



*Université de Strasbourg*  
*École doctorale des Sciences de la Vie et de la Santé*

## **THÈSE**

Présentée pour obtenir le grade de  
Docteur de l'Université de Strasbourg

Discipline : *Sciences du Vivant*  
Spécialité : *Aspects moléculaires et cellulaires de la Biologie*

Par

**Mohamed-Amin CHOUKRALLAH**

### **Partie I :**

**Etude structure-fonction de TBP dans les fibroblastes murins**

### **Partie II :**

**Etude de l'action de l'acide rétinoïque dans la différenciation neuronale  
des cellules souches embryonnaires**

Soutenue publiquement le 28/09/2010

### **Membres du jury :**

|                      |                             |
|----------------------|-----------------------------|
| Directeur de thèse : | <b>Dr. Irwin DAVIDSON</b>   |
| Rapporteur externe : | <b>Dr. Michel WERNER</b>    |
| Rapporteur externe : | <b>Dr. Martin TEICHMANN</b> |
| Examineur :          | <b>Dr. Ali HAMICHE</b>      |

Remerciement

## Remerciements

Je tiens d'abord à remercier les membres du jury Michel WERNER, Martin TEICHMANN et Ali HAMICHE pour avoir accepté d'évaluer mon travail de thèse.

Je remercie Irwin pour m'avoir accepté dans son équipe et d'avoir été toujours disponible pour discuter, écouter et proposer.

Je remercie tous les membres de l'équipe, Igor mon fidèle compagnon de beuveries et de CHIP, Dominique qui s'est occupée de ma formation quand je suis arrivé au laboratoire, Thomas "de son vrai nom : Junior" pour la bonne humeur qu'il répand toujours autour de lui, les membres de l'autre bord (4039), Daniel, Manu, Attailah et Isabelle. Enfin un grand merci à Gabrielle pour les nombreuses corrections qu'elle a apportées à ce manuscrit.

Je souhaite bonne chance aux nouveaux membres, Sylvia et Diana pour le projet ES.

Je remercie également les membres passés de l'équipe, Emilie, Martin, Raffa, Ullas et Anas avec qui j'ai passé d'agréables moments à l'IGBMC. Un merci particulier à mon autre "compagnon" de java Laurence. Je remercie aussi mes acolytes du baby-foot, Benjamin, Benoit et Nico. Merci également à Arnaud Krebs, je profite de l'occasion pour le féliciter pour la naissance de la petite Justine.

J'aimerais également remercier les membres de la plateforme de séquençage, Bernard, Serge, Stéphanie, Céline, Stéphanie et Tao.

Enfin, je remercie mes amis, Vincent, Nico, Krichou, Eric, Karine, Anto et... enfin la liste est longue, mais je devrais quand même remercier l'abruti qui me sert de meilleur ami pour son soutien indéfectible.

Pour finir, je remercie Galina pour sa patience et son affection. Cette thèse lui est naturellement dédiée.

## Table des matières

|   |           |
|---|-----------|
| <b>Remerciements .....</b>  | <b>1</b>  |
| <b>Table des matières.....</b>  | <b>2</b>  |
| <b>Liste des abréviations.....</b>  | <b>8</b>  |
| <b>Liste des figures et des tableaux, partie I .....</b>                            | <b>13</b> |
| <b>Liste des figures et des tableaux, partie II .....</b>                           | <b>15</b> |
| <br>  |           |
| <b>PARTIE I : ETUDE STRUCTURE-FONCTION DE TBP DANS LES FIBROBLASTES MURINS.....</b> | <b>17</b> |
| <br>  |           |
| <b>INTRODUCTION .....</b>   | <b>18</b> |
| <br>  |           |
| <b>CHAPITRE 1 : LA TRANSCRIPTION DES GENES DE CLASSE II .....</b>                   | <b>19</b> |
| <br>  |           |
| <b>I. Anatomie des promoteurs des gènes de classe II.....</b>                       | <b>20</b> |
| I.1. Le promoteur minimal .....   | 20        |
| I.1.1. Vision classique du promoteur minimal.....                                   | 20        |
| I.1.1.1. L'élément TATA .....   | 21        |
| I.1.1.2. L'élément Initateur (Inr) .....  | 21        |
| I.1.1.3. L'élément DPE .....  | 21        |
| I.1.1.4. Les éléments BRE.....  | 22        |
| I.1.1.5. Les îlots CpG .....  | 22        |
| I.1.2. Nouvelle vision du promoteur minimal .....                                   | 23        |
| I.1.2.1. Promoteurs à TSS unique versus promoteurs à multiples TSSs .....           | 23        |

## Table des matières

|   |           |
|---|-----------|
| I.1.2.2. Les promoteurs alternatifs .....   | 23        |
| I.2. Les éléments régulateurs .....   | 24        |
| I.2.1. Le promoteur proximal .....  | 24        |
| I.2.2. Les séquences distales .....   | 24        |
| <b>II. L'initiation de la transcription.....</b>                                  | <b>25</b> |
| II.1. La machinerie de transcription de base .....                                | 26        |
| II.1.1. L'ARN polymérase II .....   | 26        |
| II.1.1.2. La structure de l'ARN polymérase II .....                               | 27        |
| II.1.1.3. Le domaine carboxy-terminal de RPB1 .....                               | 27        |
| II.1.2. TFIIA .....   | 29        |
| II.1.3. TFIIB .....   | 31        |
| II.1.4. TFIIF .....   | 32        |
| II.1.5. TFIIE .....   | 33        |
| II.1.6. TFIIF .....   | 34        |
| II.1.7. TFIS .....  | 35        |
| II.2. Les cofacteurs généraux de la transcription .....                           | 36        |
| II.2.1. Les complexes USA .....   | 37        |
| II.2.2. Le complexe Médiateur .....   | 38        |
| II.2.3. Negative cofactor 2 .....   | 41        |
| II.3. Les modèles d'assemblage du complexe de préinitiation <i>in vitro</i> ..... | 41        |
| <b>III. La régulation de la transcription .....</b>                               | <b>42</b> |
| <b>IV. La transcription dans le contexte chromatinien.....</b>                    | <b>44</b> |
| IV.1. L'organisation de la chromatine .....                                       | 44        |
| IV.2. Remodelage et modification de la chromatine .....                           | 45        |
| IV.2.1. La méthylation de l'ADN .....   | 45        |
| IV.2.2. Les modifications des histones .....                                      | 46        |
| IV.2.2.1. L'acétylation des histones .....  | 46        |

## Table des matières

|  |           |
|--|-----------|
| IV.2.2.2. La méthylation des histones .....  | 47        |
| IV.2.2.3. La phosphorylation des histones .....  | 49        |
| IV.2.3. La reconnaissance des modifications des histones .....   | 50        |
| IV.2.4. Les complexes de remodelage de la chromatine .....   | 50        |
| <b>CHAPITRE 2 : TBP .....</b>  | <b>51</b> |
| <b>I. Organisation structurale de TBP .....</b>  | <b>52</b> |
| I.1. Le domaine N-terminal de TBP .....  | 53        |
| II.2. Le domaine C-terminal de TBP .....   | 53        |
| II.3. Cartographie des sites d'interaction de TBP avec ses partenaires.....                              | 54        |
| II.3.1. Structure cristallographique du complexe TBP-ADN .....   | 54        |
| II.3.2. Structure cristallographique du complexe TBP-TFIIB-ADN .....                                     | 55        |
| II.3.3. Structure cristallographique du complexe TBP-TFIIA-ADN.....                                      | 55        |
| II.3.4. Structure cristallographique du complexe NC2-TBP-ADN .....                                       | 56        |
| II.3.5. Analyse de TBP par mutagenèse .....  | 56        |
| II.3.5.1. La variante TBP spm3 .....   | 56        |
| II.3.5.2. Mutagenèse systématique des résidus exposés au solvant dans le domaine C-terminal de TBP ..... | 57        |
| <b>II. TBP dans la transcription des gènes de classe II.....</b>   | <b>59</b> |
| II.1. TFIID .....  | 59        |
| II.1.1. Organisation structurale de TFIID.....   | 59        |
| II.1.2. Recrutement de TFIID sur les promoteurs.....   | 60        |
| II.1.2.1. Recrutement via des interactions directes avec les activateurs.....                            | 61        |
| II.1.2.2. Recrutement via des interactions avec les modifications des histones .....                     | 62        |
| II.1.3. Les activités enzymatiques de TFIID .....  | 62        |
| II.1.4. Le rôle de coactivateur de TFIID.....  | 63        |
| II.2. B-TFIID .....  | 64        |
| II.2.1. Organisation structurale de B-TFIID .....  | 64        |

## Table des matières

|  |                |
|--|----------------|
| II.2.1.1. Organisation de la séquence primaire de B-TAF1 .....   | 64             |
| II.2.1.2. TBP interagit avec la région N-terminale de B-TAF1 .....   | 65             |
| II.2.1.3. Cartographie des surfaces de TBP impliquées dans l'interaction avec B-TAF1 .....   | 65             |
| II.2.1.4. L'architecture tridimensionnelle de B-TFIID.....   | 67             |
| II.2.1.5. Immunolocalisation de TBP et B-TAF1 dans le complexe B-TFIID.....  | 67             |
| II.2.2. La fonction de B-TAF1 et MOT1P et leur mécanisme d'action.....   | 67             |
| II.2.2.1. La dualité de la fonction de B-TAF1 : répresseur / activateur.....   | 67             |
| II.2.2.2. L'activité ATPase est essentielle pour la fonction de B-TAF1 .....   | 68             |
| II.2.2.3. MOT1P et NC2 agissent en synergie pour réguler l'activité de TBP .....   | 69             |
| <b>RESULTATS .....</b>   | <b>- 72 -</b>  |
| <b>Publication 1 :.....</b>  | <b>- 75 -</b>  |
| <b>RNA polymerase II transcription in murine cells lacking the B-TFIID complex. ....</b>   | <b>- 75 -</b>  |
| <b>Discussion.....</b>   | <b>- 116 -</b> |
| I. Les effets des mutations dans TBP sur sa fonction <i>in vivo</i> .....  | - 116 -        |
| II. Altération du complexe B-TFIID <i>in vitro et in vivo</i> .....  | - 117 -        |
| III. Dérégulation de l'expression des gènes après l'altération de B-TFIID .....  | - 117 -        |
| IV. La transcription par la Pol II en absence de B-TFIID.....  | - 117 -        |
| V. Le rôle d'élément TATA <i>in vivo</i> .....   | - 121 -        |
| <b>Publication 2 :.....</b>  | <b>- 123 -</b> |
| <b>Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. Mol Cell Biol. 30(1):231-44. ....</b> | <b>- 123 -</b> |
| <b>Publication 3 :.....</b>  | <b>- 124 -</b> |
| <b>Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. BMC Genomics, 11:530.....</b>                                     | <b>- 124 -</b> |
| <b>Publication 4 :.....</b>  | <b>- 125 -</b> |

Table des matières

|  |                |
|--|----------------|
| seqMINER: An integrated ChIP-seq data interpretation platform .....  | - 125 -        |
| <b>PARTIE II: ETUDE DE L'ACTION DE L'ACIDE RETINOÏQUE DANS LA DIFFERENCIATION NEURONALE DES CELLULES SOUCHES EMBRYONNAIRES..</b> | <b>- 126 -</b> |
| <b>AVANT-PROPOS.....</b>   | <b>- 127 -</b> |
| <b>INTRODUCTION .....</b>  | <b>- 128 -</b> |
| <b>I. Les cellules souches embryonnaires pluripotentes.....</b>  | <b>- 129 -</b> |
| <b>II. La neurulation chez l'embryon de souris.....</b>  | <b>- 130 -</b> |
| <b>III. La voie de signalisation de l'acide rétinoïque.....</b>  | <b>- 130 -</b> |
| III.1. Le cycle de l'acide rétinoïque.....   | - 131 -        |
| III.2. Les récepteurs de l'acide rétinoïque .....  | - 131 -        |
| III.2.1 Organisation structurale des récepteurs de l'acide rétinoïque .....  | - 132 -        |
| III.2.2. Le domaine de liaison à l'ADN : DBD .....   | - 132 -        |
| III.2.3. Le domaine de liaison du ligand : LBD .....   | - 133 -        |
| III.2.4. La fonction d'activation AF1 .....  | - 134 -        |
| III.3. Modèle classique de l'activation des gènes cibles de l'AR .....   | - 134 -        |
| III.3.1. Les corépresseurs des récepteurs à l'AR.....  | - 134 -        |
| III.3.2. Les Coactivateurs des récepteurs à l'AR.....  | - 135 -        |
| III.3.3. Modèle séquentiel d'activation des gènes cibles de l'AR .....   | - 135 -        |
| <b>IV. La différenciation neuronale dépendante de l'AR <i>in vitro</i> .....</b>   | <b>- 136 -</b> |
| <b>V. Les facteurs HES dans la différenciation neuronale .....</b>   | <b>- 137 -</b> |
| <b>OBJECTIFS.....</b>  | <b>- 139 -</b> |
| <b>MATERIEL ET METHODES.....</b>   | <b>- 140 -</b> |
| I. Technique de RNA-seq (3' end sequencing).....   | - 140 -        |

## Table des matières

|   |                |
|---|----------------|
| II. Génération de lignées cellulaires stables exprimant les shRNA .....                                     | - 140 -        |
| <b>RESULTATS .....</b>  | <b>- 142 -</b> |
| I. Définition de groupes de gènes ayant un profil d'expression similaire ou "clustering" .....              | - 142 -        |
| II. Génération des lignées stables dans lesquelles l'expression de HES3 est réduite.....                    | - 143 -        |
| III. L'expression de HES3 est nécessaire à l'expression des marqueurs de la différenciation neuronale. .... | -              |
| 144 -   |                |
| Discussion et perspectives .....  | - 145 -        |
| <b>Bibliographie .....</b>  | <b>- 148 -</b> |
| Bibliographie partie  |                |
| Bibliographie partie II   |                |



## Liste des abréviations

|          |  |
|----------|--|
| Aa       | Acides aminés  |
| ACE1     | <i>Transcriptional activator protein ACE1</i>              |
| ADH1     | <i>Alcohol dehydrogenase</i>                               |
| ADH7     | Alcool Déshydrogénase 7                                    |
| AdMLP    | <i>Adenovirus Major Late Promoter</i>                      |
| ADN      | Acide Désoxyribonucléique                                  |
| ADP      | Adénosine DiPhosphate                                      |
| AF2      | <i>Activation Function 2</i>                               |
| AP2      | <i>Activator Protein 2</i>                                 |
| AR       | Acide Rétinoïque   |
| ARC      | <i>Activator Recruited Cofactor</i>                        |
| ARN      | Acide RiboNucléique  |
| ARNm     | ARN messenger  |
| ARNr     | ARN ribosomique  |
| ARNt     | ARN de transfert   |
| ASCL1    | <i>Achaete-Scute Complex homolog 1</i>                     |
| ATOH1    | <i>Atonal Homolog 1</i>                                    |
| ATP      | Adénosine TriPhosphate                                     |
| BDF1     | <i>Bromodomain Factor 1</i>                                |
| bHLH     | <i>basic Helix-Loop-Helix</i>                              |
| BMP      | <i>Bone Morphogenic Proteins</i>                           |
| BRE      | <i>TFIIB Responsive Element</i>                            |
| CAGE     | <i>Cap Analysis of Gene Expression</i>                     |
| CAK      | <i>Cyclin-Activating Kinase complex</i>                    |
| CARM1    | <i>Coactivator-Associated aRginine Methyltransferase 1</i> |
| CDK      | <i>Cyclin Dependant Kinase</i>                             |
| ChIP     | <i>Chromatin ImmunoPrecipitation</i>                       |
| CK2      | <i>Casein Kinase 2</i>                                     |
| CNS      | Système nerveux central                                    |
| CoRNRbox | <i>CoRepressor Nuclear Receptor box</i>                    |
| CRABP2   | <i>Cellular Retinoic-Acid Binding Protein</i>              |
| CRSP     | <i>Cofactor Required for SP1 activation</i>                |
| CTD      | <i>C-Terminal Domain</i>                                   |
| Cyp26    | <i>Cytochrome P450, family 26</i>                          |
| D        | <i>Drosophila melanogaster</i>                             |
| Da       | <i>Daltons</i>   |

## Liste des abréviations

|         |   |
|---------|---|
| DBD     | <i>DNA Binding Domain</i>   |
| DNMT    | <i>DNA MethylTransferase</i>  |
| Dot1L   | <i>Disrupter Of Telomere silencing protein 1- Like</i>  |
| DPE     | <i>Downstream Promoter Element</i>  |
| DR      | Répétition Directe  |
| DSIF    | <i>5,6-dichloro-1-<math>\beta</math>-D-ribofuranosylbenzimidazole (DRB)-Sensitivity Inducing Factor</i> |
| EB      | <i>Embryoid Bodies</i>  |
| EMSA    | <i>Electrophoretic Mobility Shift Assay</i>   |
| ES      | <i>Embryonic Stem (cells)</i>   |
| FCP1    | <i>TFIIF-associated CTD Phosphatase 1</i>   |
| FGF     | <i>Fibroblast Growth Factor</i>   |
| FRAP    | <i>Fluorescent Recovery After Photobleaching</i>  |
| FXR     | <i>Farnesoid X Receptor</i>   |
| GAL     | Galactose   |
| GBX2    | <i>Gastrulation Brain Homeobox 2</i>  |
| GPS2    | <i>G protein Pathway Suppressor 2</i>   |
| Grg     | <i>Groucho homolog</i>  |
| GRO-seq | <i>Global Run-on Sequencing</i>   |
| GTF     | <i>General Transcription Factor</i>   |
| H       | <i>human (Homo sapiens)</i>   |
| HAT     | histone acétyltransférase   |
| HDAC    | histone désacétylases   |
| HDM     | histones déméthylases   |
| HEK     | <i>Human Embryonic Kidney</i>   |
| HES3    | <i>Hairy and Enhancer of Split 3</i>  |
| HFD     | <i>Histone Fold Domain</i>  |
| HK      | histone kinases   |
| HMG1    | <i>High Mobility Group Protein 1</i>  |
| HMT     | histone méthyltransférase   |
| HP1     | <i>Heterochromatin Protein 1</i>  |
| HTH     | <i>Helix-Turn-Helix</i>   |
| HXT     | <i>Hexose Transporter</i>   |
| ICM     | <i>Inner Cell Mass</i>  |
| IL6     | <i>Interleukine 6</i>   |
| Inr     | élément initiateur  |
| LBD     | <i>Ligand Binding Domain</i>  |
| LBP     | <i>Ligand Binding Pocket</i>  |
| LIF     | <i>Leukemia Inhibitory Factor</i>   |
| LSD1    | <i>Lysine Specific Demethylase 1</i>  |
| LXR     | <i>Liver X Receptor</i>   |
| M       | <i>mouse (Mus musculus)</i>   |

## Liste des abréviations

|         |   |
|---------|---|
| MAPt    | <i>Microtubule Associated Protein tau</i>   |
| MBD     | <i>Methyl Binding Domain</i>  |
| MED     | <i>Mediateur</i>  |
| MEF     | <i>Mouse Embryonic Fibroblast</i>   |
| MET17   | <i>Methionine and cysteine synthase (O-acetyl homoserine-O-acetyl serine sulfhydrylase)</i> |
| Mot1P   | <i>Modifier Of Transcription 1 protein</i>  |
| NAD     | nicotinamide adénine dinucléotide   |
| NC      | <i>Negative Cofactor</i>  |
| NCoA    | <i>Nuclear receptor CoActivator</i>   |
| NCoR    | <i>Nuclear receptor CoRepressor</i>   |
| NELF    | <i>Negative Elongation Factor</i>   |
| NER     | nucleotide excision repair  |
| NEUROG  | Neurogenine   |
| NF-Y    | <i>Nuclear Factor-Y</i>   |
| NGFIB   | <i>Nerves Growth Factor IB</i>  |
| NR      | <i>Nuclear Receptors</i>  |
| NRbox   | <i>Nuclear Receptor box</i>   |
| NURR1   | Nuclear Receptor Related protein 1  |
| PARP-1  | <i>Poly ADP-Ribose Polymerase-1</i>   |
| Pax6    | <i>Paired box gene 6</i>  |
| Pb      | Paires de Bases   |
| Pc      | Polycomb  |
| PC      | <i>Positive Cofactor</i>  |
| PGK1    | <i>PhosphoGlycerate Kinase</i>  |
| PHD     | <i>Plant Homeodomain</i>  |
| PIC     | <i>PreInitiation Complex</i>  |
| Pol I   | ARN polymérase I  |
| Pol II  | ARN polymérase II   |
| Pol III | ARN polymérase III  |
| PPAR    | <i>Peroxisome Proliferator-Activated Receptor</i>   |
| PRMT    | <i>Protein aRginine MethylTransferase</i>   |
| P-TEFb  | <i>Positive Transcription Elongation Factor</i>   |
| PXR     | Pregnane X Receptor   |
| qPCR    | <i>quantitative Polymerase Chain Reaction</i>   |
| Ral     | Rétinaldéhyde   |
| RALDH   | Rétinaldéhyde Déshydrogénase  |
| RAP1    | <i>Repressor/Activator Protein 1</i>  |
| RAR     | <i>Retinoic Acid Receptor</i>   |
| RARE    | <i>Retinoic Acid Response Element</i>   |
| RBP4    | <i>Retinol-Binding Protein 4</i>  |
| RD      | <i>Repression Domain</i>  |

## Liste des abréviations

|                  |   |
|------------------|---|
| RDH10            | Rétinol Déshydrogénase 10                                   |
| RNA-seq          | séquençage à haut débit des ARNs                            |
| RP               | Ribosomal Protein   |
| RPB              | <i>RNA Polymerase B</i>                                     |
| RT               | <i>Reverse Transcription</i>                                |
| RXR <sub>s</sub> | <i>Retinoid X Receptors</i>                                 |
| SAGA             | SPT ADA GCN5  |
| SANT             | Swi3/Ada2/NCoR/TFIIID                                       |
| SCP1             | <i>Small CTD phosphatase 1</i>                              |
| SET              | <i>Su(var)3-9, Enhancer of zest, Trithorax</i>              |
| SHH              | <i>Sonic HedgeHog</i>                                       |
| shRNA            | <i>short hairpin RNA</i>                                    |
| SMAD             | <i>Mothers Against Decapentaplegic Homolog 6</i>            |
| SMCC             | <i>SRB/MED-Containing Cofactor</i>                          |
| SMRT             | <i>Silencing Mediator of Retinoic and Thyroid receptors</i> |
| SNAP             | <i>Synaptosomal-Associated Protein</i>                      |
| spm3             | <i>altered SPecificity Mutant 3</i>                         |
| SRB              | <i>Suppressor of RNA Polymerase B</i>                       |
| SSU72            | <i>Suppressor of SUA7 gene 2 (SUA7= yeast TFIIIB)</i>       |
| STAT             | <i>Signal Transducer and Activator of Transcription</i>     |
| STRA6            | <i>Stimulated by Retinoic Acid gene 6</i>                   |
| SV40             | <i>Simian Virus 40</i>                                      |
| SWI/SNF          | <i>SWItch/Sucrose NonFermentable</i>                        |
| TAF              | <i>TBP Associated Factors</i>                               |
| Taspase          | threonine aspartase   |
| TBL1             | <i>Transducin Beta-Like protein 1</i>                       |
| TBLR1            | <i>Transducin Beta-Like Related protein 1</i>               |
| TBP              | <i>TATA Binding Protein</i>                                 |
| TF               | <i>Transcription Factor</i>                                 |
| TFB5             | <i>Subunit of yeast TFIIH complex</i>                       |
| TLE              | <i>Transducin-Like Enhancer of split</i>                    |
| TLF              | <i>TBP Like Factor</i>                                      |
| Topo I           | Topoisomérase-1   |
| TR               | <i>Tyrosine receptor</i>                                    |
| TRAP             | <i>TRa-Associated Protein complex</i>                       |
| TRF              | <i>TBP Related Factor 1</i>                                 |
| TSS              | <i>Transcription Start Site</i>                             |
| TTD              | TrichoThioDystrophie  |
| UAS              | <i>Upstream Activating Sequence</i>                         |
| UBAC             | <i>Ubiquitin Activating and Conjugating</i>                 |
| USA              | <i>Upstream Stimulatory Activity</i>                        |

## Liste des abréviations

|      |   |
|------|---|
| UTR  | <i>UnTranslated Region</i>                |
| UV   | Ultra-violet                              |
| VDR  | <i>Vitamin D Receptor</i>                 |
| VTC3 | <i>Vacuolar transporter chaperone</i>     |
| WDR5 | <i>WD40 Repeat protein 5</i>              |
| XP   | <i>Xeroderma Pigmentosum</i>              |
| Y    | yeast ( <i>Saccharomyces cerevisiae</i> ) |
| YY1  | <i>Ying Yang 1</i>                        |

## Liste des figures et des tableaux, partie I

Figure 1 : Vision classique d'un promoteur des gènes de classe II

Figure 2 : les promoteurs à TSS unique et les promoteurs à TSS multiples

Figure 3 : Purification des facteurs généraux de la transcription

Figure 4 : Structure cristallographique de l'ARN polymérase II

Figure 5 : La phosphorylation du CTD de la Pol II durant la transcription

Figure 6 : Structure cristallographique de TFIIA

Figure 7 : Organisation structurale de TFIIB

Figure 8 : Organisation structurale des sous-unités RAP30 et RAP74 de TFIIF

Figure 9 : Organisation structurale de TFIIE

Figure 10 : Composition et fonction des sous-unités de TFIIH

Figure 11 : Structure cristallographique de TFIIIS

Figure 12: purification des complexes USA

Figure13 : Architecture du Médiateur de levure

Figure 14 : Structure cristallographique du complexe TBP-ADN-NC2

Figure 15 : Les modèles d'assemblage du PIC in vitro

Figure 16 : Le Cycle de la transcription par la Pol II

## Liste des figures et des tableaux

Figure 17 : Les différents niveaux de compaction de la chromatine

Figure 18 : Organisation structurale du motif "histone fold "

Figure 19: Organisation structurale de TBP

Figure 20: Caractéristiques structurales de l'interaction entre TBP et l'élément TATA

Figure 21 : Modélisation du complexe TFIIB-TBP-ADN-TFIIA

Figure 22 : L'interaction entre TFIIA et le complexe TBP-ADN

Figure 23 : Structure du complexe TBP-ADN-NC2

Figure 24 : Architecture tridimensionnelle de TFIID

Figure 25 : Dynamique d'interaction entre TFIID, TFIIA, RAP1 et l'ADN

Figure 26 : Modèles de recrutement de TFIID sur le promoteur

Figure 27 : Organisation en domaines de B-TAF1 et MOT1P

Figure 28 : Architecture tridimensionnelle de B-TFIID

Tableau 1 : Nomenclature unifiée des sous-unités des complexes Médiateurs

Tableau 2 : Les enzymes impliquées dans la méthylation de l'ADN

Tableau 3 : Les différentes modifications des histones et leurs conséquences fonctionnelles

Tableau 4 : Les différentes familles d'enzymes responsables de l'acétylation des histones

Tableau 5: Les différentes enzymes responsables de la méthylation des lysines des histones

## Liste des figures et des tableaux

Tableau 6: Les différentes enzymes responsables de la méthylation des arginines des histones

Tableau 7 : Les différentes familles de cofacteurs ATP-dépendants

Tableau 8 : Résultat de l'étude de mutagenèse réalisée par O.Bryant et al.

Tableau 9 : Nomenclature unifiée des TAFs

## Liste des figures et des tableaux, partie II

Figure 1: Les lignages cellulaires du blastocyste

Figure 2 : La formation du tube neural dans la région dorsale de l'embryon

Figure 3 : Différenciation des cellules proneurales issues des cellules ES en différents types de cellules neuronales

Figure 4 : Le cycle de l'acide rétinoïque

Figure 5 : Organisation structurale des récepteurs nucléaires.

Figure 6 : Structure tridimensionnelle du DBD

Figure 7 : Structure tridimensionnelle du LBD

Figure 8 : Changements conformationnels du LBD du RXR induits par la liaison du ligand

Figure 9 : Complexe répresseur associé à un dimère RAR/RXR en absence de ligand

Figure 10 : Complexes activateur associés à un dimère RAR/RXR en présence de ligand

Figure 11 : Modèle séquentiel de l'activation de la transcription des gènes cibles de l'AR



## Liste des figures et des tableaux

Figure 12 : Protocole de différenciation neuronale dépendante de l'AR

Figure 13 : Organisation structurale des facteurs HES et mécanismes d'action

Figure 15 : Schématisation de la technique de RNA-seq

Figure 14 : Rôle des facteurs HES dans la différenciation neuronale in vivo

Figure 17: Les différents profils d'expression des gènes après traitement des EB

Figure 18 : Expression de HES3 dans les lignées sh-HES3 et dans la lignée contrôle

Figure 16: Génération de lignées stables exprimant des shRNA.

Figure 19: Expression des marqueurs Pax6, Nestin, Brn2 et Mapt dans les lignées sh-HES et dans la lignée contrôle

**PARTIE I : ETUDE STRUCTURE-FONCTION  
DE TBP DANS LES FIBROBLASTES MURINS.**

INTRODUCTION

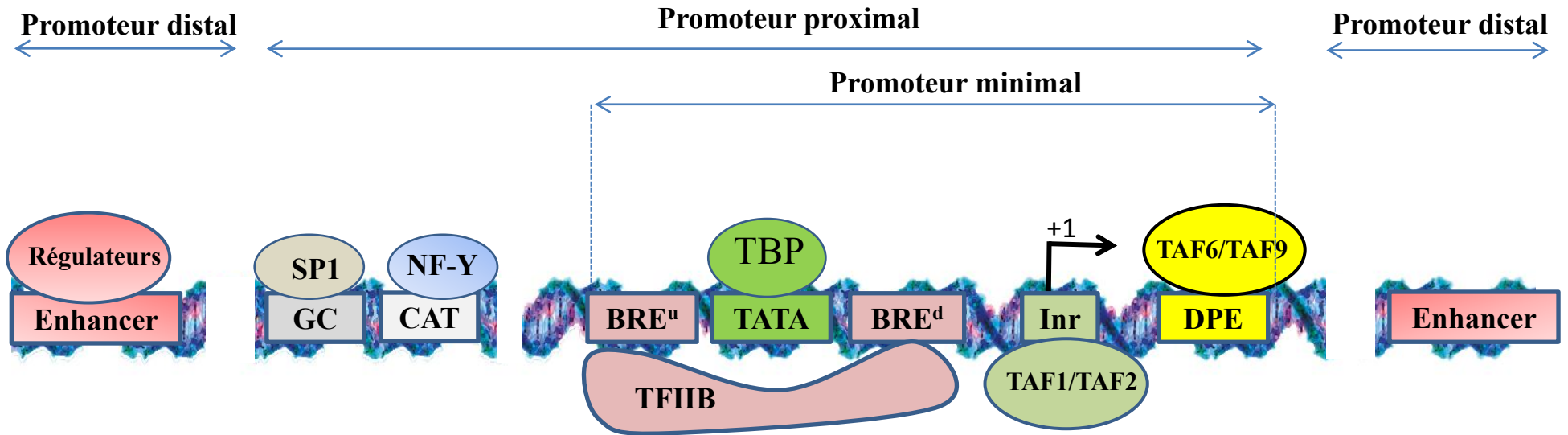
# CHAPITRE 1 : LA TRANSCRIPTION DES GÈNES DE CLASSE II

Pour assurer sa fonction au sein d'un organisme, une cellule eucaryote doit répondre aux divers stimuli de son environnement en régulant de façon précise l'expression de son patrimoine génétique. L'information génétique portée par l'ADN est d'abord transcrite en ARN. Il existe deux classes d'ARN, les ARN codants qui sont traduits en protéines et les ARN non codants tels les ARN de transferts (ARNt) ou les ARN ribosomiques (ARNr) qui sont impliqués dans d'autres processus cellulaires.

Chez les eucaryotes, la transcription est assurée par trois ARN polymérases. L'ARN polymérase I (Pol I) est responsable de la synthèse des ARNr. L'ARN polymérase II (Pol II) transcrit les ARN messagers (ARNm) et les petits ARN nucléaires à l'exception de l'ARNsn U6. L'ARN polymérase III (Pol III) transcrit les ARNt, l'ARNsn U6 et l'ARNr 5S. Les gènes sont ainsi répartis en trois classes : I, II et III, transcrits respectivement par la Pol I, la Pol II et la Pol III.

La synthèse des ARNm par la Pol II est l'une des principales étapes de la régulation de l'expression génique. La transcription des gènes de classe II nécessite l'action de la Pol II et d'une multitude de facteurs protéiques spécialisés, appelés facteurs de transcription. Le processus de transcription comporte trois étapes majeures : l'initiation, l'élongation et la terminaison. La transcription peut être régulée à chaque stade. Le contrôle de l'initiation de la transcription est l'un des principaux niveaux de la régulation de l'expression des gènes.

L'expression des gènes de classe II est contrôlée par des séquences ADN en cis appelées promoteurs. Le promoteur se définit comme la région de l'ADN qui dirige la transcription d'un gène. Il contient des séquences spécifiques reconnues par les facteurs de transcription. L'interaction entre ces facteurs et le promoteur est nécessaire pour une initiation spécifique de la transcription.



**Figure 1 : Vision classique d'un promoteur des gènes de classe II**

Le promoteur d'un gène de classe II peut être subdivisé en trois parties, un promoteur minimal, un promoteur proximal et un promoteur distal. Les éléments contenus dans chaque partie du promoteur sont indiqués: GC (élément GGGCGG); CAT (élément CCAAT); BRE<sup>u</sup> (Upstream TFIIB Response Element); BRE<sup>d</sup> (Downstream TFIIB Response Element); Inr (élément initiateur); DPE (Downstream Promoter Element).

Les facteurs de transcription interagissant avec ces éléments sont indiqués.

### **I. Anatomie des promoteurs des gènes de classe II**

La caractérisation des promoteurs a été longtemps basée sur des expériences de transcription *in vitro* réalisées à l'échelle d'un seul promoteur. Ces expériences individuelles ont conduit à l'élaboration d'un modèle selon lequel la transcription est initiée au niveau d'un nucléotide particulier appelé site d'initiation de la transcription (TSS : *Transcription Start Site*). Dans ce modèle, les facteurs généraux de transcription (GTF : *General Transcription Factor*) et la Pol II sont recrutés au niveau du TSS pour former le complexe de préinitiation (PIC : *preinitiation complex*). *In vitro*, la formation du PIC est nécessaire à une initiation spécifique de la transcription. Pendant de longues années, ce modèle a été admis comme un mécanisme général de l'initiation de la transcription par la Pol II (pour revues, Smale et Kadonaga, 2003 ; Thomas et Chiang, 2006). Cependant, des techniques récentes permettant des analyses à l'échelle du génome entier, montrent une grande diversité dans l'architecture des promoteurs et les mécanismes d'initiation de la transcription (pour revues, Frith et al., 2010 ; Sikorski et al., 2007).

Classiquement, un promoteur de classe II peut être divisé en trois parties majeures : le promoteur minimal, le promoteur proximal et les régions régulatrices distales. *Cf. Figure 1*

#### **I.1. Le promoteur minimal**

##### **I.1.1. Vision classique du promoteur minimal**

Le promoteur minimal a été défini comme la séquence d'ADN minimale nécessaire et suffisante pour permettre l'assemblage du PIC et l'initiation de la transcription *in vitro* (pour revue, Roeder 1996). Cette séquence minimale requise pour la transcription peut comporter plusieurs motifs tels l'élément TATA, l'élément initiateur (Inr), le DPE (*Downstream Promoter Element*) et les éléments BRE (*TFIIB Responsive Element*). Les promoteurs contiennent fréquemment des régions riches en dinucléotides CG appelées "îlots CpG" (pour revue, Smale et Kadonaga, 2003). *In vitro*, certains éléments régulateurs présentent des affinités pour des facteurs de transcription et sont par conséquent définis comme étant les sites de fixation de ces facteurs. Cependant, la présence de ces motifs dans les promoteurs des gènes de classe II peut être combinatoire mais n'est pas obligatoire et certains promoteurs ne contiennent donc aucun motif. Malgré cela, ces promoteurs sont fixés par les GTFs et transcrits par la Pol II. Cette observation

## INTRODUCTION PARTIE I

illustre la complexité des interactions entre les facteurs de transcription et les promoteurs et démontre que l'initiation de la transcription n'est pas due simplement à des interactions entre les facteurs de transcription et leurs motifs de liaison sur l'ADN.

### I.1.1.1. L'élément TATA

Chez les mammifères, l'élément TATA se situe entre 25 et 30 nucléotides en amont du TSS, sa séquence consensus est "5' TATA[T/A]AA[G/A] 3'". Elle constitue le site de liaison de la sous-unité TBP (*TATA Binding Protein*) du facteur général de transcription TFIID. L'introduction de mutations dans l'élément TATA réduit la transcription de plusieurs promoteurs *in vitro*. L'élément TATA serait important pour la détermination du TSS et l'orientation de la transcription. L'importance de l'interaction spécifique entre TBP et l'élément TATA a été démontrée avec le variant TGTAAG de l'élément TATA qui requiert la substitution de trois acides aminés dans la séquence de TBP (TBP spm3) pour permettre une initiation spécifique de la transcription (Strubin et Struhl, 1992). La majorité des promoteurs sont cependant dépourvus d'élément TATA (Bajic et al., 2006), définissant ainsi des promoteurs TATA-dépendants et des promoteurs TATA-indépendants. *In vivo*, les promoteurs TATA-indépendants sont également fixés par TFIID, ce qui suggère l'existence de mécanismes d'interaction de TFIID avec les promoteurs ne reposant pas sur l'interaction TBP-TATA.

### I.1.1.2. L'élément Initiateur (Inr)

L'élément initiateur contient une séquence riche en pyrimidine de consensus "YYA<sub>(+1)</sub>N[T/A]YY" où A<sub>(+1)</sub> correspond au TSS. L'Inr peut diriger la transcription seul ou en synergie avec l'élément TATA et/ou le DPE. Les sous-unités TAF1 et TAF2 (*TAF : TBP Associated Factors*) du complexe TFIID peuvent interagir avec l'élément Inr (pour revue, Thomas et Chiang., 2006)

### I.1.1.3. L'élément DPE

Initialement identifié chez la drosophile, le DPE est fréquemment trouvé dans les promoteurs TATA-indépendants. Sa séquence consensus est "[A/G]G[A/T][C/T][G/A/C]". Le DPE permet, en synergie avec l'Inr, le recrutement de TFIID au niveau du promoteur. Le DPE se situe de 28 à 34 nucléotides en aval de l'Adénine+1 de l'Inr. L'interaction entre TFIID et le DPE

## INTRODUCTION PARTIE I

s'effectue *via* TAF6 et TAF9 (Burke et Kadonaga, 1997). Chez la drosophile, NC2 (*Negative Cofactor 2*), initialement décrit comme un répresseur de la transcription, peut agir en synergie avec TFIID pour stimuler la transcription des promoteurs contenant un DPE, mais réprime les promoteurs contenant un élément TATA (Willy et al., 2000).

### I.1.1.4. Les éléments BRE

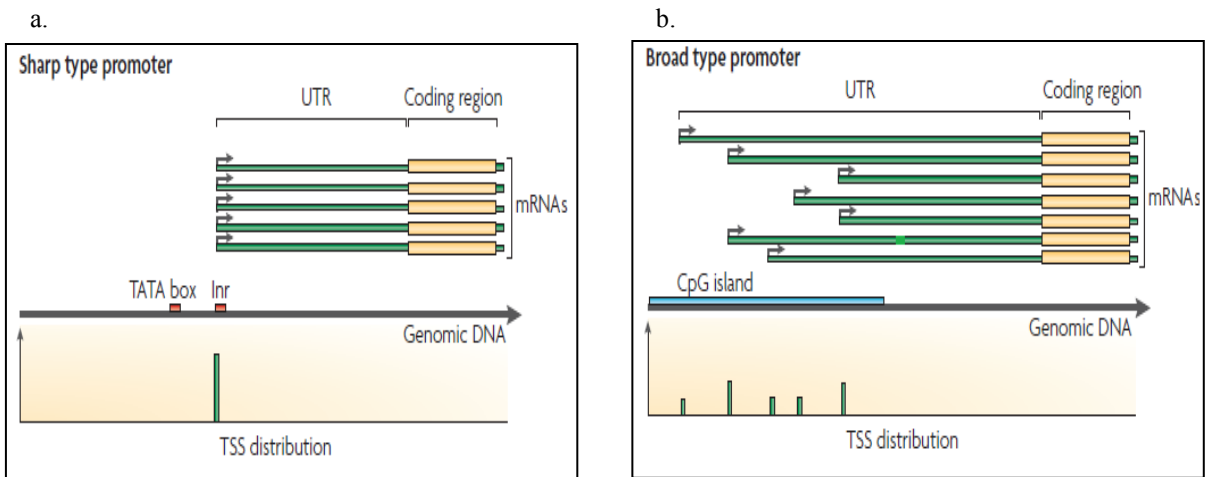
Il existe deux éléments BRE : un élément BRE<sup>u</sup> localisé en amont de l'élément TATA dont la séquence consensus est "[G/C][G/C][G/A]CGCC" et un élément BRE<sup>d</sup> localisé en aval de l'élément TATA dont la séquence consensus est "[G/A]T[T/G/A][T/G][G/T][T/G][T/G]". Ces séquences semblent augmenter l'affinité de TFIIB pour le promoteur *in vitro* (Deng and Roberts, 2005). Cependant, des analyses récentes de CHIP on chip (*Chromatin ImmunoPrecipitation on chip*) montrent que, *in vivo*, l'interaction entre TFIIB et les promoteurs ne dépend pas de la présence des éléments BRE (Albert et al., 2010).

### I.1.1.5. Les îlots CpG

Les îlots CpG correspondent à des régions enrichies en dinucléotides CG. Une région est considérée comme un îlot CpG lorsqu'elle contient un taux de CG supérieur à 10% sur une longueur de 0,5 à 2 kb. Le génome humain contiendrait environ 29 000 îlots CpG qui seraient associés à 72% des promoteurs (Saxonov et al., 2006). Les promoteurs riches en CpG sont souvent dépourvus des éléments TATA, Inr et DPE, mais peuvent contenir des sites de liaison du facteur SP1. Ces promoteurs possèdent souvent plusieurs TSS contrairement aux promoteurs contenant un élément TATA et/ou Inr qui eux ne posséderaient qu'un seul TSS bien défini (Carninci et al., 2006).

Les dinucléotides CpG sont les substrats des enzymes ADN-méthyltransférases. Lors du développement embryonnaire précoce, le taux de méthylation de l'ADN est considérablement réduit, puis, peu avant l'implantation, l'ADN est re-méthylé. Les îlots CpG sont largement exclus de cette re-méthylation de l'ADN (pour revue, Bird, 2002). La méthylation de l'ADN est associée à un état transcriptionnel réprimé alors que l'ADN déméthylé serait associé à un état transcriptionnel actif.





D'après Sandelin et al., 2007

**Figure 2 :** Les promoteurs à TSS unique et les promoteurs à TSS multiples

a) Les promoteurs à TSS unique posséderaient un élément TATA et/ou Inr

b) Les promoteurs à multiples TSS serait riches en îlots CpG et dépourvus d'éléments TATA et Inr.

### **I.1.2. Nouvelle vision du promoteur minimal**

#### I.1.2.1. Promoteurs à TSS unique versus promoteurs à multiples TSSs

Les techniques récentes d'analyses à haut débit associées aux avancées réalisées dans le domaine de la bioinformatique ont permis l'analyse des promoteurs à l'échelle du génome. Les données obtenues par ces techniques remettent en question la vision conventionnelle du fonctionnement d'un promoteur tel qu'il est présenté dans la figure 1. Deux enseignements principaux sont à retenir de ces études : 1) les éléments du promoteur minimal, tels l'élément TATA et Inr, ne sont présents que dans une petite proportion de promoteurs ; 2) la plupart des promoteurs possèdent plusieurs TSS dans une fenêtre de 50 à 100 pb (Suziki et al., 2001 ; Carninci et al., 2006).

Carninci et al. ont utilisé la technique CAGE (*Cap Analysis of Gene Expression*) pour cartographier les sites d'initiation de la transcription chez la souris et chez l'humain (Carninci et al., 2006). Les résultats obtenus montrent que les promoteurs peuvent être classés en deux catégories majeures : 1) les promoteurs à TSS unique qui possèdent souvent un élément TATA et/ou Inr et dirigent la transcription de gènes finement régulés tels les gènes spécifiquement exprimés dans un tissu ; 2) les promoteurs à multiples TSSs qui seraient enrichis en îlots CpG et dirigeraient la transcription de gènes constitutifs. Cette classification n'est pas absolue, car il existe des promoteurs de gènes constitutifs possédant un élément TATA et des gènes hautement régulés dépourvus de cet élément. Cf. *Figure 2*

#### I.1.2.2. Les promoteurs alternatifs

Un gène peut posséder plusieurs promoteurs. Les promoteurs sont dits alternatifs lorsqu'ils dirigent la transcription du même gène, mais qu'ils sont séparés par une distance supérieure à 500 pb. Selon Carninci et al., 58% des gènes humains et murins codant pour les protéines auraient au moins deux promoteurs alternatifs (Carninci et al., 2006). Les promoteurs alternatifs seraient majoritairement associés aux îlots CpG. Ils peuvent être localisés dans les régions 5' UTR (*Untranslated Region*), mais également dans les exons et les introns. Ils rajoutent ainsi un degré supplémentaire dans la régulation de l'expression des gènes et peuvent jouer un rôle dans l'expression différentielle des différents isoformes d'une protéine. Étonnamment, des promoteurs alternatifs peuvent se situer dans les régions 3' UTR de certains gènes. Les transcrits

## INTRODUCTION PARTIE I

correspondants peuvent jouer un rôle dans la régulation d'autres gènes adjacents par des mécanismes de régulation anti-sens (Carninci et al., 2006).

