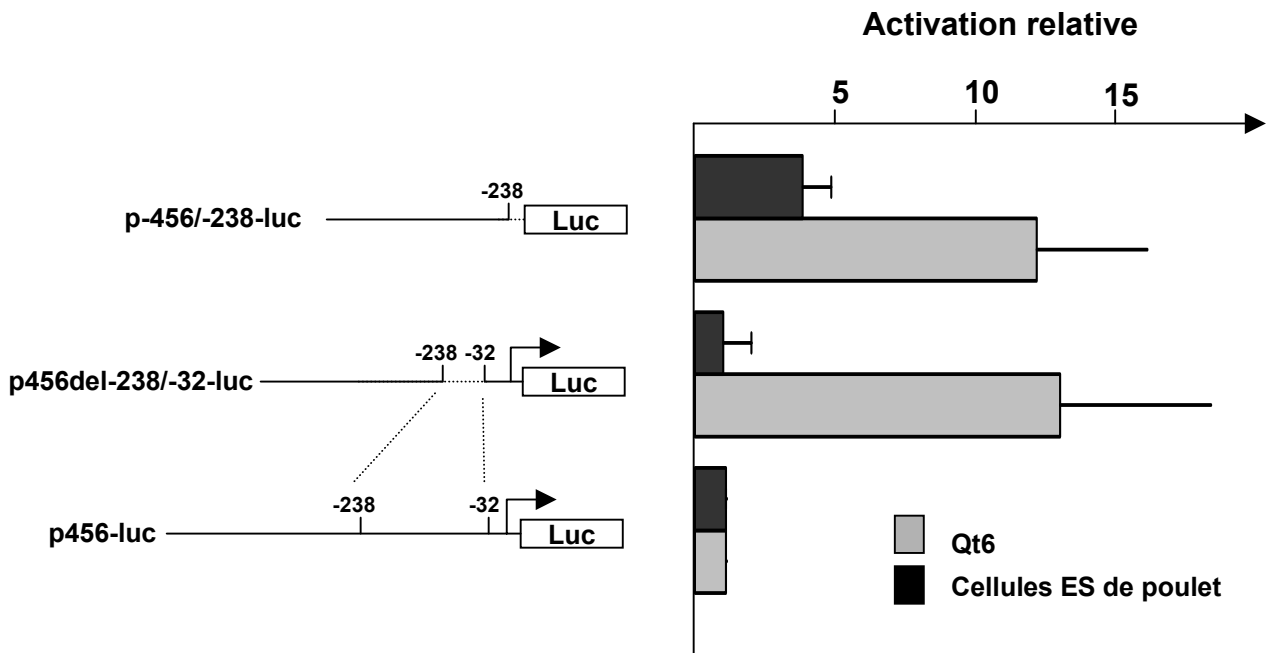
**B:** gels retards réalisés avec une sonde marquée comportant l'élément A et des extraits de cellules CES (ligne1). Le complexe ADN-protéine est indiqué par une flèche. L'antisérum CP2 et le sérum préimmun de lapin ont été incubés avec des extraits nucléaires de cellules CES avant l'addition de la sonde A (lignes 2 et 3).



**Figure 43: l'élément B fixe une protéine reconnue par un antisérum CP2.**

**A:** gels retards avec une sonde marquée seule, comportant l'élément B (ligne 1), ou en présence d'extraits nucléaires de cellules ES de poulet (ligne 2). Le complexe ADN-protéine est indiquée par une flèche noire. Des compétitions réalisées avec un excès molaire de 100 fois d'oligonucléotides mutés pour l'élément B (ligne 3) ou non muté (ligne 4). L'antisérum CP2 et le sérum préimmun de lapin (1µl) ont été incubés avec des extraits nucléaires de cellules CES avant l'addition de la sonde B marquée (respectivement ligne 5 et 6).

**B:** gels retards réalisés avec une sonde marquée comportant l'élément A et des extraits de cellules CES (ligne1). Le complexe ADN-protéine est indiqué par une flèche. L'antisérum CP2 et le sérum préimmun de lapin ont été incubés avec des extraits nucléaires de cellules CES avant l'addition de la sonde A (lignes 2 et 3).



**Figure 44: Analyse fonctionnelle du promoteur ENS par des délétions en 3'**

La construction p456del-238/-32-luc a été produite par PCR et correspond à une délétion du promoteur entre les positions -238 et -32. La boîte TATA est conservée dans cette construction. La construction p-456/-238-luc correspond à la région 5' du promoteur entre les positions -456 et -238. Ces constructions ont été transfectées dans des cellules ES de poulet et l'activité de la luciférase de luciole a été normalisée par le rapporteur interne, luciférase de renilla. Les données sont proportionnelles, indépendamment pour chaque lignée cellulaire, à celles obtenues avec la construction p456-luc.