### **I.2. Les éléments régulateurs**

Les éléments régulateurs sont capables de moduler le taux de transcription d'un gène donné. Ces séquences sont classées en deux groupes : les séquences régulatrices proximales (promoteur proximal) et distales. Elles sont reconnues par les régulateurs transcriptionnels. Ces régulateurs sont souvent organisés de façon modulaire avec un domaine de liaison à l'ADN (DBD : *DNA Binding Domain*) et un domaine régulateur (activateur ou répresseur).

#### **I.2.1. Le promoteur proximal**

Chez les métazoaires, il existe de nombreuses séquences en amont de l'élément TATA reconnues par divers régulateurs transcriptionnels. Habituellement situées entre les positions -110 et -40, elles peuvent être composées de sites de fixation de divers facteurs. Ces séquences perdent rapidement leur effet cis-régulateur quand elles sont trop éloignées du TSS (Mitchell and Tjian, 1989).

Les séquences régulatrices proximales les plus connues sont les éléments CCAAT et GGGCGG. L'élément CCAAT est souvent localisé entre 60 à 80 nucléotides en amont du TSS, il constitue un élément de liaison du facteur nucléaire Y (NF-Y : Nuclear Factor-Y) (Mantovani, 1999). L'élément GGGCGG, de séquence consensus "[G/T]GGGCGG[G/A][G/A][C/T]", est positionné en amont du promoteur minimal et est reconnu par des facteurs de transcription de la famille SP1 via leurs motifs en doigt de zinc C-terminaux (Marco et al., 2003).

#### **I.2.2. Les séquences distales**

Les séquences distales sont généralement spécifiques d'un gène donné et sont appelées "enhancers" lorsqu'elles sont des sites de liaison d'activateurs transcriptionnels ou "silencers" lorsqu'elles sont des sites de liaison de répresseurs. Les séquences distales sont le plus souvent composées d'un nombre important de séquences régulatrices regroupées, capables de réguler toute une région chromosomique. Deux modèles ont été proposés pour expliquer la régulation de la transcription par ces séquences. Un premier modèle, dit de "boucle" (*looping out*), suggère que

## INTRODUCTION PARTIE I

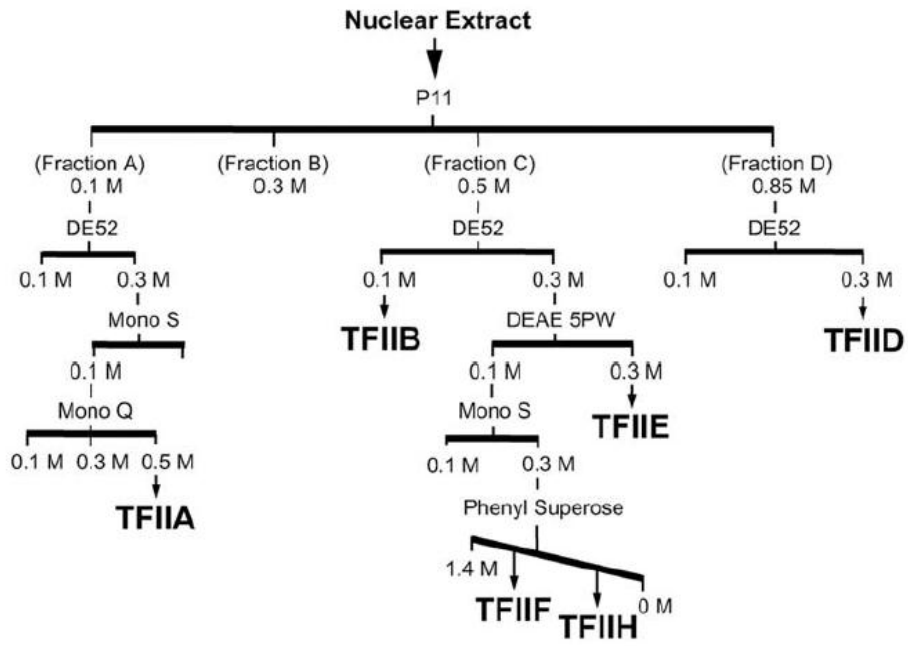
l'ADN se courbe pour permettre un contact entre les activateurs fixés sur les régions régulatrices et les GTFs fixés sur le promoteur proximal. Un deuxième modèle, dit de "liaison" (*linking*), suggère que les régions de contrôle vont permettre d'établir une structure chromatinienne, activatrice ou inhibitrice, qui se propagerait jusqu'au gène à réguler (Bulger and Groudine, 1999).

## II. L'initiation de la transcription

D'après les études biochimiques réalisés par Weil et al en 1979, il apparaissait évident que la Pol II nécessitait la présence de facteurs supplémentaires pour initier la transcription. En effet, la Pol II nécessite l'addition d'extraits cellulaires pour initier la transcription à partir du promoteur AdMLP (*Adenovirus Major Late Promoter*) (Weil et al., 1979). La purification de ces extraits cellulaires a permis d'isoler des fractions chromatographiques contenant des activités nécessaires pour l'initiation de la transcription par la Pol II (Matsui et al., 1980 ; Samuels et al., 1982 ; Sawadogo and Roeder, 1985 ; Reinberg and Roeder, 1987; Flores et al., 1989 ; Flores et al., 1992 ; Ge et al., 1996) (*Cf. Figure 3*). Les fractions purifiées correspondaient à des facteurs nucléaires qui ont été nommés TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH et désignés comme les facteurs généraux de transcription. Le terme "machinerie de transcription de base" désigne généralement ces six GTFs et la Pol II.

L'investigation des rouages de l'initiation de la transcription a été d'abord basée sur des expériences *in vitro*, en utilisant comme matrice des promoteurs bien caractérisés possédant des éléments TATA et Inr. Ces expériences ont conduit logiquement à un modèle selon lequel les GTFs et la Pol II sont recrutés séquentiellement au niveau du promoteur minimal pour former le PIC, la formation de ce complexe multiprotéique est une étape essentielle à l'initiation spécifique de la transcription *in vitro*. Le modèle basé sur la formation du PIC est admis comme un mécanisme général de l'initiation de la transcription et logiquement, tous les GTFs ont été pendant longtemps considérés comme indispensables à la transcription par la Pol II (pour revue, Thomas et Chiang, 2006).

L'émergence de techniques nouvelles permettant l'investigation d'un nombre croissant de promoteurs et la cartographie à l'échelle du génome des sites de liaison des facteurs de transcription a remis en cause le caractère général des GTFs et du PIC. Il s'avère que le PIC peut avoir des compositions différentes selon les gènes et que d'autres facteurs tels TFIIIS, le



D'après Thomas et Chiang, 2006

**Figure 3** : Purification des facteurs généraux de la transcription

## INTRODUCTION PARTIE I

Médiateur, Mot1P ou NC2 peuvent faire partie du PIC. Les derniers facteurs cités sont considérés comme des cofacteurs ou des corégulateurs de la transcription, mais leur présence sur un grand nombre de promoteurs actifs rend difficile la distinction entre les facteurs généraux et les cofacteurs. Afin de faciliter la lecture de ce manuscrit, les six GTFs classiques, TFIIS et la Pol II seront considérés comme des composants de la machinerie de transcription de base. Les autres facteurs seront considérés comme des cofacteurs de la transcription.

Outre les GTFs et les cofacteurs, l'initiation de la transcription nécessite l'intervention des activateurs et/ou des répresseurs de la transcription qui régulent spécifiquement la transcription d'un ou plusieurs gènes.

La transcription des gènes dépend également de la structure de la chromatine. La chromatine est dynamiquement modifiée par les complexes de remodelages et les enzymes de modification des histones. Ce niveau de régulation supplémentaire est désigné par le terme contrôle épigénétique.

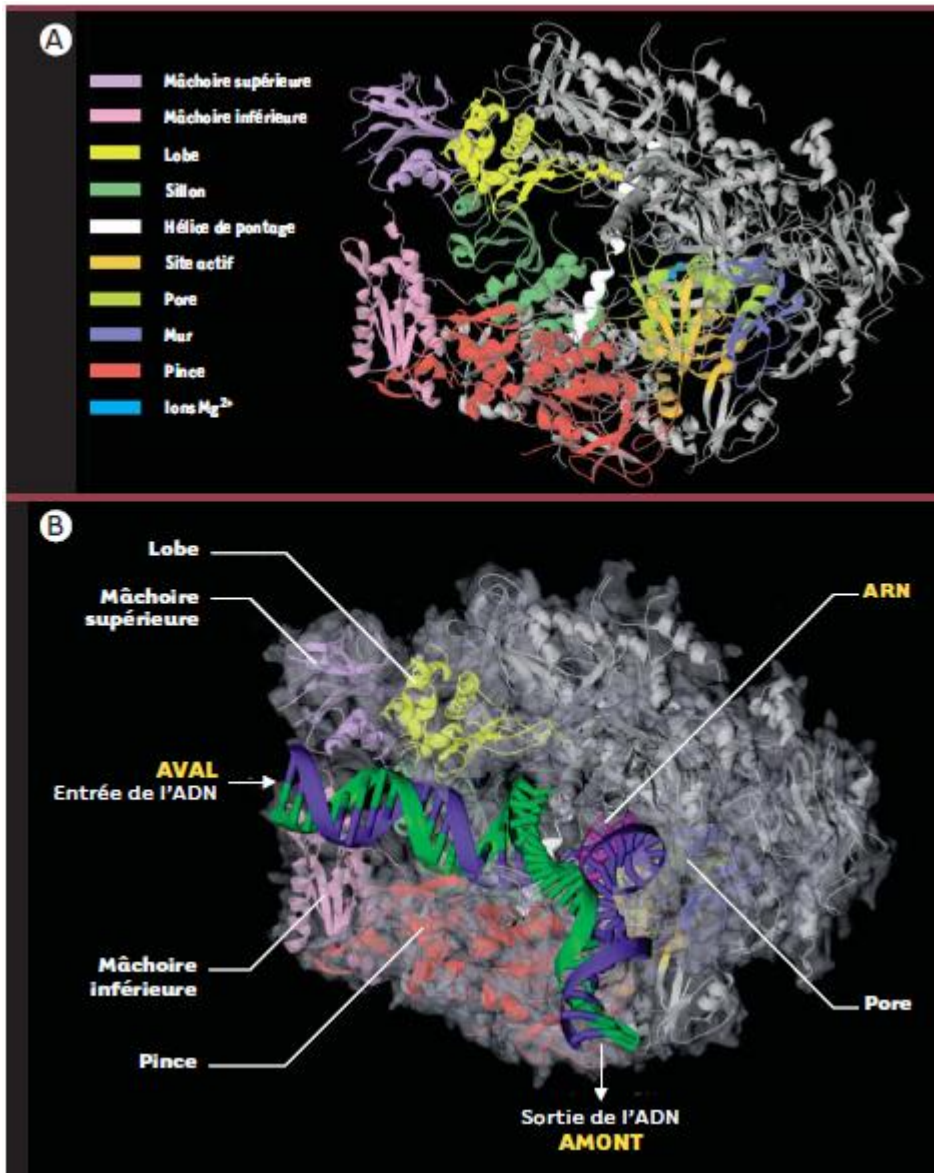
Ce chapitre comportera une description des acteurs majeurs impliqués dans l'initiation de la transcription, exceptés les complexes TFIID et B-TFIID qui seront abordés dans un chapitre consacré aux complexes contenant TBP.

### **II.1. La machinerie de transcription de base**

La machinerie de transcription de base est composée de la Pol II et des facteurs généraux de la transcription TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH et TFIIS.

#### **II.1.1. L'ARN polymérase II**

La Pol II est un complexe multiprotéique de poids moléculaire 0,5 MDa composé de 12 sous-unités conservées de la levure à l'homme, dont certaines sont communes aux trois ARN polymérases (Pol I, II et III), alors que d'autres sont spécifiques à la Pol II. Les gènes codant pour ces différentes sous-unités sont en général essentiels pour la viabilité cellulaire. Les sous-unités de la Pol II sont appelées "RPB" (*RNA Polymerase B*), leur masse varie de 10 à 220 kDa (Sawadogo and Sentenac, 1990 ; Young, 1991).



Adapté de Langelier et al., 2002,

**Figure 4** : Structure cristallographique de l'ARN polymérase II

## INTRODUCTION PARTIE I

Les sous-unités RPB5, RPB6, RPB8, RPB10 et RPB12 sont communes aux trois ARN polymérase eucaryotes (Woychik et al., 1990 ; Carles et al., 1991 ; Young, 1991 ; Hampsey, 1998). Les sous-unités RPB1, RPB2, RPB3 et RPB11 présentent des homologies de séquence avec des sous-unités équivalentes de la Pol I et la Pol III. La conservation de ces sous-unités au cours de l'évolution suggère qu'elles possèdent certaines fonctions communes à toutes les ARN polymérase. Seules les sous-unités RPB4, RPB7 et RPB9 sont spécifiques à la Pol II.

### II.1.1.2. La structure de l'ARN polymérase II

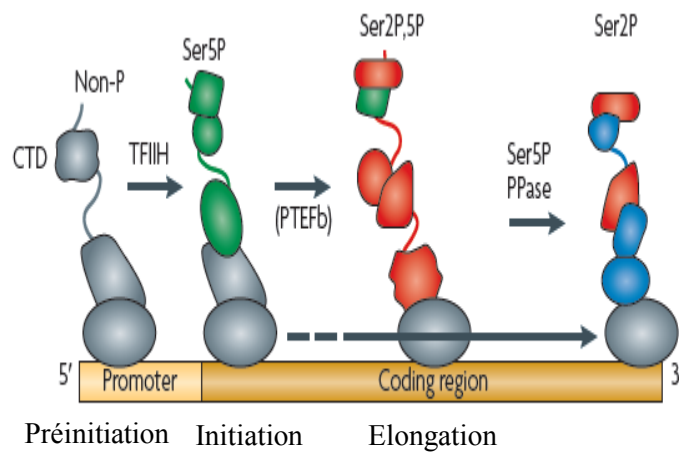
Lors des étapes de purification, la Pol II peut être séparée en deux sous-complexes. Un sous-complexe mineur qui consiste en un dimère RPB4/RPB7 et un sous-complexe majeur contenant les dix sous-unités restantes et nommé "partie catalytique".

D'après les études cristallographiques, RPB1 et RPB2 forment un sillon qui entoure l'ADN. Au fond de ce sillon, se trouvent deux ions  $Mg^{2+}$  qui marquent l'emplacement du site actif. L'entrée du sillon est bordée par une paire de mâchoires formée en partie de RPB1, RPB9 et RPB5. Cette paire de mâchoires interagit avec l'ADN en aval du site actif. Près du site actif, une hélice de RPB1, l'hélice de pontage (bridge helix), traverse le sillon et rejoint RPB2. Cette hélice est impliquée dans la translocation de l'enzyme le long de l'ADN lors de l'élongation. Une pince formée par des parties de RPB1, RPB2 et RPB6 se referme sur l'ADN pendant l'élongation (Cramer et al., 2001 ; Gnatt et al., 2001). Cf. *Figure 4*

### II.1.1.3. Le domaine carboxy-terminal de RPB1

RPB1 possède un domaine C-terminal (CTD : *C-Terminal Domain*) spécifique à la Pol II qui est composé de multiples répétitions d'un heptapeptide de séquence consensus YSPTSPS (Allison et al., 1985 ; Corden et al., 1985). Le nombre de répétitions est variable selon les espèces et augmente avec la complexité du génome. La levure comprend 26 à 29 répétitions, la drosophile 42 et l'homme 52. Le CTD n'est pas essentiel à l'activité catalytique de l'enzyme *in vitro*, mais sa délétion est létale chez la levure (West and Corden, 1995), la drosophile (Zehring et al., 1988) et dans des cellules murines (Bartolomei et al., 1988).





apté de Phatnani et al. , 2006

**Figure 5** : La phosphorylation du CTD de la Pol II durant la transcription

La Pol II est recrutée sur le promoteur sous sa forme hypophosphorylée (Non-P). Lors de l'initiation de la transcription, elle est phosphorylée par TFIID au niveau de la serine 5 (Ser5P). Une phosphorylation supplémentaire par le facteur pTEFb au niveau de la sérine 2 (Ser2P) lui permet de s'engager dans la phase d'élongation. Dans la région 3' du gène, la serine 5 est dephosphorylée.

## INTRODUCTION PARTIE I

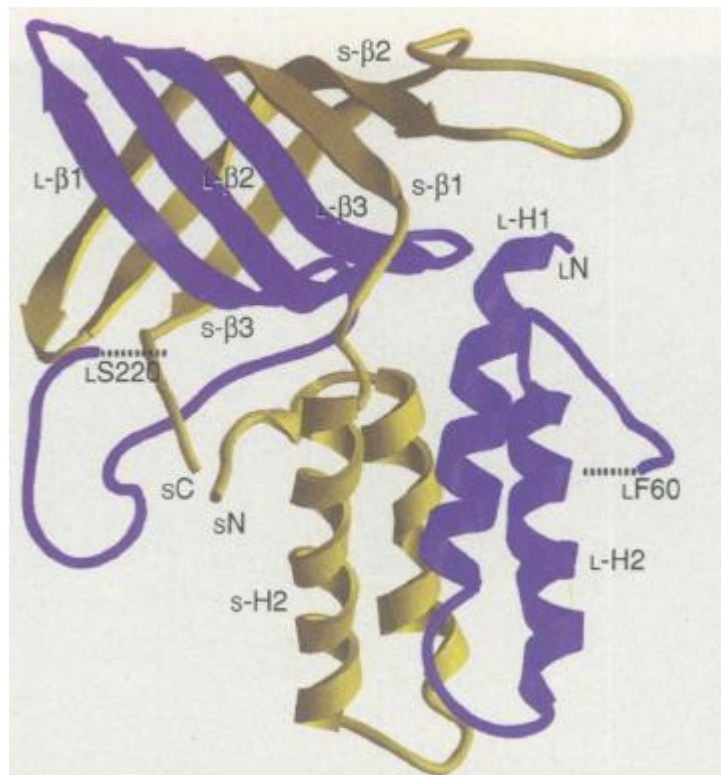
Le CTD de la Pol II est le site de plusieurs modifications post-traductionnelles dont la phosphorylation, la glycosylation et l'ubiquitination. Ces modifications jouent un rôle important dans la régulation de l'activité de la Pol II.

La modification du CTD la plus caractérisée est la phosphorylation. Exceptées les prolines, tous les résidus du CTD peuvent être phosphorylés. En fonction de la phosphorylation de ces résidus, la Pol II peut se trouver sous deux formes, IIO et IIA (Cadena and Dahmus, 1987). La forme IIA correspond à un état hypophosphorylé, impliqué dans l'assemblage du PIC (Lu et al., 1991 ; Serizawa et al., 1993). La forme IIO est hyperphosphorylée et serait impliquée dans l'élongation et la terminaison (Bartholomew et al., 1986 ; Payne et al., 1989). Une troisième forme nommée IIB correspond à la Pol II tronquée, dépourvue du CTD (pour revue, Thomas et Chiang., 2006). Le CTD de la Pol II est phosphorylé dès l'entrée de cette dernière dans le PIC (Laybourn and Dahmus, 1990).

La phosphorylation des sérines 2 et 5 des heptapeptides joue un rôle central dans la régulation de l'activité de la Pol II. Ces deux résidus subissent des cycles de phosphorylation/déphosphorylation pendant le processus de transcription (*Cf. Figure 5*).

La sérine 5 est phosphorylée par la protéine kinase CDK7, une sous-unité de TFIIH (Coin and Egly., 1998). Les expériences d'immunoprécipitation de la chromatine (*ChIP*) utilisant des anticorps reconnaissant spécifiquement la phosphosérine 5 ont montré que la Pol II phosphorylée sur la sérine 5 (Ser5P-Pol II) est majoritairement associée aux promoteurs et non aux régions codantes (Komarnitsky et al., 2000). Ces résultats suggèrent que la Pol II est phosphorylée sur la sérine 5 par TFIIH lors du démarrage de la transcription puis est déphosphorylée peu après l'initiation de la transcription. La phosphorylation du CTD par TFIIH peut déstabiliser certaines interactions entre la Pol II et les GTFs et, par conséquent, faciliter la libération de la Pol II du PIC et son entrée en phase d'élongation. La phosphorylation de la sérine 5 joue également un rôle dans le recrutement des enzymes responsables de la coiffe des ARN pré-messagers (Ho and Shuman, 1999).

La Pol II phosphorylée en sérine 2 du CTD (Ser2P-Pol II) serait associée à l'étape d'élongation, et par conséquent, est présente dans la région codante du gène (Komarnitsky et al.,



**Figure 6** : Structure cristallographique de TFIIA

## INTRODUCTION PARTIE I

2000). Il a été suggéré que la phospho-sérine 2 est importante pour l'activité des enzymes impliquées dans le processus d'épissage des ARN. La sérine 2 est phosphorylée par la kinase CDK9, une sous-unité du facteur d'élongation P-TEFb (*Positive Transcription Elongation Factor*) (Komarnitsky et al., 2000).

Ces deux formes de la Pol II ne sont cependant pas mutuellement exclusives. En effet l'analyse de données de ChIP suivies de séquençage à haut débit (ChIP-seq) montre que, pour une population de gènes, la Ser5P-Pol II peut colocaliser avec la Ser2P-Pol II dans la région codante et surtout à l'extrémité 3' des gènes. En revanche, la Ser2P-Pol II est absente des promoteurs de la majorité des gènes.

Des analyses récentes montrent que la phosphorylation de la sérine 7 du CTD est également associée à la Pol II en phase élongation (Chapman et al., 2007).

Les phosphorylations du CTD sont des réactions réversibles. Plusieurs phosphatases ont été décrites comme responsables de sa déphosphorylation.  $\gamma$ Ssu72 chez la levure et SCP1 (*Small CTD phosphatase 1*) chez l'homme sont capables de déphosphoryler la sérine 5 *in vitro*. FCP1 (*TFIIF-associated CTD Phosphatase 1*) déphosphoryle préférentiellement la sérine 2 (Meinhart et al., 2005). FCP1 interagit avec TFIIB (Kobor et al., 2000) et la sous-unité RPB4 de la Pol II (Kimura et al., 2002) et jouerait un rôle dans le recyclage de la Pol II et la réinitiation de la transcription.

### II.1.2. TFIIA

Chez les mammifères TFIIA est composé de 3 sous-unités  $\alpha$ ,  $\beta$  et  $\gamma$ . Ces 3 sous-unités sont codées par deux gènes, TFIIA $\alpha/\beta$  et TFIIA $\gamma$ . TFIIA $\alpha/\beta$  code pour un précurseur de 55 kDa qui, après clivage par la Taspase1, génère TFIIA $\alpha$  et  $\beta$ . Chez la levure,  $\gamma$ TFIIA est constitué de deux sous-unités TOA1 et TOA2. TOA1 est l'homologue de hTFIIA $\alpha/\beta$ , mais chez la levure cette protéine n'est pas clivée. TOA2 est l'homologue de hTFIIA $\gamma$ . Le rôle du clivage de TFIIA $\alpha/\beta$  chez les mammifères n'est pas élucidé, vu que TFIIA $\alpha/\beta$  non clivé, dans le contexte de l'inactivation de la Taspase 1, semble être fonctionnel (pour revue, Høiby et al., 2007).

## INTRODUCTION PARTIE I

Structuralement, TFIIA s'organise en deux domaines, un domaine en forme de tonneau constitué de 12 feuillets  $\beta$  ("12 strand  $\beta$ -barrel") et un domaine comportant 4 hélices  $\alpha$  ("4 helix bundle"). TFIIA stabilise l'interaction entre TBP et l'ADN en établissant des contacts directs avec ces deux molécules (Geiger et al., 1996 ; Tan et al., 1996). Cf. *Figure 6*

Les données concernant le rôle de TFIIA dans la transcription par la Pol II sont contradictoires. En effet, TFIIA a été initialement décrit comme un facteur nécessaire pour la transcription par la Pol II *in vitro* (Matsui et al., 1980 ; Reinberg et al., 1987). Mais par la suite, d'autres travaux ont montré que TFIIA n'est pas indispensable pour initier la transcription basale (Van Dyke et al., 1988 ; Wu et al., 1998).

En réalité, TFIIA n'est pas nécessaire pour la transcription *in vitro* quand les composants du PIC (fractions purifiées ou protéines recombinantes) utilisés sont hautement purifiés. En revanche, TFIIA devient indispensable dans un système où la purification des composants n'est que partielle. L'explication probable est qu'un système de transcription constitué à partir de composants partiellement purifiés contiendrait des répresseurs de la transcription comme NC2 ou B-TAF1 qui interagissent avec TBP et inhibent la transcription. TFIIA rentrerait en compétition avec ces répresseurs en augmentant l'affinité de TBP pour l'ADN et stimulerait ainsi la transcription. En revanche, dans un système non contaminé par des répresseurs, TFIIA devient inutile. Ces observations ont conduit à considérer TFIIA comme un anti-répresseur et non un GTF.

Il a été démontré récemment que TFIIA agit en synergie avec le régulateur de la transcription RAP1 (*repressor/activator protein 1*) pour moduler le recrutement de TFIID sur le promoteur (Papai et al., 2010).

Les résultats des études de TFIIA *in vivo* sont également contradictoires. Chez la levure, l'introduction d'une mutation dans TBP (TBP<sub>N2-1</sub>) affectant l'interaction avec TFIIA, inhibe la transcription induite par les activateurs à domaine d'activation acide tel GAL4. En revanche, cette mutation n'affecte pas la transcription basale, mais cause une baisse modérée de la prolifération. Dans les cellules TBP<sub>N2-1</sub>, la transcription activée est restaurée par l'expression d'une protéine de fusion TBP<sub>N2-1</sub>-TOA2 (TOA2 : petite sous-unité de  $\gamma$ TFIIA) (Stargell et

## INTRODUCTION PARTIE I

Struhl., 1995). Ces résultats suggèrent que TFIIA joue un rôle dans la transcription activée via son interaction avec TBP et qu'il est nécessaire pour une prolifération normale.

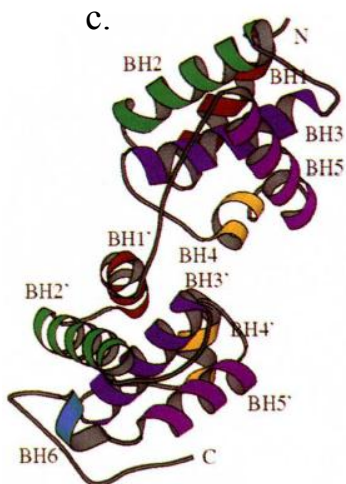
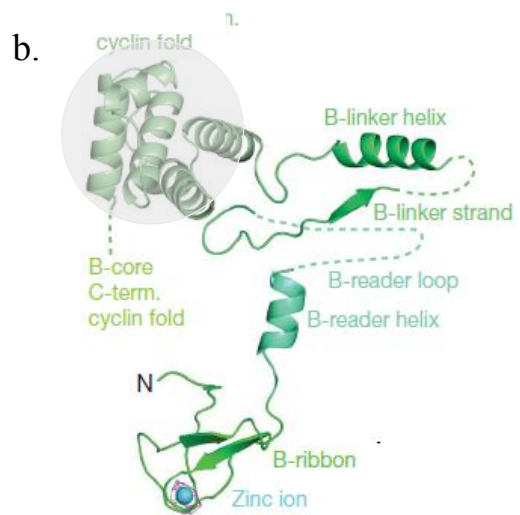
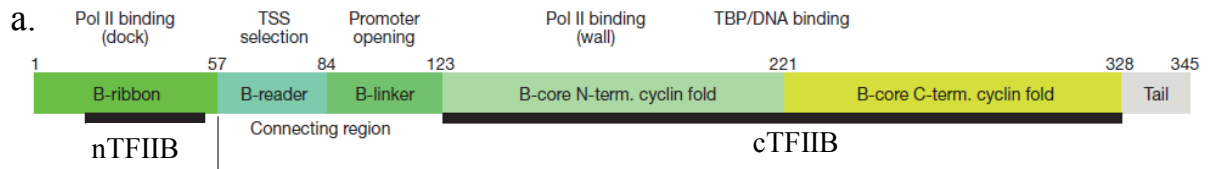
Etonnamment, la réduction significative (>99%) du taux de TFIIA dans la cellule n'a pas le même effet que la perte de l'interaction entre TFIIA et TBP. En effet, l'inactivation de TFIIA dans la levure affecte modérément la transcription basale, provoque un arrêt du cycle cellulaire lors de la transition G2/M, mais n'a aucun effet sur la transcription activée par GAL4 et ACE1 (Chou et al., 1999). Cette étude contrairement à la précédente, démontre que TFIIA n'est pas indispensable pour la transcription activée, mais qu'il joue un rôle dans la transcription des gènes impliqués dans la progression du cycle cellulaire. Cette contradiction peut être expliquée par le fait que la mutation TBP<sub>N2-1</sub> pourrait affecter d'autres interactions importantes pour l'activation de la transcription indépendamment de TFIIA.

### II.1.3. TFIIIB

Contrairement à TFIIA, l'inactivation de TFIIIB abolit complètement la transcription par la Pol II (Moqtaderi et al., 1996). Chez l'homme, TFIIIB est constitué d'un seul polypeptide de 33 kDA. Il comprend deux domaines : un domaine N-terminal (nTFIIIB) et un domaine C-terminal (cTFIIIB).

nTFIIIB interagit avec les sous-unités RPB1 et RPB2 de la Pol II et avec TFIIF via un motif de liaison du zinc "interacting zinc-ribbon domain" (Bushnell et al., 2004 ; Ha et al., 1993). Ces interactions facilitent le recrutement de la Pol II et de TFIIF au niveau du PIC. Immédiatement après ce motif de liaison du zinc se trouve une région hautement conservée appelée "B-finger" constituée de deux hélices  $\alpha$  séparées par une boucle et un feuillet  $\beta$  (Kostrewa et al., 2009) (*Cf. figure 7a*). Cette région joue un rôle important dans la détermination du TSS, l'ouverture du promoteur et la libération de TFIIIB après le démarrage de la phase d'élongation.

cTFIIIB contient deux répétitions directes imparfaites. Ces deux répétitions s'organisent en deux domaines globulaires de type "Cyclin fold" reliés par un court bras. Chaque répétition est constituée de 5 hélices  $\alpha$  numérotées BH1 à BH5 pour la première répétition et BH1' à BH5' pour la deuxième répétition (*Cf. Figure 7b, 7c*). cTFIIIB établit des contacts directs avec TBP et



D'après Nikolov et al., 2006; et Kostrewa et al., 2009

**Figure 7** : Organisation structurale de TFIIB

- a) Organisation en domaines de TFIIB. Les différents domaines ainsi que leur fonctions sont indiqués
- b) Structure cristallographique de TFIIB. Les motif Cyclin fold sont indiqués
- c) Structure cristallographique du domaine C-terminal de TFIIB (cTFIIB)

## INTRODUCTION PARTIE I

lie l'ADN dans le grand sillon en amont de l'élément TATA et dans le petit sillon en aval de celle-ci. Cette interaction asymétrique joue un rôle dans l'orientation du PIC et stabilise TFIID sur l'ADN (Nikolov et al., 1995). cTFIIB établit également des contacts avec la Pol II (Kostrewa et al., 2009).

*In vitro*, TFIIB peut interagir spécifiquement avec l'élément BRE<sup>d</sup> du promoteur en absence de TBP (Lagrange et al., 1998). Cette interaction se fait via un motif HTH (*Helix-Turn-Helix*) composé par BH3', BH4' et BH5'. Étonnamment, ce motif n'est pas conservé chez la levure et les plantes. Il a été démontré récemment que TFIIB peut interagir avec une autre séquence en amont de l'élément TATA, nommée BRE<sup>u</sup> (Deng et al., 2005). Cette interaction se fait via un motif conservé formé par la boucle entre BH2 et BH3. Cependant, des analyses récentes de CHIP on chip dans les cellules B humaines montrent que les éléments BRE n'influencent pas l'interaction entre TFIIB et les promoteurs, en revanche l'élément TATA est enrichi dans les promoteurs fortement liés par TFIIB (Albert et al., 2010).

### II.1.4. TFIIF

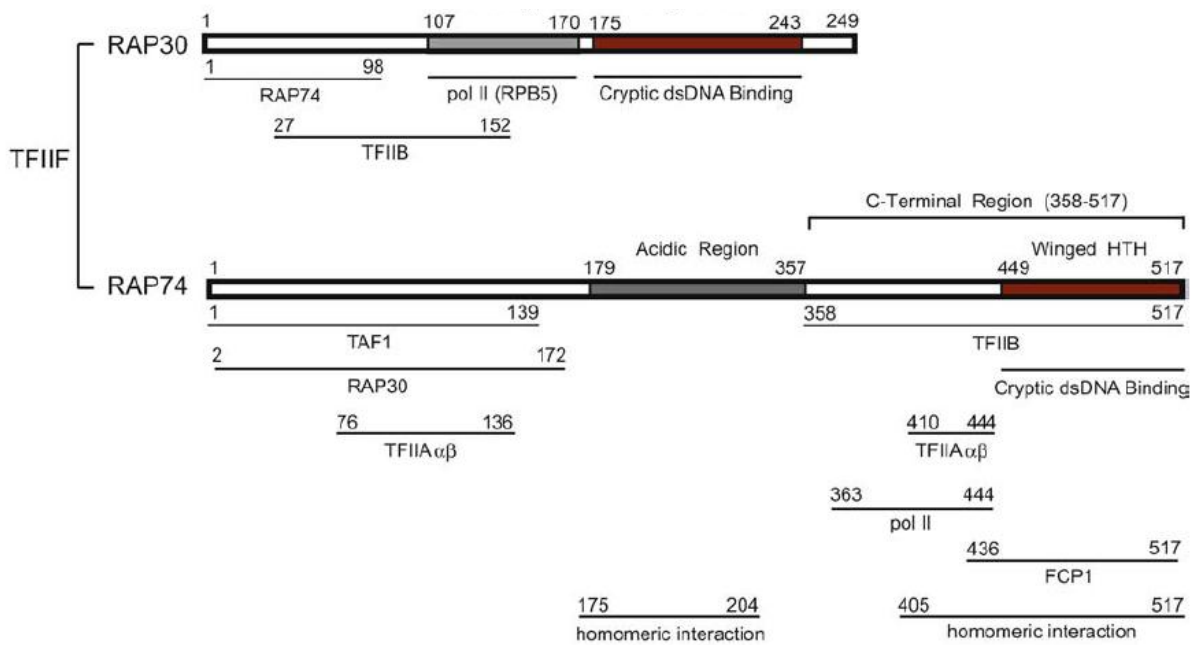
TFIIF joue un rôle important dans l'assemblage du PIC et l'initiation de la transcription. En effet, il est impliqué dans le recrutement de la Pol II (Flores et al., 1991), de TFIIE et de TFIIH (Orphanides et al., 1996) sur le promoteur.

Le TFIIF humain est un hétérotétramère composé de deux molécules de la sous-unité RAP74 et deux molécules de la sous-unité RAP30 (Flores et al., 1990).

RAP30 comprend trois régions : la région N-terminale responsable de l'interaction avec la région N-terminale de RAP74 (Fang and Burton, 1996), la région centrale qui interagit avec la sous-unité RPB5 de la Pol II (Wei et al., 2001) et la région C-terminale qui forme un motif structural appelé "winged-helix" similaire à celui de l'histone H5 (Groft et al., 1998) et qui est impliqué dans l'interaction avec l'ADN.

RAP74 comporte également trois régions fonctionnelles : la région N-terminale responsable de l'interaction avec RAP30, TAF1 (Ruppert and Tjian, 1995) et TFIIA $\alpha/\beta$





D'après Thomas et Chiang, 2006

**Figure 8 :** Organisation structurale des sous-unités RAP30 et RAP74 de TFIIF

Les différents domaines et leurs partenaires d'interaction sont indiqués

## INTRODUCTION PARTIE I

(Langelier et al., 2001), la région centrale hautement chargée et dépourvue de résidus hydrophobes et la région C-terminale très conservée qui contient un motif "winged helix". Cette région est également nécessaire pour l'interaction avec TFIIA $\alpha/\beta$ , TFIIB, la Pol II et la phosphatase FCP1 qui déphosphoryle la sérine 2 du CTD de la Pol II. Cf. *Figure 8*

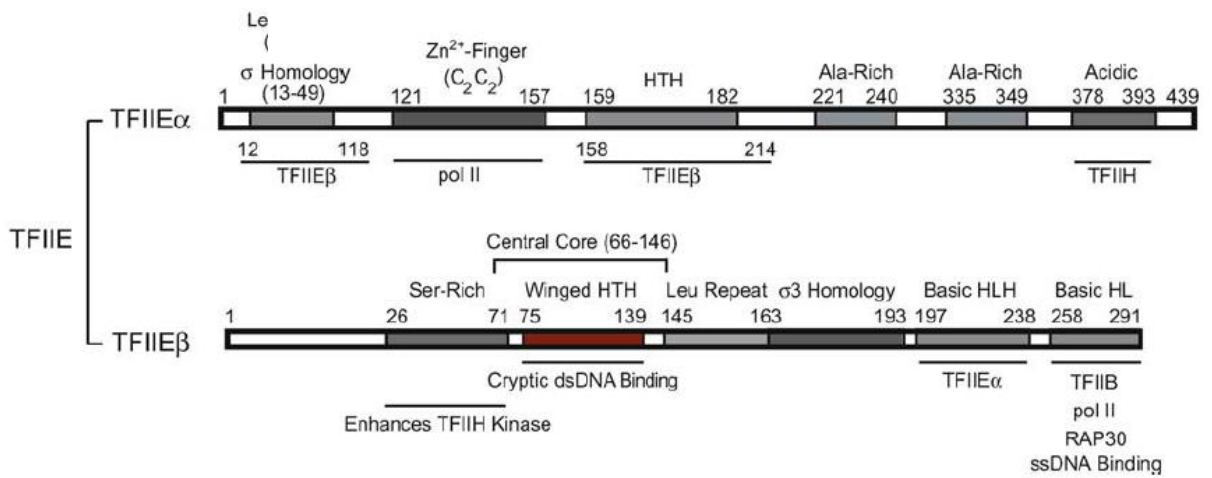
### II.1.5. TFIIE

Dans le modèle séquentiel de l'assemblage du PIC, TFIIE rejoint le PIC après le recrutement de la Pol II et TFIIF mais avant celui de TFIIH. TFIIE joue un rôle essentiel dans le recrutement et dans la régulation des activités enzymatiques de TFIIH. De plus, TFIIE est nécessaire pour la formation du complexe ouvert et la transition de la phase d'initiation à celle de l'élongation.

TFIIE existe majoritairement sous forme de dimère composé d'une sous-unité TFIIE $\alpha$  de 56 kDa et d'une sous-unité TFIIE $\beta$  de 35 kDa (Jawhari et al., 2006). TFIIE peut également exister sous forme de tétramère comportant 4 sous-unités (2 $\alpha$ /2 $\beta$ ). Cf. *Figure 9*

TFIIE $\alpha$  contient plusieurs motifs structuraux dont un motif en doigt de zinc et un motif HTH. Le motif en doigt de zinc est constitué d'une hélice  $\alpha$  et cinq feuillets  $\beta$  (Okuda et al., 2004). Ce motif augmente l'affinité de TBP pour le promoteur AdMLP (Yokomori et al., 1998). TFIIE $\alpha$  est impliqué dans l'interaction avec la Pol II et est nécessaire pour la stimulation de l'activité de TFIIH. Des expériences de pontages covalents ont mis en évidence que TFIIE $\alpha$  interagit avec l'ADN en aval de l'élément TATA au niveau du nucléotide -10 (le TSS étant le nucléotide +1) et serait important pour l'ouverture de l'ADN (Forget et al., 2004).

TFIIE $\beta$  contient également plusieurs domaines fonctionnels. Son domaine N-terminal contient une région riche en sérines qui peut stimuler la phosphorylation du CTD par TFIIH. Le domaine central contient un motif HTH qui interagit avec l'ADN double brin. Le domaine C-terminal contient deux régions basiques, la première contient un motif bHLH (*basic Helix-Loop-Helix*) qui contacte TFIIE $\alpha$ , la deuxième contient un motif "basic Helix-Loop" qui interagit avec l'ADN simple brin, la Pol II, TFIIB et RAP30 (Watanabe et al., 2003). La capacité de TFIIE à interagir avec l'ADN double brin et simple brin supporte son rôle dans l'ouverture de l'ADN.



D'après Thomas et Chiang, 2006

**Figure 9** : Organisation structurale de TFIIIE

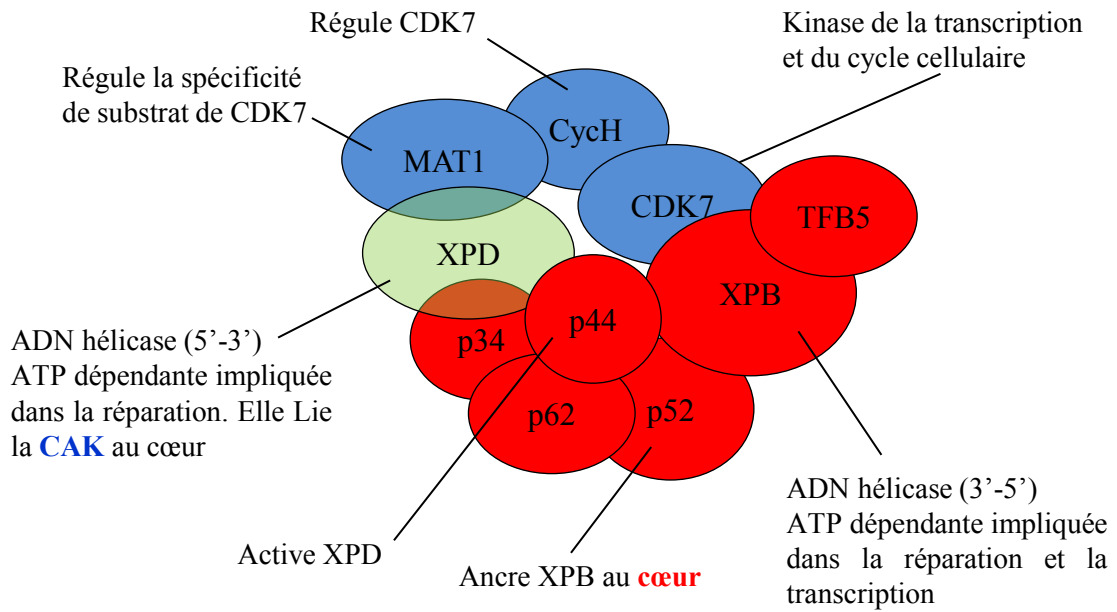
Les différents domaines et leurs partenaires d'interaction sont indiqués

### II.1.6. TFIIH

TFIIH est un complexe multiprotéique composé de dix sous-unités de masse moléculaire comprise entre 8 et 89 kDa (Cf. Figure 10). Il possède plusieurs activités enzymatiques nécessaires pour la transcription, une activité ATPase ADN-dépendante, des activités hélicases ATP-dépendante et une activité CTD-kinase. Les sous-unités XPB et XPD jouent également un rôle dans la réparation de l'ADN par excision de nucléotides (NER : nucleotide excision repair). Les mutations de XPB ou XPD sont associées à plusieurs pathologies humaines telles, le Xeroderma Pigmentosum (XP) et la Trichothiodystrophie (TTD). Ces pathologies sont caractérisées par des défauts de réparation des lésions de l'ADN causées par les rayonnements UV.

TFIIH peut être subdivisé en deux sous-complexes : le noyau et le complexe CAK (*Cyclin-Activating Kinase complex*). Le noyau contient les sous-unités XPB, p62, p52, p44 et p34, alors que le complexe CAK contient CDK7, MAT1 et la cycline H. Lors des étapes de purification, XPD peut être purifié avec le noyau ou avec CAK (Cf. Figure9). Il joue probablement un rôle de pontage entre les deux sous-complexes. La dixième sous-unité de TFIH n'a été découverte que récemment, il s'agit de la protéine TFB5 impliquée dans l'activité de réparation de l'ADN par TFIH chez la levure (Ranish et al., 2004) et chez l'homme (Giglia-Mari et al., 2004). En effet, l'inactivation du gène TFB5 chez la levure provoque une augmentation du taux de mortalité après exposition aux radiations UV. De même, des mutations dans le gène codant pour TFB5 sont fréquentes chez les patients atteints de TTD. Ces patients présentent également un taux élevé de dégradation de TFIH. Cela suggère que TFB5 est essentielle pour la stabilité de TFIH (Giglia-Mari et al., 2004). Des expériences *in vitro* montrent que TFB5 stimule l'activité ATPase de XPB et facilite l'ouverture de l'ADN double brin autour de la lésion lors des réactions de réparation par TFIH (Coin et al., 2006).

Au sein du complexe TFIH, le sous-complexe CAK phosphoryle la sérine 5 du CTD de la Pol II via l'activité kinase de sa sous-unité CDK7. Cette activité est stimulée par TFIIE, le médiateur et XPD (Svejstrup et al., 1996 ; Keriél et al., 2002). En revanche, cette activité CTD kinase est réduite quand CAK n'est pas associé à TFIH (Yankulov and Bentley, 1997). L'activité kinase de CAK n'est pas spécifique à la transcription. En effet, sous sa forme libre non



**Figure 10** : Composition et fonction des sous-unités de TFIIH

## INTRODUCTION PARTIE I

complexée à TFIIH, CAK peut phosphoryler d'autres protéines comme les CDKs (*cyclin dependant kinase*) impliquées dans la régulation du cycle cellulaire. CAK serait donc une kinase "mercenaire" recrutée par des facteurs spécialisés dans des processus différents.

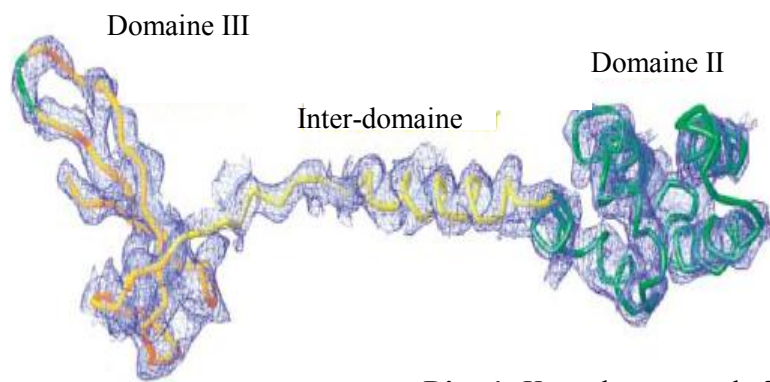
TFIIH possède deux activités hélicases, une activité hélicase 3'-5' nécessaire pour la réparation et la transcription assurée par XPB et une activité hélicase 5'-3' nécessaire uniquement pour la réparation assurée par XPD (Schaeffer et al., 1994 ; Zurita and Merino, 2003). Dans des expériences de transcription *in vitro* utilisant une matrice d'ADN dénaturée ou super-enroulée négativement, TFIIH n'est pas indispensable pour la transcription (Parvin and Sharp, 1993 ; Pan and Greenblatt, 1994). Cela suggère un rôle de TFIIH dans le déroulement de l'ADN et la formation du complexe ouvert.

### II.1.7. TFIIIS

TFIIIS est un facteur d'initiation et d'élongation de la transcription conservé au cours de l'évolution. Il est surtout connu pour son rôle dans le processus d'élongation de la transcription puisqu'il stimule le clivage de l'ARN naissant par la Pol II ce qui permet à cette dernière de passer d'un état d'arrêt à un état d'élongation. Des études récentes ont révélé que TFIIIS est également impliqué dans le processus d'initiation de la transcription par les Pol II et III (Prather et al., 2004 ; Guglielmi et al., 2007 ; Ghavi-Helm et al., 2008).

TFIIIS est un facteur monomérique qui comporte trois domaines. Le domaine I n'est pas structuré alors que les domaines II et III ainsi que la région qui les relie forment la partie structurée de la protéine. Le domaine II est composé de trois hélices  $\alpha$  ( $\alpha_1$ ,  $\alpha_2$  et  $\alpha_3$ ) formant une structure dite "three-helix bundle" suivies de trois courtes hélices supplémentaires ( $\alpha_4$ ,  $\alpha_5$  et  $\alpha_6$ ). L'hélice  $\alpha_3$  est essentielle à l'interaction entre TFIIIS et RPB1. Le bras entre les domaines II et III contient une seule hélice  $\alpha_7$ . Le domaine III contient trois feuilletts  $\beta$  qui forment un ruban de zinc (*zinc ribbon*) qui contient le motif conservé RSADE responsable de la stimulation du clivage de l'ARN (Kettenberger et al., 2003). Cf. *Figure 11*.

Les trois régions structurées de TFIIIS sont requises pour assurer son rôle dans l'élongation de la transcription. En effet, l'interaction entre le domaine II et la mâchoire formée par les sous-unités RPB1/RPB9 de la Pol II permet au domaine III de contacter le site actif de la



D'après Kettenberger, et al., 2003

**Figure 11** : Structure cristallographique de TFIIIS

## INTRODUCTION PARTIE I

Pol II afin de stimuler l'activité de clivage de l'ARN (Kettenberger et al., 2003). Par contre, seuls le domaine II et le bras sont impliqués dans l'initiation de la transcription (Guglielmi et al., 2007). Ce résultat démontre que les deux activités de TFIIIS (initiation/élongation) sont indépendantes mais nécessitent néanmoins l'interaction entre TFIIIS et la Pol II.

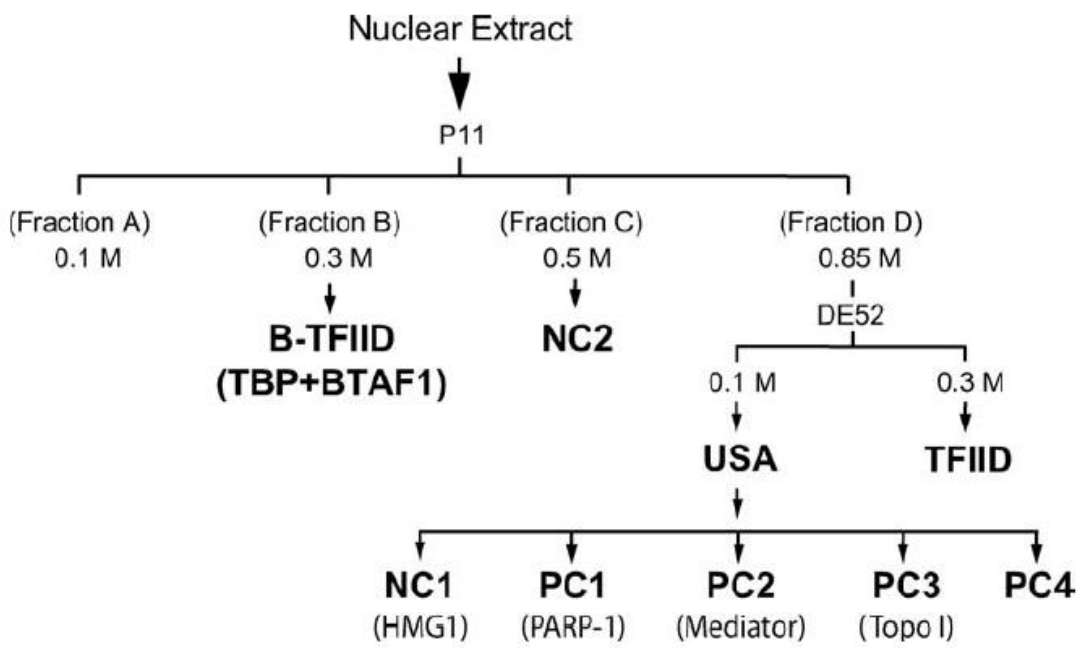
Chez la levure, TFIIIS interagit avec l'élément UAS (*Upstream Activating Sequence*) du gène GAL1 et, en synergie avec le complexe SAGA et le Médiateur, il stimule la transcription de ce gène (Prather et al., 2004). Il a également été démontré que TFIIIS est recruté sur les promoteurs et les régions codantes des gènes ADH1, MET17 et VTC3. Sur le gène MET17, le recrutement de TFIIIS sur le promoteur est plus élevé que dans la région codante (Guglielmi et al., 2007). Ce profil de liaison est caractéristique des facteurs d'initiation de la transcription. L'ensemble de ces études indique que TFIIIS est bel et bien un facteur d'initiation de la transcription.

Afin de lui attribuer le titre de GTF, il restait à savoir si TFIIIS est recruté sur un grand nombre de promoteurs. Des expériences de ChIP on chip montrent que TFIIIS est recruté au niveau de la majorité des gènes transcrits par la Pol II, mais également des gènes transcrits par la Pol III (Ghavi-Helm et al., 2008). Cette étude démontre clairement que TFIIIS est bel et bien un facteur de transcription impliqué dans la transcription par la Pol II et la Pol III et qu'il mérite de faire partie du clan restreint des GTFs.

### **II.2. Les cofacteurs généraux de la transcription**

Les cofacteurs de transcription jouent le rôle d'intermédiaires entre les régulateurs de la transcription et la machinerie de transcription de base. Ils peuvent être classés en trois classes majeures : les TAFs (*TBP Associated Factors*), les complexes Médiateurs et les complexes USA (*Upstream Stimulatory Activity*). Outre ces trois classes, le régulateur de la transcription NC2 (*Negative Cofactor 2*), les complexes de remodelage de la chromatine et les enzymes de modifications des histones peuvent également être considérés comme des cofacteurs de la transcription. Ce chapitre abordera uniquement les complexes USA, les complexes Médiateurs et le régulateur NC2. Le rôle des TAFs comme cofacteurs sera abordé dans le chapitre consacré aux





**Figure 12:** purification des complexes USA

## INTRODUCTION PARTIE I

complexes contenant TBP. Les modifications de la chromatine seront également abordées dans un chapitre indépendant.

### II.2.1. Les complexes USA

Les complexes USA correspondent à une fraction chromatographique initialement définie comme un coactivateur de la transcription activée (Meisterernst et al., 1991 a). Cette fraction contient en réalité cinq cofacteurs distincts : PC1, PC2, PC3, PC4 et NC1 (*PC* : *Positive Cofactor* et *NC* : *Negative Cofactor*) (Cf. Figure 12). Il s'est avéré par la suite que PC1 est l'enzyme PARP-1 (*Poly ADP-Ribose Polymerase-1*), PC2 contient une forme du complexe Médiateur, PC3 correspond à l'ADN Topoisomérase-1 (Topo I) et NC1 correspond à la protéine HMG1 (*High Mobility Group Protein 1*). De manière générale, les complexes USA agissent comme coactivateurs en présence des activateurs de la transcription, mais en absence de ces derniers ils agissent comme répresseurs.

PC1/PARP-1 est une enzyme plus connue pour son rôle dans la réparation de l'ADN que pour son rôle de coactivateur de la transcription. Elle modifie les protéines impliquées dans le métabolisme des acides nucléiques en catalysant le transfert du groupe ADP-ribose du donneur NAD<sup>+</sup> vers les protéines cibles. PARP-1 peut modifier une grande variété de protéines dont des histones, des GTFs ou encore des suppresseurs de tumeurs comme P53. PARP-1 interagit avec de nombreux activateurs tels TR $\alpha$ , RAR $\alpha$  (Pavri et al., 2005) et NF $\kappa$ B (Hassa et al., 2003), mais également avec le Médiateur (Pavri et al., 2005). PARP1 joue un rôle important dans l'activation de la transcription du gène RAR $\beta$ 2 par RAR. Lors de cette activation, PARP-1 est co-recruté avec le Médiateur sur le promoteur de RAR $\beta$ 2 et stimule la dissociation de la sous-unité CDK8 du Médiateur. Cette dissociation est nécessaire à l'activation de la transcription (Pavri et al., 2005).

Afin de supprimer les surenroulements de l'ADN, PC3/Topo-I introduit des coupures transitoires dans l'ADN et catalyse le passage des brins d'ADN à travers ces coupures avant de les refermer. Cette action est importante pour la transcription et la réplication. Topo-I interagit avec des activateurs comme AP2 (Kannan et al., 1999) et c-Jun (Mialon et al., 2005), mais également avec TBP (Merino et al., 1993). Topo-I peut donc être recrutée par les activateurs afin de modifier la topologie de l'ADN pour faciliter la transcription.

## INTRODUCTION PARTIE I

PC4 est une protéine multifonctionnelle de 127 acides aminés, elle peut être subdivisée en un domaine N-terminal et un domaine C-terminal. Via son domaine N-terminal, PC4 peut interagir avec les activateurs de la transcription, avec TBP, mais également avec l'ADN double brin. Son domaine C-terminal peut interagir avec l'ADN simple brin. Le domaine N-terminal de PC4 peut être phosphorylé par la kinase CK2. Cette modification inhibe son potentiel coactivateur, car elle inhibe l'interaction entre PC4 et TBP (Kretzschmar et al., 1994). En absence de coactivateurs, PC4 peut agir comme un répresseur de la transcription (Wu and Chiang, 1998). PC4 est également impliqué dans la réparation de l'ADN : des analyses récentes montrent l'accumulation rapide de PC4 au niveau des lésions d'ADN introduites par des agents chimiques ou par des irradiations. Cela suggère que PC4 joue un rôle dans les étapes précoces du processus de réparation. Cette activité dépend de sa capacité à lier l'ADN simple brin (Mortusewicz et al., 2008).

NC1/HMG1 est un membre de la famille des protéines associées à la chromatine (*Chromatin-Associated Proteins*). HMG1 contient deux domaines de liaisons à l'ADN, un domaine HMG-box de type A et un domaine HMG-box de type B, ainsi qu'un troisième domaine C-terminal constitué principalement de résidus acides. Ses domaines HMG-box lui permettent d'interagir avec l'ADN de manière non spécifique (Thomas, 2001 ; Thomas and Travers, 2001). Le domaine C-terminal acide semble moduler cette liaison. HMG1 agit comme un coactivateur en présence d'activateurs tels P53 ou les récepteurs à l'hormone stéroïdienne (Thomas and Travers, 2001 ; Agresti and Bianchi, 2003). En absence d'activateur, HMG1 agit comme un corépresseur de la transcription. L'activité de répression de HMG1 semble être liée à sa capacité de lier TBP et d'empêcher ainsi son interaction avec TFIIB (Ge and Roeder, 1994b).

### **II.2.2. Le complexe Médiateur**

Initialement découvert chez la levure, le Médiateur est un cofacteur de la transcription composé d'une vingtaine de sous-unités. Il est capable d'interagir avec la Pol II, de stimuler l'activité CTD kinase de TFIIF et de stimuler la transcription basale et activée (Kim et al., 1994).

Chez l'homme, l'équivalent du Médiateur a été initialement décrit comme un cofacteur spécifique du récepteur des hormones thyroïdiennes TR $\alpha$ . Ce complexe a été appelé TRAP (TR $\alpha$ -

## INTRODUCTION PARTIE I

*Associated Protein complex*) (Fondell et al., 1996). Le lien ne fut pas fait immédiatement entre TRAP et le Médiateur de levure.

Ce n'est qu'en 1999 qu'un complexe appelé SMCC (*SRB/MED-Containing Cofactor*) a été décrit comme l'équivalent humain du Médiateur de levure (Gu et al., 1999). Il s'est avéré par la suite que TRAP et SMCC partageaient un grand nombre de sous-unités et étaient probablement un seul et unique complexe (Ito et al., 1999). Le complexe Médiateur humain est encore fréquemment appelé TRAP/SMCC.

La purification du complexe Médiateur chez d'autres espèces a montré l'hétérogénéité de ce complexe. En effet, la composition du médiateur peut varier entre les espèces et plusieurs formes du complexe Médiateur peuvent exister chez la même espèce. Chez l'homme, le complexe TRAP/SMCC partage plusieurs sous-unités avec d'autres complexes jouant des rôles similaires. Parmi lesquels se trouvent ARC (*Activator Recruited Cofactor*) (Näär et al., 1999), CRSP (*Cofactor Required for SP1 activation*) (Ryu et al., 1999) et PC2 (Malik et al., 2000). Il est donc plus approprié de parler des complexes Médiateurs. Cf. *Tableau 1*

Malgré cette diversité, une classification basée sur la taille des complexes a été proposée. Dans les cellules humaines, deux formes majeures des complexes Médiateurs peuvent être purifiées séparément dans une colonne échangeuse d'ions de type phosphocellulose. La première forme correspond à la fraction éluée à la concentration saline de 0,5 M KCl, appelée Médiateur-P.5. La deuxième forme est isolée à 0,85 M KCl et est nommée Médiateur-P.85. La forme Médiateur-P.5 comporte les grands complexes Médiateurs comme TRAP et ARC, alors que Médiateur-P.85 comporte les petits complexes Médiateurs comme CRSP et PC2. Les grands complexes Médiateurs se caractérisent par la présence d'un module labile composé de MED12, MED13, CDK8 et CycC (Wu et al., 2003). Ce module, souvent appelé "module CDK8" ou "module CDK", joue un rôle central dans l'activité des complexes Médiateur.

Une trentaine de protéines appartenant aux complexes Médiateurs ont été identifiées. De nombreuses protéines sont présentes dans plusieurs complexes et portent des noms différents créant ainsi une mosaïque complexe de cofacteurs. Afin de simplifier cette mosaïque, une nomenclature unifiée a été proposée en 2004. Toutes les sous-unités des complexes médiateurs

| New name | <i>C. elegans</i>                 |                 |                            |          | <i>H. sapiens</i> <sup>d</sup>      |                 |             |         |                 |            |
|----------|-----------------------------------|-----------------|----------------------------|----------|-------------------------------------|-----------------|-------------|---------|-----------------|------------|
|          | <i>S. cerevisiae</i> <sup>a</sup> | <i>S. pombe</i> | Previous name <sup>b</sup> | New name | <i>D. melanogaster</i> <sup>c</sup> | TRAP/SMCC       | ARC/DRIP    | CRSP    | PC2             | OTHERS     |
| MED1     | Med1                              | Pmc2            | SOP-3*                     | MDT-1.1  | Trap220*                            | TRAP220         | ARC/DRIP205 | CRSP200 | TRAP220         | PBP        |
| MED1L    |                                   |                 | T23C6.1*                   | MDT-1.2  |                                     |                 |             |         |                 |            |
| MED2     | Med2                              |                 |                            |          |                                     |                 |             |         |                 |            |
| MED3     | Pgd1/Hrs1/Med3                    |                 |                            |          |                                     |                 |             |         |                 |            |
| MED4     | Med4                              | Pmc4/SpMed4     | ZK546.13*                  | MDT-4    | Trap36                              | TRAP36          | ARC/DRIP36  |         | TRAP36          | p34        |
| MED5     | Nut1                              |                 |                            |          |                                     |                 |             |         |                 |            |
| MED6     | Med6                              | Pmc5/SpMed6     | LET-425/MED-6              | MDT-6    | Med6                                | hMed6           | ARC/DRIP33  |         | hMed6           | p32        |
| MED7     | Med7                              | SpMed7          | LET-49/MED-7               | MDT-7    | Med7*                               | hMed7           | ARC/DRIP34  | CRSP33  | hMed7           | p36        |
| MED8     | Med8                              | Sep15/SpMed8    | Y62F5A.1b*                 | MDT-8    | Arc32*                              |                 | ARC32       |         |                 | mMed8      |
| MED9     | Cse2/Med9                         |                 |                            |          | CG5134*                             |                 |             |         |                 | Med25      |
| MED10    | Nut2/Med10                        | SpNut2          | T09A5.6                    | MDT-10   | Nut2*                               | hNut2           | hMed10      |         | hNut2           |            |
| MED11    | Med11                             |                 | R144.9*                    | MDT-11   | Med21                               |                 |             |         |                 | HSPC296    |
| MED12    | Srb8                              | SpSrb8          | DPY-22/SOP-1*              | MDT-12   | Kto*                                | TRAP230         | ARC/DRIP240 |         |                 |            |
| MED12L   |                                   |                 |                            |          |                                     |                 |             |         |                 | TRALPUSH*  |
| MED13    | Ssn2/Srb9                         | SpTrap240       | LET-19*                    | MDT-13   | Skd/Pap/Bli*                        | TRAP240         | ARC/DRIP250 |         |                 |            |
| MED13L   |                                   |                 |                            |          |                                     |                 |             |         |                 | PROSIT240  |
| MED14    | Rgr1                              | Pmc1/SpRgr1     | RGR-1*                     | MDT-14   | Trap170                             | TRAP170         | ARC/DRIP150 | CRSP150 | TRAP170         | p110       |
| MED15    | Gal11                             | SpGal11*        | R12B2.5b*                  | MDT-15   | Arc105*                             |                 | ARC105      |         | PCQAP           | TIG-1      |
| MED16    | Sin4                              |                 |                            |          | Trap95*                             | TRAP95          | DRIP92      |         | TRAP95          | p96b       |
| MED17    | Srb4                              | SpSrb4          | Y113G7B.18*                | MDT-17   | Trap80                              | TRAP80          | ARC/DRIP77  | CRSP77  | TRAP80          | p78        |
| MED18    | Srb5                              | Pmc6/Sep11      | C55B7.9*                   | MDT-18   | p28/CG14802                         |                 |             |         |                 | p28b       |
| MED19    | Rox3                              | SpRox3          | Y71H2B.6*                  | MDT-19   | CG5546*                             |                 |             |         |                 | LCMR1      |
| MED20    | Srb2                              | SPAC17G8.05*    | Y104H12D.1*                | MDT-20   | Trfp                                | hTRFP           |             |         | hTRFP           | p28a       |
| MED21    | Srb7                              | SpSrb7          | C24H11.9*                  | MDT-21   | Trap19                              | hSrb7           | hSrb7       |         | hSrb7           | p21        |
| MED22    | Srb6                              | SpSrb6          | ZK970.3*                   | MDT-22   | Med24                               |                 |             |         |                 | Surf5      |
| MED23    |                                   |                 | SUR-2*                     | MDT-23   | Trap150 $\beta$ *                   | TRAP150 $\beta$ | ARC/DRIP130 | CRSP130 | TRAP150 $\beta$ | hSur2      |
| MED24    |                                   |                 |                            |          | Trap100*                            | TRAP100         | ARC/DRIP100 | CRSP100 | TRAP100         |            |
| MED25    |                                   |                 |                            |          | Arc92*                              |                 | ARC92       |         |                 | ACID1      |
| MED26    |                                   |                 |                            |          | Arc70*                              |                 | ARC70       | CRSP70  |                 |            |
| MED27    |                                   | Pmc3            | T18H9.6*                   | MDT-27   | Trap37*                             | TRAP37          |             | CRSP34  | TRAP37          |            |
| MED28    |                                   |                 | W01A8.1*                   | MDT-28   | Med23                               |                 |             |         |                 | Fksg20     |
| MED29    |                                   |                 | K08E3.8*                   | MDT-29   | Intersex*                           |                 |             |         |                 | Hinterssex |
| MED30    |                                   |                 |                            |          | Trap25                              | TRAP25          |             |         |                 |            |
| MED31    | Soh1*                             | SpSoh1/Sep10*   | F32H2.2*                   | MDT-31   | Trap18                              | hSoh1           |             |         | hSoh1           |            |
| CDK8     | Srb10/Ssn3/Ume5                   | SpSrb10         | CDK-8*                     |          | Cdk8                                | hSrb10          | CDK8        |         |                 |            |
| CycC     | Srb11/Ssn8/Ume3                   | SpSrb11         | H14E04.5*                  | CIC-1    | CycC                                | hSrb11          | CycC        |         |                 |            |

D'après Bourbon et al., 2004

**Tableau 1** : Nomenclature unifiée des sous-unités des complexes Médiateurs

## INTRODUCTION PARTIE I

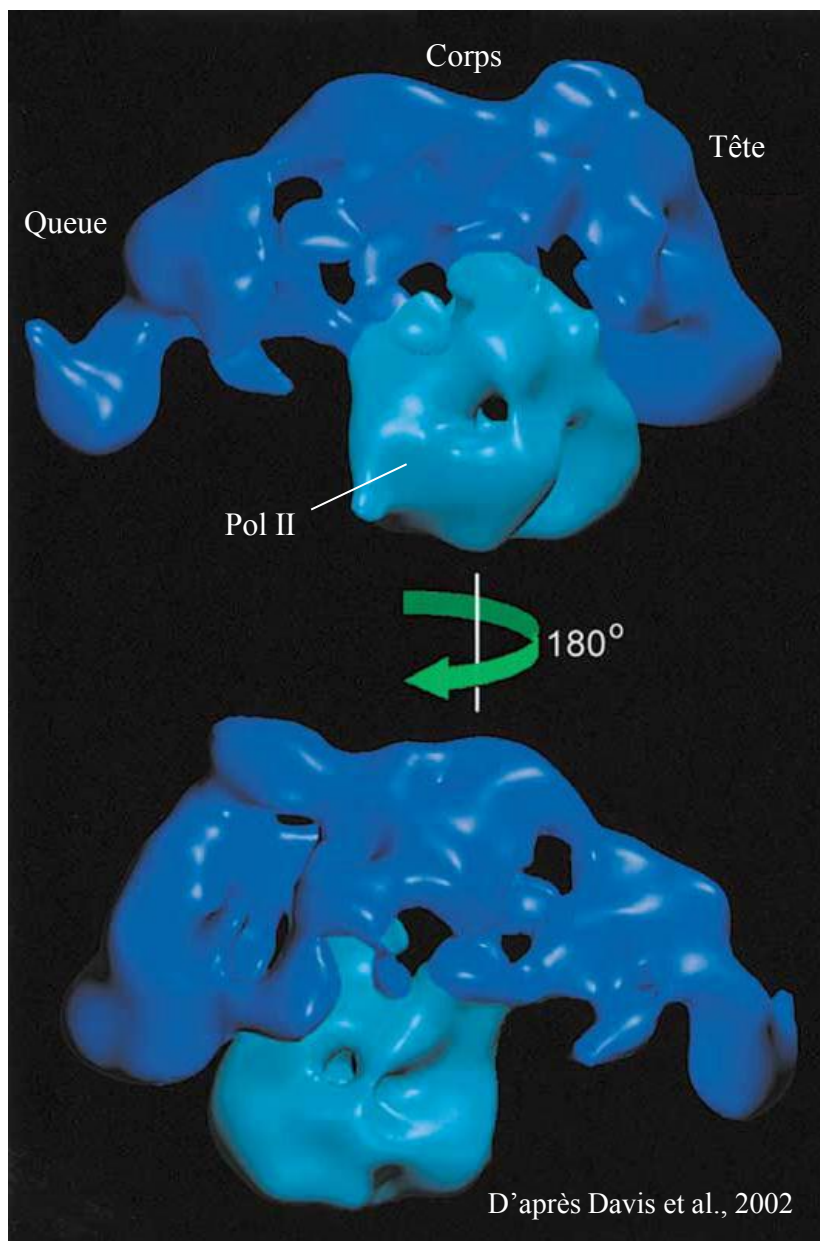
ont été renommées MEDn (n = chiffre allant de 1 à 31) à l'exception de CDK8 et CycC (Bourbon et al., 2004). Cf. *Tableau 1*

Les structures de plusieurs complexes Médiateurs chez différentes espèces ont été résolues par microscopie électronique avec une résolution allant de 30 à 40 Å. La comparaison de ces structures révèle une organisation générale commune. Les complexes Médiateurs se divisent en 3 modules : la tête, le corps et la queue. La tête et le corps interagissent avec la Pol II alors que la queue serait impliquée dans les interactions avec les activateurs (Asturias et al., 1999 ; Dotson et al., 2000 ; Davis et al., 2002). Cf. *Figure 13*

Plusieurs activités ont été associées aux complexes Médiateurs. La principale fonction qui leur a été attribuée est le rôle d'adaptateur entre les composants du PIC et les activateurs. Mais le Médiateur peut également stimuler la transcription basale indépendamment des activateurs et se comporte ainsi comme un GTF (pour revue, Taatjes, 2010). Il a été démontré récemment que le Médiateur joue un rôle central dans le recrutement de TFIID, TFIIE et de la Pol II sur les promoteurs chez la levure (Esnault et al., 2008).

Dans l'activité du Médiateur, le module labile CDK8 joue un rôle important d'inhibiteur de la transcription en phosphorylant le CTD de la Pol II sous sa forme libre (non engagée dans le PIC). Sachant que la Pol II entre dans le PIC sous sa forme non phosphorylée, cette modification l'empêcherait de rejoindre le PIC (Hengartner et al., 1998). Il a également été démontré que la dissociation de CDK8 du médiateur était nécessaire pour l'activation du gène RAR $\beta$ 2 après traitement à l'acide rétinoïque (Pavri et al., 2005).

Le module CDK8 peut également influencer la transcription en participant aux modifications des histones. En effet, sa sous-unité MED12 joue un rôle dans le recrutement de la méthyltransférase G9a sur les promoteurs des gènes neuronaux réprimés dans les cellules HeLa. La méthyltransférase G9a catalyse la méthylation de la lysine 9 de l'histone H3 (H3K9). Cette modification représente une marque de répression. Cependant la réduction de l'expression de MED12 n'affecte pas le taux global de méthylation de H3K9. Cela suggère que l'activité de MED12 est spécifique à certains gènes (Ding et al., 2008). Une autre étude montre que CDK8



**Figure13 : Architecture du Médiateur de levure**

Reconstruction chez la Levure, par microscopie électronique de l'holoenzyme Pol II-Médiateur

## INTRODUCTION PARTIE I

peut phosphoryler la sérine 10 de l'histone H3 (Meyer et al., 2008). Cette modification joue un rôle central dans la régulation du cycle cellulaire.

### **II.2.3. Negative cofactor 2**

Negative cofactor 2 "NC2" est un hétérodimère composé des sous-unités NC2 $\alpha$  et NC2 $\beta$  qui interagissent entre elles via des motifs de repliement de type histone (HFD : *Histone Fold Domain*). NC2 interagit avec TBP et l'ADN en agissant comme une pince fixant les deux molécules (Kamada et al., 2001) (*Cf . Figure 14*). NC2 a été initialement décrit comme un répresseur global de la transcription (Meisterernst et al., 2001; Gadbois et al., 1997), mais d'autres études suggèrent qu'il joue également un rôle coactivateur de la transcription en facilitant l'assemblage du PIC (Geisberg et al., 2001 ; Gilfillan et al., 2005).

*In vitro*, le NC2 de drosophile (dNC2) inhibe la transcription des promoteurs TATA-dépendants mais stimule la transcription des promoteurs TATA-indépendants possédant un élément DPE. Les deux activités d'activation et de répression de la transcription semblent être distinctes. En effet, un mutant de NC2 déficient dans la répression de la transcription des promoteurs TATA-dépendant peut stimuler la transcription à partir de promoteurs contenant un DPE (Willy et al., 2000).

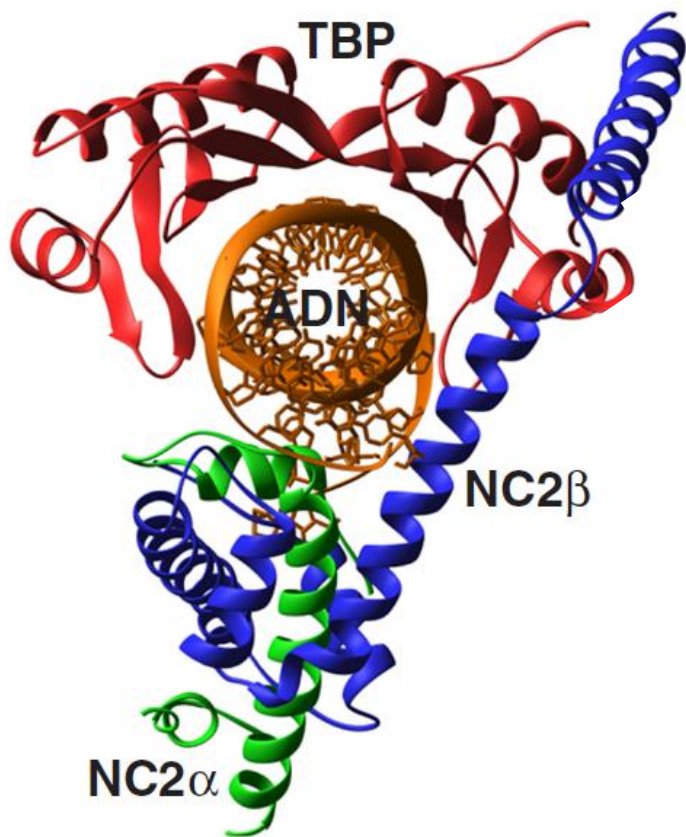
Si le rôle répresseur de NC2 est attribué à sa compétition avec TFIIA et TFIIB pour lier TBP et ainsi empêcher l'assemblage du PIC, le mécanisme de sa fonction de coactivateur reste à élucider. Il est important de noter que TFIIB et NC2 colocalisent sur un grand nombre de gènes dans les cellules B humaines (Albert et al., 2010). Cette colocalisation peut traduire un équilibre dynamique entre TFIIB et NC2 pour la liaison de TBP mais un rôle activateur de NC2 n'est pas à exclure, il est possible que l'action de NC2 soit dépendante du gène considéré.

### **II.3. Les modèles d'assemblage du complexe de préinitiation *in vitro***

*In vitro*, Il existe deux modèles d'assemblage du PIC : un modèle séquentiel et un modèle de complexe pré-assemblé dit holoenzyme.

Le modèle séquentiel se base sur l'observation que l'addition des GTFs doit se faire dans un ordre spécifique pour initier la transcription *in vitro* à partir d'une matrice d'ADN nu. Dans ce





**Figure 14** : Structure cristallographique du complexe TBP-ADN-NC2

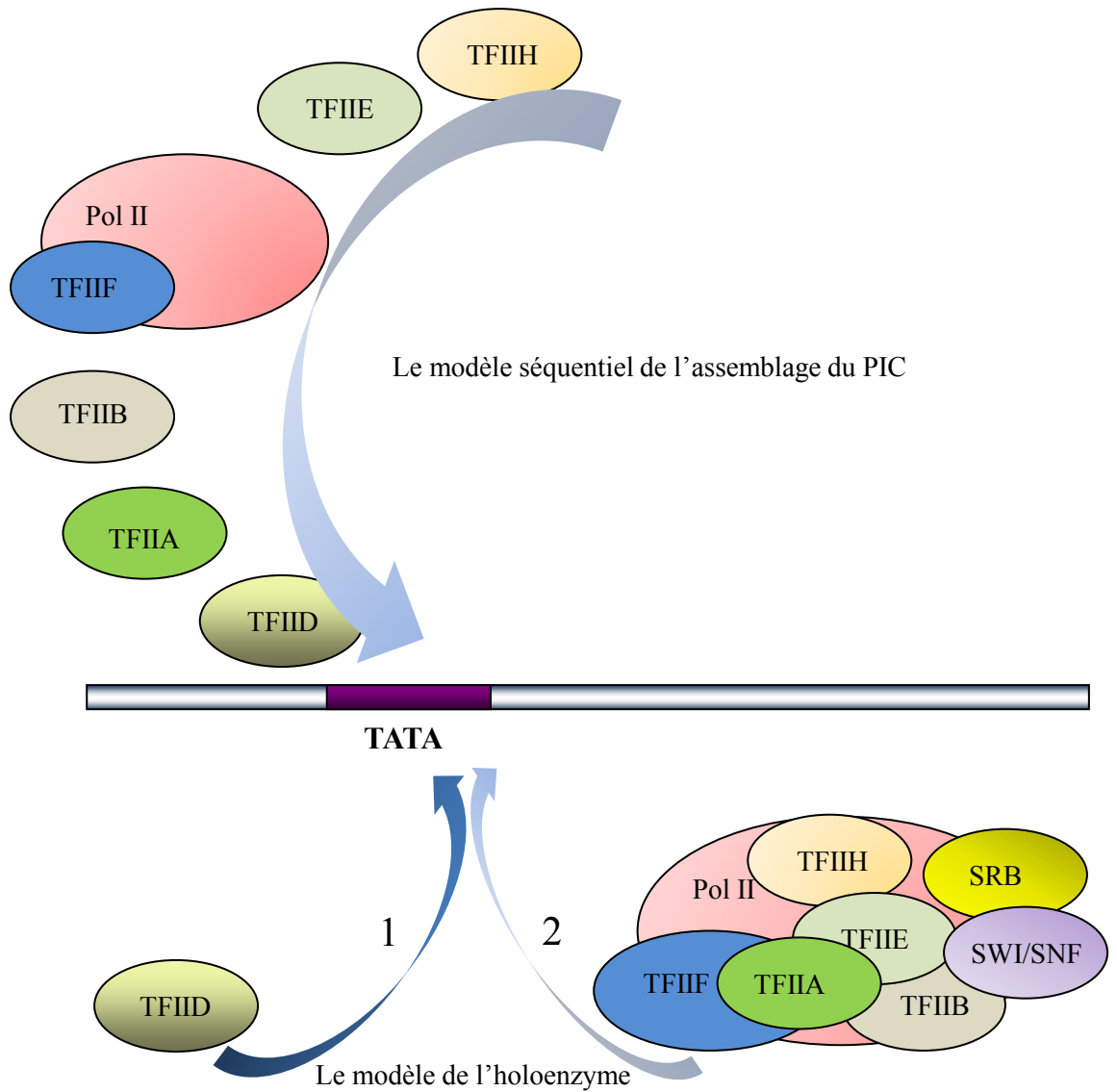
## INTRODUCTION PARTIE I

modèle, TFIID serait le premier GTF à interagir avec le promoteur via la liaison de sa sous-unité TBP à l'élément TATA. TFIIA et TFIIB se fixent ensuite sur le complexe TFIID/ADN et le stabilisent. Puis, la Pol II escortée par TFIIF rejoint le PIC. Enfin, sont recrutés les facteurs TFIIIE et TFIIH pour achever l'assemblage du PIC. L'action conjointe de TFIIH, TFIIIE et TFIIF permet la séparation des brins de l'ADN autour du site d'initiation de la transcription (pour revue, Thomas et Chiang., 2006)

Dans le modèle de l'holoenzyme, un complexe pré-assemblé dit " ARN polymérase II holoenzyme " contient la Pol II et certains facteurs de transcription. L'holoenzyme a été décrite pour la première fois en 1994 chez la levure en tant que complexe contenant la Pol II, TFIIB, TFIIF, TFIIH ainsi que certaines protéines SRB (*Suppressor of RNA Polymerase B*). Cette holoenzyme est cependant inapte à initier la transcription *in vitro* sans l'addition de TBP et de TFIIA (Koleske and Young, 1994). Plusieurs complexes de composition différente de celle proposée par Koleske ont été identifiés par d'autres équipes. Parmi les composants trouvés dans la préparation d'holoenzyme se trouvaient par exemple GAL11, certains cofacteurs positifs et négatifs ainsi que le facteur de remodelage de la chromatine SWI/SNF. Cette diversité peut être expliquée par des conditions de purification différentes. Mais l'absence de TFIID était une caractéristique commune à toutes les préparations de l'holoenzyme. Ceci suggère que le recrutement de TFIID sur le promoteur précède celui de l'holoenzyme. Cf. *Figure 15*

### **III. La régulation de la transcription**

L'activité de la Pol II est régulée individuellement au niveau de chaque gène. Cette régulation est le produit de combinaisons complexes d'interactions entre les facteurs de transcription, les séquences d'ADN, l'environnement chromatinien ainsi que les enzymes qui modifient la chromatine. Dans un environnement chromatinien permissif à la transcription, la réaction de transcription consiste en trois étapes majeures : l'initiation, l'élongation et la terminaison. Le cycle transcriptionnel commence par le recrutement des facteurs d'initiation de la transcription et de la Pol II sur le promoteur. Après la séparation des deux brins d'ADN, la Pol II se dissocie du promoteur et entre en phase d'élongation le long de la phase codante. Les signaux de terminaison portés par l'ADN lui indiquent la fin de la région à transcrire. Ces trois étapes majeures peuvent être divisées en plusieurs sous-étapes. La régulation de la transcription peut intervenir à chaque stade. Le recrutement de la Pol II et des facteurs d'initiation de la



**Figure 15** : Les modèles d'assemblage du PIC *in vitro*

Dans le modèle séquentiel, les GTF et la Pol II sont recrutés sur le promoteur dans un ordre bien défini (TFIID, TFIIA, TFIIIB, TFIIF/Pol II, TFIIIE puis TFIIH)

Dans le modèle de l'holoenzyme, un complexe pré-assemblé contient la Pol II et les GTFs (sauf TFIID). TFIID est recruté d'abord sur le promoteur suivi par l'holoenzyme.

l'holoenzyme peut contenir d'autres facteurs comme les protéines SRB ou les complexes de remodelage de la chromatine SWI/SNF

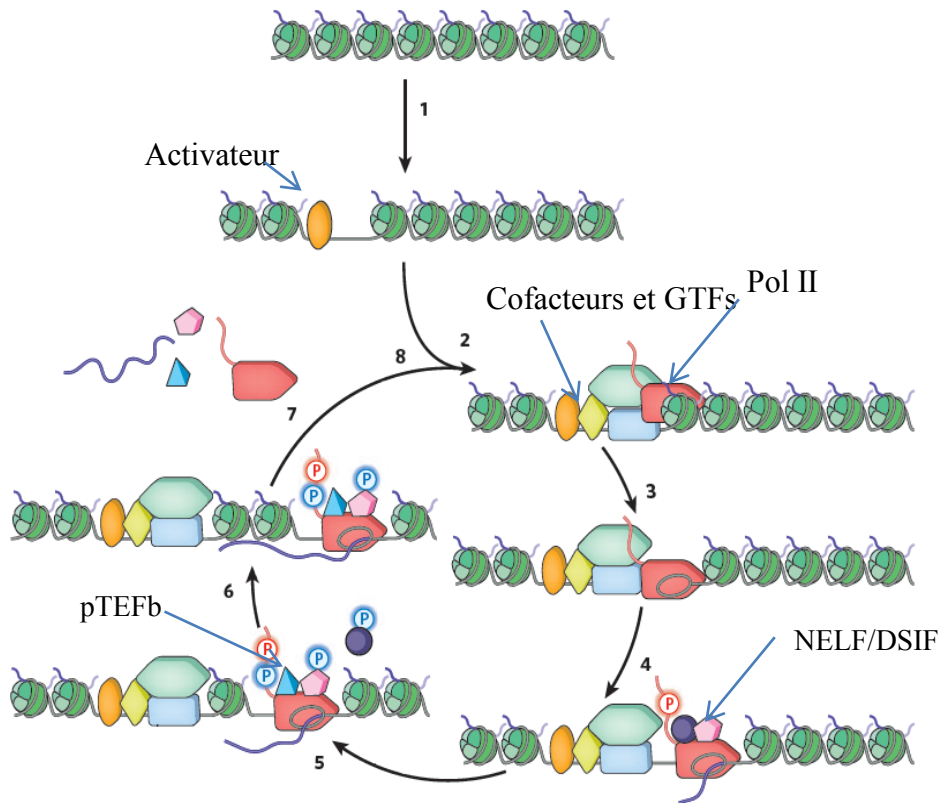
## INTRODUCTION PARTIE I

transcription sur les promoteurs est certainement une étape clé de cette régulation. Mais des étapes postérieures au recrutement de la Pol II peuvent également être régulées. En effet, après le recrutement de la Pol II sur le promoteur, cette dernière peut avoir plusieurs devenir en fonction des signaux de régulation qui sont spécifiques à chaque gène dans une situation physiologique donnée (pour revues, Fuda et al., 2009 ; Core and Lis, 2008) (*Cf. figure 16*).

Les expériences de ChIP-seq montrent que la Pol II est enrichie dans les régions 5' de plusieurs de gènes et non dans leurs régions codantes. Cette observation suggère que pour certains gènes, l'étape limitante de la transcription intervient après le recrutement de la Pol II. Pour ces gènes, la Pol II peut être bloquée au niveau du PIC ou être en état de pause après la transition entre l'initiation et l'élongation précoce (Core et al., 2008). Une technique récente basée sur le séquençage de l'ARN naissant (GRO-seq : *Global Run-on Sequencing*) montre que, pour un grand nombre de gènes, la Pol II est engagée en phase d'élongation mais reste bloquée au niveau du promoteur proximal entre +20 et +50 nucléotides du TSS (Core et al., 2008). Étonnamment, cette étude montre également que, pour certains gènes, la Pol II peut transcrire dans les deux sens à partir du même promoteur. Cet état de pause appelé "promoter-proximal pausing ou stalling" présente probablement une étape cruciale de régulation de la transcription pour un grand nombre de gènes.

L'état de pause de la Pol II peut être favorisé par des régulateurs négatifs comme DSIF (*DRB-Sensitivity Inducing Factor*) ou NELF (*Negative Elongation Factor*) (Wada et al., 1998a ; Yamaguchi et al., 1999) alors que des régulateurs positifs comme P-TEFb peuvent stimuler la reprise de l'élongation et la sortie de l'état de pause (Barboric et al., 2001 ; Core and Lis, 2008 ; Eberhardy and Farnham, 2001, 2002 ; Kanazawa et al., 2003 ; Peterlin and Price, 2006). Il est important de noter que P-TEFb phosphoryle NELF, DSIF et la sérine 2 du CTD de la Pol II via sa sous-unité CDK9 (Kim and Sharp, 2001; Marshall et al., 1996; Marshall and Price, 1995; Wada et al., 1998b; Yamada et al., 2006).

Dans les cellules souches embryonnaires murines (mES), la Pol II est fortement enrichie dans les régions 5' de la majorité des gènes (Rahl et al., 2010). Sur ces gènes, la Pol II colocalise avec les facteurs NELF et DSIF, ce qui confirme le rôle de ces deux facteurs dans le contrôle de l'état de pause de la Pol II. Dans ces cellules, le facteur d'élongation P-TEFb est nécessaire à la



**Figure 16 :** Le Cycle de la transcription par la Pol II

- 1) Ouverture de la chromatine
- 2) Assemblage du PIC
- 3) Initiation de la transcription et formation du complexe ouvert
- 4) La Pol II se dissocie du PIC et entre en phase précoce d'élongation, elle est phosphorylée en serine 5. Des facteurs tels NELF ou DSIF peuvent stimuler la pause de la Pol II après que cette dernière ait transcrit 20 à 30 nucléotides.
- 5) La reprise de l'élongation peut être stimulée par des facteurs positifs qui recrutent pTEFb, ce dernier phosphoryle la Pol II en serine 2 et permet une élongation productive.
- 6) Elongation
- 7) Terminaison

## INTRODUCTION PARTIE I

reprise de l'élongation de la Pol II. Il a été démontré que le facteur c-Myc interagit directement avec P-TEFb (Rahl et al., 2010). L'ensemble de ces données suggère un modèle dans lequel, c-Myc recrute P-TEFb au niveau du promoteur proximal où la Pol II est en état de pause, l'activité kinase de la sous-unité CDK9 de P-TEFb stimulerait alors la reprise de l'élongation en phosphorylant la sérine 2 de la Pol II, DSIF et NELF (pour revue, Fuda et al., 2009).

## IV. La transcription dans le contexte chromatinien

### IV.1. L'organisation de la chromatine

Dans le noyau d'une cellule eucaryote, l'ADN s'associe avec les histones pour former la chromatine. Cette structure nucléoprotéique adopte plusieurs niveaux de repliement pour permettre l'empaquetage de l'ADN dans un petit compartiment comme le noyau. L'organisation de la chromatine influence l'accessibilité des séquences régulatrices aux facteurs de transcription.

L'unité fondamentale de la chromatine est le nucléosome. Le nucléosome est composé d'un octamère d'histones [(H2A/H2B) - (H3/H4)<sub>2</sub> - (H2A/H2B)] autour duquel s'enroule 146 paires de bases d'ADN. La succession de nucléosomes forment la fibre de 10 nm observée par microscopie électronique en 1974 et qui a l'apparence d'un collier de perles (Olins et Olins, 1974). La chromatine pourrait adopter une structure plus compactée appelée la fibre de 30 nm (Marsden et Laemmli, 1979). Des niveaux de compaction supérieurs conduisent à la formation des chromosomes métaphasiques. *Cf. Figure 17*

Les histones sont des petites protéines basiques conservées au cours de l'évolution. Elles possèdent un domaine central structuré appelé HFD organisé en 3 hélices  $\alpha$  reliées entre elles par deux boucles flexibles. *Cf. Figure 18*. Les extrémités N-terminales des histones sont non structurées, pointent en dehors de la structure du nucléosome et sont les sites d'un grand nombre de modifications post-traductionnelles.

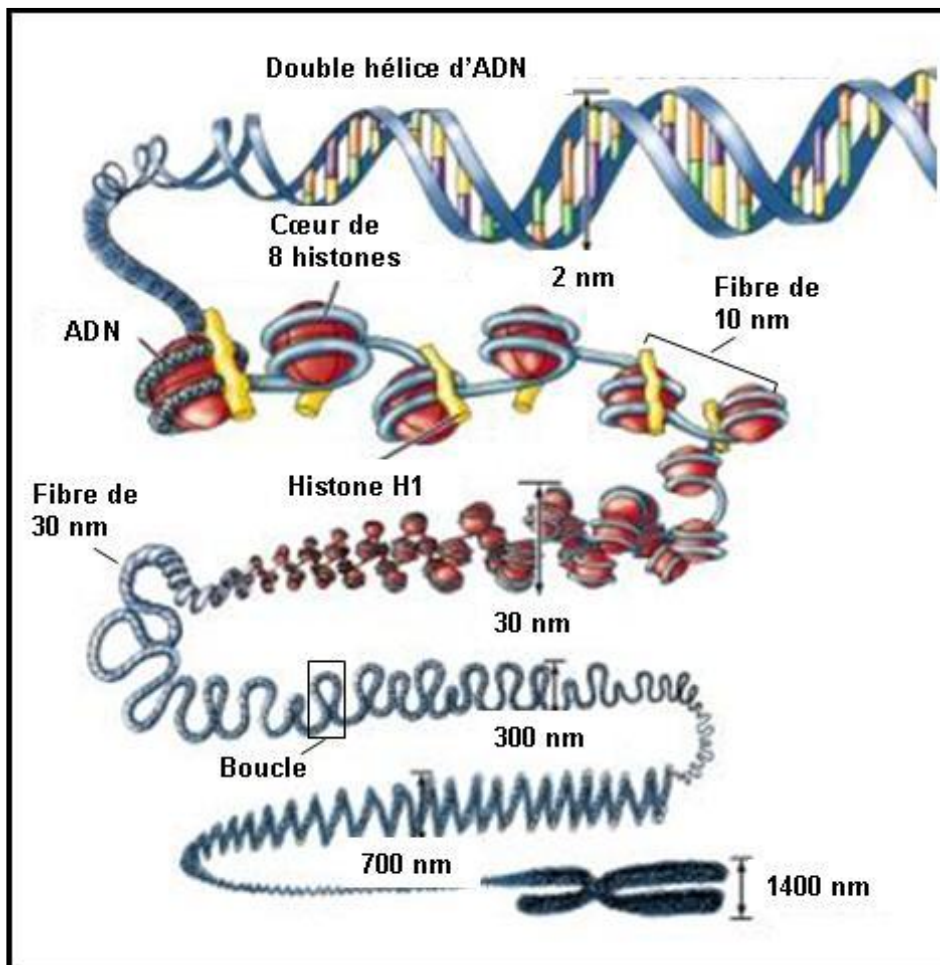


Figure 17 : Les différents niveaux de compaction de la chromatine.

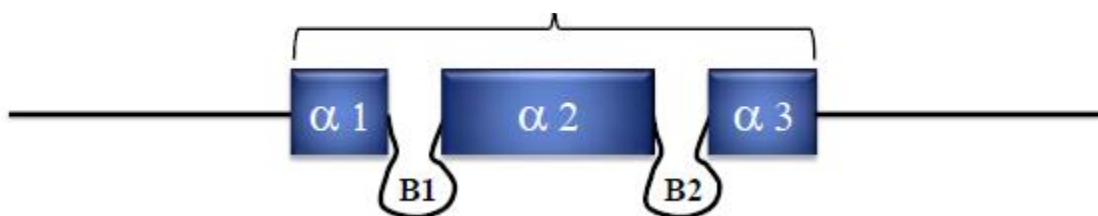


Figure 18 : Organisation structurale du motif "histone fold"

Les trois hélices  $\alpha$  du domaine "histone fold" sont reliées par deux boucles (B1 et B2).

## IV.2. Remodelage et modification de la chromatine

### IV.2.1. La méthylation de l'ADN

Chez les mammifères, l'ADN peut être méthylé au niveau des cytosines localisées en 5' d'une guanine "5'CpG3'" (Hermann et al., 2004). Les dinucléotides CpG sont souvent regroupés en régions riches en C/G appelées " îlots CpG" (pour revue, Bird, 2002). Plus de la moitié des îlots CpG sont localisés au niveau des promoteurs de classe II. La méthylation de l'ADN est associée à un état réprimé de la transcription. Deux mécanismes sont proposés pour expliquer cette répression : 1) la méthylation des cytosines empêche la fixation des facteurs de transcription sur les promoteurs (Watt et Molloy, 1988) ; 2) les CpG méthylés sont reconnus par des protéines telles les membres de la famille MBD (*Methyl Binding Domain*) qui recrutent d'autres répresseurs de la transcription (Bogdanovic and Venstra, 2009).

Les enzymes capables de méthyler l'ADN peuvent être regroupées en deux grandes classes: 1) les méthylases de maintenance telle que DNMT1 (*DNA MethylTransferase 1*). Ces enzymes copient les motifs méthylés préexistant dans le nouveau brin d'ADN au cours de la réplication (Leonhardt et al., 1992) ; 2) les méthylases de novo telles que DNMT3a et DNMT3b qui méthylent des sites CpG non méthylés (Okano et al., 1999) et coopèrent avec DNMT1 pour propager la méthylation au cours de la division cellulaire (Liang et al., 2002). D'autres méthylases de l'ADN ont été identifiées telles que DNMT2 qui a une faible activité méthyltransférase in vitro (Hermann et al., 2003), DNMT3L qui module l'activité de DNMT3a et DNMT3b (Suetake et al., 2004), ainsi que DNMT1o spécifique des oocytes (Howell et al., 2001).

Cf. *Tableau 2*

| <b>Méthyltransférase</b> | <b>Fonction</b>  | <b>Phénotype mutant</b>         |
|--------------------------|--|---------------------------------|
| Dnmt1                    | Maintenance de la méthylation                                | Embryonnaire létal              |
| Dnmt1o                   | Spécifique des oocytes                                       | Perte de l'empreinte maternelle |
| Dnmt2                    | Faible activité  | Pas de phénotype                |
| Dnmt3a, Dnmt3b           | Etablissement de la méthylation                              | Embryonnaire létal              |
| Dnmt3L                   | Pas d'activité catalytique, colocalise avec Dnmt3a et Dnmt3b | Empreinte maternelle anormale   |

**Tableau 2** : Les enzymes impliquées dans la méthylation de l'ADN



| Modification covalente   | Résidus modifiés   | Rôle dans la transcription | Fonctions régulées  |
|--------------------------|--------------------|----------------------------|---|
| <b>Acétylation</b>       | H3 (9, 14, 18, 23) | Activation                 | <i>Transcription, Réparation, Réplication, Condensation</i> |
| Lysine (K)               | H4 (5, 8, 13, 16)  | Activation                 |   |
|                          | H2A, H2B           | Activation                 |   |
| <b>Phosphorylation</b>   | H3 (3, 10, 11, 28) | Activation                 | <i>Transcription, Réparation, Condensation</i>              |
| Sérine/Thréonine (S/T)   | H2A, H2B           | Activation                 |   |
| <b>Méthylation</b>       | H3 (2, 17, 23)     | Activation                 | <i>Transcription</i>  |
| Arginine (R)             | H4 (3)             | Activation                 |   |
| <b>Méthylation</b>       | H3 (4, 36, 79)     | Activation                 | <i>Transcription, Réparation</i>                            |
| Lysine (K)               | H3 (9, 27)         | Répression                 |   |
|                          | H4 (20)            | Répression                 |   |
| <b>Ubiquitine</b>        | H2A (119)          | Répression                 | <i>Transcription, Réparation, Condensation</i>              |
| Lysine (K)               | H2B (120)          | Activation                 |   |
| <b>ADP-ribosylation</b>  | H1                 | Activation                 | <i>Transcription Réparation, Réplication</i>                |
| Glutamate/Arginine (E/R) | H2B (120)          | Activation                 |   |
| <b>SUMOylation</b>       | H2A (126)          | Répression                 | <i>Transcription</i>  |
| Lysine (K)               | H2B (6, 7)         | Répression                 |   |
| <b>Citrullination</b>    | H3, H4             | Répression                 | <i>Transcription</i>  |
| Arginine (R)             |                    |                            |   |
| <b>Biotinylation</b>     | H3, H4, H2A        | Répression                 | <i>Transcription, Réparation, Condensation</i>              |
| Lysine (K)               |                    |                            |   |
| <b>Isomérisation</b>     | H3 (30, 38)        | Activation/Répression      | <i>Transcription</i>  |
| Proline (P)              |                    |                            |   |

Adapté de Kouzarides, 2007

**Tableau 3** : Les différentes modifications des histones et leurs conséquences fonctionnelles

### **IV.2.2. Les modifications des histones**

Les histones constituant l'octamère du nucléosome sont les cibles d'une multitude de modifications post-traductionnelles. Ces modifications sont localisées principalement sur les extrémités N-terminales des histones (queues des histones). Il a été rapporté que le HFD et les extrémités C-terminales peuvent également être les cibles de modifications (Hyland et al., 2005).

Les résidus des histones peuvent subir plusieurs types de modifications covalentes dont la méthylation (me), l'acétylation (ac), la phosphorylation (ph), l'ubiquitination (ub), la SUMOylation (su), l'ADP-ribosylation (ar) (pour revue, Kouzarides, 2007). Les lysines et arginines sont d'un intérêt particulier car elles peuvent être sujettes à différents types de modifications. De plus, dans le contexte de la méthylation, ces résidus peuvent être mono-, di- ou triméthylés. Les modifications des histones jouent un rôle important dans plusieurs processus dont la transcription, la réplication, la réparation et la compaction de la chromatine (*Cf. Tableau 3*).

Parmi les modifications les mieux étudiées figurent l'acétylation, la méthylation et la phosphorylation. Seules ces trois modifications seront abordées dans ce chapitre. Afin de simplifier la rédaction, les localisations des modifications sont symbolisées : "histone-résidu-modification". Par exemple, la triméthylation de la lysine 4 de l'histone H3 est notée H3K4me3.

#### IV.2.2.1. L'acétylation des histones

Les résidus lysines des histones peuvent être acétylés par des enzymes appelées histone acétyltransférases (HAT) (*Cf. Tableau 4*). Les HAT catalysent le transfert d'un groupement acétyle (CH<sub>3</sub>-CO-) à partir de l'acétyl-Coenzyme A (Roth et al., 2001). L'acétylation annule la charge positive des lysines, ce qui diminue l'affinité des histones pour l'ADN (Wolffe et Pruss, 1996). La chromatine adopte alors une structure plus relâchée qui facilite l'accès de l'ADN aux facteurs de transcription et favorise ainsi l'activation de la transcription. Les HAT sont donc considérées comme des coactivateurs de la transcription. L'acétylation des histones est un processus dynamique et réversible. En effet, les histones peuvent être déacétylées par les enzymes histone désacétylases (HDAC). Les HDAC interviennent essentiellement dans la répression de la transcription.

## INTRODUCTION PARTIE I

Des expériences de Chip on chip et de ChIP-seq réalisées chez plusieurs espèces montrent que les histones acétylées sont localisées au niveau des régions promotrices et des séquences codantes des gènes (Schübeler et al., 2004 ; Bernstein et al., 2005). Ces modifications jouent un rôle important dans la transcription, la réplication et la réparation de l'ADN. La modification H3K56ac est par exemple impliquée dans la réparation, la préservation de la stabilité du génome ainsi que dans la réplication de l'ADN (Maas et al., 2006 ; Han et al., 2007). H4K8ac intervient dans la réplication (Doyon et al., 2006) et H4K16ac joue un rôle dans la décondensation de la chromatine (Shogren-Knaak et al., 2006).

| <b>Famille</b> | <b>Domaines structuraux</b>                            | <b>Enzyme</b>       | <b>Complexes associés</b> | <b>Substrat</b>         | <b>Fonction</b>   |
|----------------|--|---------------------|---------------------------|-------------------------|---|
| GNAT/<br>PCAF  | <i>bromodomaine</i>                                    | Gcn5                | SAGA, ATAC, TFTC          | H3K9, 14, 18, 36<br>H2B | <i>Co-activateur</i>  |
|                |  | PCAF                | STAGA                     | H3K14                   | <i>Co-activateur</i>  |
| MYST           | <i>chromodomaine</i><br><i>PHDfinger</i>               | Tip60               | NuA4                      | H4K5, 8, 12, 16         | <i>Co-activateur, Réparation de l'ADN, Cycle cellulaire</i> |
|                |  | Sas2, MOF           | SAS                       | H4K16                   | <i>Répression / Compensation de dosage</i>                  |
|                |  | Sas3                | NuA3                      | H3K14, 23               | <i>Répression / Compensation de dosage</i>                  |
|                |  | MOZ                 | MOZ                       | H3K14                   | <i>Co-activateur</i>  |
|                |  | HBO1                | HBO1                      | H3/H4                   | <i>Réplication de l'ADN</i>                                 |
| CBP/<br>p300   | <i>Bromodomaine</i><br><i>Motifs riches en Cys/His</i> | CBP, p300<br>Rtt109 | <i>nombreux</i>           | <i>tous histones</i>    | <i>Co-activateur, Réplication et réparation</i>             |

**Tableau 4** : Les différentes familles d'enzymes responsables de l'acétylation des histones

### IV.2.2.2. La méthylation des histones

Les histones peuvent être méthylées par les histone méthyltransférases (HMT) (Cf. Tableau 5). Les HMT catalysent le transfert d'un groupement méthyle (-CH<sub>3</sub>) à partir d'une molécule donneuse vers un résidu lysine ou arginine. Les arginines peuvent être mono- et di-méthylées par des PRMT (*Protein aRginine MethylTransferase*) (Cf. Tableau 6), alors que les lysines peuvent être mono-, di- et tri-méthylées par des protéines contenant un domaine SET (*Su(var)3-9, Enhancer of zest, Trithorax*) ainsi que par la protéine Dot1L (*Disrupter Of Telomere silencing protein 1- Like*) (pour revue, Zhang et Reinberg, 2001).

## INTRODUCTION PARTIE I

| <b>Substrat</b> | <b>Histone méthyltransférase</b>                       | <b>Fonctions</b>  |
|-----------------|--|---|
| H3K4            | SET1, MLL, SET7/9, ASH1L, SMYD3, PRDM9, SETMAR         | Activation  |
| H3K9            | SUV39h1/2, EHMT1/2, SETDB1, ASH1L, ESET, G9a, GLP, RIZ | Répression, hétérochromatinisation<br>Inactivation du Chr X |
| H3K27           | EZH1/2, EHMT2  | Répression, Inactivation Chr X                              |
| H3K36           | NSD1, SETD2/HYPB, SETMAR                               | Elongation transcriptionnelle                               |
| H3K79           | DOT1L  | Activation  |
| H4K20           | SUV4-20H1/2, SET8, NSD1, ASH1L                         | Répression, hétérochromatinisation                          |

**Tableau 5:** Les différentes enzymes responsables de la méthylation des lysines des histones

| <b>PRMT</b> | <b>Classe</b> | <b>Substrat</b>                        | <b>Fonctions</b>   |
|-------------|---------------|--|--|
| PRMT1       | 1             | H4R3                                   | Activation, réparation de l'ADN                          |
| PRMT4/CARM1 | 1             | H3R2, R17, R26, R128, R129, R131, R134 | Activation, différenciation, développement, tumorigénèse |
| PRMT5/JPB1  | 2             | H3R8, H4R3                             | Répression   |
| PRMT6       | 1             | H3R2, R17, R26, R128,                  | Répression, réparation de l'ADN                          |
| PRMT7       | 2             | H2A, H4R3                              | Empreinte parentale des cellules germinales mâles        |
| PRMT8       | 1             | H4R3                                   | Indéterminés   |
| PRMT9       | 2             | H2A, H4                                | Indéterminés   |

**Tableau 6:** Les différentes enzymes responsables de la méthylation des arginines des histones

De manière analogue à l'acétylation, la méthylation des histones est une réaction réversible. La déméthylation des histones est catalysée par les histones déméthylases (HDM). Par exemple, la protéine LSD1 (*Lysine Specific Demethylase 1*) déméthyle spécifiquement les lysines 4 et 9 de l'histone H3 lorsqu'elles sont mono- ou diméthylées.

La méthylation des histones joue un rôle complexe car elle peut être associée à l'activation ou à la répression de la transcription, et ce en fonction du résidu méthylé et du degré de méthylation. Par exemple, les modifications H3K4me3, H3K36me3 et H3K79me3 sont associées à des gènes transcrits et sont donc considérées comme des marques d'activation. En revanche, H3K9me3 et H3K27me3 sont associées à des gènes réprimés et sont donc considérées comme des marques de répression (pour revue, Martin et Zang, 2005).

## INTRODUCTION PARTIE I

Différents degrés de méthylation sur un même résidu peuvent avoir des rôles différents. Par exemple, H3K4me3 est particulièrement enrichie dans les régions promotrices actives puis sa concentration diminue au fur et à mesure le long de la région codante, alors que H3K4me2 est faiblement enrichie au niveau des promoteurs puis augmente dans la région codante. H3K4me serait quant à lui enrichi à la fin de la phase codante des gènes et dans les régions régulatrices distales "Enhancers" (Zhang et al., 2009).

La modification H3K9me3 favorise en général la formation d'une structure chromatinienne condensée, donc inactive. Cette modification stimulerait l'activité des ADN méthyltransférases (Smallwood et al., 2007) et jouerait un rôle important dans le recrutement de HP1 (*Heterochromatin Protein 1*). Cependant, H3K9me3 a également été observée au niveau de gènes activement transcrits dans des cellules de mammifères (Vakoc et al., 2005). Cela suggère que l'effet de cette modification peut dépendre de sa position sur le gène (région promotrice ou codante) ou de son association avec une autre modification.

Les modifications H3K9me2, H3K27me3 et H4K20me3 sont retrouvées au niveau des gènes soumis à l'empreinte parentale (Fournier et al., 2002 ; Delaval et al., 2007) et servent au maintien de l'inactivation du chromosome X (Fang et al., 2004). H4K20me3 est également impliquée dans la progression du cycle cellulaire et la réparation de l'ADN (Martin et Zang, 2005).

### IV.2.2.3. La phosphorylation des histones

La phosphorylation des histones est catalysée par les histone kinases (HKs) et consiste en une réaction de transfert d'un groupement phosphate à partir d'une molécule d'ATP vers un résidu sérine ou thréonine. La phosphorylation des histones est importante pour l'activation de la transcription, la réparation, la progression du cycle cellulaire et l'apoptose.

Le niveau global de phosphorylation des histones varie de manière très marquée au cours du cycle cellulaire (Bradbury et al., 1973). La phosphorylation de l'histone H3 par la kinase mitotique Aurora B atteint un pic lors de la mitose (Gurley et al., 1978). Cette phosphorylation affecte le résidu H3S10 et semble être liée de manière très directe au processus de condensation des chromosomes (Wei et al., 1999). Il a été démontré que la modification H3S10ph est

## INTRODUCTION PARTIE I

nécessaire pour la dissociation de HP1 de la chromatine pendant la métaphase (Fischle et al., 2005).

La phosphorylation des histones est également importante pour la régulation de la transcription. Par exemple, la modification H3T11ph est associée à l'activation des gènes dépendants des récepteurs aux androgènes (Metzger et al., 2008). Cette modification intervient aussi au cours de la réparation, car la thréonine 11 est déphosphorylée au niveau des lésions de l'ADN (Shimada et al., 2008).

La phosphorylation est également impliquée dans l'apoptose car la phosphorylation de la sérine 10 de l'histone H2B chez la levure et de la sérine 14 de l'histone H2B chez les mammifères sont importantes pour l'induction de l'apoptose (Cheung et al., 2003 ; Ahn et al., 2006).

### **IV.2.3. La reconnaissance des modifications des histones**

Les modifications post-traductionnelles des histones peuvent remplir leur fonction du fait de leur seule présence sur les histones, mais également grâce à l'intervention de protéines intermédiaires capables de reconnaître et de « lire » les marques épigénétiques présentes sur la chromatine. Ces protéines possèdent des domaines spécifiques capables de reconnaître des modifications bien précises et recrutent d'autres facteurs de régulation de la transcription ou de remodelage de la chromatine, qui vont dicter l'activité transcriptionnelle (pour revues, Featherstone, 2002 ; Martin et Zhang, 2005 ; Taverna et al., 2007).

A titre d'exemple, les bromodomains reconnaissent les lysines acétylées, les chromodomains présents dans les protéines HP1 et la protéine Polycomb (Pc) reconnaissent H3K9me3 et H3K27me3 respectivement, alors que les répétitions WD40 de l'activateur transcriptionnel WDR5 (WD40 Repeat protein 5) reconnaissent H3R2me et H3K4me (Wysocka et al., 2005 ; Couture et al., 2006).

### **IV.2.4. Les complexes de remodelage de la chromatine**

En plus des modifications post-traductionnelles des histones, des complexes de remodelage de la structure de la chromatine participent aussi à la régulation de la transcription.

## INTRODUCTION PARTIE I

Ces complexes utilisent l'énergie libérée par l'hydrolyse de l'ATP pour déplacer, insérer ou restructurer les nucléosomes, de façon à faciliter ou réduire l'accessibilité de l'ADN à des facteurs de transcription, de réplication ou de réparation (pour revues, Mohrmann et Verrijzer, 2005 ; Saha et al., 2006). Ils sont caractérisés par la présence d'une sous-unité catalytique ATPase et peuvent être classés en quatre familles selon la présence d'autres domaines structuraux caractéristiques : les familles SWI/SNF, ISWI, NuRD/Mi2 et INO (pour revue, Tsukiyama, 2002) (Cf. Tableau 7)

| <b>Famille</b> | <b>Complexes</b>    | <b>Fonctions</b>  |
|----------------|---------------------|---|
| SWI/SNF        | SWI/SNF             | <i>Elongation, Différenciation, Développement</i>           |
| ISWI           | WCRF,<br>CHRAC, RSF | <i>Activation/Répression<br/>Espacement des nucléosomes</i> |
| Mi-2/CHD       | Nurd                | <i>Répression / Remodelage chromatine</i>                   |
| INO            | SWR1                | <i>Réparation ADN</i>                                       |

**Tableau 7** : Les différentes familles de cofacteurs ATP-dépendants

## CHAPITRE 2 : TBP

La TBP est la protéine responsable de la reconnaissance de l'élément TATA du promoteur. Elle possède un domaine N-terminal variable et un domaine C-terminal hautement conservé au cours de l'évolution. Initialement, l'activité de TBP était attribuée au facteur TFIID qui portait également le nom « TATA-box factor ». En 1988, deux équipes ont montré qu'une activité présente dans les extraits de levures pouvait remplacer le TFIID de mammifère dans la transcription *in vitro* (Buratowski et al., 1988 ; Cavallini et al., 1988). La purification de la protéine portant cette activité a permis le clonage de la TBP de levure puis le clonage de ses orthologues chez plusieurs espèces mammifères (Hernandez 1993 ; Hoey et al., 1993 ; Peterson et al., 1990). Cependant, contrairement à TFIID (750 kDA), la TBP recombinante (38 kDA) était incapable de supporter la transcription activée (Pugh and Tjian, 1990, 1991). Il a été démontré par la suite que TFIID était un complexe multiprotéique composé de TBP et de plusieurs protéines associées nommées TAFs (*TBP-Associated Factors*).

D'autres complexes contenant TBP ont été identifiés par la suite. En effet, TBP est une sous-unité commune à plusieurs complexes protéiques impliqués dans la transcription. Parmi ces

## INTRODUCTION PARTIE I

complexes se trouvent : SL1, B-TFIID et TFIIB impliqués dans la transcription par les Pol I, II et III respectivement. TBP est donc impliquée dans la transcription par les trois ARN polymérase chez les eucaryotes. Elle est donc considérée comme un facteur universel de la transcription.

Cependant, ce caractère universel a été remis en question après la découverte des protéines apparentées à TBP. Elles sont au nombre de trois : TRF1 (*TBP Related Factor 1*), TRF2 également appelé TLF (*TBP Like Factor*) et TRF3. La similarité entre ces protéines est limitée au domaine C-terminal conservé. TRF1 est exprimé uniquement chez la drosophile pendant l'embryogenèse puis, de façon spécifique, dans le système nerveux central et les cellules germinales chez l'adulte (pour revue, Davidson et al., 2004). Il agit comme un facteur de transcription spécifique pour l'expression d'un sous-ensemble de gènes dont Tudor. TRF1 dirige la transcription de Tudor à partir d'un deuxième site d'initiation en amont de celui dirigé par TBP (Holmes et al., 2000). TRF2 est essentiel à l'embryogénèse chez *Caenorhabditis elegans*, *Xenopus laevis* et le poisson zèbre où il est responsable de l'expression d'un ensemble de gènes (pour revue, Davidson et al., 2004). Chez la souris, TRF2 n'est pas requis pendant l'embryogenèse, mais joue un rôle plus spécifique dans la spermatogénèse (Martianov et al., 2002). Enfin, TRF3 également appelé TBP2 présente 92% de similarité avec le domaine C-terminal conservé de TBP, il interagit avec TFIIA et TFIIB et peut diriger la transcription par la Pol II *in vitro* (Bartfai et al. 2004). Durant l'ovogénèse chez la souris, TRF3 est exprimée uniquement dans les ovocytes. L'inactivation de TRF3 provoque la stérilité en altérant la folliculogénèse. Cette altération est due à une désorganisation de la chromatine et un défaut d'expression de gènes spécifiques aux ovocytes (Gazdag et al., 2009)

Ce chapitre abordera la structure de TBP et son activité dans le contexte de la transcription par la Pol II.

### **I. Organisation structurale de TBP**

La comparaison des séquences protéiques de TBP chez plusieurs espèces a mis en évidence l'existence de deux régions : une région N-terminale non conservée et une région C-terminale très conservée au cours de l'évolution (Hernandez, 1993). Cf. *Figure 19a*



### **I.1. Le domaine N-terminal de TBP**

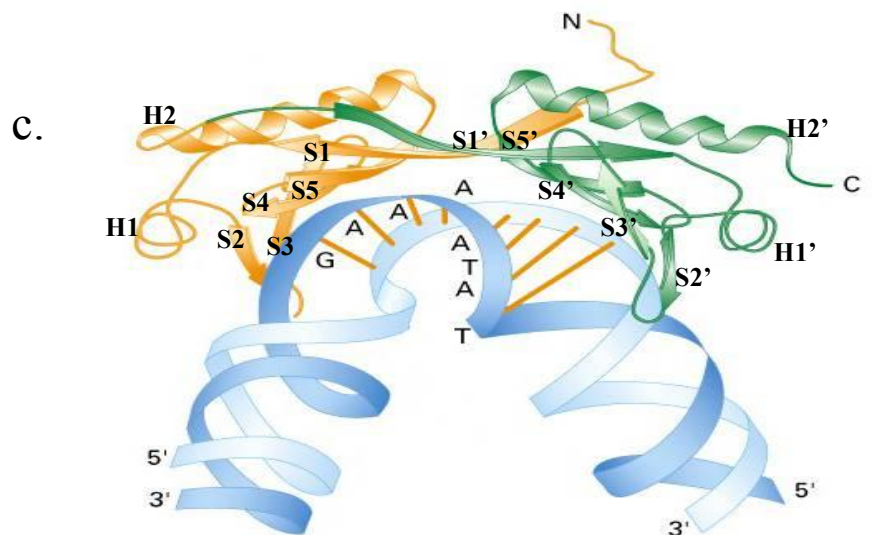
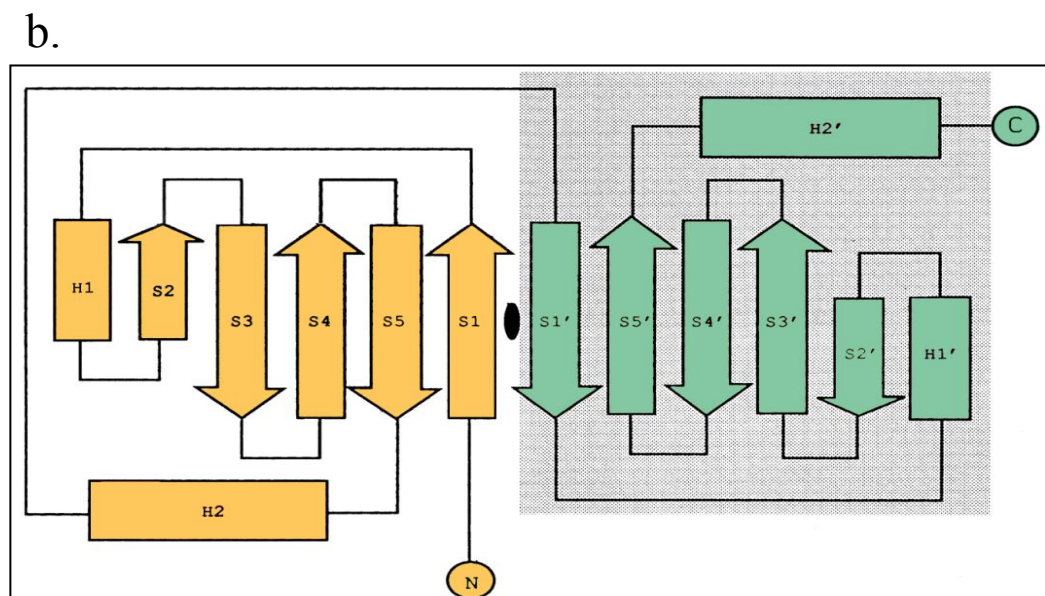
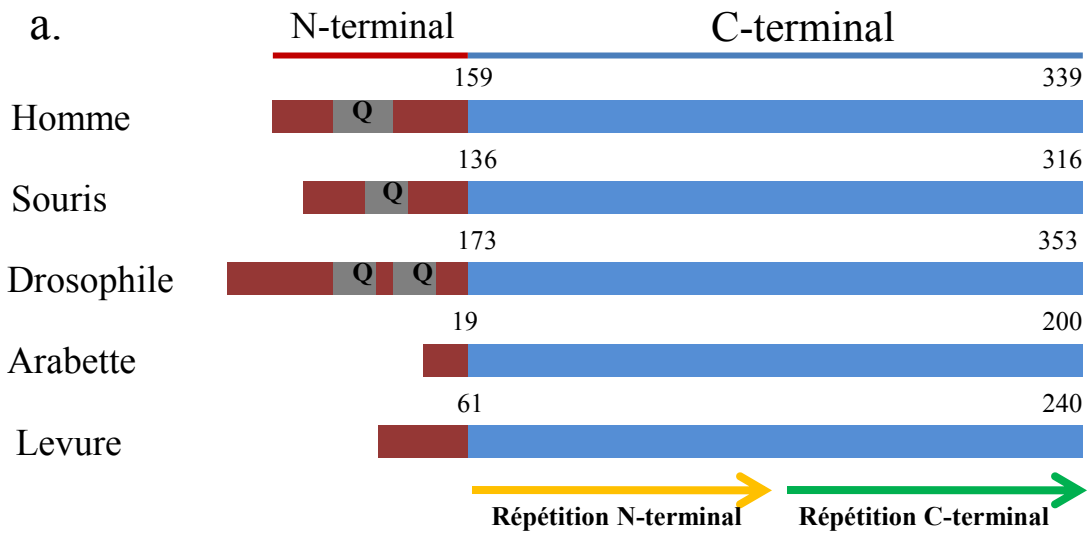
Le domaine N-terminal de TBP présente une importante variabilité de séquence et de taille entre les espèces. Ce domaine ne contient que 19 résidus chez *Arabidopsis thaliana* mais atteint 173 résidus chez la drosophile. Chez l'homme, la souris et la drosophile, ce domaine se caractérise par la présence d'une région polyglutamine. Chez l'homme, il existe plusieurs allèles codant pour des régions polyglutamines de longueurs différentes. Des expansions anormales des répétitions polyglutamines de plusieurs protéines sont souvent associées avec des maladies neurodégénératives. Dans le cas de TBP, une expansion de cette région est associée à une forme d'ataxie spinocérébelleuse (Nolte et al., 2010).

Les données concernant la fonction du domaine N-terminal de TBP sont peu nombreuses et contradictoires. Certaines études démontrent qu'il n'est pas indispensable pour la transcription *in vitro* (Keveaney et al. 1993 ; Tansey et al. 1994; Zhou et al. 1993), alors qu'une seule étude suggère qu'il joue un rôle dans la transcription des promoteurs possédant un élément TATA (Lescure et al., 1994). Chez la levure, la délétion de la région N-terminale est létale seulement en combinaison avec d'autres mutations dans le domaine conservé affectant la fixation à l'ADN (Lee and Struhl, 2001). D'autres études ont montré que le domaine N-terminal est essentiel pour l'interaction entre TBP et SNAPc. Cette interaction augmenterait la transcription du gène U6 par l'ARN Pol III (Mittal and Hernandez, 1997). Malgré ces résultats, le rôle du domaine N-terminal dans la transcription reste énigmatique.

La délétion du domaine N-terminal de TBP par recombinaison homologue chez la souris provoque un rejet immunologique maternel du fœtus (Hobbs et al., 2002). La majorité des embryons TBP<sup>ΔN/ΔN</sup> meurent *in utero* du fait d'une anomalie du placenta impliquant un défaut de l'établissement de la tolérance foeto-maternelle. Le mécanisme moléculaire de cet effet inhibiteur du rejet embryonnaire n'est pas clairement établi.

### **II.2. Le domaine C-terminal de TBP**

Le domaine C-terminal de TBP est très conservé entre les espèces. Il comprend deux répétitions directes imparfaites riches en acides aminés basiques. *In vitro*, ce domaine se lie à l'ADN et est suffisant pour initier la transcription en combinaison avec les autres GTFs et la Pol



**Figure 19:**

- a) Conservation de TBP entre les espèces
- b) Topologie du domaine C-terminal de TBP
- c) Structure du complexe TBP-ADN

## INTRODUCTION PARTIE I

II (Zhou et al., 1993). Ce domaine C-terminal interagit également avec TFIIA, TFIIB et avec le domaine CTD non phosphorylé de la Pol II (Usheva et al., 1992).

La structure cristallographique du domaine C-terminal de plusieurs orthologues de TBP a été résolue. La première structure résolue était celle de la TBP d'*Arabidopsis thaliana* (atTBP) (Nikolov et al., 1992). Suivirent celles de la levure (Chasman et al., 1993), l'homme (Nikolov et al., 1996) et l'archéobactérie *Sulfolobus acidocaldarius* (Koike et al., 2004).

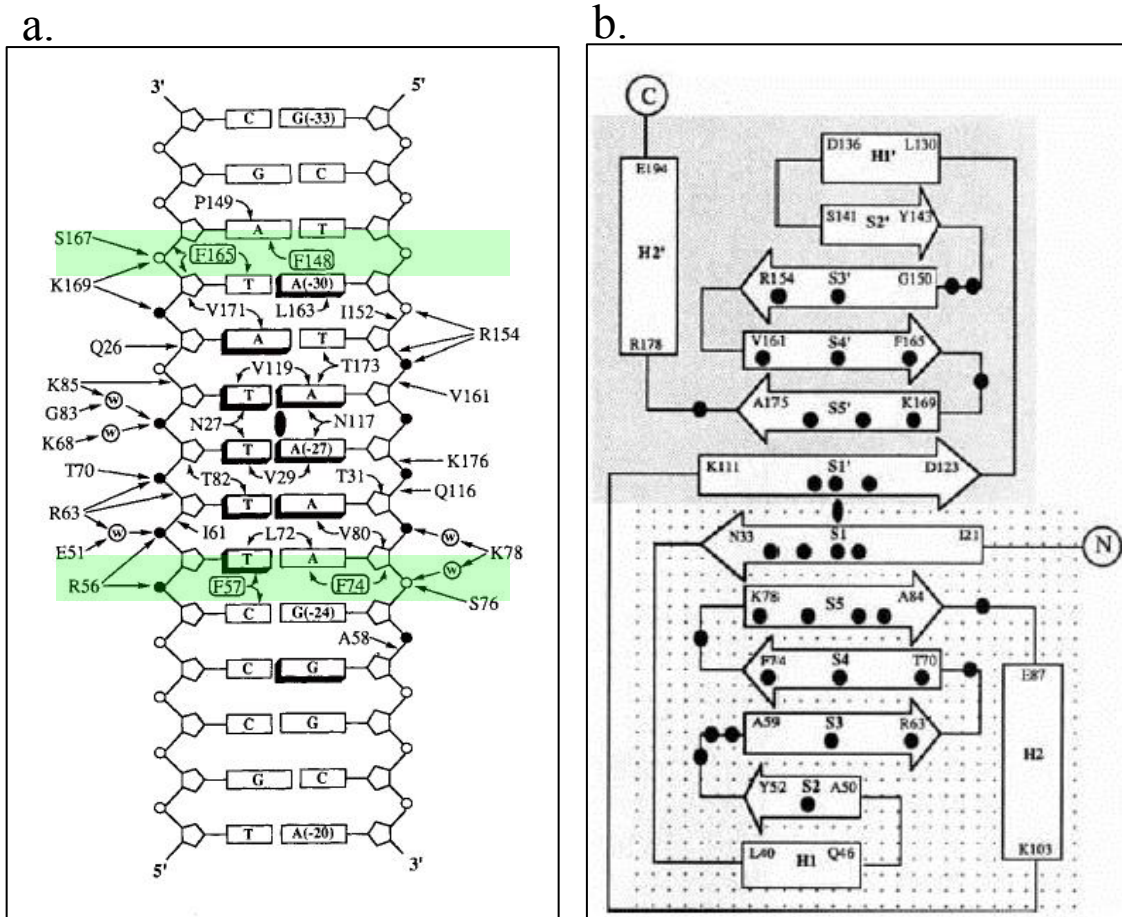
La conservation de la séquence primaire de la région C-terminale de TBP se traduit également par une conservation structurale. En effet, tous les orthologues de TBP adoptent une forme de "selle à cheval" (Cf. *Figure 19b, 19c*) avec une surface concave hydrophobe et une surface convexe hydrophile. La surface concave est constituée de 10 feuillets  $\beta$  (S) antiparallèles alors que la surface convexe est composée en partie par 4 hélices  $\alpha$  (H). Chaque répétition directe est composée de deux hélices  $\alpha$  et de cinq feuillets  $\beta$  antiparallèles. Une boucle en forme d'étrier est présente à chaque extrémité du grand axe de TBP, lui donnant l'apparence d'une selle à cheval. La répétition N-terminale dite aussi "première répétition" se compose des éléments S1-H1-S2-S3-S4-S5-H2. La répétition C-terminale dite aussi "deuxième répétition" se compose des éléments S1'-H1'-S2'-S3'-S4'-S5'-H2'.

### **II.3. Cartographie des sites d'interaction de TBP avec ses partenaires**

La détermination des sites d'interaction de TBP avec ses partenaires est basée sur deux approches majeures : la cristallographie et la mutagénèse. Les structures cristallographiques de plusieurs complexes contenant TBP ont été résolues, ce qui a permis de déterminer les propriétés des interactions entre TBP et ses partenaires. Les études par mutagénèse ont également contribué à définir les surfaces et les résidus importants pour ces interactions.

#### **II.3.1. Structure cristallographique du complexe TBP-ADN**

Les structures cristallographiques de plusieurs orthologues de TBP en complexe avec l'élément TATA ont été déterminées avec des résolutions allant de 1,9 à 2,5 Å (Kim J.L et al., 1993 ; Kim Y et al., 1993 ; Nikolov et al., 1996). Ces études montrent que la surface concave hydrophobe de TBP interagit avec l'élément TATA au niveau du petit sillon de la double hélice



D'après Kim et al., 1993

**Figure 20:** Caractéristiques structurales de l'interaction entre TBP et l'élément TATA

- a) Deux paires de phénylalanines particulièrement conservées interagissent avec le petit sillon de l'ADN. Dans la TBP d'*Arabidopsis thaliana*, les résidus F57 et F74 de l'étrier N-terminal (première répétition) s'insèrent entre les bases A7 et G8 de l'élément TATA, alors que les résidus F148 et F165 de l'étrier C-terminal (seconde répétition) s'insèrent entre les bases T1 et A2. Ces interactions sont indiqués par les rectangles verts
- b) Les résidus de TBP qui contactent l'ADN sont indiqués (●)

## INTRODUCTION PARTIE I

d'ADN et entraîne une distorsion de l'ADN avec une courbure de 80° orientée vers le grand sillon. Cette interaction est plutôt étonnante car la majorité des facteurs qui interagissent avec une séquence spécifique d'ADN établissent des contacts avec le grand sillon. Cf. *Figure 19c ; figure 20a*

Deux paires de phénylalanines particulièrement conservées interagissent notamment avec le petit sillon d'ADN. Dans la TBP d'*Arabidopsis thaliana*, les résidus F57 et F74 de l'étrier N-terminal (première répétition) s'insèrent entre les bases A7 et G8 de l'élément TATA, alors que les résidus F148 et F165 de l'étrier C-terminal (seconde répétition) s'insèrent entre les bases T1 et A2. Cf. *Figure 20a*

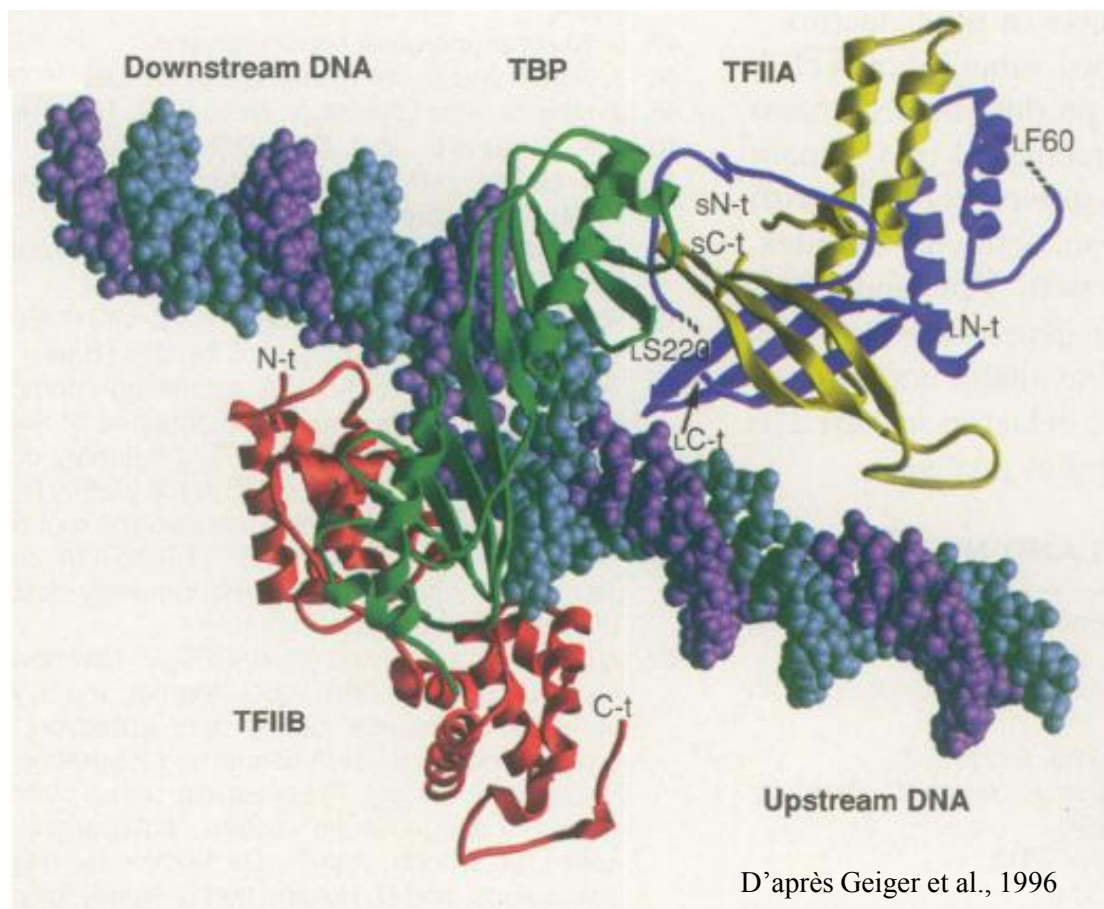
Malgré la symétrie quasi-parfaite du domaine C-terminal de TBP, ce dernier interagit avec l'élément TATA de manière directionnelle. En effet, la première et la seconde répétition interagissent respectivement avec l'élément TATA en 3' et en 5' (Kim et al., 1993).

### **II.3.2. Structure cristallographique du complexe TBP-TFIIB-ADN**

La structure cristallographique du complexe TFIIB-TBP-ADN a également été déterminée à une résolution de 2,7 Å (Nikolov et al., 1995). Dans cette structure cristallographique, TFIIB établit des contacts directs avec TBP et avec l'ADN. TFIIB lie TBP en trois sites distincts : au niveau de l'étrier S2'-S3', de l'hélice H1' et de la lysine 197 qui se situe dans le domaine N-terminal non conservé. TFIIB lie également l'ADN et agit ainsi comme une pince pour stabiliser le complexe TBP-ADN. La formation du complexe TFIIB-TBP-DNA provoque un léger changement dans la conformation de TBP. Ce changement se traduit par une modeste rotation (twist) intramoléculaire des deux répétitions de TBP l'une par rapport à l'autre. Cf. *Figure 21*

### **II.3.3. Structure cristallographique du complexe TBP-TFIIA-ADN**

La structure cristallographique du complexe TFIIA-TBP-ADN montre que TFIIA lie la face latérale de la répétition N-terminale de TBP au niveau des feuillets S2 et S3 et de l'hélice H1, donc du côté opposé à celui lié par TFIIB (Geiger et al., 1996 ; Tan et al., 1996). TFIIA lie également l'ADN en amont de l'élément TATA. L'interaction entre TFIIA et le complexe TBP-ADN provoque un repositionnement de TBP sur l'ADN. En effet, dans ce complexe TBP se



**Figure 21** : Modélisation du complexe TFIIB-TBP-ADN-TFIIA



**Figure 22** : L'interaction entre TFIIA et le complexe TBP-ADN provoque un repositionnement de TBP sur l'ADN. En rouge, les nucléotides liés par TBP dans le complexe TBP-ADN. Le carré bleu indique les nucléotides liés par TBP dans le complexe TBP-ADN-TFIIA

positionne deux bases en amont par rapport à sa position initiale dans le complexe TBP-ADN. Cf. *Figure 22*

### **II.3.4. Structure cristallographique du complexe NC2-TBP-ADN**

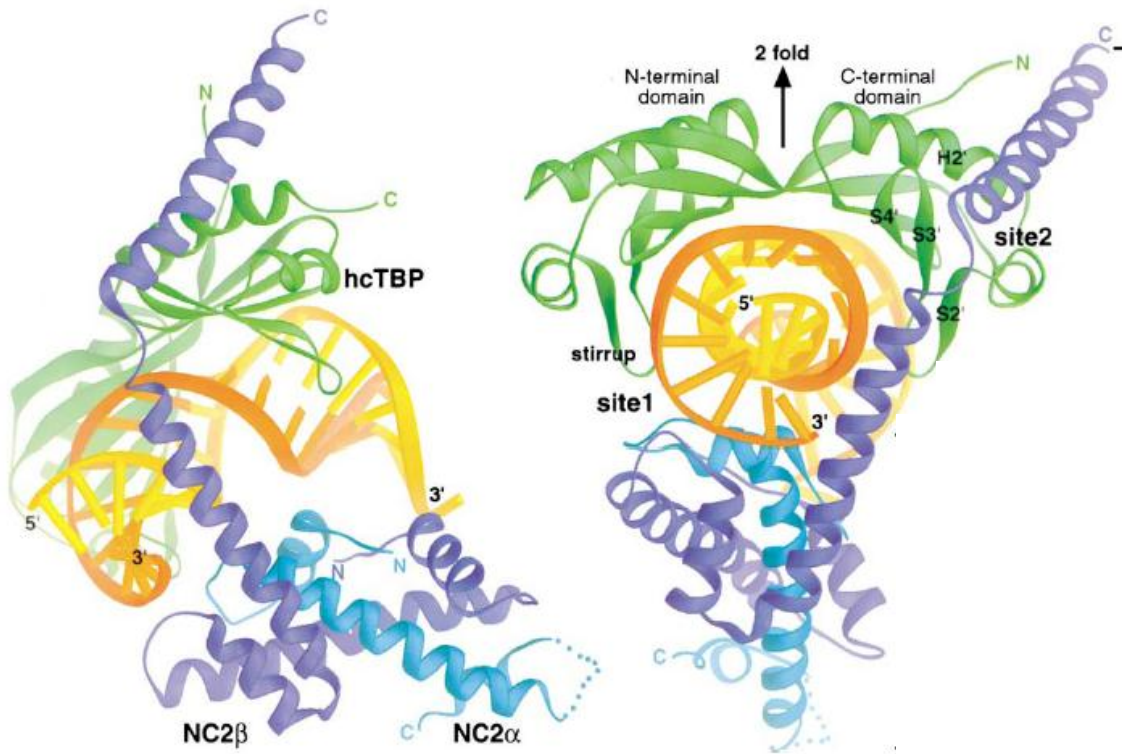
L'hétérodimère NC2 (NC2 $\alpha$ /NC2 $\beta$ ) agit comme une pince moléculaire qui lie les faces supérieure et inférieure du complexe TBP-ADN. La sous-unité NC2 $\alpha$  lie TBP au niveau de l'étrier N-terminal alors que NC2 $\beta$  recouvre la face convexe de TBP d'une longue hélice  $\alpha$  qui interagit avec les éléments H2', S3' et S2' (Kamada et al., 2001). Cf. *Figure 23*

### **II.3.5. Analyse de TBP par mutagenèse**

Plusieurs études basées sur la mutagenèse systématique des résidus du domaine C-terminal de TBP ont été réalisées par plusieurs équipes. Ces études étant réalisées sur des orthologues différents de TBP et dans des systèmes variés, ont produit des résultats parfois contradictoires. Il est donc difficile de faire une synthèse de l'ensemble de ces études. Ce chapitre résumera les études les plus pertinentes ayant utilisé la mutagenèse pour analyser la relation structure/fonction de TBP.

#### **II.3.5.1. La variante TBP spm3**

Dans le cadre de l'étude par mutagenèse de TBP, il est important de citer un mutant particulier appelé spm3 (*altered SPecificity Mutant 3*). Chez la levure, où il a été caractérisé pour la première fois, ce mutant porte la triple substitution yI194F, yV203T et yL205V. Ces résidus sont situés sur la surface concave de TBP. Cette triple mutation permet à la variante TBP spm3 de lier l'élément TATA canonique mais également la variante TGTA AAA (Strubin and Struhl, 1992). TBP spm3 est fréquemment utilisée dans des études de transcription *in vivo*, grâce à l'utilisation d'un gène rapporteur, placé sous le contrôle d'un promoteur dont l'élément TATA a été remplacé par la séquence TGTA AAA. La transcription de ce gène rapporteur ne pourra se faire que *via* un TFIID néoformé ayant intégré la TBP spm3, étant donné que le promoteur modifié n'est alors plus reconnu par la TBP endogène (Lavigne et al., 1999). L'introduction de mutations supplémentaires dans la TBP spm3 permet d'évaluer l'impact de ces mutations sur la transcription activée dans un système cellulaire. La variante spm3 chez l'homme porte les mutations I292F, V301T et L303V.



D'après Kamada et al., 1996

**Figure 23** : Structure du complexe TBP-ADN-NC2

Les sites d'interaction entre TBP et NC2 sont indiqués (sites 1 et 2)



## INTRODUCTION PARTIE I

### II.3.5.2. Mutagenèse systématique des résidus exposés au solvant dans le domaine C-terminal de TBP

Dans l'étude réalisée par O.Bryant et al., les résidus exposés au solvant sur la surface convexe et les faces latérales de hTBP ont été systématiquement mutés (O.Bryant et al., 1996). Les acides aminés chargés ont été remplacés par des acides aminés de charge opposée, alors que les résidus non chargés ont été remplacés par des résidus plus volumineux (l'acide glutamique ou l'arginine) afin de générer un encombrement stérique. Une librairie de 89 mutants a été ainsi générée. Ces mutants ont été testés pour leur capacité à interagir avec l'ADN en présence ou en absence de TFIIA et/ou TFIIB.

La capacité de ces mutants à supporter la transcription activée a également été évaluée. Pour cela, les auteurs ont utilisé la stratégie suivante :

Les mutations étudiées ont été introduites dans la variante hTBP *spm3*, qui porte déjà la triple mutation I292F, V301T et L303V. Des cellules COS ont été transfectées par les vecteurs d'expression codant pour ces protéines mutées et par un vecteur codant pour le gène rapporteur de la luciférase. Ce dernier a été placé sous le contrôle du promoteur du gène *c-fos* ou du gène *RARβ2* dont les éléments TATA ont été remplacés par TGTA, ce qui implique que le gène rapporteur ne peut être transcrit que via la TBP *spm3* et non la TBP sauvage endogène. Ces deux promoteurs contenaient également la séquence de liaison de GAL4 qui permet aux protéines de fusion GAL4-VP16 ou GAL4-E1A d'activer la transcription du gène de la luciférase. L'activité de la luciférase reflétait donc la capacité des mutants à supporter la transcription activée. La capacité de ces mutants à supporter la transcription basale a été évaluée *in vitro*. L'ensemble de ces résultats est présenté dans le tableau 8.

Parmi les 89 mutants étudiés, 13 sont incapables d'interagir avec l'ADN en présence ou en absence de TFIIA et/ou TFIIB, malgré le fait que ces mutations ne soient pas localisées sur la face concave de TBP engagée dans l'interaction avec l'ADN. Les auteurs ont conclu que ces mutations altèrent la structure de TBP et la rendent par conséquent incapable de reconnaître l'élément TATA et de contacter ses partenaires. La majorité de ces mutants se caractérise par une capacité réduite voire nulle à initier la transcription basale et activée.

## INTRODUCTION PARTIE I

Cependant, le mutant E227R qui est incapable d'interagir avec l'élément TATA dans les expériences d'EMSA (*Electrophoretic Mobility Shift Assay*) même en présence de TFIIA et TFIIB, peut initier la transcription à partir du promoteur de RAR $\beta$ 2 en présence de Gal4-VP16. Cet exemple montre que les résultats obtenus *in vitro* ne reproduisent pas toujours l'effet réel des mutations *in vivo* et qu'une fonction altérée *in vitro* peut être restaurée *in vivo*.

Parmi les 75 mutants restants, 13 présentent une capacité réduite voire nulle à supporter la transcription. Parmi eux, les mutants A184E, N189E, E191R et R205E sont déficients dans l'interaction avec TFIIA. Cela peut expliquer leur incapacité à supporter la transcription. De même les mutations E284R et L287E affectent l'interaction avec l'ADN uniquement en présence de TFIIB. Le mutant F250E affecte l'interaction avec TFIIA et est inapte à supporter la transcription, mais il est capable de former un complexe TBP-TFIIA-ADN en EMSA. L'ensemble de ces données démontre la complexité des interactions entre TBP et ses partenaires.

Il est important de noter que lors de cette étude, les auteurs n'ont pas pris en considération la possibilité que les effets observés ne soient pas dus uniquement aux mutations étudiées mais à la combinaison de ces dernières avec la triple mutation *spm3*. A titre d'exemple, le mutant R188E dans un contexte *spm3* est incapable d'interagir avec l'ADN et de supporter la transcription, alors que dans une autre étude (Klejman et al., 2005), cette même mutation introduite dans une hTBP sauvage, non *spm3*, peut interagir avec l'ADN.

La mutagénèse systématique a également été utilisée par Tang et al. pour identifier les aminoacides du domaine C-terminal de TBP nécessaires pour son interaction avec TFIIA, TFIIB, TFIIF, TFIIE et la Pol II. Dans cette étude, les résidus du domaine C-terminal de TBP ont été systématiquement remplacés par une alanine (*Alanine scanning*), puis la capacité de ces mutants à lier les facteurs cités plus haut a été évaluée (Tang et al., 1996). Étonnamment, les expériences d'alanine scanning révèlent que peu de résidus sont essentiels pour les interactions entre TBP et ses partenaires. Selon ces résultats, seuls les résidus E228 et R235 sont décrits comme essentiels pour l'interaction avec TFIIA. De même, seuls les résidus E284, E286 et L287 sont critiques pour l'interaction avec TFIIB. Un seul résidu, E206 est essentiel pour l'interaction avec la Pol II et aucune mutation ne présente un effet significatif sur l'interaction de TBP avec TFIIF et TFIIE.

1 2 3 4 5 6 7 8 9 10 11 12

|        |       | c-fos<br>GAL4-E1A | c-fos<br>GAL4-VP16 | β-retinoic AF<br>GAL4-E1A | β-retinoic AF<br>GAL4-VP16 | <i>in vitro</i><br>Txn | TATA | TBP +<br>TFIIA | TBP +<br>TFIIB | TBP +<br>TFIIA&B | TFIIA<br>column |
|--------|-------|-------------------|--------------------|---------------------------|----------------------------|------------------------|------|----------------|----------------|------------------|-----------------|
|        | spm3  | 100               | 100                | 100                       | 100                        | 100                    | +    | +              | +              | +                | 100             |
|        | S156E | 69±23             | 79±17              | 85                        |                            | 29±6                   | +    | +              | +              | +                | 119             |
|        | E157R | 26±6              | 35±20              | 60±19                     | 118±31                     | 78±6                   | +    | +              | +              | +                | 83              |
|        | S158E | 64±30             | 51±11              | 65±17                     | 104±18                     | 112                    | +    | +              | +              | +                | 218             |
|        | S159E | 117±71            | 115±25             | 91±25                     | 121±26                     | 69±17                  | +    | +              | +              | +                | 55              |
|        | I161R | 29±14             | 22±27              | 39±15                     | 47±1                       | 16±22                  | -    | +/-            | +/-            | +/-              | 38±7            |
|        | V162R | 105±24            | 89±35              | 92±26                     | 126±54                     | 78±44                  | +    | +              | +              | +                | 49              |
| S1     | L165E | 7±4               | 4±5                | 18±8                      | -5±10                      | -7±3                   | -    | -              | -              | -                | 22±15           |
| S1     | N173E | 52±13             | 58±23              | 94±35                     | 147±01                     | 44                     | +    | +              | +              | +                | 58              |
|        | G175R | 18±11             | 23±9               | 41±18                     | 78±14                      | 19±12                  | -    | +              | +              | +                | 67±19           |
|        | C176R | 9±4               | 6±7                | 19±5                      | 28±7                       | 15±15                  | -    | +              | +              | +                | 23±9            |
|        | K177E | 72±35             | 123±42             | 94±14                     | 102±34                     | 68±9                   | +    | +              | +              | +                | 81±33           |
|        | D179R | 60±13             | 100±45             | 54±8                      | 102±10                     | 2±3                    | -    | +/-            | -              | +/-              | 64              |
| H1     | K181E | 39±4              | 56±12              | 57±9                      | 118±32                     | 118±17                 | +    | +              | +              | +                | 69              |
| H1     | A184E | 1±3               | 5±4                | 4±5                       | -2±12                      | -5±9                   | +/-  | -              | +              | +                | 21±11           |
| H1     | L185E | 67±17             | 77±12              | 104±31                    | 149±14                     | 114                    | +    | +              | +              | +                | 135±75          |
| H1     | R186E | 160               | 126±23             | 105±19                    | 113±1                      | 75±42                  | +    | +              | +              | +                | 49              |
|        | R188E | -1±2              | 1±5                | -1±3                      | -5±9                       | 3±7                    | -    | -              | -              | -                | 28±27           |
|        | N189E | -1±0              | -4±2               | -2±1                      | -4±2                       | 46±8                   | -    | -              | +'             | +'               | 19±8            |
| S2     | A190E | 6±2               | 14±2               | 9±6                       | 14±19                      | 0±8                    | -    | -              | -              | -                | 15±4            |
| S2     | E191R | 1±1               | 2±2                | -3±3                      | -13±4                      | 68±34                  | +    | -              | +              | +                | 31±1            |
| S2     | Y192E | 20±2              | 26±26              | 11±9                      | 7±14                       | -5±6                   | -    | -              | -              | -                | 15±2            |
| étrier | N193R | 26±26             | 53±12              | 50±21                     | 145±51                     | 47±14                  | +    | +              | +              | +                | 71              |
| étrier | P194E | 38±15             | 48±23              | 69±31                     | 103±2                      | 67±23                  | +    | +              | +              | +                | 82              |
| étrier | K195E | 72±6              | 66±26              | 99±33                     | 85±3                       | 88±8                   | +    | +              | +              | +                | 120±12          |
| S3     | I201D | 2±1               | 3±4                | 7±7                       | -5±3                       | 0±18                   | -    | +/-            | -              | -                | 21±3            |
| S3     | R203E | -3±2              | -2±6               | -5±2                      | -6±5                       | 1±6                    | -    | -              | -              | +/-              | 74±15           |
|        | R205E | 1±2               | 5±2                | 1±4                       | 5±3                        | 98±36                  | -    | -              | +'             | +'               | 16±6            |
|        | E206R | 139±45            | 119±36             | 124±16                    | 75±11                      | 88±47                  | +    | +              | +              | +                | 141             |
|        | R208E | 78±36             | 72±24              | 76±13                     | 114±22                     | 123±11                 | -    | +              | +/-            | +                | 63              |
|        | S215E | 66±26             | 73±30              | 116±36                    | 197±50                     | 58±30                  | +    | +              | +              | +                | 81              |
|        | S216R | 36±12             | 60±10              | 68±31                     | 151±48                     | 6±10                   | -    | +/-            | -              | +/-              | 59              |
|        | G223R | 9±5               | 2±2                | 11                        |                            | 10±6                   | -    | -              | -              | -                |                 |
| H2     | E227R | 39±19             | 50±31              | 37±6                      | 70±11                      | 0±9                    | -    | -              | -              | -                | 19±8            |
| H2     | E228A | 78±36             | 139±45             | 139±45                    | 139±45                     | 0±10                   | -    | +              | +/-            | +                | 45              |
| H2     | R231E | 19±9              | 25±22              | 23±8                      | 102±31                     | 87±23                  | +    | +              | +'             | +'               | 26±9            |
| H2     | L232E | 64±6              | 50±7               | 62±4                      | 100±19                     | 30±21                  | -    | +              | +              | +                | 81              |
| H2     | R235E | 1±0               | 3±2                | 6±4                       | 18±2                       | 127±22                 | -    | +              | +'             | +'               | 27±25           |
| H2     | K236Q | 37±9              | 103±17             | 49±13                     | 76±25                      | 98±25                  | -    | +              | +'             | +'               | 73±29           |
| H2     | R239S | 3±1               | 5±1                | 8±4                       | 31±16                      | 100±64                 | -    | +              | +'             | +                | 26±12           |
| H2     | V240Q | 57±19             | 69±32              | 74±10                     | 137±19                     | 115±2                  | +    | +              | +              | +                | 64              |
|        | L244E | 32±1              | 38±5               | 61±23                     | 59±5                       | 3±6                    | -    | +/-            | -              | +/-              | 33±5            |
|        | G245E | 50±10             | 71±31              | 73±10                     | 53±13                      | 1±1                    | -    | +              | -              | +'               | 33±16           |
|        | F246E | 35±4              | 39±13              | 52±26                     | 57±1                       | 13±4                   | -    | +/-            | -              | +/-              | 31±1            |
|        | P247E | -1±1              | -5±1               | 0±2                       | 10±6                       | 87±72                  | +    | +              | +              | +                | 65±18           |
|        | K249E | 120±46            | 92±46              | 69±21                     | 89±28                      | 62±29                  | -    | +              | +'             | +                | 31±13           |
|        | F250E | 2±2               | -1±0               | 3±2                       | 19±3                       | 0±5                    | +    | +              | +              | +                | 32±25           |
| S1'    | L251E | 31±12             | 34±15              | 75±43                     | 72±9                       | 123±10                 | +    | +              | +              | +                | 54±9            |
| S1'    | D252R | 41±12             | 40±11              | 74±40                     | 95±9                       | 67±38                  | -    | +              | +              | +                | 46              |

|        |       | c-fos<br>GAL4-E1A | c-fos<br>GAL4-VP16 | o-retinoic AR<br>GAL4-E1A | o-retinoic AR<br>GAL4-VP16 | <i>in vitro</i><br>Txn | TATA | TBP +<br>TFIIA | TBP +<br>TFIIB | TBP +<br>TFIIA&B | TFIIA<br>column |
|--------|-------|-------------------|--------------------|---------------------------|----------------------------|------------------------|------|----------------|----------------|------------------|-----------------|
|        | spm3  | 100               | 100                | 100                       | 100                        | 100                    | +    | +              | +              | +                | 100             |
| S2'    | S282E | 45±8              | 51±23              | 59±12                     | 61±4                       | 72±56                  | +    | +              | +              | +                | 102             |
| S2'    | Y283E | 3±3               | 6±4                | 15±5                      | 22±9                       | 4±7                    | -    | -              | -              | -                | 16±1            |
| étrier | E284R | 5±2               | 7±5                | 8±11                      | -12±3                      | 2±8                    | +    | +              | -              | +                | 113±54          |
| étrier | P285E | 1±1               | 5±4                | -1±3                      | 0±3                        | 1±0                    | -    | -              | -              | -                | 33±22           |
| étrier | E286R | 52±15             | 52±36              | 61±5                      | 65±33                      | 4±5                    | +    | +              | -              | +                | 166             |
| étrier | L287E | 11±3              | 12±5               | 8±2                       | 19±3                       | 12±13                  | +    | +              | -              | +                | 69              |
|        | K297E | 108±12            | 120±17             | 111±30                    | 84±43                      | 58±34                  | +    | +              | +              | +                | 66              |
|        | R299E | 84±6              | 96±20              | 113±16                    | 93±1                       | 68±40                  | -    | +              | +              | +                | 56              |
|        | F305A | 52±6              | 60±22              | 72±13                     | 28±6                       | -1±0                   | -    | +/-            | -              | -                | 46              |
|        | V306E | 77±10             | 99±5               | 81±6                      | 96±16                      | 71±21                  | +    | +              | +              | +                | 144±19          |
|        | S307F | 27±22             | 18±9               | 19±17                     | 62±13                      | -1±5                   | -    | +/-            | -              | +/-              | 36±12           |
|        | K316E | 29±17             | 21±8               | 30±6                      | 42±11                      | 0±11                   | -    | -              | -              | -                | 35±8            |
|        | V317E | 79±11             | 78±18              | 72±30                     | 52±30                      | 146±18                 | +    | +              | +              | +                | 91              |
| H2'    | R318E | 93±47             | 137±32             | 193±28                    | 267±48                     | 77±121                 | -    | +              | +              | +                | 67±14           |
| H2'    | A319R | 85±16             | 70±25              | 100±15                    | 99±11                      | 96±35                  | +    | +              | +              | +                | 57              |
| H2'    | E320R | 40±4              | 19±21              | 32±15                     | 30±22                      | -8±11                  | -    | -              | -              | -                | 24±3            |
| H2'    | Y322E | 58±9              | 47±15              | 98±3                      | 121±1                      | 99±3                   | +    | +              | +              | +                | 89±62           |
| H2'    | E323R | 71±10             | 52±30              | 80                        |                            | 25±32                  | +/-  | +              | +              | +                | 80              |
| H2'    | E326A | 104±10            | 100±15             | 152±2                     | 115±31                     | 64±33                  | +    | +              | +              | +                | 64              |
| H2'    | N327E | 89±18             | 60±8               | 95±23                     | 69±6                       | 42±4                   | +    | +              | +              | +                | 87              |
| H2'    | Y329E | 33±10             | 19±6               | 32±8                      | 36±30                      | 3±5                    | -    | -              | -              | -                | 34±4            |
| H2'    | P330R | 110±70            | 67±17              | 85±24                     | 96±32                      | 9±11                   | +/-  | +              | +/-            | +                | 36±27           |
| H2'    | K333E | 96±24             | 93±17              | 100±6                     | 99±22                      | 80±36                  | +    | +              | +              | +                | 145             |
| H2'    | G334R | 131±26            | 153±61             | 127±18                    | 102±12                     | 102±8                  | +    | +              | +              | +                | 109             |
|        | R336E | 115±47            | 114±46             | 201±48                    | 161±7                      | 114±25                 | +    | +              | +              | +                | 157±29          |
|        | K337E | 39±3              | 55±21              | 42±4                      | 84±11                      | 41±14                  | +    | +              | +              | +                | 120             |
|        | T338R | 110±54            | 97±37              | 92±4                      | 124±30                     | 104±37                 | +    | +              | +              | +                | 110±90          |
|        |       | 0±0               | 0±0                | 0±0                       | 0±0                        | 0±0                    | -    | -              | -              | -                | 0±0             |

**Tableau 8** : Résultat de l'étude de mutagenèse réalisée par O.Bryant *et al.*

Colonne 1 : Localisation de la mutation dans la structure de TBP (Cf. Figure 19).

Colonne 2 : Résidu muté (en noir, les mutations désorganisant la structure du TBP).

Colonnes 3, 4, 5 et 6 : Les pourcentages d'activité de la luciférase à partir des promoteurs indiqués. Ce pourcentage est calculé par rapport à l'activité de la TBP spm3 (100%) et reflète la capacité de ces mutants à assurer la transcription.

Colonne 7 : La colonne Txn indique le taux de transcription basal observé pour chaque mutant, *in vitro*.

Colonnes 8, 9, 10 et 11: Les résultats d'interaction en EMSA entre TBP et les molécules indiquées.

Colonne 12 : Indique la capacité des mutants à interagir avec TFIIA.

## INTRODUCTION PARTIE I

Le nombre de résidus identifiés dans les expériences d'alanine scanning est beaucoup moins important que celui rapporté dans l'étude de O.Bryant et al. Ces contradictions peuvent être expliquées par le fait qu'une substitution d'un résidu par une alanine aurait moins d'effet que la substitution de ce même résidu par un résidu de charge opposée. La substitution N189A, par exemple, n'a que peu d'influence sur les partenaires de TBP alors que la mutation N189E empêche la formation du complexe TBP/TFIIA.

D'autres études par mutagenèse des résidus de TBP ont été réalisées dans l'équipe du Dr Timmers afin d'identifier les surfaces de TBP impliquées dans l'interaction avec B-TAF1 au sein du complexe B-TFIID. Ces données seront abordées dans le chapitre consacré au complexe B-TFIID.

## II. TBP dans la transcription des gènes de classe II

### II.1. TFIID

#### II.1.1. Organisation structurale de TFIID

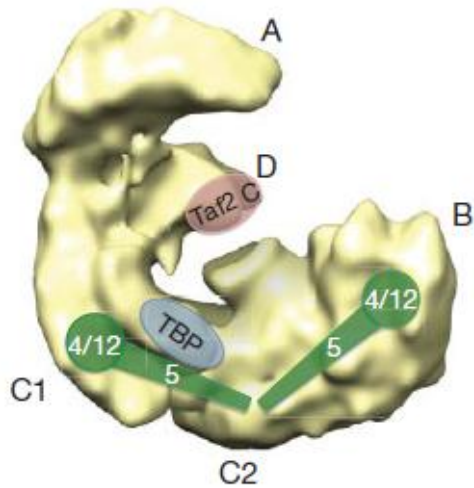
TFIID est un complexe macromoléculaire composé de TBP et de 13 à 14 TAFs. Les TAFs ont été isolés et caractérisés d'abord chez la drosophile (Dynlacht et al., 1991). Puis les gènes codant pour ces TAFs ont été isolés et clonés chez plusieurs espèces, permettant ainsi l'identification des domaines structuraux et fonctionnels conservés. L'alignement de leurs séquences et la comparaison de leurs propriétés biochimiques et fonctionnelles ont mis en évidence une importante conservation (pour revue, Cler et al., 2009). Initialement les TAFs étaient nommés selon leur poids moléculaires, mais devant la multitude de paralogues et d'orthologues des TAFs, une nouvelle nomenclature a été adoptée, où les TAFs sont numérotés de 1 à 14 (Tora 2002). Cf. *Tableau 9*

L'analyse des séquences peptidiques a montré que certains TAFs possèdent une similarité importante avec les histones. En effet, neuf TAFs contiennent le domaine HFD (*Histone Fold Domain*), impliqué dans l'interaction entre les histones au sein du nucléosome. TAF6 et TAF9 présentent des similarités de séquences avec les histones H4 et H3, alors que TAF4 et TAF12 sont similaires aux histones H2A et H2B. Les structures cristallographiques des hétérodimères hTAF(11/13) (Birck et al., 1998), hTAF(4/12) (Werten et al., 2002) et de l'hétérotétramère

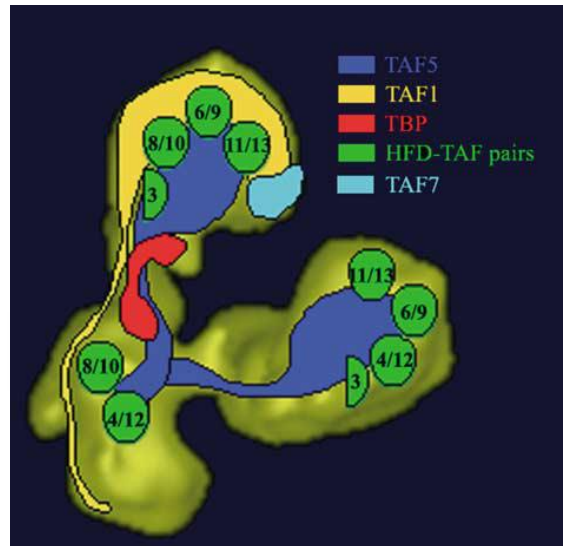
| New name       | <i>H. sapiens</i> (hs)           | <i>D. melanogaster</i> (dm)   | <i>C. elegans</i> (ce)      |                 | <i>S. cerevisiae</i> (sc) | <i>S. pombe</i> (sp)      |
|----------------|----------------------------------|-------------------------------|-----------------------------|-----------------|---------------------------|---------------------------|
|                |                                  |                               | previous name               | new name        |                           |                           |
| TAF1           | TAF <sub>II</sub> 250            | TAF <sub>II</sub> 230         | <i>taf-1</i> (W04A8.7)      | <i>taf-1</i>    | Taf145/130                | TAF <sub>II</sub> 111     |
| TAF2           | TAF <sub>II</sub> 150            | TAF <sub>II</sub> 150         | <i>taf-2</i> (Y37F11B.4)    | <i>taf-2</i>    | Taf150 or TSM1            | (T38673)                  |
| TAF3           | TAF <sub>II</sub> 140            | TAF <sub>II</sub> 155 or BIP2 | (C11G6.1)                   | <i>taf-3</i>    | Taf47                     |                           |
| TAF4           | TAF <sub>II</sub> 130/135        | TAF <sub>II</sub> 110         | <i>taf-5</i> (R119.6)       | <i>taf-4</i>    | Taf48 or MPT1             | (T50183)                  |
| TAF4b          | TAF <sub>II</sub> 105            |                               |                             |                 |                           |                           |
| TAF5           | TAF <sub>II</sub> 100            | TAF <sub>II</sub> 80          | <i>taf-4</i> (F30F8.8)      | <i>taf-5</i>    | Taf90                     | TAF <sub>II</sub> 72      |
| TAF5b          |                                  |                               |                             |                 |                           | TAF <sub>II</sub> 73      |
| TAF5L          | PAF65β                           | Cannonball                    |                             |                 |                           |                           |
| TAF6           | TAF <sub>II</sub> 80             | TAF <sub>II</sub> 60          | <i>taf-3.1</i> (W09B6.2)    | <i>taf-6.1</i>  | Taf60                     | (CAA20756)                |
| TAF6L          | PAF65α                           | (AAF52013)                    | <i>taf-3.2</i> (Y37E11AL.8) | <i>taf-6.2</i>  |                           |                           |
| TAF7           | TAF <sub>II</sub> 55             | (AAF54162)                    | <i>taf-8.1</i> (F54F7.1)    | <i>taf-7.1</i>  | Taf67                     | TAF <sub>II</sub> 62/PTR6 |
| TAF7L          | TAF2Q                            |                               | <i>taf-8.2</i> (Y111B2A.16) | <i>taf-7.2</i>  |                           |                           |
| TAF8           | (BAB71460)                       | Prodos                        | (ZK1320.12)                 | <i>taf-8</i>    | Taf65                     | (T40895)                  |
| TAF9           | TAF <sub>II</sub> 32/31          | TAF <sub>II</sub> 40          | <i>taf-10</i> (T12D8.7)     | <i>taf-9</i>    | Taf17                     | (S62536)                  |
| TAF9L          | TAF <sub>II</sub> 31L (AAG09711) |                               |                             |                 |                           |                           |
| TAF10          | TAF <sub>II</sub> 30             | TAF <sub>II</sub> 24          | <i>taf-11</i> (K03B4.3)     | <i>taf-10</i>   | Taf25                     | (T39928)                  |
| TAF10b         |                                  | TAF <sub>II</sub> 16          |                             |                 |                           |                           |
| TAF11          | TAF <sub>II</sub> 28             | TAF <sub>II</sub> 30β         | <i>taf-7.1</i> (F48D6.1)    | <i>taf-11.1</i> | Taf40                     | (CAA93543)                |
| TAF11L         |                                  |                               | <i>taf-7.2</i> (K10D3.3)    | <i>taf-11.2</i> |                           |                           |
| TAF12          | TAF <sub>II</sub> 20/15          | TAF <sub>II</sub> 30α         | <i>taf-9</i> (Y56A4.3)      | <i>taf-12</i>   | Taf61/68                  | (T37702)                  |
| TAF13          | TAF <sub>II</sub> 18             | (AAF53875)                    | <i>taf-6</i> (C14A4.10)     | <i>taf-13</i>   | Taf19 or FUN81            | (CAA19300)                |
| TAF14          |                                  |                               |                             |                 | Taf30                     |                           |
| TAF15          | TAF <sub>II</sub> 68             |                               |                             |                 |                           |                           |
| <i>B-TFIID</i> |                                  |                               |                             |                 |                           |                           |
| BTAF1          | TAF <sub>II</sub> 170/TAF-172    | Hel89B                        | (F15D4.1)                   | <i>btaf-1</i>   | Mot1                      | (T40642)                  |

**Tableau 9** : Nomenclature unifiée des TAFs

a.



b.



Leurent et al., 2004 ; Papai et al., 2010

**Figure 24 :** Architecture tridimensionnelle de TFIID

- a) Localisation des lobes A, B, C1, C2 et D.
- b) Distribution des TAFs dans les différents lobes. Les TAFs interagissant via des motifs HFD sont indiqués.

## INTRODUCTION PARTIE I

dTAF(6/9)<sub>2</sub> (Xie et al., 1996) confirment l'importance des HFDs dans les interactions entre les TAFs.

En absence d'une structure cristallographique de TFIID, les informations les plus pertinentes sur son architecture proviennent des expériences de microscopie électronique associée à l'immunolocalisation (immunomicroscopie). Cette technique combine la microscopie électronique et l'immunomarquage avec des anticorps reconnaissant des régions précises des différentes sous-unités d'un complexe protéique. L'analyse informatique des images issues de ces immunomarquages permet de réaliser des modèles tridimensionnels (3D) des complexes étudiés. Ces modèles 3D d'une résolution allant de 23 à 30 Å, révèlent la forme des complexes, leur dimension et la localisation approximative de leurs sous-unités.

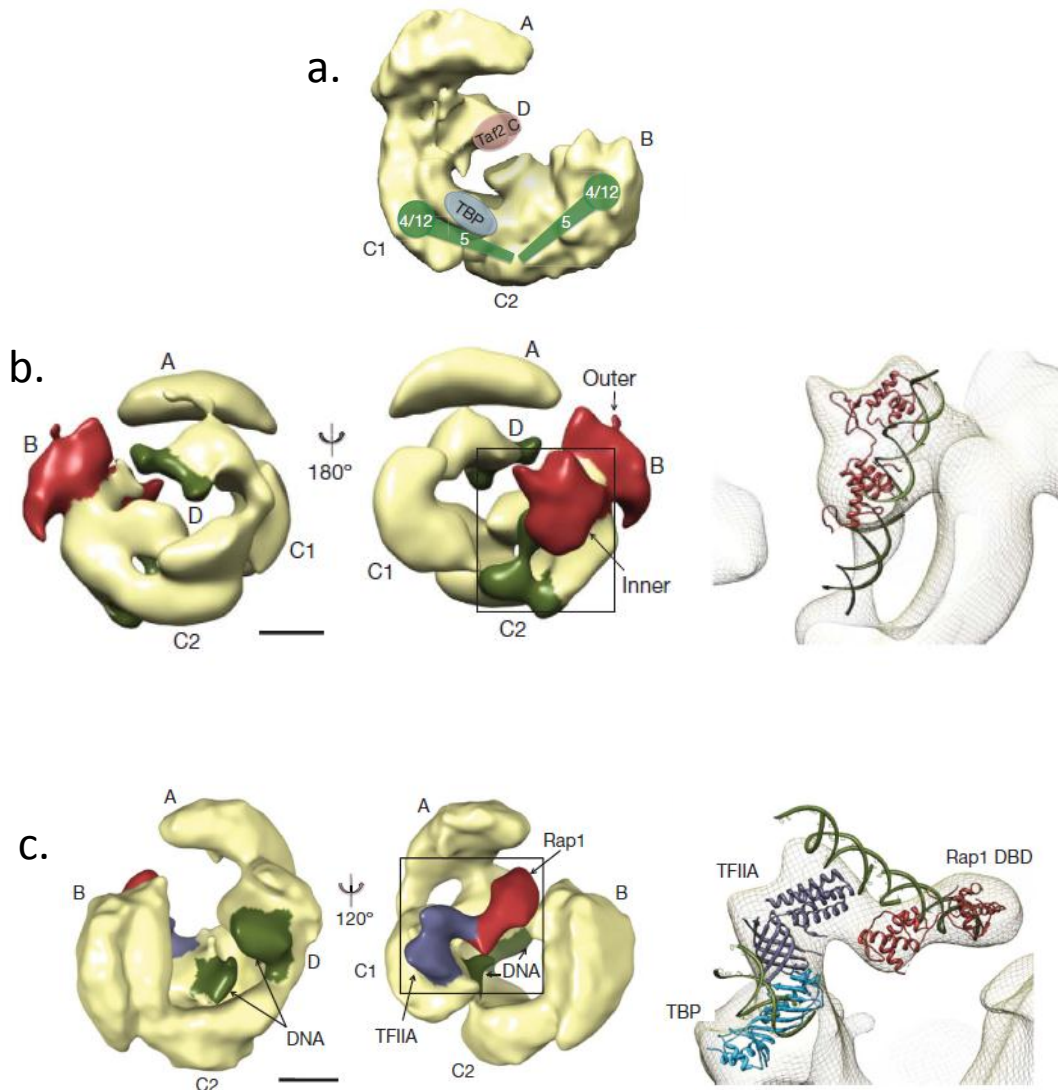
Ces études montrent que le TFIID humain et celui de la levure adoptent une organisation similaire. Cette organisation consiste en cinq lobes nommés A, B, C1, C2 et D. Ces lobes dessinent une pince dont les mâchoires sont constituées par les lobes A et B. La cavité de la pince est accessible au solvant et peut donc constituer une surface d'interaction pour les partenaires de TFIID. Les expériences d'immunomarquage ont permis de localiser TBP et les TAFs dans cette structure (*Cf. Figure 24*) (Leurent et al., 2004 ; Papai et al., 2010) (pour revue, Cler et al., 2009).

Ces études montrent également que la structure de TFIID est dynamique. En effet, la pince formée par les lobes peut adopter une conformation ouverte ou fermée. Ces deux conformations joueraient probablement un rôle dans l'interaction entre TFIID et ses partenaires (Grob et al., 2006).

### **II.1.2. Recrutement de TFIID sur les promoteurs**

Dans le modèle d'assemblage du PIC *in vitro*, TFIID interagit avec l'élément TATA du promoteur via sa sous-unité TBP et initie ainsi la formation du PIC. Deux exemples dans la littérature suggèrent que TFIID peut interagir avec l'élément initiateur via l'hétérodimère TAF1/TAF2 et avec le DPE via l'hétérodimère TAF6/TAF9. Or, la majorité des promoteurs sont dépourvus des éléments TATA, Inr et DPE. Deux questions se posent donc : 1) Comment TFIID interagit avec ces promoteurs? 2) Comment est défini le TSS en absence de tels éléments du promoteur?





**Figure 25** : Dynamique d'interaction entre TFIID, TFIIA, RAP1 et l'ADN

a) Architecture de TFIID libre, les lobes A, B, C1, C2 et D sont indiqués ainsi que la localisation des sous-unités de TFIID dans les différents lobes.

b) Dans le complexe, RAP1-TFIID-ADN, RAP1 interagit avec le lobe B de TFIID. L'ADN est localisé au niveau du lobe D (TAF2) et du lobe C2 en contact avec RAP1.

c) Dans le complexe TFIIA-TFIID-RAP1-ADN, l'ADN établit un contact supplémentaire avec TFIID au niveau du lobe C1 (TBP).

## INTRODUCTION PARTIE I

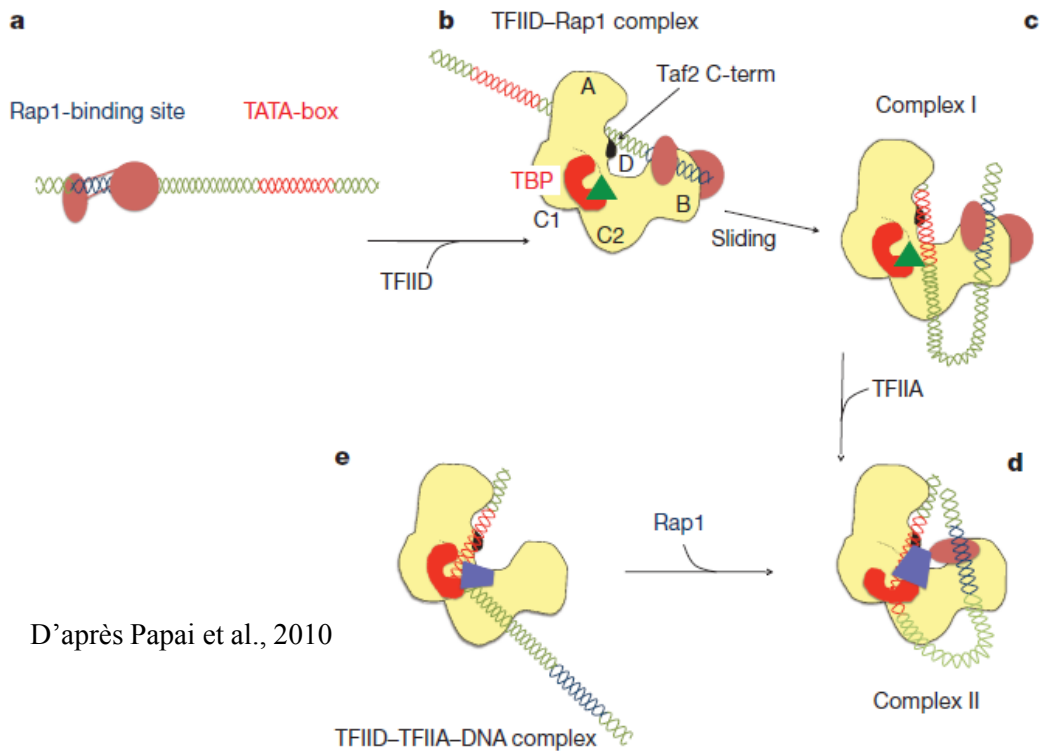
### II.1.2.1. Recrutement via des interactions directes avec les activateurs

Certaines études ont démontré que les activateurs de la transcription peuvent interagir directement avec les TAFs pour recruter TFIID sur le promoteur. Par exemple, SP1 interagit directement avec dTAF4, cette interaction est nécessaire pour l'activation de la transcription *in vitro* par SP1 (Gill et al., 1994). Le régulateur de la transcription RAP1 interagit également avec TAF5 et avec le dimère TAF4/TAF12 au sein de TFIID pour activer la transcription du gène RP codant pour une protéine ribosomale chez la levure (Garbett et al., 2007).

Une étude récente basée sur des expériences d'immunomicroscopie électronique propose un mécanisme d'activation de la transcription par RAP1 via TFIID et TFIIA (Papai et al., 2010). Dans cette étude, RAP1, TFIIA et TFIID ont été mis en présence d'un ADN chimérique contenant le promoteur minimal de PGK1 et une séquence de liaison de RAP1. Dans la solution contenant ces 4 molécules, les auteurs ont pu distinguer deux complexes : le premier complexe contient RAP1, TFIID et l'ADN, alors que le deuxième complexe contient RAP1, TFIID, TFIIA et l'ADN. Dans les deux complexes, RAP1 contacte TFIID au niveau du lobe B contenant TAF5 et le dimère TAF4/TAF12. Cela confirme les résultats obtenus précédemment.

Etonnamment, dans le premier complexe (RAP1-TFIID-ADN), le contact entre TFIID et l'ADN se fait via le lobe D à proximité de TAF2 et non via le lobe C1 qui contient TBP. En revanche, dans le complexe quaternaire (TFIIA-RAP1-TFIID-ADN), TFIID contacte l'ADN via le lobe C1 (TBP) et le lobe D (TAF2). Dans ce complexe, TFIIA interagit avec RAP1 et le lobe C1 de TFIID formant ainsi un pont qui relie RAP1 à TBP. TFIIA stabilise donc le complexe RAP1-TFIID et favoriserait des contacts supplémentaires entre TFIID et l'ADN. *Cf. Figure 25*

Les auteurs proposent que les deux complexes détectés correspondent à deux intermédiaires fonctionnels lors de l'activation de la transcription par RAP1. Ils proposent un modèle selon lequel, RAP1 interagit avec son élément de reconnaissance situé en amont du promoteur proximal et recrute TFIID via des interactions directes avec TAF5 et le dimère TAF4/TAF12. TFIID interagit à la fois avec RAP1 et avec l'ADN via TAF2 induisant ainsi une boucle entre le site de liaison de RAP1 et le promoteur proximal. TFIIA stabilise le complexe et induit probablement un changement de conformation permettant l'interaction entre TBP et l'ADN. *CF. Figure 26*



D'après Papai et al., 2010

Figure 26 : Modèles de recrutement de TFIID sur le promoteur

- RAP1 interagit avec son site de reconnaissance sur le promoteur
- RAP1 recrute TFIID sur le promoteur. Ce dernier interagit avec l'ADN via TAF2
- L'ADN se courbe et forme d'une boucle
- TFIIA est recruté et ferme la boucle d'ADN en interagissant avec les lobes B et C1 de TFIID
- Modèle alternatif, où RAP1 est recruté en second pour stabiliser le complexe TFIIA-TFIID-ADN

### II.1.2.2. Recrutement via des interactions avec les modifications des histones

Les interactions entre les TAFs et les histones jouent probablement un rôle important dans le recrutement et/ou la stabilisation de TFIID sur les promoteurs. Nous avons vu dans le chapitre consacré aux modifications des histones que certaines modifications sont caractéristiques des régions promotrices actives. Ces modifications sont des marques d'activation reconnues par les facteurs de transcription via des domaines protéiques spécifiques. Certains TAFs contiennent ces domaines de reconnaissance. Par exemple, hTAF1 contient un double bromodomaine, TAF5 contient un domaine WD40 et mTAF3 possède un domaine PHD (*Plant Homeodomain*). De plus, il a été démontré que TAF1 interagit préférentiellement avec l'histone H4 hyperacétylé (Jacobson et al., 2000) et TAF3 interagit spécifiquement avec l'histone modifiée H3K4me3 (Vermeulen et al., 2007). Cependant, ces domaines ne sont pas présents dans tous les orthologues des TAFs, par exemple yTAF1 ne possède pas le double bromodomaine et yTAF3 ne possède pas le domaine PHD. En revanche, chez la levure, yTAF1 est associé à la protéine BDF1 qui, elle, possède un bromodomaine et qui pourrait donc jouer un rôle dans le recrutement de TFIID. L'ensemble de ces observations suggère une certaine hétérogénéité dans les mécanismes d'interaction entre les orthologues des TAFs et les promoteurs.

### **II.1.3. Les activités enzymatiques de TFIID**

TFIID possède trois activités enzymatiques portées par TAF1 : 1) une activité histone acétyltransférase ; 2) une activité kinase ; 3) une activité de conjugaison et d'activation de l'ubiquitine (UBAC : *Ubiquitin Activating and Conjugating*).

*In vitro*, TAF1 peut acétyler les histones H3 et H4 (Mizzen et al., 1996). Cette activité est spécifique à certains promoteurs, car l'inactivation de TAF1 dans les cellules "ts13" de hamster provoque une hypoacétylation de H3 au niveau du promoteur de la cycline D1 et non du promoteur *c-fos* (Hilton et al., 2005). TAF1 peut également acétyler TFIIE $\beta$  (Imhof et al., 1997).

TAF1 possède deux activités kinases indépendantes, l'une située dans sa partie N-terminale (NTK), alors que la deuxième se trouve dans la partie C-terminale (CTK). TAF1 peut s'autophosphoryler (Auty et al., 2004) et phosphoryler plusieurs protéines dont l'histone H2B (Maile et al., 2004), la sous-unité RAP74 de TFIIF (Dikstein et al., 1996) et TFIIA $\beta$  (Solow et al.,

## INTRODUCTION PARTIE I

2001). *In vivo*, ces fonctions kinases sont également importantes puisque la délétion du CTK est létale chez la Drosophile (Wassarman and Sauer, 2001).

La troisième fonction enzymatique portée par TAF1 est une activité de conjugaison et d'activation de l'ubiquitine dirigée contre l'histone H1 (Pham and Sauer, 2000). TAF1 ne possède pourtant pas d'activité ubiquitine ligase apte à générer une chaîne polyubiquitine sur la protéine cible. La signification de cette monoubiquitination de H1 reste encore à découvrir.

### **II.1.4. Le rôle de coactivateur de TFIID**

Certains TAFs peuvent être directement contactés par des activateurs de la transcription. Nous avons vu que TAF4 peut interagir avec le domaine activateur de SP1 chez la drosophile. Étonnamment, le domaine de liaison à l'ADN (DBD) de SP1 peut également contacter TAF7 (Chiang and Roeder, 1995). TAF7 peut aussi interagir avec les activateurs YY1, USF et E1A. Les TAFs qui interagissent avec les activateurs sont considérés comme des coactivateurs. Ces interactions expliquent pourquoi TFIID peut permettre la transcription activée contrairement à la TBP seule (pour revue, Thomas et Chiang., 2006)

Il a également été démontré chez plusieurs espèces que la transcription de certains gènes ne nécessite pas tous les TAFs (Moqtaderi et al., 1996 ; Chen and Manley, 2000 ; Hayashida et al., 1994). Cette observation remet en cause l'idée selon laquelle le TFIID complet (TBP + 13-14 TAFs) est un GTF nécessaire pour la transcription de tous les gènes. De plus, chez la levure, la TBP peut être purifiée indépendamment de TFIID (Kuras et al., 2000), ce qui suggère que TBP pourrait agir indépendamment des TAFs pour assurer la transcription de certains gènes. Cette hypothèse a été vérifiée *in vitro*, en utilisant un système reconstitué de facteurs hautement purifiés, il a été démontré que TBP peut supporter la transcription activée en absence des TAFs, mais en présence des autres GTFs et du cofacteur PC4 (Oelgeschläger et al., 1998). Ces résultats suggèrent que les TAFs agissent comme des coactivateurs spécifiques à certains gènes et que plusieurs sous-complexes formés par les TAFs peuvent exister. En revanche, chez les mammifères, TBP est majoritairement associée aux TAFs dans les complexes TFIID et B-TFIID.

### **II.2. B-TFIID**

TBP peut également influencer la transcription par la Pol II en interagissant avec B-TAF1 pour former le complexe B-TFIID. B-TFIID est un hétérodimère stable de 300 kDa composé de TBP et de B-TAF1, plus abondant, au moins dans les cellules HeLa, que TFIID. Ce complexe a été purifié pour la première fois dans les cellules HeLa (Timmers et al., 1992). B-TFIID possède une activité ATPase et est capable de supporter la transcription basale *in vitro* avec une efficacité équivalente à celle de TFIID. B-TAF1 n'est pas une sous-unité de TFIID. TFIID et B-TFIID sont donc deux complexes distincts.

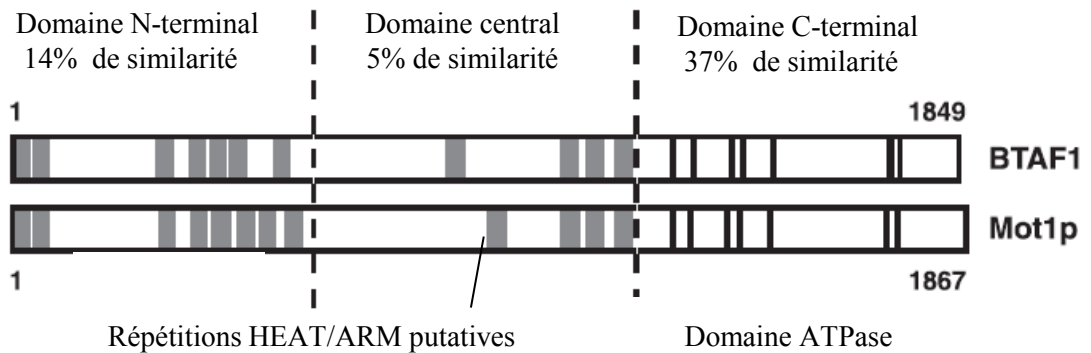
B-TAF1 est un membre conservé de la famille des ATPases SNF2/SWI2. Les analyses d'homologie de séquence révèlent que tous les génomes eucaryotes séquencés contiennent un orthologue unique de B-TAF1. Ces orthologues ont été caractérisés chez plusieurs espèces, dont la levure, la drosophile et la souris. Chez toutes ces espèces, B-TAF1 interagit avec TBP pour former le complexe B-TFIID (pour revue, Pereira et al., 2003).

MOT1P (*Modifier Of Transcription 1 protein*) est l'orthologue de B-TAF1 chez la levure. Le gène *mot1* a été identifié dans des expériences de criblage génétique. MOT1P a été initialement considéré comme un répresseur global de la transcription car son inactivation provoque l'augmentation de la transcription basale et non de la transcription activée de plusieurs gènes. Le mutant thermosensible *mot1-1* (mutation non-sens provoquant une faible expression de MOT1P) se caractérise par un défaut de prolifération à température permissive et une létalité à température non-permissive. Le mutant *mot1-1* a été utilisé dans la majorité des études traitant la fonction de MOT1P et sera fréquemment cité dans ce manuscrit.

#### **II.2.1. Organisation structurale de B-TFIID**

##### II.2.1.1. Organisation de la séquence primaire de B-TAF1

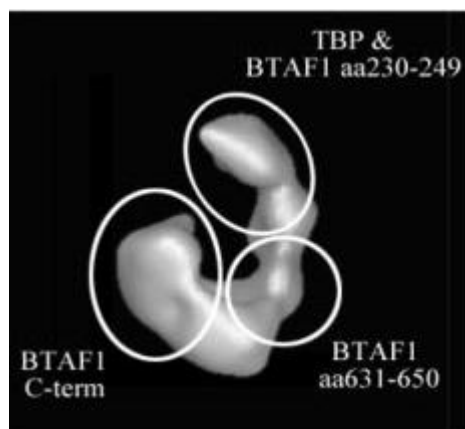
La compilation des bases de données génomiques montre que tous les eucaryotes dont le génome a été séquencé possèdent un orthologue putatif de B-TAF1. L'alignement des séquences primaires de plusieurs orthologues de B-TAF1 montre que ces protéines sont formées de trois régions majeures de taille identique : une région N-terminale modérément conservée, une région centrale faiblement conservée et une région C-terminale hautement conservée. Cf. *Figure 27*



D'après, Pereira et al., 2003

**Figure 27 :** Organisation en domaines de B-TAF1 et MOT1P

La conservation des différents domaines entre les espèces est indiquée (espèces considérées : Homme, Souris, Rat, Drosophile, *A. thaliana*, *S. cerevisiae* et *S. pombe*). Les répétitions HEAT/ARM putatives (barres grises) sont enrichies dans le domaine central et le domaine N-terminal. Le domaine C-terminal contient la signature du motif ATPase



**Figure 28 :** Architecture tridimensionnelle de B-TFIID

L'architecture tridimensionnelle de B-TFIID a été déterminée par microscopie électronique. L'utilisation d'anticorps reconnaissant spécifiquement des régions distinctes de B-TAF1 et de TBP, a permis de déterminer la localisation approximative de ces deux protéines dans le complexe B-TFIID.

## INTRODUCTION PARTIE I

La région C-terminale contient le domaine ATPase, alors que les régions centrale et N-terminale contiennent des répétitions HEAT/ARM putatives (Neuwald et Hirano, 2000; Andrade et al., 2001). Les répétitions HEAT/ARM sont composées de séquences riches en résidus hydrophobes. Leur structure consiste en deux hélices  $\alpha$  reliées par une courte boucle flexible. Les éléments HEAT/ARM se trouvent dans une variété de protéines de fonctions différentes telles les importines, les phosphatases, les protéine kinases ou encore des facteurs de transcriptions (Andrade et al., 2001; Perry et Kleckner, 2003). Le rôle potentiel de ces répétitions dans la fonction de B-TAF1 sera abordé plus loin. *Cf. Figure 27*

### II.2.1.2. TBP interagit avec la région N-terminale de B-TAF1

Dans l'équipe du Dr Timmers, la technique du double hybride a été utilisée pour identifier les régions de B-TAF1 impliquées dans l'interaction avec TBP pour former le complexe B-TFIID. Plusieurs régions de B-TAF1 ont été fusionnées au domaine de liaison à l'ADN LexA, puis leur capacité à interagir avec TBP fusionnée au domaine activateur B42 a été évaluée. Ces expériences montrent que la région N-terminale de B-TAF1 (résidus 1-504) est nécessaire et suffisante pour interagir avec TBP. Le complexe formé par le tiers N-terminal de B-TAF1 et TBP est stable en concentrations salines élevées, ce qui suggère que cette interaction est soutenue par des interactions hydrophobes. Trois parties distinctes du tiers N-terminal sont importantes pour l'interaction avec TBP. Il s'agit des régions 2-137, 290-381 et 380-460. La délétion d'une de ces régions altère l'interaction entre TBP et B-TAF1 (Pereira et al., 2001).

Chez la drosophile, la région N-terminale (de 1 à 800 aa) de dB-TAF1 est également nécessaire pour l'interaction avec TBP (Auble et al 1997 ; Adamkewicz, et al 2001). La surexpression de cette partie de MOT1P dans la levure est létale, car elle agit probablement comme un dominant négatif. Cette létalité peut être complétée par une surexpression de TBP (Auble et al 1997 ; Adamkewics et al 2001).

### II.2.1.3. Cartographie des surfaces de TBP impliquées dans l'interaction avec B-TAF1

Les études de mutagenèse réalisée dans l'équipe du Dr Timmers (Pereira et al., 2001 ; Klejman et al., 2005), indiquent que deux régions majeures de TBP sont impliquées dans l'interaction avec B-TAF1. La première région est située au niveau de la surface supérieure et la



## INTRODUCTION PARTIE I

surface latérale de la répétition N-terminale. La deuxième région est située au niveau de la surface concave hydrophobe formée par les dix feuillets  $\beta$  et qui interagit avec l'ADN.

Au niveau de la surface convexe, les résidus exposés au solvant R235, K236, K243 et R239 de l'hélice H2 sont particulièrement sensibles à la mutagenèse, leur substitution par une alanine ou par un résidu de charge opposée altère l'interaction avec B-TAF1. Une partie du feuillet S1' et la boucle qui le relie à l'hélice H2 sont aussi sensibles à la mutation. Dans cette zone, la substitution des résidus L244, K249, F250 et F253 altère l'interaction avec B-TAF1. Les mutations des résidus dans l'hélice H1 et la boucle la reliant au feuillet S2 (résidus R186 à N189) affectent aussi l'interaction entre TBP et B-TAF1. Les résidus I161 et V162 localisés sur la face latérale de la répétition C-terminale sont également sensibles à la mutagenèse.

La deuxième région qui s'est révélée importante pour l'interaction avec B-TAF1 est la région concave qui est impliquée dans la liaison de TBP à l'ADN. Plusieurs résidus qui interagissent avec l'ADN (selon la structure cristallographique) sont sensibles à la mutation. Il s'agit notamment des résidus R203, L212, F214 et K218 situés dans la première répétition de TBP et les résidus F288, R294, F305 et K309 de la seconde répétition de TBP. Cette étude n'a identifié que quatre résidus interagissant avec l'ADN dont la mutation n'a aucun effet sur l'interaction avec B-TAF1. Il s'agit des résidus R196 dans le premier étrier, R208 et T210 dans le feuillet S4 et R299 dans le feuillet S4'.

L'interaction entre B-TAF1 et la face concave de TBP est étonnante, car cette même surface est impliquée dans l'interaction avec l'ADN. Il a été proposé que B-TAF1 inhibe l'interaction entre TBP et l'ADN par compétition. Cette hypothèse est contradictoire avec le fait que TBP, B-TAF1 et l'ADN peuvent former un complexe stable.

B-TAF1 interagit donc avec des surfaces distinctes et étendues de TBP. Cette caractéristique peut être expliquée par la présence des motifs HEAT/ARM dans le domaine N-terminal de B-TAF1. En effet, ces motifs permettent la formation d'une large structure super-hélicoïdale, générant ainsi une surface d'interaction étendue.

Il est intéressant de noter que six des onze répétitions HEAT/ARM dans B-TAF1 coïncident avec les régions de liaison à TBP (Pereira et al., 2001). De même, huit des 12

## INTRODUCTION PARTIE I

répétitions HEAT/ARM de MOT1P se situent dans les régions impliquées dans la liaison à TBP (Adamkewicz et al., 2001; Darst et al., 2003).

### II.2.1.4. L'architecture tridimensionnelle de B-TFIID

L'architecture tridimensionnelle du B-TFIID humain a été résolue par la technique d'immunomicroscopie électronique dans l'équipe du Dr. Schultz à l'IGBMC (Pereira et al., 2004). Selon cette étude, B-TFIID est composé d'un large domaine globulaire de masse approximative de 170 kDa et qui représente 68% du volume total du complexe. De cette forme globulaire se projette une extension en forme de pouce formée par 3 sous-domaines de 25, 15 et 35 kDa, à la fin du pouce se trouve un autre domaine globulaire. Cf. *Figure 28*

### II.2.1.5. Immunolocalisation de TBP et B-TAF1 dans le complexe B-TFIID

Des anticorps dirigés contre des régions distinctes de TBP et de B-TAF1 ont permis la localisation et l'orientation approximatives des deux protéines dans le complexe B-TFIID.

Le domaine C-terminal de B-TAF1 contenant le motif ATPase est localisé dans le large domaine globulaire, alors que le domaine N-terminal de B-TAF1 est situé à l'extrémité du pouce à proximité de TBP. Ces observations confortent les résultats de mutagenèse qui suggéraient que TBP interagit avec la région N-terminale de B-TAF1.

Cependant, l'anticorps dirigé contre TBP utilisé dans cette étude reconnaît un épitope situé dans le domaine N-terminal non conservé. Ce domaine est flexible, ce qui implique que la localisation de TBP est hypothétique car le domaine C-terminal de TBP peut interagir avec d'autres régions de B-TAF1 (Pereira et al., 2004).

## **II.2.2. La fonction de B-TAF1 et MOT1P et leur mécanisme d'action**

### II.2.2.1. La dualité de la fonction de B-TAF1 : répresseur / activateur

MOT1P a été initialement considéré comme un répresseur global de la transcription par la Pol II, car son inactivation (mutation *mot1-1*) augmente la transcription basale et non la transcription activée de plusieurs gènes impliqués dans des processus différents chez la levure (Davis et al., 1992). Des expériences de transcription *in vitro* montrent également que MOT1P réduit la transcription basale et non la transcription activée de plusieurs promoteurs. Des résultats

## INTRODUCTION PARTIE I

similaires ont été obtenus *in vivo* en évaluant la transcription à partir de gènes rapporteurs. L'action de MOT1P est spécifique à la Pol II, car les gènes transcrits par la Pol I et la Pol III ne sont que peu affectés par MOT1P (Auble et al., 1994). MOT1P et B-TAF1 peuvent également dissocier TBP de l'ADN en utilisant l'énergie fournie par l'hydrolyse de l'ATP et ainsi inhiber la transcription (Auble et al., 1994, Chicca et al., 1998).

Etonnamment, le complexe B-TFIID purifié à partir de cellules HeLa, stimule la transcription basale mais avec une efficacité légèrement inférieure à celle de TFIID (Timmers et al., 1991). Cette différence d'efficacité dans la stimulation de la transcription peut être expliquée par le fait que l'association entre B-TFIID et l'ADN est moins stable que celle de TFIID (Timmers et al., 1991). Néanmoins, cette observation suggère que B-TFIID peut également jouer un rôle activateur.

Des études ultérieures ont démontré que l'effet répressif ne représente qu'une facette de la fonction de MOT1P. En effet l'inactivation de MOT1P réduit également l'expression de plusieurs gènes (Collart 1996 ; Madison and Winston 1997 ; Prelich 1997 ; Li et al., 2000).

Des expériences de ChIP ont montré que MOT1P peut être recruté rapidement aux promoteurs des gènes HXT2 et HXT4 (transporteurs de glucose) lors de leur activation dans un environnement pauvre en glucose. Le recrutement de MOT1P à ces promoteurs coïncide avec celui de TBP, de TAF1 et de la Pol II. L'interaction de MOT1P avec le promoteur est transitoire contrairement à celle de TBP et de TAF1. L'inactivation de MOT1P inhibe le recrutement de TBP sur les promoteurs des gènes HXT2 et HXT4. Les auteurs de cette étude proposent que pour certains promoteurs tels HXT2 et HXT4, MOT1P escorte TBP durant la phase précoce d'activation et que suite à la compétition avec d'autres facteurs de transcription tels TFIIA et TFIIB, MOT1P se dissocie de TBP et du promoteur (Andrau et al., 2002).

### II.2.2.2. L'activité ATPase est essentielle pour la fonction de B-TAF1

Le principal mécanisme d'action connu de B-TAF1 et de MOT1P concerne leur capacité de dissocier TBP de l'ADN en utilisant l'énergie fournie par l'hydrolyse de l'ATP (Auble et al., 1994, Pereira et al., 2004).

Le domaine C-terminal de MOT1P est suffisant pour assurer son activité ATPase. Ce domaine contient le motif DEGH nécessaire pour la liaison et l'hydrolyse de l'ATP. L'altération

## INTRODUCTION PARTIE I

de ce motif par mutagenèse inhibe l'activité ATPase de MOT1P *in vitro* et altère sa fonction *in vivo*. L'activité ATPase de MOT1P ne dépend pas de l'ADN (Auble et al., 1994). Au contraire, l'addition de l'ADN réduit de deux à trois fois l'activité ATPase de B-TAF1 (Timmers et al., 1992).

### II.2.2.3. MOT1P et NC2 agissent en synergie pour réguler l'activité de TBP

Dans l'équipe du Dr Timmers, la technique de ChIP on chip a été utilisée pour cartographier les sites de liaison de TBP, MOT1P, NC2 $\alpha$ , NC2 $\beta$ , TAF1, SPT20 (sous-unité de SAGA) et la Pol II dans la levure (Van Werven et al., 2008). Cette étude montre que MOT1P, NC2 et TBP colocalisent avec la Pol II, TAF1 et SPT20 sur un grand nombre de promoteurs de gènes activement transcrits. Afin d'expliquer ces observations, les auteurs proposent un modèle dans lequel les promoteurs peuvent à la fois être liés par des régulateurs positifs comme SAGA et TFIID et des régulateurs négatifs comme MOT1P et NC2. Cela crée un équilibre dynamique qui peut réguler la transcription positivement ou négativement en réponse à différents stimuli. Néanmoins, l'hypothèse d'un rôle activateur de MOT1P et de NC2 n'est pas à exclure.

Le grand nombre de promoteurs liés conjointement par NC2 et MOT1P indique clairement que ces deux facteurs agissent en synergie pour réguler l'activité de TBP. Des analyses biochimiques confortent cette idée. En effet, TBP, MOT1P et NC2 s'associent en complexe stable avec l'ADN. Etonnamment, en présence d'ATP, MOT1P dissocie le complexe MOT1P - TBP-NC2-ADN et provoque également une dissociation partielle du dimère NC2 $\alpha$ /NC2 $\beta$  (Van Werven et al., 2008).

Afin d'évaluer le rôle de MOT1P dans l'interaction de TBP et de NC2 avec les promoteurs *in vivo*, l'expression du gène HXT2 a été induite en baissant le taux de glucose dans le milieu, puis réprimée en remettant de nouveau les cellules dans un milieu contenant du glucose. Cette expérience permet d'évaluer le recrutement des facteurs de transcription sur le promoteur HXT2 lors de l'induction ainsi que leur comportement lors de la répression du gène. Cette expérience a été réalisée dans les cellules sauvages et dans les cellules dépourvues de MOT1P (mutant *mot1-1*). La même stratégie a été utilisée par Andrau et al. en 2002 pour évaluer le rôle de Mot1P dans la régulation de l'activité de TBP.

## INTRODUCTION PARTIE I

Les résultats de cette dernière étude montrent que le recrutement de TBP sur le promoteur HXT2 lors de l'induction n'est pas affecté par l'inactivation de MOT1P (Van Werven et al., 2008). Ce résultat contredit celui publié par Andrau et al. en 2002, qui démontrait que l'inactivation de MOT1P inhibe le recrutement de TBP sur le promoteur HXT2. Le rôle de MOT1P dans le recrutement de TBP sur les promoteurs des gènes actifs reste donc à éclaircir.

En revanche, l'inactivation de MOT1P affecte la dissociation de TBP et de NC2 du promoteur HXT2 lors de la répression de ce gène. Cela suggère que MOT1P est nécessaire pour dissocier le complexe TBP-NC2-ADN *in vivo* lors de la répression de la transcription (Van Werven et al., 2008).

Il est important de noter que dans les cellules sauvages, NC2 n'interagit que modérément avec le promoteur HXT2 lors de son activation, alors que dans les cellules *mot1-1*, NC2 est plus fortement recruté sur ce promoteur (Van Werven et al., 2008). Ce résultat suggère que MOT1P peut réguler négativement le recrutement de NC2 sur les promoteurs. Cette hypothèse est supportée par des expériences de CHIP réalisées dans les cellules HeLa et qui montrent que la réduction d'expression (Knock Down) de B-TAF1 augmente la liaison de NC2 sur plusieurs promoteurs (Graaf et al., 2010).

Des analyses de FRAP (*Fluorescent Recovery After Photobleaching*) montrent que le "knock-down" de B-TAF1 dans les cellules HeLa provoque une baisse de la mobilité de TBP dans la cellule et augmente son association avec la chromatine. Alors que la surexpression de B-TAF1 a un effet opposé. Les auteurs ont conclu que B-TAF1 est impliqué dans la dissociation de TBP de la chromatine et non dans son association à cette dernière. Il est important de noter, que le "knock-down" de B-TAF1 n'affecte pas la mobilité de TAF1 et TAF5, ce qui suggère que l'action de B-TAF1 sur TBP se fait indépendamment de TFIID (de Graaf et al., 2010). De plus, "le knock-down" de B-TAF1 semble augmenter la liaison de TBP sur la majorité des promoteurs dans les cellules HeLa (Johannes et al., 2010). Ces études supportent donc le modèle selon lequel B-TAF1 est un répresseur global de la transcription.

Contrairement à B-TAF1, NC2 semble favoriser l'association de TBP à la chromatine (de Graaf et al., 2010).

## INTRODUCTION PARTIE I

A la lumière de ces résultats souvent contradictoires, il semblerait que B-TAF1 et NC2 agissent de concert pour réguler l'activité de TBP. Si le potentiel répresseur de B-TAF1/ MOT1P et NC2 a été clairement démontré *in vitro*, la situation *in vivo* semble plus complexe. Le rôle activateur de B-TAF1 reste à éclaircir. Des investigations supplémentaires sont nécessaires pour comprendre les mécanismes d'action des régulateurs généraux de la transcription *in vivo*. Tout en sachant que le terme "général" est à utiliser avec précaution.

## RÉSULTATS

Mon principal projet de thèse concerne l'étude structure-fonction de TBP dans les fibroblastes murins. Etant donné l'importance de TBP dans la régulation de la transcription et l'absence d'études pertinentes sur le rapport structure/fonction de TBP *in vivo* chez les eucaryotes supérieurs, l'équipe d'accueil a mis en place un système cellulaire permettant d'étudier la fonction et la structure de TBP *in vivo*. Ce système est basé sur des fibroblastes embryonnaires *Tbp<sup>lox/-</sup>* dérivés de souris génétiquement modifiées permettant une inactivation conditionnelle du gène codant pour la TBP endogène. Cette inactivation est létale mais peut être complétée par l'expression stable d'une TBP exogène sauvage (WT) ou portant des mutations ponctuelles d'un seul acide aminé dans le domaine C-terminal conservé. L'expression stable de mutants de TBP dans ces cellules permet d'évaluer l'impact de ces mutations sur l'interaction entre TBP et ses partenaires *in vivo* et d'étudier l'effet de ces mutations sur l'expression des gènes.

Mes travaux de thèse se sont focalisés sur l'étude de deux mutants affectant spécifiquement l'interaction entre TBP et son partenaire B-TAF1 dans le complexe B-TFIID. Il s'agit des mutants R188E et K243E. Les deux lignées cellulaires exprimant ces deux mutants présentent un excellent modèle pour étudier la fonction de B-TFIID *in vivo*.

Nos résultats montrent que l'altération du complexe B-TFIID dérégule l'expression d'un nombre réduit de gènes. Certains de ces gènes sont réprimés alors que d'autres sont induits. Nous avons utilisé les techniques de ChIP et de ChIP-seq pour disséquer les mécanismes moléculaires impliqués dans cette dérégulation.

L'ensemble des résultats obtenus a permis la rédaction d'une publication qui sera soumise prochainement. Cf. *Publication 1, (RNA polymerase II transcription in murine cells lacking the B-TFIID complex)*

Parallèlement à mes travaux sur le projet TBP, j'ai également participé à d'autres projets au sein du laboratoire. Durant la dernière année de ma thèse, j'ai notamment initié un projet dont le but est de mieux comprendre les réseaux transcriptionnels régulant la différenciation neuronale des cellules souches embryonnaires (ES). Les données concernant ce projet seront détaillées dans une partie indépendante de ce manuscrit (Cf. *Partie II*).



J'ai également collaboré avec Igor Martianov pour un projet concernant la caractérisation des gènes cibles des facteurs de transcription CREM (*cAMP Response Element Modulator*) et CREB (*cAMP Response Element Binding protein*). (Martianov et al., 2010)

J'ai aussi participé au projet traitant du rôle des récepteurs de l'acide rétinoïque (RARs) dans les fibroblastes embryonnaires murins et dans les cellules souches embryonnaires (Delacroix et al., 2010).

Ma participation à ces deux projets s'est limitée à l'analyse bioinformatique des données obtenues, par conséquent, ils ne seront pas discutés dans ce manuscrit. Enfin, j'ai participé au développement d'un outil informatique dédié à l'analyse des données de ChIP-seq. Nous avons baptisé cet outil SeqMiner. La publication concernant cet outil a été soumise au journal *Nucleic Acids Research*. Cf. Publications, (seqMINER: An integrated ChIP-seq data interpretation platform)

**Publication 1 :**

**RNA polymerase II transcription in murine cells lacking the B-TFIID complex.**

Mohamed-Amin Choukrallah<sup>1</sup>, Dominique Kobi<sup>1</sup>, Igor Martianov<sup>1</sup>, W.W.M. Pim Pijnappel<sup>2,4</sup>, Nikolai Mischerikow<sup>3,4</sup>, Tao Ye<sup>1</sup>, Albert J.R. Heck<sup>3,4,5</sup>, H.Th. Marc Timmers<sup>2,4</sup> and Irwin Davidson<sup>1,#</sup>

Manuscrit en préparation

**RNA polymerase II transcription in murine cells lacking the B-TFIID complex.**

Mohamed-Amin Choukrallah<sup>1</sup>, Dominique Kobi<sup>1</sup>, Igor Martianov<sup>1</sup>, W.W.M. Pim Pijnappel<sup>2,4</sup>, Nikolai Mischerikow<sup>3,4</sup>, Tao Ye<sup>1</sup>, Albert J.R. Heck<sup>3,4,5</sup>, H.Th. Marc Timmers<sup>2,4</sup> and Irwin Davidson<sup>1,#</sup>

1. Institut de Génétique et de Biologie Moléculaire et Cellulaire. CNRS/INSERM/ULP. 1 Rue Laurent Fries, 67404 Illkirch Cédex. France.

2 Department of Physiological Chemistry, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

3 Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

4 Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

5 Centre for Biomedical Genetics, Padualaan 8, 3584 CH Utrecht, The Netherlands

# To whom correspondence should be addressed

Institut de Génétique et de Biologie Moléculaire et Cellulaire. CNRS/INSERM/ULP. 1 Rue Laurent Fries, 67404 Illkirch Cédex. France.

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

E mail : irwin@igbmc.fr

Running Title : Function of B-TFIID. Key words; BTAF1, CHIP-seq, proliferation, transcription regulation.

**Abstract.**

The conserved C-terminal domain of the TATA binding protein (TBP) interacts with multiple partners to form the complexes required for transcription by RNA Polymerases I, II and III. To analyse these interactions in mammalian cells we have used genetically modified mouse embryonic fibroblasts (MEFs) where loss of endogenous TBP induced by Cre recombinase can be complemented by exogenous wild-type or mutant TBPs with single amino acid substitutions in the C-terminal domain. We show that many mutant TBPs can complement loss of the endogenous TBP, but induce a slow growth phenotype. Tandem immunopurifications and mass-spectrometry analysis identifies two TBP mutations, R188E and K243E, that completely disrupt the TBP/BTAF1 interaction and formation of the B-TFIID complex. Transcriptome analysis shows that the R188E mutation affects the expression of only 474 genes. In agreement with this, ChIP-seq analysis does not show major changes in genomic TBP and Pol II distribution in cells expressing TBP R188E. However, at affected promoters, mutations in TBP either result in its *de novo* recruitment or an exchange of the B-TFIID complex in wild-type cells for TFIID in R188E mutant cells promoting Pol II recruitment and gene activation. These data show that formation of a stable B-TFIID complex is not essential for cell viability, but is required for normal proliferation. The TBP-BTAF1 interaction is not required for normal regulation of the vast majority of genes, but many be important for the specific regulation of a subset of promoters.

**Introduction.**

Accurate initiation of transcription by RNA polymerase II (Pol II) requires the assembly of the multiprotein preinitiation complex on the core promoter around the mRNA start site (Fuda et al., 2009; Hampsey, 1998; Thomas and Chiang, 2006). Amongst the general transcription factors in this process, is the TFIID complex comprising the TATA binding protein (TBP) and a set of 13-14 TBP associated factors (TAFs) (Cler et al., 2009; Tora, 2002). Two additional TBP-containing complexes involved in Pol II transcription have been described, the B-TFIID complex, where TBP is associated with BTAF1 an ATPase domain containing protein that can remove TBP from DNA in an ATP dependent manner (Auble et al., 1994; Timmers et al., 1992; Van Der Knaap et al., 2000), and TAC, a complex specific to undifferentiated embryonic stem cells, where TBP is associated with an unprocessed form of general transcription factor TFIIA (Mitsiou and Stunnenberg, 2000). However, the unique aspect of TBP is its additional presence in complexes required for Pol I (TIF1B/SL1) and Pol III (TFIIIB) transcription. In each of these complexes, TBP associates with a distinct set of TAFs (Hernandez, 1993).

TBP has a bipartite structure with a highly conserved C-terminal core domain that is 80% identical between yeast and mammals. The core folds as a molecular saddle responsible for binding DNA via the concave underside and interaction with general transcription factors and with a host of other regulatory proteins via the solvent exposed convex surface (Burley and Roeder, 1996).

The structure of TBP bound to TATA-containing DNA, as well as with the general transcription factors TFIIA, TFIIIB, NC2, BRF, and the N-terminal domain of TAF1 have all been described (Juo et al., 2003; Kamada et al., 2001; Nikolov et al., 1995; Tan et al., 1996). The surfaces involved in these interactions have been mutated as part of a systematic

mutagenesis of all the solvent exposed residues of the core domain (Bryant et al., 1996; Klejman et al., 2005; Tang et al., 1996; Zhao et al., 2003). The properties of the mutant TBPs have been evaluated *in vitro* with respect to transcription using Pol II and Pol III promoters, binding to DNA, interaction with cofactors and the ability to form higher order complexes with its partners in EMSA. The properties of the mutant TBPs have also been evaluated for their ability to support activated transcription by transfection in mammalian cells (Bryant et al., 1996; Tansey et al., 1994).

While the above studies provided considerable insight into the structure-function relationships of TBP, their scope has been limited to *in vitro* assays or transfections with artificial promoters that do not reproduce the complexity of the real *in vivo* situation with the diversity of promoters in the cell. We have developed a unique approach to study TBP function *in vivo* using mouse embryonic fibroblasts (MEFs) where one allele of the TBP gene has been inactivated and the other has been floxed such that endogenous TBP can be inactivated by Cre-recombinase. Upon inactivation of TBP, the *Tbp*<sup>-/-</sup> cells die, but expression of wild-type human TBP following retroviral infection restores cell viability. We show that mutations affecting critical interactions *in vitro* can complement loss of endogenous TBP *in vivo*, but lead to impaired cell proliferation. We have characterised MEFs expressing TBP R188E. We show that this mutation leads to a selective loss of interaction with BTAF1 and deregulates a small set of genes. Chromatin immunoprecipitation (ChIP)-sequencing shows that this mutation disrupting the B-TFIID complex does not affect the global genomic distribution of TBP, but positively or negatively affects TBP and/or Pol II recruitment at a small subset of promoters.

## **Experimental procedures.**

### **Isolation of the *Tbp*<sup>lox/-</sup> MEFs**

Mouse ES cells with a null allele of TBP (Martianov et al., 2002) were re-targetted with a second vector to introduce LoxP sites around exon 3 of the *Tbp* gene. These ES cells were used to generate mice with the corresponding genotype. MEFs were derived from E10.5 embryos by standard procedures and immortalised using large T antigen from SV40.

### **Generation and characterisation of cells expressing mutant TBPs.**

*Tbp*<sup>lox/-</sup> MEFs were first infected with pBABE retroviruses expressing WT or mutant N-terminal Flag-HA tagged TBP. After puromycin selection, cell extracts were prepared and expression of the exogenous TBP verified by immunoblot. Cells were reinfected with a second retrovirus vector expressing the tamoxifen inducible Cre-ER<sup>T2</sup>. After blasticidine selection, cells were treated with hydroxy-tamoxifen (OHT) and clonal lines established from limiting dilutions of cells. The deletion of the *Tbp* alleles was verified first by triplex PCR genotyping and then cell extracts were prepared from the selected cell clones and TBP expression verified by immunoblot. Cell proliferation analysis, Affymetrix microarrays, qPCR and immunoblot were all performed by standard methods as previously described (Mengus et al., 2005). TBP was detected using the 3G3 antibody that recognises an epitope conserved at the extreme N-terminus of mouse and human TBP (Lescure et al., 1994).

**ChIP-sequencing.**

ChIP and ChIP-seq experiments were performed according to standard protocols as previously described (Delacroix et al., 2010; Kobi et al.). Briefly ChIP-seq was performed using an Illumina GAIx sequencer and the raw data analysed by the Illumina Eland pipeline V1.6. Peak detection was performed using the MACS software (<http://liulab.dfci.harvard.edu/MACS/>) (Zhang et al., 2008) under settings where the HA-ChIP from untagged 501Mel cells was used as a negative control. Peaks were then annotated using GPAT ((Krebs et al., 2008), [http://bips.u-strasbg.fr/GPAT/Gpat\\_home.html](http://bips.u-strasbg.fr/GPAT/Gpat_home.html)) using a window of +/- 20kb with respect to the coordinates of the beginning and end of RefSeq transcripts. Global clustering analysis was performed using seqMINER (<http://bips.u-strasbg.fr/seqminer/> ]. Further details are provided in the Supplemental text. ChIP was performed with the following antibodies; TBP (Abcam, AB28175), Pol II, MED6, XPB, CDK7 (Santa Cruz SC-9001, SC9434, SC-293, SC-529, respectively), H3K4me3 (Upstate 07-473). The TFIIA, and TAF1 antibodies were gifts from Dr H. Stunenberg and L. Tora.

**Proteomics analysis.**

MEF cell nuclear extracts were prepared as previously described (Mengus et al., 2005) and subjected to tandem Flag-HA affinity immunopurification using Flag M2 beads, epitope peptide elution followed by reprecipitation with HA-beads. The precipitated proteins were digested with trypsin, and the tryptic peptides separated by tandem strong cation exchange. Peptides were then analysed by nanoflow-LC coupled to an LTQ-FTICR mass spectrometer (Thermo Fisher Scientific). Spectra were processed with Bioworks (version 3.1, Thermo Fisher Scientific) and the subsequent data analysis was carried out using Mascot (version



2.2.1, Matrix Science). Mascot results files were filtered to contain only peptides above a Mascot ion score of 15 using in-house written software. Scaffold (version 2.01.02, Proteome Software) was used to validate protein identifications. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least 2 identified peptides. Human TBP and its variants were identified following essentially the same procedure but using a manually generated database containing sequences of mouse TBP (Swiss-Prot accession number P29037), human TBP (Swiss-Prot accession number P20226), and those of the TBP mutants used in this study. Further details are described in the supplemental text.

## Results

### **A mouse embryonic fibroblast model for functional analysis of TBP.**

Using homologous recombination, we have generated ES cells and mice harbouring a null allele in the *Tbp* gene by insertion of a hygromycin resistance cassette in exon III and a floxed allele, in which exon III is surrounded by LoxP sites [(Martianov et al., 2002) and see Fig. 1A]. From these mice, we generated an embryonic fibroblast *Tbp*<sup>lox/-</sup> cell line in which the presence of these alleles can be detected by a triplex PCR reaction using three primers A-C (Fig. 1A and B).

These cells were transfected with vectors expressing the Cre recombinase and the green fluorescent protein to detect the transfected cells. After 24 hours, transfected cells were subjected to immunostaining with an anti-TBP antibody and cells where TBP expression was lost from the nuclei were detected (Supplemental Fig. 1A). Similarly, expression of the Cre recombinase leads to the appearance of specific short PCR fragment generated upon deletion of the floxed allele. This fragment is abundant in the transfected cells between 2-4 days after

transfection, but then rapidly disappears (Supplemental Fig.1B, lanes 1-5). These observations indicate that expression of Cre recombinase leads to deletion of the floxed allele, a loss of TBP expression, and the subsequent death and loss of the *Tbp*<sup>-/-</sup> cells from the population. However, when wild-type (wt) human (h) TBP is co-expressed with Cre, *Tbp*<sup>-/-</sup> cells persist in the population for at least up to 10 days showing that loss of endogenous mouse (m)TBP could be complemented by exogenous hTBP (supplemental Fig.1B, lanes 6-10).

Similarly, we co-expressed the altered specificity mutant hTBP<sub>spm</sub> (Strubin and Struhl, 1992) with Cre. As observed for wt hTBP, the resulting *Tbp*<sup>-/-</sup> cells persist in the transfected cell population (Supplemental Fig. 1B, lanes 11-15). The hTBP<sub>spm3</sub> mutant can thus rescue the lethality resulting from loss of endogenous mTBP. The *Tbp*<sup>-/-</sup> cells can therefore be used to investigate the ability of mutated human TBPs to complement loss of endogenous mTBP.

### **Mutations in TBP induce a slow growth phenotype.**

We adopted a two step strategy to generate *Tbp*<sup>-/-</sup> cell lines expressing mutant hTBPs. Cells were first infected with pBABE retrovirus vectors expressing Wt hTBP or a series of mutants spanning the two repeats of the TBP core region (see Fig. 2A). All hTBPs carry a Flag-HA tag on their N-terminus. Cells expressing endogenous mTBP and the exogenous hTBP were then infected with a second retrovirus expressing the 4 hydroxytamoxifen (OHT) inducible Cre-ER<sup>T2</sup>. In the absence of OHT, a low level of spontaneous deletion of the floxed allele was observed (Fig. 1B, lanes 2, 4), that was strongly stimulated in the presence of OHT (lanes 3 and 4). We were however, unable to directly generate homogenous *Tbp*<sup>-/-</sup> cell populations due to incomplete defloxed. We therefore resorted to picking multiple clonal populations from the OHT treated cells and genotyping them by PCR to identify the *Tbp*<sup>-/-</sup> clones (Supplemental Fig. 1C).

After the first infection with retroviral vectors expressing Wt or mutant hTBPs, the infected cell population was analysed by Western blot. The presence of endogenous mTBP and the hTBP could be observed in most lines (Fig. 1C lanes 2, 3 and 5). In contrast, for several mutants, we were reproducibly unable to observe stable expression of the mutant hTBP, (lanes 1 and 4 and data not shown, summarised in Fig. 2A). These mutants correspond to radical mutations on the DNA binding surface of the first repeat (R203E and T210K), the H2 helix (R239E) or the TFIIB interaction surface on the C-terminal stirrup (E286R) (Fig. 2A). These mutants could therefore not be further analysed.

Cells expressing the other mutant hTBPs were infected with the Cre-ER<sup>T2</sup> retrovirus and then subsequently treated with OHT for 5 days before isolation of clones. In this way, we analysed at least 2 independent clones where expression of the endogenous mTBP was lost and replaced by the exogenous hTBP (see Fig. 1D). Each mutant hTBP was expressed at levels comparable, but not identical to that of the endogenous mTBP in the *Tbp*<sup>lox/-</sup> cells, with V162A, Q242A and R318A showing higher expression. These results indicate that all of these mutants are capable of supporting cell viability (summarised in Fig. 2A).

In contrast, we were unable to isolate *Tbp*<sup>-/-</sup> cells expressing the compound SPM3/N189E and SPM3/L270A/E271A mutants (Lavigne et al., 1999). These two mutants could be expressed in the *Tbp*<sup>lox/-</sup> cells (data not shown), however after addition of OHT only clones expressing the mTBP and hTBP or mTBP alone could be isolated showing that they are unable to complement loss of mTBP. We also asked whether TBP2/TRF3 could replace mTBP (Gazdag et al., 2009). Cells expressing TBP2 and mTBP were generated, however after OHT treatment and isolation of more than 20 clonal lines, no cells expressing only TBP2 were obtained (data not shown). This observation shows that TBP2 cannot replace TBP in these cells.

Although many mutant hTBPs can complement loss of mTBP, the majority of cell lines show a slow growth phenotype (Fig. 2B and summarised in Fig. 2A). The R186E mutation leads to increased proliferation, while the majority of the other mutations moderately or strongly reduce proliferation. At least two independent clones for each mutant were analysed to ensure the reproducibility of this phenotype. Slow growth is particularly severe for the SPM3, L212A, K243E and R294A mutants. Analysis of the R188E and K243E mutant cells by flow analysis did not reveal specific G1/S or G2/M checkpoint arrest (data not shown). The slow growth thus rather results from a general lengthening of all phases of the cycle. We also grew each cell line at 39.5°C to ask if the mutations would show a temperature sensitive phenotype resulting in cell death, however all cells remained viable at the elevated temperature (data not shown).

The above results show that most of the tested mutant hTBPs are capable of supporting cell viability, but show defects in cell proliferation. To determine whether this is due to defects in Pol I and/or Pol III transcription, we investigated the effect of the TBP mutations on expression of transcripts for each of these polymerases. To measure Pol I transcription, we performed RT-qPCR using primers specific for the short lived pre-45S rRNA. In this analysis, most lines showed no significant differences with the exception of R188E and Q242A where a two-fold increase was observed and K265A, where a 2.5 fold decrease was seen (Supplemental Fig. 2A). Similarly, we measured Pol III transcription by RT-qPCR using primers specific for the pre-5S RNA (Supplemental Fig. 2C) or U6 RNA (data not shown), but no significant effects were observed. These results show that these TBP mutations do not have a strong effects on Pol I and Pol III transcription that would account for the slow growth.

### **TBP mutations that specifically disrupt the B-TFIID complex**

The properties of the hTBP mutants used here have been previously evaluated in various *in vitro* assays for their ability to interact with DNA, TFIIA, TFIIB, NC2 and BTAF1 and/or for their ability to form ternary complexes comprising TBP, DNA and one of the additional cofactors in EMSA [(Bryant et al., 1996; Klejman et al., 2005; Tang et al., 1996; Zhao et al., 2003), summarised in Fig. 2A]. In these experiments, complexes were formed from recombinant proteins *in vitro*. We took advantage of the Flag-HA tag on the N-terminus of the mutant TBPs to perform tandem affinity purification of the Wt and mutant hTBPs from nuclear extracts of the corresponding cell lines to address how these mutations may affect interactions with the known partners *in vivo*. Following purification and tryptic digestion, TBP and its partners were directly identified by mass-spectrometry.

TBP and all of the known TFIID TAFs were identified in purifications from cells expressing Wt hTBP, but not in control purifications using the Flag and HA antibodies from the *Tbp<sup>lox/-</sup>* cells with untagged TBP (Fig. 3A and Supplemental Table 1 showing the results of two independent experiments, the first comparing Wt hTBP and R188E with a negative control and the second, Wt hTBP, a negative control and a series of mutants). It should be noted that we did not identify the human orthologue of yTAF14 in these experiments. Similarly, BTAF1, BRF and all 4 SL1/TIF1B subunits were present in the Wt hTBP samples, including the more recently identified TAFID (JOSD3) subunit (Gorski et al., 2007).

In these experiments, spectra for  $\approx 600$  proteins were identified (Supplemental Table 1), however in addition to the known TAF partners, only a small number of other proteins showed a pattern consistent with a specific association with TBP, i.e. not present in the control samples, but present in the Wt and in several of the mutant samples (Supplemental Table 1 and data not shown). Amongst the proteins that have TAF-characteristics is the transcriptional regulator SPEN, a protein family is found in worms, flies, and mammals, and

implicated in diverse biological processes from embryogenesis to aging. The mammalian SPEN proteins MINT and its human ortholog SHARP are nuclear matrix proteins that interact with the Notch-signaling mediator RBP-J as well as MSX2 and several unliganded nuclear hormone receptors, and impart transcription-repressing activity to these molecules by recruiting corepressors through the SPOC domain (Kuroda et al., 2003; Sierra et al., 2004; Tsuji et al., 2007). Similarly, enhancer of decapping 3 and 4 [EDC3 and 4] (Kshirsagar and Parker, 2004; Simon et al., 2006) both appear to specifically associate with TBP (Fig. 3A). Further studies will be required to determine whether these interactions are specific and have functional significance.

Analysis of the results from the mutant TBP samples showed that none of the analysed mutations significantly affected interaction of TBP with the TFIID TAFs. Several mutations appeared to affect interaction with the SL1 subunits in particular TAFIC, however, the number of recorded spectra for these subunits is in general low and a significant loss of this interaction was not confirmed by Western blot (Supplemental Fig. 2A and see below). We also detected the TFIIAa/b and g subunits in the Wt hTBP samples, and this interaction appeared to be affected by several of the mutations. Again, with the low number of observed spectra, we cannot conclude whether these are significant effects. Moreover, under the purification conditions used we did not detect any spectra specific for NC2.

In contrast, two mutations (R188E and K243E) clearly led to a complete loss of interaction with BTAF1 as no spectra corresponding to this protein were found in these samples, while abundant spectra were observed with Wt hTBP. Loss of BTAF1 interaction with these mutants was confirmed by western blot analysis. TBP and TAF4 and TAF6 were precipitated by Flag antibody from cells expressing Wt hTBP, R188E, K243E, and R294A, but not from *Tbp*<sup>lox/-</sup> cells (Fig. 3B). BTAF1 was precipitated with Wt hTBP and R294A, but

not with R188E and K243E. These two mutations therefore lead to a specific loss of interaction with BTAF1 and disruption of the B-TFIID complex defining a surface on TBP critical for this interaction.

### **Changes in gene expression upon disruption of the B-TFIID complex.**

To determine the effect of loss of the TBP-BTAF1 interaction on gene expression, we made a more detailed transcriptome analysis on two independent Wt hTBP and R188E mutant lines. Compared to MEFs expressing Wt hTBP, 274 and 200 genes were up or down-regulated 2-fold or more respectively in R188E MEFs (Fig. 4A and Supplemental Table 2) corresponding to only around 4% of the total expressed genes. Amongst these, are genes that are either strongly induced or repressed and for several of these, the changes in expression were confirmed in RT-qPCR (Fig. 4B and data not shown). Thus, the R188E mutation that compromises B-TFIID stability does not have a general effect on transcription, but rather affects positively or negatively the expression of a small number of genes.

For example, *Rplp0* is a highly expressed TATA-containing gene encoding a acidic ribosomal protein whose expression is not affected by the R188E mutation and is used as a control (Fig. 4B). In agreement with the microarray analysis, R-spondin 2 (*Rspo2*) and potassium channel tetramerisation domain containing 12 (*Kctd12*) were up-regulated in the cells expressing the R188E mutant. In contrast, histocompatibility 2, K1 (H2-k1) and plakophilin 2 (*Pkp2*) were down-regulated. Amongst the up-regulated genes, we noted a 3-fold increase in expression of the cell cycle inhibitor *Cdkn1a* encoding p21<sup>cip1</sup> (Fig. 4B). Increased CDKN1A expression was also seen by immunoblot in cells expressing the R188E mutant (Fig. 4C). The up-regulation of *Cdkn1a* expression may at least in part account for the diminished proliferation of these cells.

**Mutations in TBP affect recruitment to target promoters.**

To understand the molecular basis underlying the changes in gene expression that are seen when B-TFIID integrity is compromised, we performed ChIP-seq experiments on cells expressing Wt or R188E mutant TBPs. ChIP was performed against TBP, Pol II and trimethylated lysine 4 of histone H3 (H3K4me3) and the ChIPed DNA analysed by Illumina sequencing. Peak detection (see Materials and methods and Supplemental material), identified 12 954 TBP binding sites in the genome the majority of which are located at the TSS of annotated genes, although TBP binding sites can be found in intergenic regions with or without associated Pol II (Supplemental Figs. 3 and 4).

We first examined the *Rplp0* locus that is highly expressed and whose expression is not affected by the R188E mutation. At this locus, comparable peaks corresponding to TBP occupancy can be seen in the Wt and R188E-expressing cells along with Pol II occupancy at the promoter and transcribed regions and H3K4me3 at the 5' region (Fig. 5A). The specificity of the ChIPs were verified by ChIP-seq with an anti-GFP antibody that showed no significant signal. Analogous profiles for TBP, Pol II and H3K4 are seen at the majority of the expressed genes in each cell line. A global comparison of the ChIP-seq data for each cell type by read density clustering analysis, shows that TBP and Pol II are present at >10 000 major annotated start sites corresponding to the expressed genes (Fig. 5B). In addition, there is no gross qualitative nor quantitative re-distribution of TBP or Pol II at the TSS in cells expressing the hTBP R188E mutation (Fig. 5C), consistent with the observation that only a small number of genes show altered expression. Thus, loss of B-TFIID complex integrity does not result in a global redistribution of TBP at TSS.

The above results address the localisation of TBP Wt and R188E at the TSS. To determine whether the R188E mutation had an effect on TBP distribution at other sites, we



used clustering analysis to compare global TBP distribution. Comparison of TBP Wt and R188E at the 12 954 Wt TBP occupied loci revealed a group of more than 4000 sites that appeared enriched in the R188E expressing cells (Fig. 6A). However, after additional re-clustering of this group, we identified 1334 loci that are enriched in the R188E cells and 397 loci that are weakly occupied in the R188E, but show no signal in Wt cells. This is confirmed by a quantitative comparison showing that these are sites weakly occupied in by TBP R188E, but show no significant occupancy by Wt TBP (Fig. 6B). In addition, the vast majority of these sites are located in inter and intragenic regions, rather than at the TSS of annotated genes.

We next examined TBP and Pol II occupancy at the loci of differentially regulated genes. As shown above, *Kctd12* and *Rspo2* are two genes that are not significantly expressed in cells expressing Wt TBP, but are strongly induced in cells expressing the R188E mutant (Fig. 4B). At the *Kctd12* locus, no significant H3K4me3, Pol II or TBP occupancy are observed, while in the R188E expressing cells, robust signals could be observed for all three (Fig. 7A). The ChIP-seq data were verified by independent ChIP-qPCR experiments showing Pol II, H3K4 and TBP occupancy only in the R188E-expressing cells (Fig. 7C). At this locus the single amino acid change in TBP allows its recruitment along with the rest of the transcription machinery to initiate expression of this gene.

At the *Rspo2* promoter, both H3K4 trimethylation and Wt TBP occupancy can be clearly observed, but no Pol II is observed at this promoter (Fig. 7B). In the R188E-expressing cells, similar signals for TBP and H3K4me3 are observed and Pol II is also present. Analogous observations were made in independent ChIP-qPCR experiments, where significant TBP occupancy is seen in the Wt hTBP cells that is increased in the R188E-expressing cells, but Pol II is recruited only in the cells expressing mutant TBP (Fig. 7C).

Thus, in contrast to *Kctd12* whose expression activated by *de novo* recruitment of TBP and the transcription machinery, Wt hTBP is present on the *Rspo2* promoter but does not recruit Pol II that is recruited only with the R188E mutant TBP.

One interpretation of these observations is that the R188E mutation is required for TBP to recruit Pol II at *Rspo2* promoter. Alternatively, it is possible that in cells expressing Wt TBP, the TBP present at this promoter represents the B-TFIID complex, while in the R188E-expressing cells the promoter is occupied by TBP present in TFIID. To test this possibility we performed ChIP-qPCR experiments with antibodies against other components of the basal transcription complex. None of the existing antibodies allow significant ChIP of mouse BTAF1 and we have not succeeded in generating ChIP grade antibodies for the mouse protein (data not shown). Similarly, amongst a panel of antibodies that ChIP the general transcription machinery in human cells, we saw significant ChIP-signals using only antibodies against the ERCC3 subunit of TFIID, the MED6 subunit of the mediator complex and the TAF1 subunit of the TFIID complex. Each of these antibodies show significant enrichment at the *Rplp0* promoter compared to the -1kb control region (Fig. 5D), as well as other non-promoter regions (data not shown).

At the *Kctd12* promoter, the presence of these factors is observed only in the R188E-expressing cells consistent with what is seen with TBP and Pol II confirming the *de novo* recruitment of the transcription machinery and gene activation (Fig. 7C). At the *Rspo2* promoter, significant occupancy for TFIID and MED6 are observed in cells expressing Wt TBP showing that TBP and at least some of the general machinery is present at this promoter. However, unlike TBP that is present at this promoter in Wt hTBP cells, TAF1 is recruited only in the cells expressing the R188E mutant. These results are in agreement with a model where Wt TBP is present in the form of B-TFIID that initiates PIC formation, but does not

allow recruitment of Pol II, while in the cells expressing the R188E mutant that does not form B-TFIID, TFIID is present allowing recruitment of Pol II and transcription.

Further support for this idea comes from analysis of TBP and TAF1 occupancy of the *Rspo2* promoter in cells expressing the K243E mutation that also disrupts B-TFIID integrity. Expression of *Rspo2* is strongly activated by this mutation (Fig. 4B). ChIP indicates that TBP K243E is present at this promoter and like TBP R188E recruits TAF1 and Pol II (Fig. 7D). Thus, in cells expressing mutant TBPs that do not form the B-TFIID complex, TFIID along with Pol II are recruited to activate *Rspo2* expression.

At genes showing down-regulation such as *H2-k1*, the TBP seen at its promoter in Wt cells is strongly reduced in the R188E mutant cells, while there is a complete loss of Pol II although H3K4me3 persists (Fig. 8A and B). These results were confirmed by ChIP-qPCR and show that the R188E mutation leads to reduced TBP recruitment and loss of TBP at this promoter. A similar situation was seen at the promoters of other down-regulated genes for example, plakophilin 2 (*Pkp2*) or coxsackie virus and adenovirus receptor (*Cxadr*). At each promoter, while TBP and Pol II are no longer present H3K4me3 persists (Supplemental Fig. 5).

The differential recruitment of Wt and mutant TBP does not always have functional consequences. For example, TBP-R188E is strongly recruited to the promoter of the divergent potassium voltage-gated channel, shaker-related subfamily, beta member 2 (*Kcnab2*) and nephronophthisis 4 (*Nphp4*) genes, but there is no recruitment of Pol II and no expression of these genes (Supplemental Fig. 6). Thus, TBP recruitment does not necessarily lead to gene activation.

## **Discussion.**

**TBP surfaces required for function in living cells.**

Using genetically modified MEFs, we have investigated the ability of TBPs bearing for the most part single amino acid substitutions to complement lethality upon endogenous TBP inactivation. Several mutant TBPs could not be stably expressed even in cells expressing Wt TBP. Lack of expression was not due to spurious causes, but most likely reflects the dominant negative activity of the mutant TBPs. Mutation R203E has been shown to affect interaction with DNA (Bryant et al., 1996; Klejman et al., 2005) and TFIIA (Bryant et al., 1996), while T210K is a radical mutation in the first half of the DNA binding surface. These mutant TBPs interact normally with the TAFs partners, but cannot bind TATA-containing promoter DNA and thus act as dominant negatives. The other mutations that we analysed on the DNA binding surfaces are alanine substitutions rather than radical mutations and while, they complement loss of endogenous TBP, all lead to significantly reduced proliferation.

On the other hand, E286R is a radical mutation in the TFIIB interaction surface (Bryant et al., 1996). This mutant interacts normally with TAFs and binds promoter DNA, but fails to interact with TFIIB thus acting as a dominant negative. TBP-TFIIB interaction is thus essential for cell viability. The R239A mutation on the H2 helix has been previously shown to strongly suppress transcription from the U1 promoter (Klejman et al., 2005), while the R239S mutation essentially abolishes activated transcription in transfected cells (Bryant et al., 1996). Similar effects with the radical mutation R239E would explain the dominant negative effect of this mutant. Note however that loss of the ability to support activated transcription is not due to a failure to form the TFIID complex.

Several other mutations were unable to complement loss of endogenous TBP, but did not act as dominant negatives. This is the case for the combination of the N189E or L270A/E271A mutations with Spm3. The N189E mutation affects formation of the

TBP/TFIIA/DNA complex (Klejman et al., 2005) and has a strongly detrimental effect on U1 and U6 transcription that could account for its inability to complement (Zhao et al., 2003). Apart from its effect on RAR activity (Lavigne et al., 1999), the L270A/E271A mutation is less well characterised. It should be noted however that the Spm3 mutation on its own, while viable leads to a severe slow growth phenotype. It is thus possible that additional mutations in this already compromised background lead to a complete loss of function. This has previously been noted in the R188E and N189E mutations that their own have no effect on transcription of the AdMLP, but are strongly detrimental when combined with the Spm3 mutation (Bryant et al., 1996; Zhao et al., 2003)

Our data show that mutations previously shown to strongly compromise specific TBP interactions or activated transcription in transfected cells are nevertheless capable of complementation. This is the case for V162A, R188E, L212A and K249A. V162A has been shown to prevent formation of the TBP-TFIIB-DNA complex *in vitro* (Klejman et al., 2005). R188E affects formation of the TBP-NC2-DNA complex, strongly reduces transcription from the U1 and U6 promoters *in vitro* (Zhao et al., 2003), affects interaction with TFIIA and abolishes activated transcription (Bryant et al., 1996; Tansey et al., 1994). L212A strongly down-regulates U6 transcription *in vitro* (Zhao et al., 2003), while K249A inhibits formation of the NC2/TFIIA/TFIIB-TBP-DNA complexes (Klejman et al., 2005). Despite the effects of these mutants *in vitro*, they support MEF cell viability, and while reduced proliferation is seen for the V162A, R188E and L212A mutants, K249 expressing cells show normal proliferation. Moreover, while R188E affects TFIIA interactions *in vitro*, ChIP analysis shows no effect of TFIIA recruitment to the *Rplp0* promoter in MEFs. These results indicate that while mutations may affect a specific interaction *in vitro*, in the context of the living cell additional interactions such as those between the bromo domains of TAF1, the PHD domain of TAF3 and modified histone lysine residues (Jacobson et al., 2000; Vermeulen et al., 2007) or

interactions with TAFs such as that between TFIIA and TAF4 (Guermah et al., 2001) may stabilise the PIC and overcome weakend interactions of TBP with DNA or one of its partners. On the other hand, mutations such as V240A, and Q242A that have no described effect on interactions *in vitro*, strongly reduce cell proliferation *in vivo* suggesting that they affect interactions with as yet unidentified cofactors.

### **Loss of B-TFIID integrity does not have a major effect on transcription.**

Tandem immunopurification and proteomics analysis showed that TBP mutants that support MEF viability associate normally with the TFIID and SL1 TAFs as well as with BRF1. This is not surprising as loss of these critical interactions would be expected to compromise cell survival. Our results clearly show however that, under the same conditions, two mutations R188E and K243E lead to a selective disruption of interaction with BTAF1 and hence of the B-TFIID complex. These two mutations were previously shown to affect TBP-BTAF1 interactions *in vitro* (Klejman). However, the R188E mutation had a milder effect *in vitro* than other mutations, such as V162A, R186E, L212A, or R294A that do not lead to loss of B-TFIID complex assembly in MEFs. These differences show that assembly of the complex *in vitro* from recombinant proteins has different requirements from the chaperone assisted assembly that takes place within the cell. It has also been shown that the presence of DNA can stabilise the TBP-BTAF1 interaction as several mutants defective for the TBP-BTAF1 interaction can form a ternary complex in the presence of DNA. However this is not the case for the R188E and K243E mutations in agreement with the observation that they have the most potent effects in cells.

Transcriptome analysis shows that loss of the B-TFIID complex does not have a general effect on transcription, but affects the expression of only a small subset of genes. Both activation and repression is observed similar to what is observed upon loss of yBTAF1

(Andrau et al., 2002; Dasgupta et al., 2005). In yeast, BTAF1 has also been shown to affect Pol I transcription (Dasgupta et al., 2007), but this does not seem to be the case in mammalian cells. In agreement with the limited effects on gene expression, loss of B-TFIID does not result in a global change in TBP occupancy. We identified only a limited number of mostly inter and intragenic loci showing higher occupancy in the R188E expressing cells than in Wt cells.

These observations are unexpected given previous studies involving knockdown of BTAF1 expression. BTAF1 knockdown has been shown to lower TBP mobility in the cell and increase TBP occupancy at several promoters (de Graaf et al., 2010) and maybe also more globally in the genome (Johannes et al.). It was also previously suggested that yBTAF1 acts to mobilise TBP from promoter to non-promoter sites or amongst promoter sites (Auble, 2009; Collart, 1996; Pereira et al., 2003). In contrast, we show that loss of interaction with BTAF1 has only a minor effect on TBP distribution at promoters and rather promotes TBP accumulation at non-promoter sites suggesting that BTAF1 normally removes TBP from these non-specific sites mobilising it for use at promoter sites as previously suggested (Muldrow et al., 1999). These differences may be explained in several ways. The R188E and K243E mutations, while disrupting TBP-BTAF1 interactions *in vitro* and when we extract TBP from cells, do not disrupt interaction with TBP in the living cell. Alternatively, the documented effects of BTAF1 do not all require its interaction with TBP, implying that BTAF1 can act on TBP and promoter dynamics independently of formation of a stable BTFIID complex.

While we do not observe a general increase in promoter TBP occupancy as observed upon BTAF1 knock-down, the R188E mutation does affect TBP occupancy both positively and negatively at a limited subset of promoters. The affected promoters for the most part

show low TBP occupancy that is often at the limit of detection. Analysis of TBP occupancy at the differentially regulated genes revealed at least two distinct situations. At genes like *Kcdt12*, Wt TBP is not recruited and the locus appears inactive with no H3K4me3 (Fig. 9). In mutant R188E cells, the locus is activated with recruitment of TBP and Pol II. Many other up-regulated promoters show a similar profile of TBP and Pol II occupancy. These results imply that the R188E mutation affects interaction with a repressor that normally inhibits recruitment of TBP to these promoters. This inhibitor may be BTAF1 which would normally dissociate TBP from these promoters leading to their repression, but in the absence of this interaction TBP is now free to stably associate with the promoter and initiate PIC assembly. Alternatively, the R188E mutation may affect interaction with another as yet un-identified repressor.

At the *Rspo2* promoter, on the other hand, our results suggest that in Wt cells TBP is recruited in the form of B-TFIID that allows formation of a partial PIC, but not Pol II recruitment. In the R188E and K243E mutant cells, TBP is now recruited to the promoter in TFIID that is competent to recruit Pol II and activate transcription (Fig. 9). Amongst the up-regulated genes, a similar situation may also exist at the *Eya4*, and *Vegfc* promoters, where H3K4me3 and low levels of TBP can be seen in the Wt cells, but the genes are activated only in the R188E mutant cells. These promoters are not characterised by the presence or absence of a TATA element that is present at *Rspo2*, but not at *Eya4* and *Vegfc*, and we have not observed other characteristics of these promoters that would confer preferential recruitment of B-TFIID rather than TFIID.

In contrast to the above, we also identify genes that are down-regulated by the R188E mutation. At many of these loci, we observe a total or partial loss of TBP, a total loss of Pol II recruitment, but not a complete silencing as H3K4me3 persists at these loci. These



observations suggest that the TBP-BTAF1 interaction is required for normal TBP recruitment and PIC formation at these loci. Nevertheless, it should be noted that the R188E (and K243E) mutations may affect interaction with other cofactors such as NC2, TFIIA or even transcriptional activators. The loss of activity of the mutant TBPs at these promoters may reflect a requirement for any one or combination of these interactions. Also, not all promoters are similarly regulated by the R188E and K243E mutations, for example *Kctd12* is activated by R188E, but not K243E. It is likely that the effects of these mutations are due to changes in interactions not only with BTAF1, but also other cofactors.

### **Conclusion.**

Together our results provide novel insights into TBP interactions with its multiple cofactors and the requirement for these interactions for cell viability and transcription in living cells. Our results show that formation of a stable B-TFIID complex is not required for cell viability, but is required for normal proliferation. The TBP-BTAF1 interaction is not required for normal regulation of the vast majority of genes, but has very specific effects at a subset of promoters. Analysis of these promoters shows that the TBP mutations can affect either recruitment of TBP or promote exchange between B-TFIID and TFIID to elicit changes in gene expression.

### **Acknowledgements**

We thank, C.Thibault for the Affymetrix arrays, H. Stunnenberg for gift of anti-TFIIA antibody, L. Tora for TAF1 antibodies. This work was supported by grants from the CNRS, the INSERM, the Association pour la Recherche contre le Cancer (ARC), the Ligue Nationale et Départementale Région Alsace contre le Cancer. ID is an 'équipe labéllisée' of the Ligue Nationale contre le Cancer. The ANR grant No. 07-0043-01 M-A. C. and D.K. were supported by a fellowships from the ARC and ANR respectively.

**References:**

- Andrau, J. C., Van Oevelen, C. J., Van Teeffelen, H. A., Weil, P. A., Holstege, F. C., and Timmers, H. T. (2002). Mot1p is essential for TBP recruitment to selected promoters during in vivo gene activation. *Embo J* 21, 5173-5183.
- Auble, D. T. (2009). The dynamic personality of TATA-binding protein. *Trends Biochem Sci* 34, 49-52.
- Auble, D. T., Hansen, K. E., Mueller, C. G., Lane, W. S., Thorner, J., and Hahn, S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *GenesDev* 8, 1920-1934.
- Bryant, G. O., Martel, L. S., Burley, S. K., and Berk, A. J. (1996). Radical mutations reveal TATA-box binding protein surfaces required for activated transcription in vivo. *GenesDev* 10, 2491-2504.
- Burley, S. K., and Roeder, R. G. (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *AnnuRevBiochem* 65, 769-799.
- Cler, E., Papai, G., Schultz, P., and Davidson, I. (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci* 66, 2123-2134.
- Collart, M. A. (1996). The NOT, SPT3, and MOT1 genes functionally interact to regulate transcription at core promoters. *MolCell Biol* 16, 6668-6676.
- Dasgupta, A., Juedes, S. A., Sprouse, R. O., and Auble, D. T. (2005). Mot1-mediated control of transcription complex assembly and activity. *Embo J* 24, 1717-1729.

- Dasgupta, A., Sprouse, R. O., French, S., Aprikian, P., Hontz, R., Juedes, S. A., Smith, J. S., Beyer, A. L., and Auble, D. T. (2007). Regulation of rRNA synthesis by TATA-binding protein-associated factor Mot1. *Mol Cell Biol* 27, 2886-2896.
- de Graaf, P., Mousson, F., Geverts, B., Scheer, E., Tora, L., Houtsmuller, A. B., and Timmers, H. T. Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *J Cell Sci*.
- Delacroix, L., Moutier, E., Altobelli, G., Legras, S., Poch, O., Choukrallah, M. A., Bertin, I., Jost, B., and Davidson, I. (2010). Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Mol Cell Biol* 30, 231-244.
- Fuda, N. J., Ardehali, M. B., and Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186-192.
- Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M. E. (2009). TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev* 23, 2210-2223.
- Gorski, J. J., Pathak, S., Panov, K., Kasciukovic, T., Panova, T., Russell, J., and Zomerdijk, J. C. (2007). A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *Embo J* 26, 1560-1568.
- Guermah, M., Tao, Y., and Roeder, R. G. (2001). Positive and negative TAF(II) functions that suggest a dynamic TFIID structure and elicit synergy with traps in activator-induced transcription. *Mol Cell Biol* 21, 6882-6894.

- Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology & Molecular Biology Review* (Washington, DC) *62*, 465-503.
- Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *GenesDev* *7*, 1291-1308.
- Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module [see comments]. *Science* *288*, 1422-1425.
- Johannes, F., Wardenaar, R., Colome-Tatche, M., Mousson, F., de Graaf, P., Mokry, M., Guryev, V., Timmers, H. T., Cuppen, E., and Jansen, R. C. Comparing genome-wide chromatin profiles using ChIP-chip or ChIP-seq. *Bioinformatics* *26*, 1000-1006.
- Juo, Z. S., Kassavetis, G. A., Wang, J., Geiduschek, E. P., and Sigler, P. B. (2003). Crystal structure of a transcription factor IIIB core interface ternary complex. *Nature* *422*, 534-539.
- Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R. G., Meisterernst, M., and Burley, S. K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell* *106*, 71-81.
- Klejman, M. P., Zhao, X., van Schaik, F. M., Herr, W., and Timmers, H. T. (2005). Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. *Nucleic Acids Res* *33*, 5426-5436.

- Kobi, D., Steunou, A. L., Dembele, D., Legras, S., Larue, L., Nieto, L., and Davidson, I. (2008). Genome-wide analysis of POU3F2/BRN2 promoter occupancy in human melanoma cells reveals Kitl as a novel regulated target gene. *Pigment Cell Melanoma Res.*
- Krebs, A., Frontini, M., and Tora, L. (2008). GPAT: retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* 9, 533.
- Kshirsagar, M., and Parker, R. (2004). Identification of Edc3p as an enhancer of mRNA decapping in *Saccharomyces cerevisiae*. *Genetics* 166, 729-739.
- Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity* 18, 301-312.
- Lavigne, A. C., Gangloff, Y. G., Carr, L., Mengus, G., Birck, C., Poch, O., Romier, C., Moras, D., and Davidson, I. (1999). Synergistic transcriptional activation by TATA-binding protein and hTAFII28 requires specific amino acids of the hTAFII28 histone fold [In Process Citation]. *MolCell Biol* 19, 5050-5060.
- Lescure, A., Lutz, Y., Eberhard, D., Jacq, X., Krol, A., Grummt, I., Davidson, I., Chambon, P., and Tora, L. (1994). The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *EMBO J* 13, 1166-1175.
- Martianov, I., Viville, S., and Davidson, I. (2002). RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science* 298, 1036-1039.

- Mengus, G., Fadloun, A., Kobi, D., Thibault, C., Perletti, L., Michel, I., and Davidson, I. (2005). TAF4 inactivation in embryonic fibroblasts activates TGFbeta signalling and autocrine growth. *Embo J* 24, 2753-2767.
- Mitsiou, D. J., and Stunnenberg, H. G. (2000). TAC, a TBP-sans-TAFs complex containing the unprocessed TFIIAalpha precursor and the TFIAGamma subunit. *Mol Cell* 6, 527-537.
- Muldrow, T. A., Campbell, A. M., Weil, P. A., and Auble, D. T. (1999). MOT1 can activate basal transcription in vitro by regulating the distribution of TATA binding protein between promoter and nonpromoter sites. *Mol Cell Biol* 19, 2835-2845.
- Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377, 119-128.
- Pereira, L. A., Klejman, M. P., and Timmers, H. T. (2003). Roles for BTAF1 and Mot1p in dynamics of TATA-binding protein and regulation of RNA polymerase II transcription. *Gene* 315, 1-13.
- Sierra, O. L., Cheng, S. L., Loewy, A. P., Charlton-Kachigian, N., and Towler, D. A. (2004). MINT, the Msx2 interacting nuclear matrix target, enhances Runx2-dependent activation of the osteocalcin fibroblast growth factor response element. *J Biol Chem* 279, 32913-32923.
- Simon, E., Camier, S., and Seraphin, B. (2006). New insights into the control of mRNA decapping. *Trends Biochem Sci* 31, 241-243.

- Strubin, M., and Struhl, K. (1992). Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell* 68, 721-730.
- Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex [see comments]. *Nature* 381, 127-151.
- Tang, H., Sun, X., Reinberg, D., and Ebricht, R. H. (1996). Protein-protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex. *ProcNatlAcadSciUSA* 93, 1119-1124.
- Tansey, W. P., Ruppert, S., Tjian, R., and Herr, W. (1994). Multiple regions of TBP participate in the response to transcriptional activators in vivo. *GenesDev* 8, 2756-2769.
- Thomas, M. C., and Chiang, C. M. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 41, 105-178.
- Timmers, H. T., Meyers, R. E., and Sharp, P. A. (1992). Composition of transcription factor B-TFIID. *ProcNatlAcadSciUSA* 89, 8140-8144.
- Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev* 16, 673-675.
- Tsuji, M., Shinkura, R., Kuroda, K., Yabe, D., and Honjo, T. (2007). Msx2-interacting nuclear target protein (Mint) deficiency reveals negative regulation of early thymocyte differentiation by Notch/RBP-J signaling. *Proc Natl Acad Sci U S A* 104, 1610-1615.
- Van Der Knaap, J. A., Van Den Boom, V., Kuipers, J., Van Eijk, M. J., Van Der Vliet, P. C., and Timmers, H. T. (2000). The gene for human TATA-binding-protein-associated

factor (TAFII) 170: structure, promoter and chromosomal localization. *Biochem J* 345 Pt 3, 521-527.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007). Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* 131, 58-69.

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

Zhao, X., Schramm, L., Hernandez, N., Herr, W., Reilly, P. T., Wysocka, J., Kuball, J., Schuler, M., Antunes Ferreira, E., Neumann, M., *et al.* (2003). A shared surface of TBP directs RNA polymerase II and III transcription via association with different TFIIB family members. *Mol Cell* 11, 151-161.

### **Legends to figures.**

**Figure 1.** Exogenous human TBP complements loss of endogenous murine TBP. **A.** Schematic representation of null, floxed and deleted alleles of the mouse TBP gene showing exons II, III, and IV, LoxP (L) sites and the neomycin and hygromycin resistance cassettes. The localisation of PCR primers A-C used for genotyping are shown. **B.** Example of PCR genotyping of MEF cells shown the null, floxed and deleted allele PCR products. Cells were treated with OHT as indicated above each lane. **C.** Western blot analysis of expression of endogenous mouse and exogenous human TBP in MEFs infected with the retroviruses expressing the TBP mutants shown above each lane. Endogenous mouse TAF4 is shown as



control. **D.** Western blot analysis of cloned cell lines expressing the indicated exogenous hTBP mutants. The \* indicates that in some extracts the Flag-HA tag has been cleaved from TBP. Endogenous mouse TAF4 is shown as control.

**Figure 2.** Viability and proliferation properties of MEFs expressing mutant TBPs. **A.** Summary of cell lines expressing the analysed mutant TBPs, showing the mutated amino acid, its location in the core domain structure, viability, proliferation, the interaction that have been shown to be affected by the mutation in vitro assays, whether it affects Pol II transcription from an assayed promoter, its ability to support activation in transient expression assays from a TGTA A-containing promoter when combined with the SPM3 mutation. NA is not applicable and DN is dominant negative. **B.** Representative proliferation assay performed by counting cell numbers over a 6 day period for the indicated mutants and the initial *Tbp*<sup>Lox/-</sup> line.

**Figure 3.** Effect of TBP mutations on interaction with partner proteins. **A.** Results of mass-spectrometry analysis showing the number of unique spectra obtained for the each partner protein after purification of the indicated mutant TBP. C is the negative control immunoprecipitation from the un-tagged cells. Two independent series of experiments are shown along with corresponding negative control. All spectra for TAF4 and TAF4b, TAF9 and TAF9b are grouped together. Partner proteins are grouped by complexes. **B.** Western blot analysis of TBP, BTAF1, TAF4, and TAF6 in immunoprecipitates from cells expressing the indicated TBP mutants.

**Figure 4.** Analysis of changes in gene expression in TBP R188E expressing cells. **A.** Most affected genes from the Affymetrix table showing the MGI gene symbol and the log<sub>2</sub> change in expression. **B.** RT-qPCR analysis of gene expression. The relative expression of the indicated genes in the cells expressing TBP R188E and K243E are shown. **C.** Western

blot showing enhanced expression of the CDKN1A protein in two independent clones of cells expressing TBP R188E.

**Figure 5.** Genome wide distribution of TBP in Wt and R188E expressing cells. **A.** Screen shot of .Wig files of the TBP, Pol II and H3K4me3 ChIP-seq from cells expressing Wt and R188E TBPs in the UCSC browser at the *Rplp0* locus. **B.** Comparative clustering of ChIP-seq data at the TSS of > 30 000 annotated Ensembl transcripts in Wt and R188E expressing cells. **C.** Quantitative comparison of the TBP ChIP-seq data in Wt and R188E cells. **D.** ChIP-qPCR with the indicated antibodies in the Wt and R188E expressing cells and qPCR with primers covering the *Rplp0* TSS and the region 1Kb upstream of the TSS.

**Figure 6.** Loci enriched in TBP in R188E expressing cells. Comparative clustering of Wt and R188E ChIP-seq data reveals a set of sites that appear to show lower occupancy in Wt than in R188E cells. These sites were selected and subjected to two additional round of clustering to show a set of sites that are occupied in both cell types, but enriched in R188E (group A) and those that are occupied only in R188E cells (group B).

**Figure 7.** Differing mechanisms of TBP and Pol II recruitment at the *Kctd12* and *Rspo2* promoters. **A-B.** UCSC screen shot of TBP, Pol II and H3K4me3 ChIP-seq at the *Kctd12* and *Rspo2* loci respectively. **C.** ChIP-qPCR with the indicated antibodies in the Wt and R188E expressing cells and qPCR with primers covering the TSS of the *Kctd12* and *Rspo2* genes with the region 1Kb upstream of the *Rplp0* TSS as negative control. **D.** ChIP-qPCR with the indicated antibodies in the Wt and K243E expressing cells.

**Figure 8.** Loss of TBP recruitment at the *H2k1* locus. **A.** UCSC screen shot of TBP, Pol II and H3K4me3 ChIP-seq at the *H2k1* locus. **B.** ChIP-qPCR with the indicated antibodies

in the Wt and R188E expressing cells and qPCR with primers covering the TSS of the *H2k1* genes with the region 1Kb upstream of the *Rplp0* TSS as negative control.

**Figure 9.** Cartoons illustrating models for the differential effect of mutations in TBP on TBP recruitment and gene expression. Genes may be activated by do novo recruitment of TBP or by promoting exchange of B-TFIID for TFIID. The models also illustrate that TBP recruitment is not always a productive event at non-promoter regions.

## **Supplemental text**

### **Materials and methods.**

#### **ChIP.**

Cells were crosslinked with formaldehyde 0.4% for 10 min at room temperature and the reaction was stopped by adding glycine to final concentration 0.2M for 10 minutes at room temperature. Fixed cells were rinsed twice with PBS and resuspended in 100 ul of lysis buffer (10 mM EDTA pH 8, 50 mM Tris-HCl pH 8, SDS 1 %). Lysate was sonicated 30 min (30 sec on / 30 sec off) in Diagenode water bath-sonicator and centrifuged at 14000 rpm for 10 min. The cleared supernatant was used immediately in ChIP experiments or stored at -80°C.

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

hours minimum and DNA was submitted to RNase and proteinase K digestion and extracted by phenol-chloroform.

### **ChIP-sequencing.**

ChIP-seq experiments were performed according to standard protocols as previously described (Delacroix et al., 2010; Kobi et al.). ChIP-seq was performed using an Illumina GAIIX sequencer and the raw data analysed by the Illumina Eland pipeline V1.6. Peak detection was performed using the MACS software (<http://liulab.dfci.harvard.edu/MACS/>) (Zhang et al., 2008). Peaks were then annotated using GPAT [(Krebs et al., 2008), [http://bips.u-strasbg.fr/GPAT/Gpat\\_home.html](http://bips.u-strasbg.fr/GPAT/Gpat_home.html)] using a window of +/- 20kb with respect to the coordinates of the beginning and end of RefSeq transcripts. Global clustering of the ChIP-seq data was performed using the in house developed seqMINER programme (Ye et al., 2010 and <http://bips.u-strasbg.fr/seqminer/> ).

### **Proteomics analysis.**

#### **Sample digestion and desalting**

Precipitated affinity-purified protein complexes were dissolved in 25  $\mu$ L buffer A (50 mM ammonium bicarbonate, 8 M urea) and digested with 0.35  $\mu$ g Lys-C for 4 hours at room temperature. Disulfide bridges were reduced with 2 mM DTT in buffer A for 15 min at 50°C, followed by alkylation with 4 mM iodoacetamide in buffer A for 15 min at room temperature. The mixtures were then diluted to a urea concentration of 2 M using 50 mM ammonium bicarbonate and digested with 0.7  $\mu$ g Trypsin for 16 hours at 37°C. Samples were acidified with 0.5% (v/v) formic acid and purified using Empore C18 extraction disks (3M) packed into pipette tips as described previously (Rappsilber et al., 2007).

### **Strong cation exchange (SCX) chromatography**

Eluted peptides were dried to completion in a vacuum centrifuge, dissolved in solvent A (0.05% (v/v) formic acid in 1:4 acetonitrile:water), and fractionated by strong cation exchange (SCX), which was performed using a Shimadzu LC-9A binary pump with two ZORBAX BioSCX-Series II columns (50 mm x 0.8 mm, 3.5  $\mu$ m particle size, Agilent) in series, connected to a SPD-6A UV-detector (Shimadzu) and a FAMOS autosampler (Dionex). Peptide solutions were loaded onto the columns at 50  $\mu$ L/min solvent A for 10 min, followed by linear gradient elution of 1.3% (v/v) 1/min solvent B (500 mM NaCl in solvent A) at 50  $\mu$ L/min. 40 fractions of 1 min, 1 min each, were collected.

### **Liquid chromatography (LC) and mass spectrometry (MS)**

SCX fractions were dried to completion in a vacuum centrifuge, dissolved in 5% (v/v) formic acid and analyzed by nanoflow-LC coupled to an LTQ-FTICR mass spectrometer (Thermo Fisher Scientific). The LC was set up on an Agilent 1200 series HPLC system equipped with a 20 mm Aqua C18 (Phenomenex) trapping column (100  $\mu$ m inner diameter, 5  $\mu$ m particle size) and a 200 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH) analytical column (50  $\mu$ m inner diameter, 3  $\mu$ m particle size). Trapping was performed at 5  $\mu$ L/min solvent C (0.1 M acetic acid in water) for 10 min and elution was achieved with 10 to 40% (v/v) solvent D (0.1 M acetic acid in 1:4 acetonitrile:water in either 22 or 35 min, followed by 38 to 100% (v/v) solvent D in 3 min and 100% solvent D for 2 min. The flow rate was passively split from 0.45 mL/min to 100 nL/min as described previously (Pinkse MW, 2004). Electrospray was achieved using a distally coated fused silica emitter (360  $\mu$ m outer diameter, 20  $\mu$ m inner diameter, 10  $\mu$ m inner diameter of emitter tip, New Objective) biased to 1.7 kV. The LTQ-FT was operated in the data dependent mode to automatically switch between MS and MS/MS. MS spectra in the range of 150 or 350 to 1500 Th were acquired in the FT analyzer with a

resolution of 100,000 at  $m/z$  400 after accumulation to an automatic gain control (AGC) target value of 2,000,000 in the linear ion trap. The two most intense precursor ions with a charge bigger 1 and above a threshold of 500 were selected for fragmentation by collision-induced dissociation (CID) under dynamic exclusion for 30 seconds. CID was performed in the linear ion trap after accumulation to an AGC target value of 10,000.

### **Data processing and analysis**

Spectra were processed with Bioworks (version 3.1, Thermo Fisher Scientific) and the subsequent data analysis was carried out using Mascot (version 2.2.1, Matrix Science). Mascot was set up to search the IPI mouse database (version 3.34, [www.ebi.ac.uk/ipi/](http://www.ebi.ac.uk/ipi/)) with carbamidomethyl cysteines as fixed modification and oxidation of methionines and carbamylation of peptide N-termini as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 15 ppm, and that of fragment ions was set to 0.9 Da. Mascot results files were filtered to contain only peptides above a Mascot ion score of 15 using in-house written software. Scaffold (version 2.01.02, Proteome Software) was used to validate protein identifications. Protein identifications were accepted if they could be established at greater than 99.9% probability and contained at least 2 identified peptides. Human TBP and its variants were identified following essentially the same procedure but using a manually generated database containing sequences of mouse TBP (Swiss-Prot accession number P29037), human TBP (Swiss-Prot accession number P20226), and those of the TBP mutants used in this study.

### **Supplemental References.**

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

Kobi, D., Steunou, A. L., Dembele, D., Legras, S., Larue, L., Nieto, L., and Davidson, I. Genome-wide analysis of POU3F2/BRN2 promoter occupancy in human melanoma cells reveals Kitl as a novel regulated target gene. *Pigment Cell Melanoma Res.*

Krebs, A., Frontini, M., and Tora, L. (2008). GPAT: retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* 9, 533.

Pinkse MW, U. P., Hilhorst MJ, Ooms B, Heck AJ. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Annal Chem* 76, 3935-3943.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2, 1896-1906.

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

**Legends to Supplemental Figures and Tables.**



**Supplemental Figure 1.** Characteristics of *Tbp*<sup>lox/-</sup> MEFs. **A.** Immunofluorescence labelling with anti-TBP antibody of cells transfected with vectors encoding the Cre recombinase and GFP. A GFP positive cell lacking nuclear TBP staining is indicated. **B.** PCR genotyping of *Tbp*<sup>lox/-</sup> MEFs transfected with vectors encoding Cre recombinase and Wt or SPM3 mutant TBP. The PCR product from the deleted allele is detected at the indicated number of days post-transfection. **C.** PCR genotyping of clonal populations of *Tbp*<sup>lox/-</sup> MEFs expressing the indicated exogenous hTBPs and Cre-ER<sup>T2</sup> following OHT treatment.

**Supplemental Figure 2.** Effect of TBP mutations on Pol I and Pol III transcription. **A.** RT-qPCR quantification of pre-45S rRNA in MEFs expressing the indicated mutant TBP. A control on Wt cells in presence of actinomycin D is shown in the first 3 lanes. **B.** Western blot showing coprecipitation of TAF4 and TAFIC from MEFs expressing the indicated TBP mutants following anti-Flag immunoprecipitation. **C.** RT-qPCR quantification of pre-5S rRNA in MEFs expressing the indicated mutant TBP.

**Supplemental Figure 3. A.** TBP binding profile relative to the TSS

**Supplemental Figure 4.** TBP occupancy of intergenic regions. **A-B.** UCSC screen shot of TBP, Pol II and H3K4me3 ChIP-seq at the indicated loci.

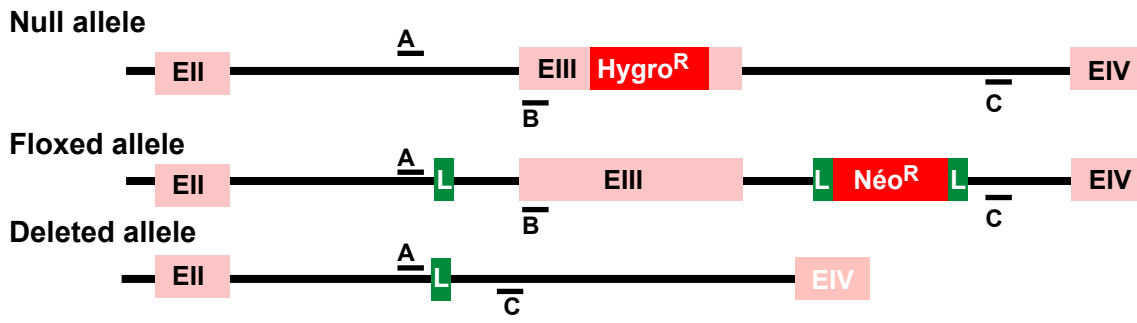
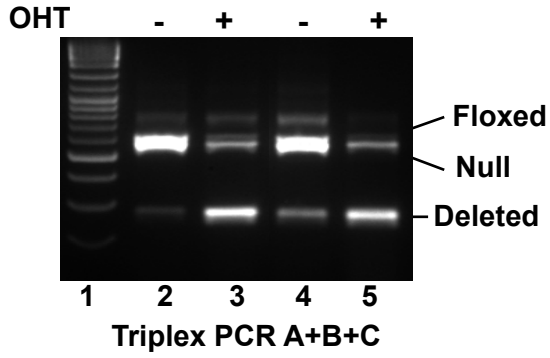
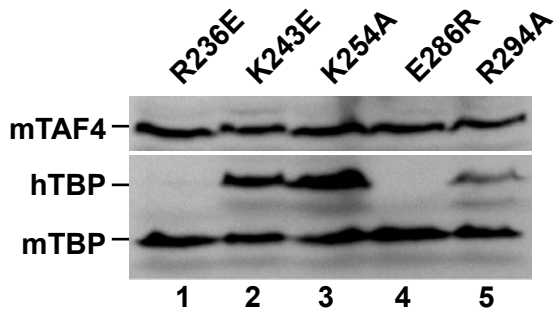
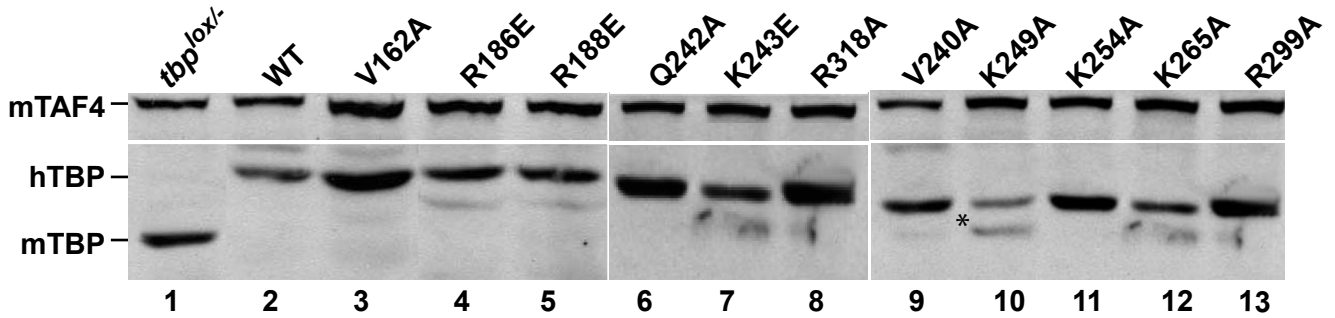
**Supplemental Figure 5.** Persistent H3K4 trimethylation at down-regulated genes. UCSC screen shot of TBP, Pol II and H3K4me3 ChIP-seq at the promoters of down-regulated genes.

**Supplemental Figure 6.** Non-productive TBP recruitment. **A.** UCSC screen shot of TBP, Pol II and H3K4me3 ChIP-seq at the divergent *Kcnab2-Nphp4* promoter showing selective recruitment of TBP R188E and H3K4me3. **B.** RT-qPCR of *Kcnab2* and *Nphp4*

expression in Wt and R188E TBP expressing cells.

**Supplemental Table 1.** Mass-spectrometry identification of TBP partners following tandem affinity purification of Wt and Mutant TBPs. Excel table. Page 1 shows refined the mass-spectrometry data for proteins with TAF-like characteristics with 0 spectra in the two negative controls and spectra in the Wt and/or mutant TBPs. C is the negative control on untagged *Tbp*<sup>lox/-</sup> cells. It should be noted that the data for the R294A mutant are not fully representative as there was much less TBP in this immunoprecipitate. Page 2 shows the raw data for all 602 identified proteins.

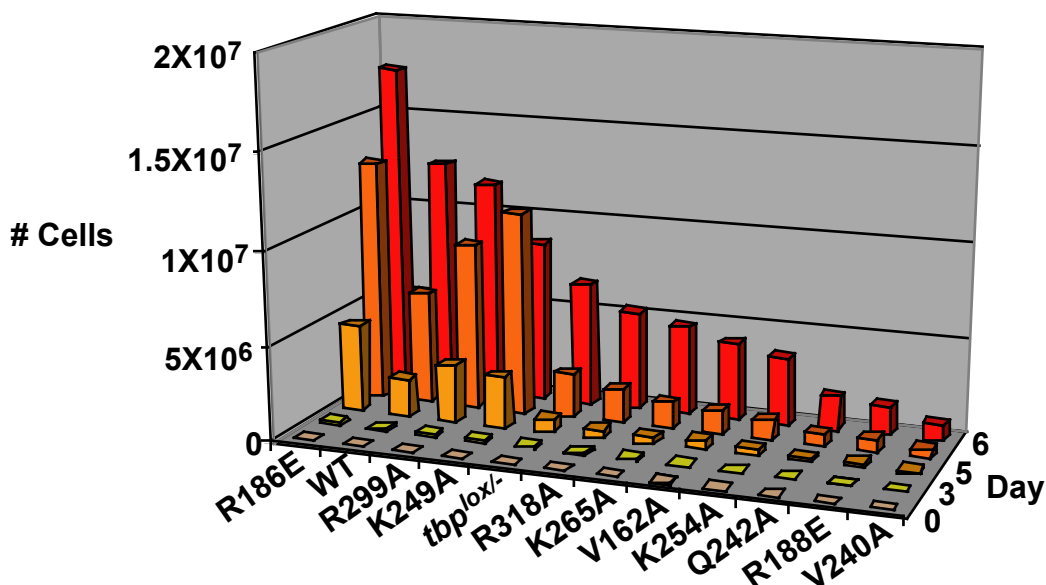
**Supplemental Table 2.** Excel table of Affymetrix data for Wt and R188E TBP expressing MEFs. Page 1 up-regulated genes, page 2 down-regulated genes.

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

A

| Mutant                  | Location       | Viability | Proliferation | Affected interactions in vitro  | Transcription in vitro | Activated transcripton                     |
|-------------------------|----------------|-----------|---------------|---------------------------------|------------------------|--|
| WT                      | NA             | Yes       | 100%          | NA                              | NA                     | NA   |
| SPM3                    | S3' S4'        | Yes       | 8,2           | Altered DNA binding specificity | AdMLP, U6              | 100%                                       |
| V162A                   | N-ter-S1       | Yes       | 29            | DNA , BTAF1, TFIIB              | NA                     | NA   |
| R186E                   | H1             | Yes       | 200           | BTAF1, TFIIB, BRF               | NA                     | >100%                                      |
| R188E                   | H1-S2          | Yes       | 17            | BTAF1, NC2                      | U1, U6                 | <6%  |
| R203E                   | S3             | No/DN     | NA            | DNA, BTAF1                      | U1, U6, VAI            | <7%  |
| T210K                   | S4             | No/DN     | NA            | DNA                             | VAI                    | NA   |
| L212A                   | S4             | Yes       | 5             | DNA, BTAF1                      | AdMLP, U6              | NA   |
| R239E                   | H2             | No/DN     | NA            | BTAF1, TAF1                     | U1                     | 3-30*                                      |
| V240A                   | H2             | Yes       | 10            | NA                              | NA                     | NA   |
| Q242A                   | H2             | Yes       | 13            | NA                              | NA                     | NA   |
| K243E                   | H2             | Yes       | 5             | BTAF1, TFIIA, BRF, TAF1         | NA                     | NA   |
| K249A                   | H2-S1'         | Yes       | 133           | DNA, BTAF1                      | NA                     | NA   |
| K254A                   | S1'            | Yes       | 23            | DNA                             | NA                     | NA   |
| K265A                   | S1'-H1'        | Yes       | 32            | DNA                             | NA                     | NA   |
| E286R                   | H1'-S2'        | No/DN     | NA            | TFIIB                           | Ad MLP, U1, U6, VAI    | >50%                                       |
| R294A                   | S3'            | Yes       | 10            | DNA, BTAF1                      | NA                     | NA   |
| R299A                   | S3'-S4'        | Yes       | 152           | DNA                             | NA                     | NA   |
| R318A                   | H2'            | Yes       | 34            | DNA                             | NA                     | NA   |
| SPM3/<br>N189E          | S3' S4'<br>S2  | No        | NA            | DNA, TFIIA                      | NA                     | <5%  |
| SPM3<br>L270A/<br>E271A | S3' S4'<br>H1' | No        | NA            | DNA                             | NA                     | Affects activation<br>by nuclear receptors |
| TRF3                    | NA             | No        | NA            | NA                              | NA                     | NA   |

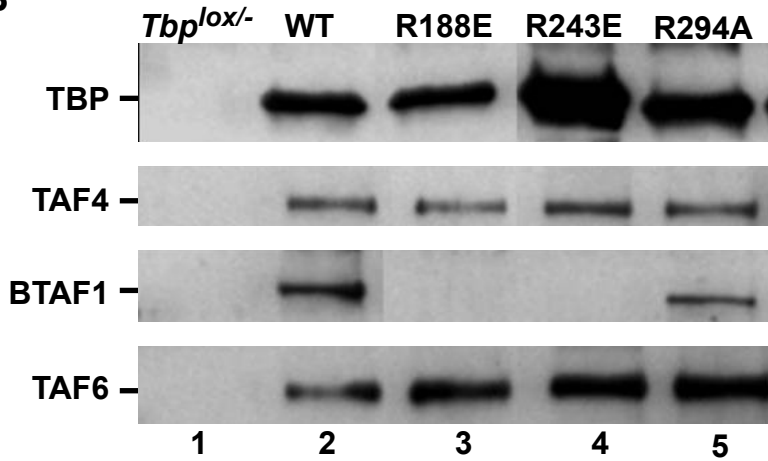
B



Choukrallah et al., Fig. 2

**A**

|                       | Protein  | No. Unique Spectra |     |       |   |     |       |       |       |       |       |      |
|-----------------------|----------|--------------------|-----|-------|---|-----|-------|-------|-------|-------|-------|------|
|                       |          | C                  | WT  | R188E | C | WT  | V162A | Q242A | K243E | K265A | R318A | SPM3 |
| TFIID                 | TBP      | 0                  | 40  | 27    | 0 | 11  | 11    | 10    | 32    | 21    | 14    | 17   |
|                       | TAF1     | 0                  | 209 | 168   | 0 | 99  | 79    | 77    | 169   | 138   | 133   | 73   |
|                       | TAF2     | 0                  | 99  | 82    | 0 | 40  | 38    | 43    | 64    | 67    | 56    | 26   |
|                       | TAF3     | 0                  | 118 | 104   | 0 | 60  | 52    | 44    | 83    | 56    | 66    | 24   |
|                       | TAF4(b)  | 0                  | 261 | 215   | 0 | 105 | 79    | 99    | 158   | 172   | 151   | 74   |
|                       | TAF5     | 0                  | 168 | 127   | 0 | 68  | 54    | 92    | 139   | 154   | 129   | 81   |
|                       | TAF6     | 0                  | 204 | 137   | 0 | 117 | 90    | 79    | 85    | 143   | 130   | 63   |
|                       | TAF7     | 0                  | 77  | 71    | 0 | 45  | 38    | 26    | 63    | 49    | 44    | 33   |
|                       | TAF8     | 0                  | 44  | 42    | 0 | 25  | 18    | 12    | 18    | 5     | 5     | 6    |
|                       | TAF9(b)  | 0                  | 103 | 103   | 0 | 50  | 40    | 21    | 34    | 31    | 27    | 17   |
|                       | TAF10    | 0                  | 48  | 30    | 0 | 17  | 14    | 15    | 23    | 20    | 18    | 7    |
|                       | TAF11    | 0                  | 24  | 18    | 0 | 10  | 12    | 8     | 19    | 14    | 9     | 11   |
|                       | TAF12    | 0                  | 50  | 33    | 0 | 13  | 10    | 13    | 13    | 17    | 14    | 5    |
|                       | TAF13    | 2                  | 22  | 30    | 0 | 9   | 11    | 7     | 13    | 13    | 9     | 8    |
| BTFIID                | BTAF1    | 0                  | 52  | 0     | 0 | 29  | 20    | 20    | 0     | 4     | 14    | 31   |
| TFIIIB                | BRF1     | 0                  | 18  | 17    | 0 | 7   | 5     | 3     | 4     | 27    | 26    | 20   |
| SL1                   | TAFIA    | 0                  | 4   | 4     | 0 | 9   | 5     | 5     | 4     | 0     | 0     | 10   |
|                       | TAFIB    | 0                  | 6   | 5     | 0 | 5   | 7     | 3     | 2     | 6     | 4     | 16   |
|                       | TAFIC    | 0                  | 12  | 12    | 0 | 15  | 17    | 3     | 2     | 0     | 4     | 16   |
|                       | TAFID    | 0                  | 7   | 8     | 0 | 2   | 6     | 4     | 1     | 4     | 2     | 13   |
| TFIIA                 | TFIIAa/b | 0                  | 4   | 0     | 0 | 1   | 0     | 0     | 0     | 3     | 2     | 2    |
|                       | TFIIAg   | 0                  | 12  | 1     | 0 | 4   | 3     | 0     | 0     | 4     | 1     | 0    |
| Potential interactors | SPEN     | 0                  | 37  | 6     | 0 | 6   | 1     | 0     | 2     | 3     | 0     | 0    |
|                       | EDC4     | 0                  | 23  | 25    | 0 | 7   | 6     | 0     | 0     | 0     | 0     | 6    |

**B**

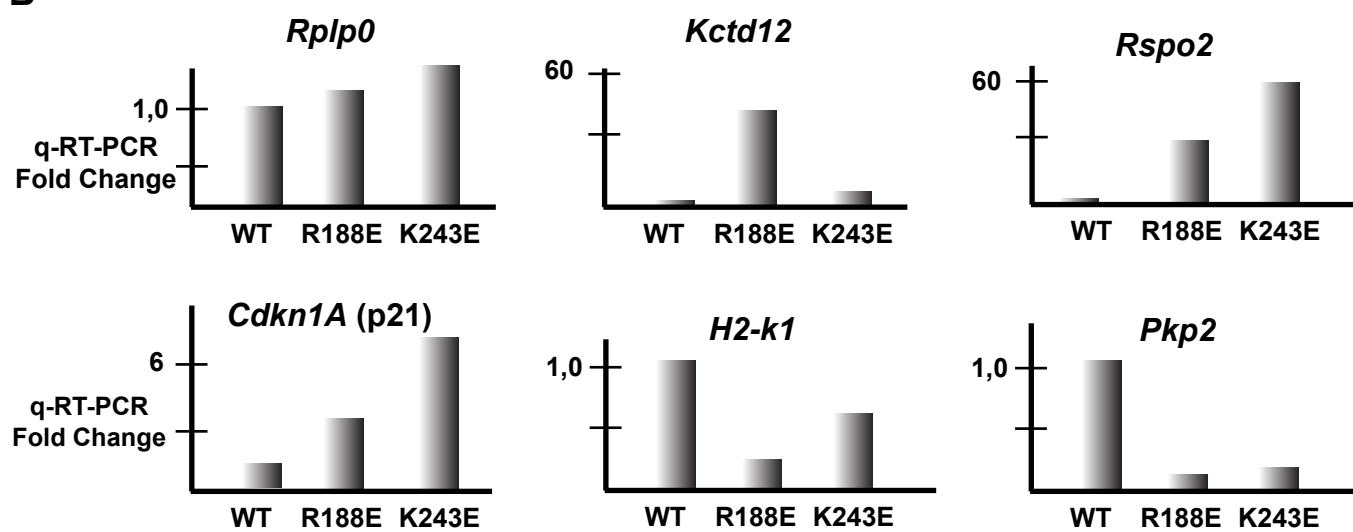
Choukrallah et al., Fig. 3

A

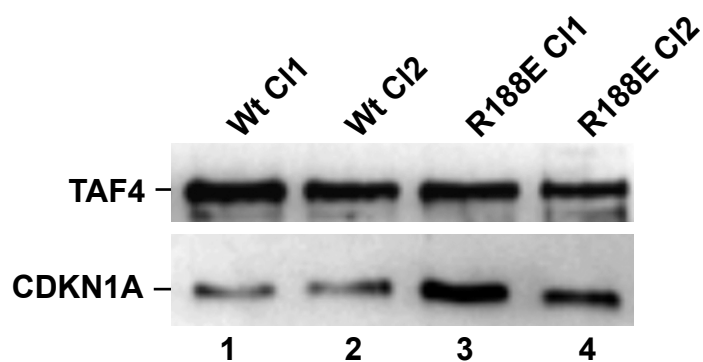
R188E

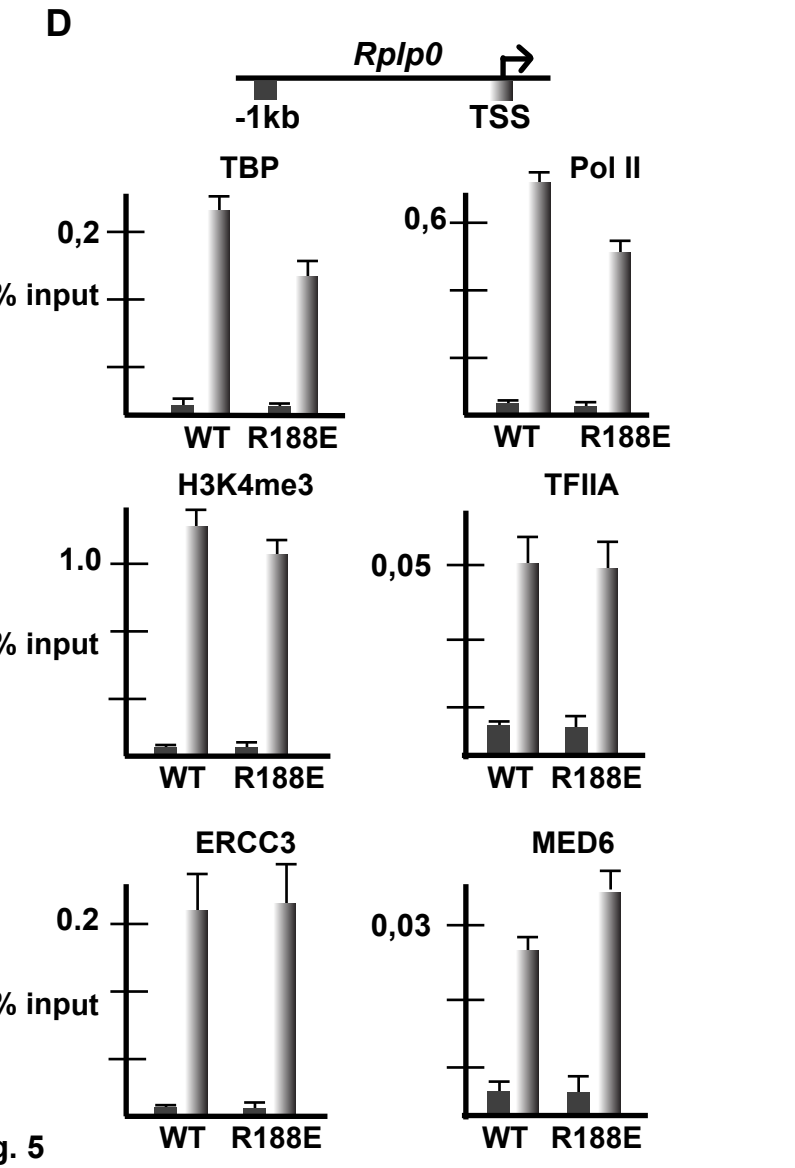
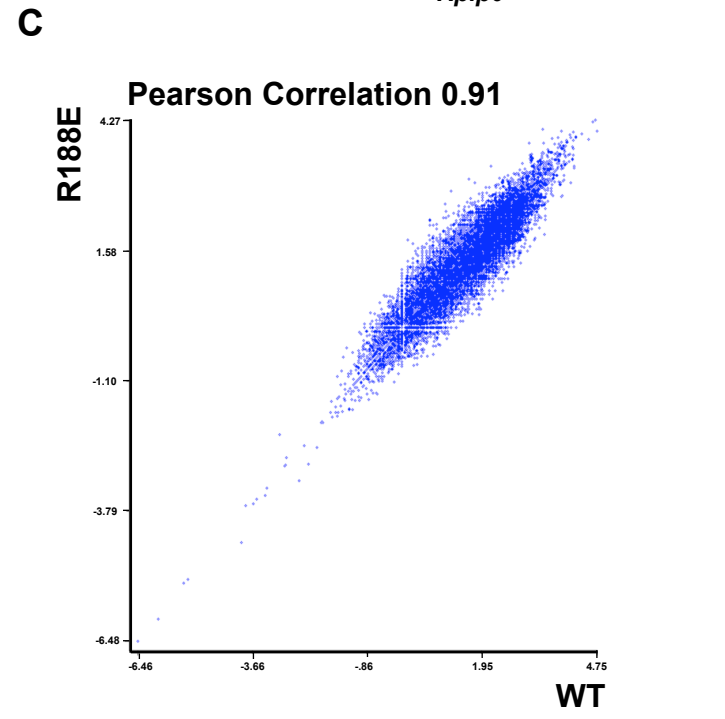
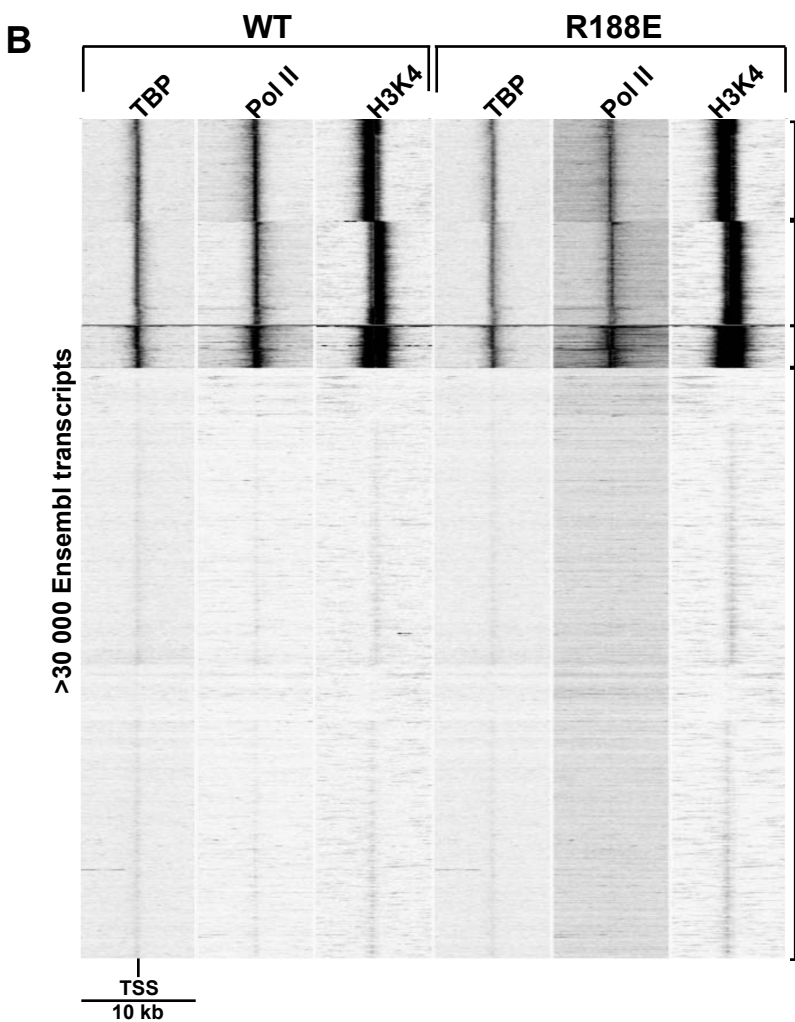
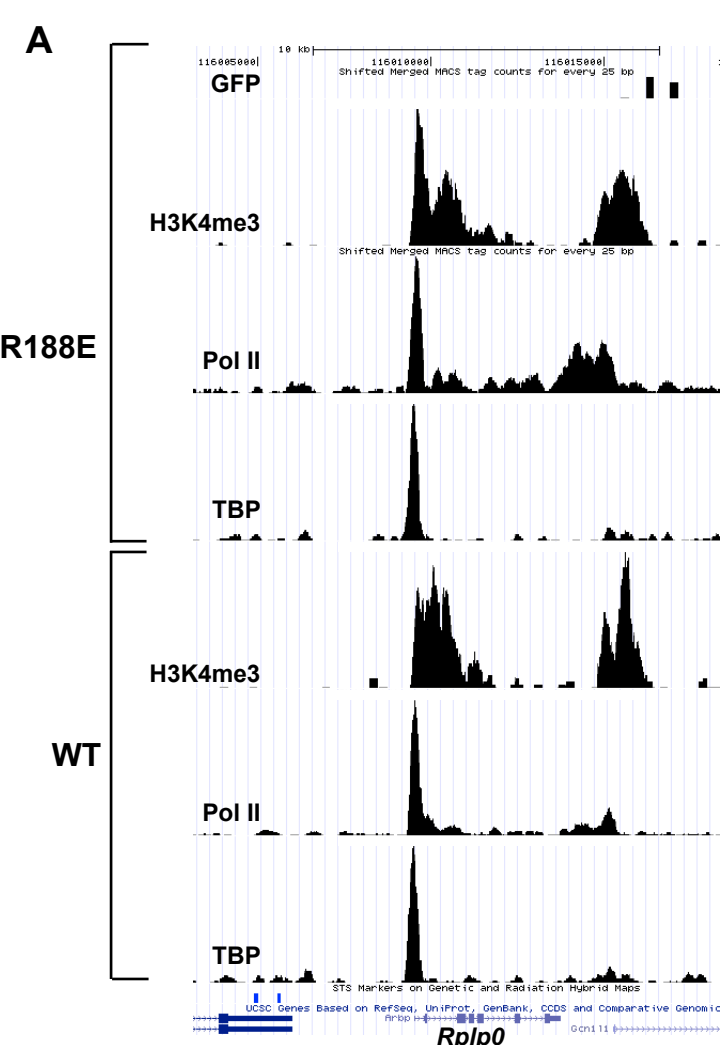
| Up (274)      |     | Down (200)     |      |
|---------------|-----|----------------|------|
| Gene          | Log | Gene           | Log  |
| <i>Cdh11</i>  | 5,4 | <i>Stmn2</i>   | -4,3 |
| <i>Kctd12</i> | 5,0 | <i>Krt1-19</i> | -4,2 |
| <i>Marcks</i> | 4,5 | <i>Grb10</i>   | -3,2 |
| <i>Aspn</i>   | 3,8 | <i>H2-k1</i>   | -3,2 |
| <i>Mmp2</i>   | 4,4 | <i>Pkp2</i>    | -3,1 |

B

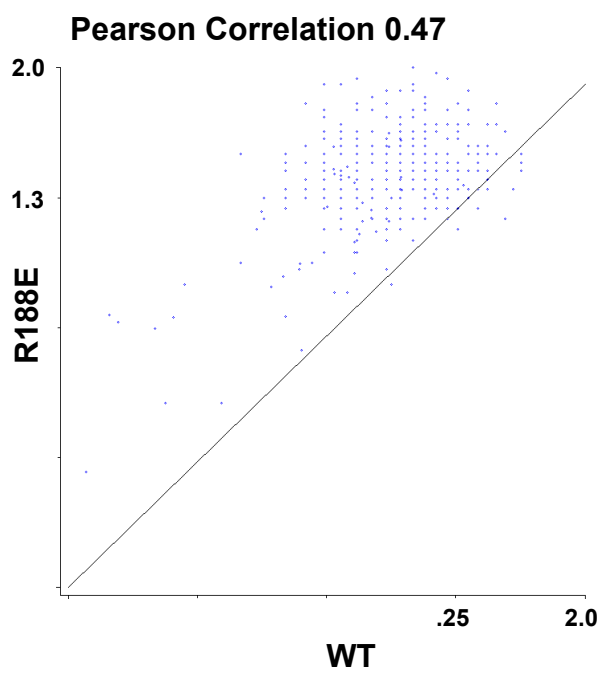
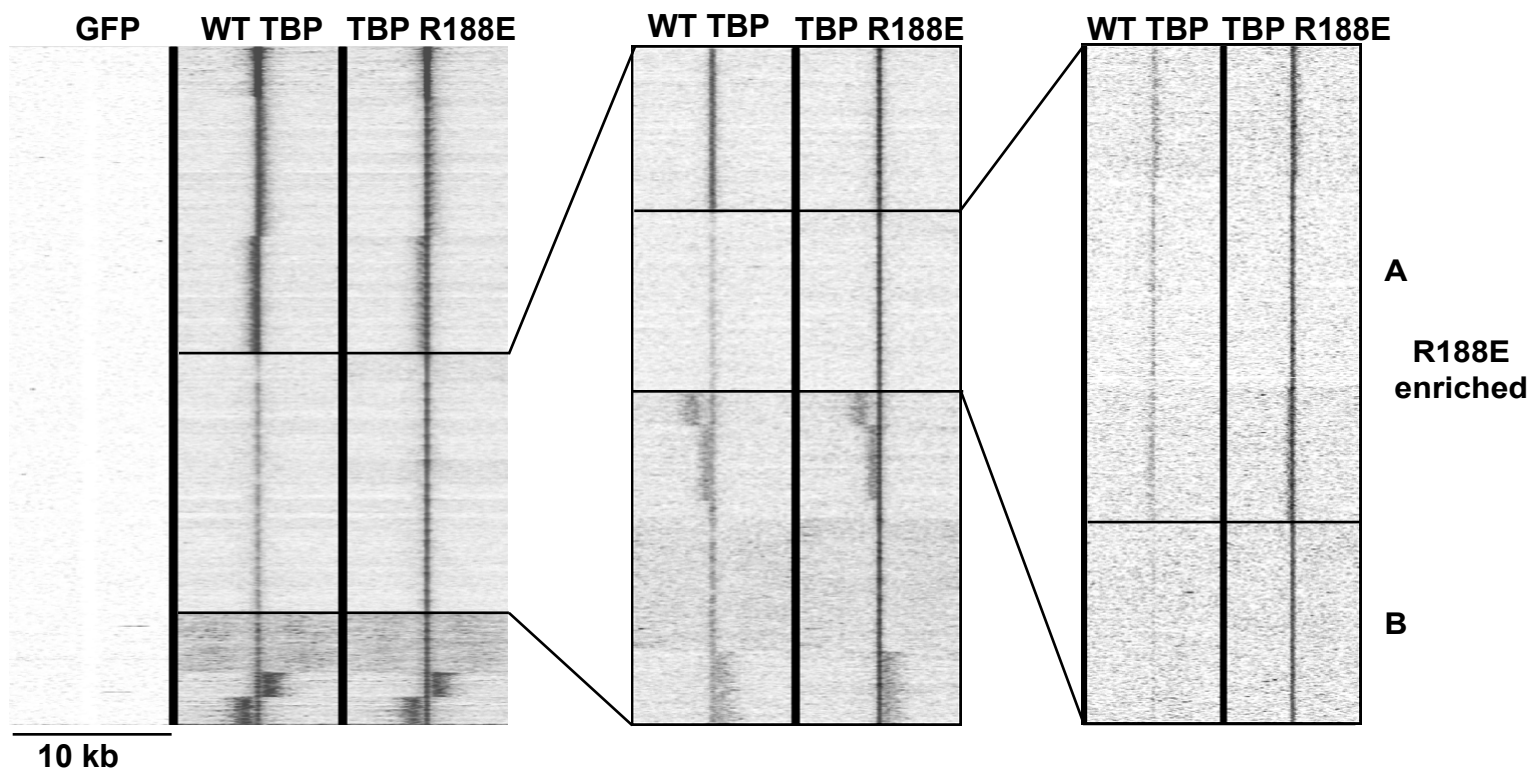


C



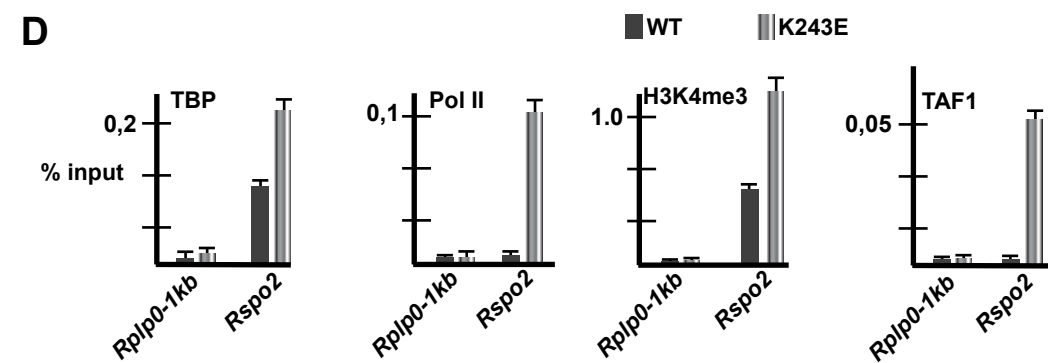
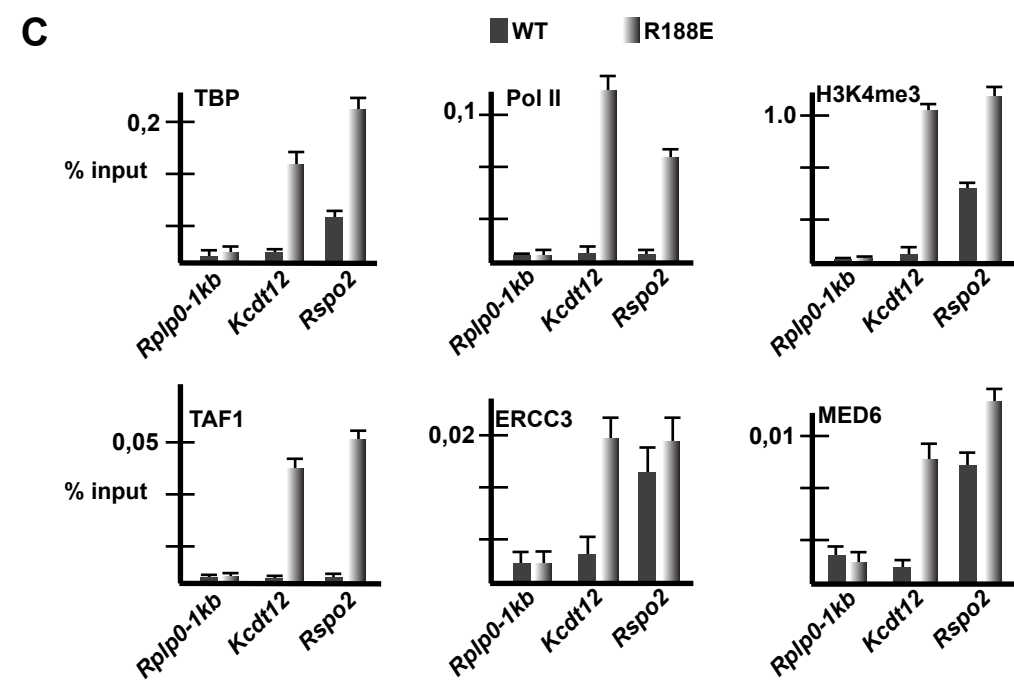
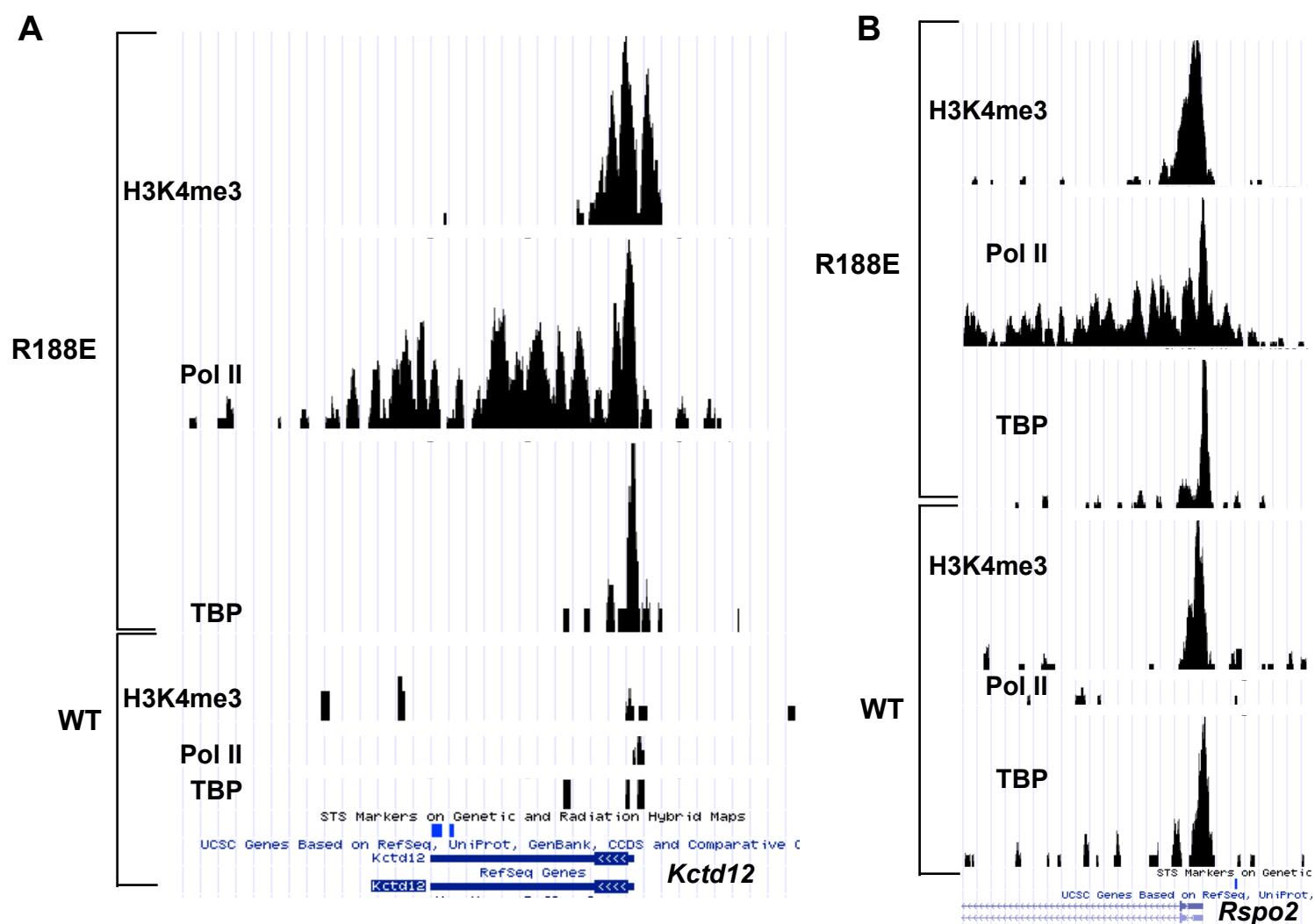


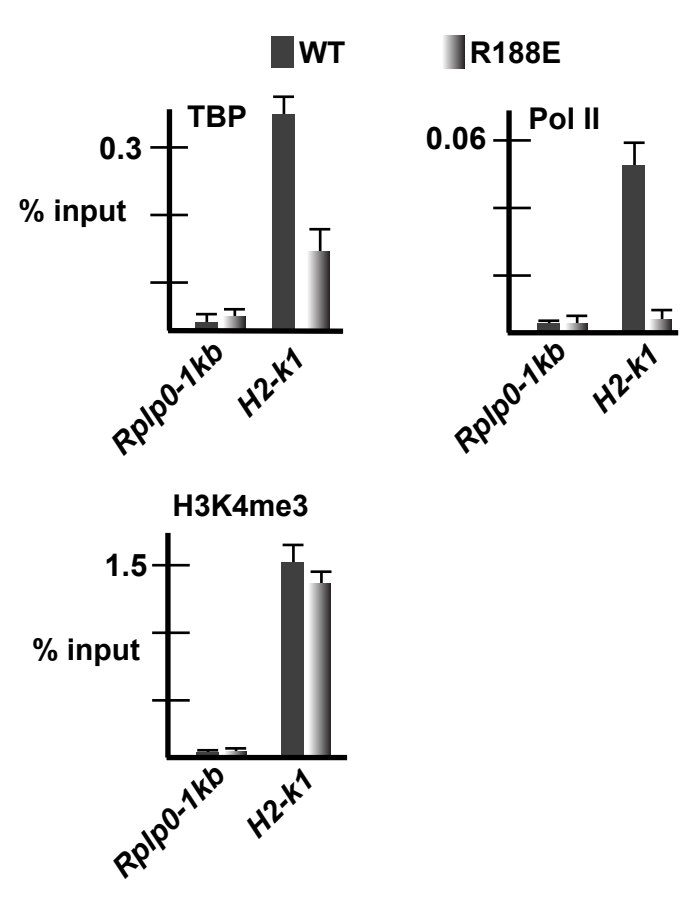
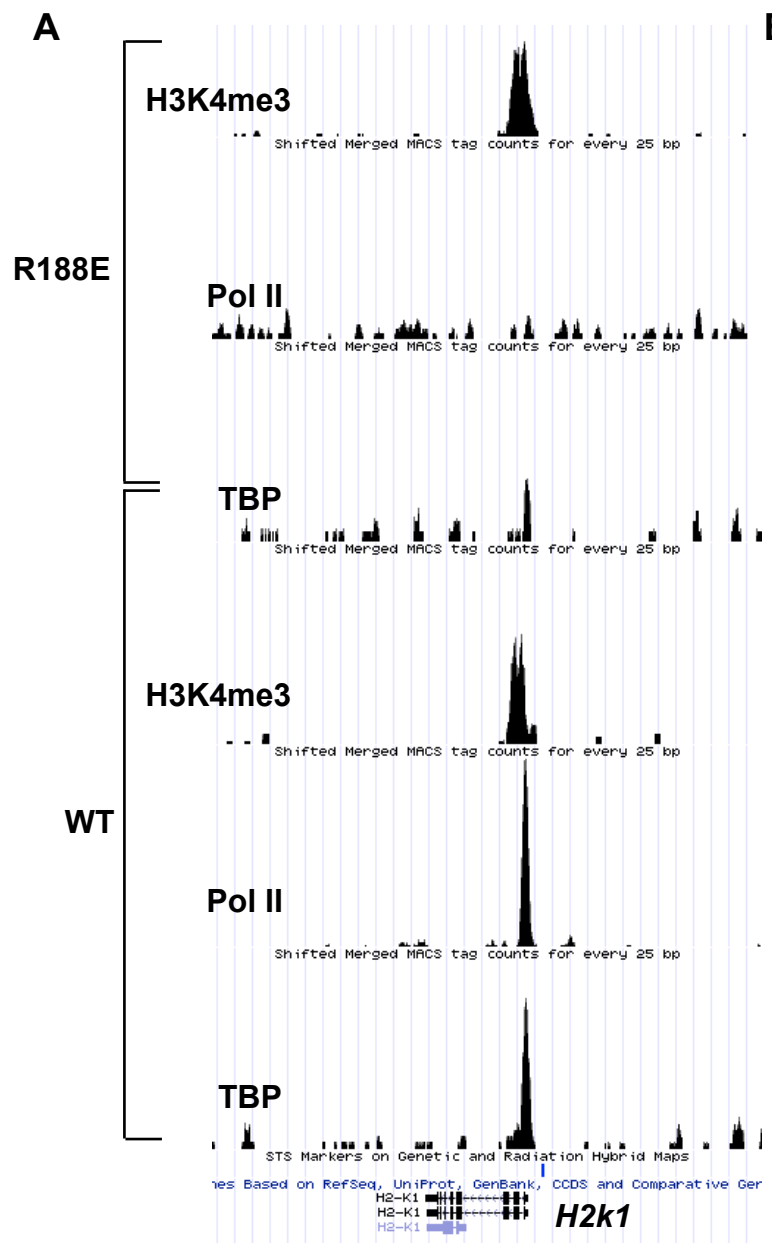
Choukrallah et al., Fig. 5



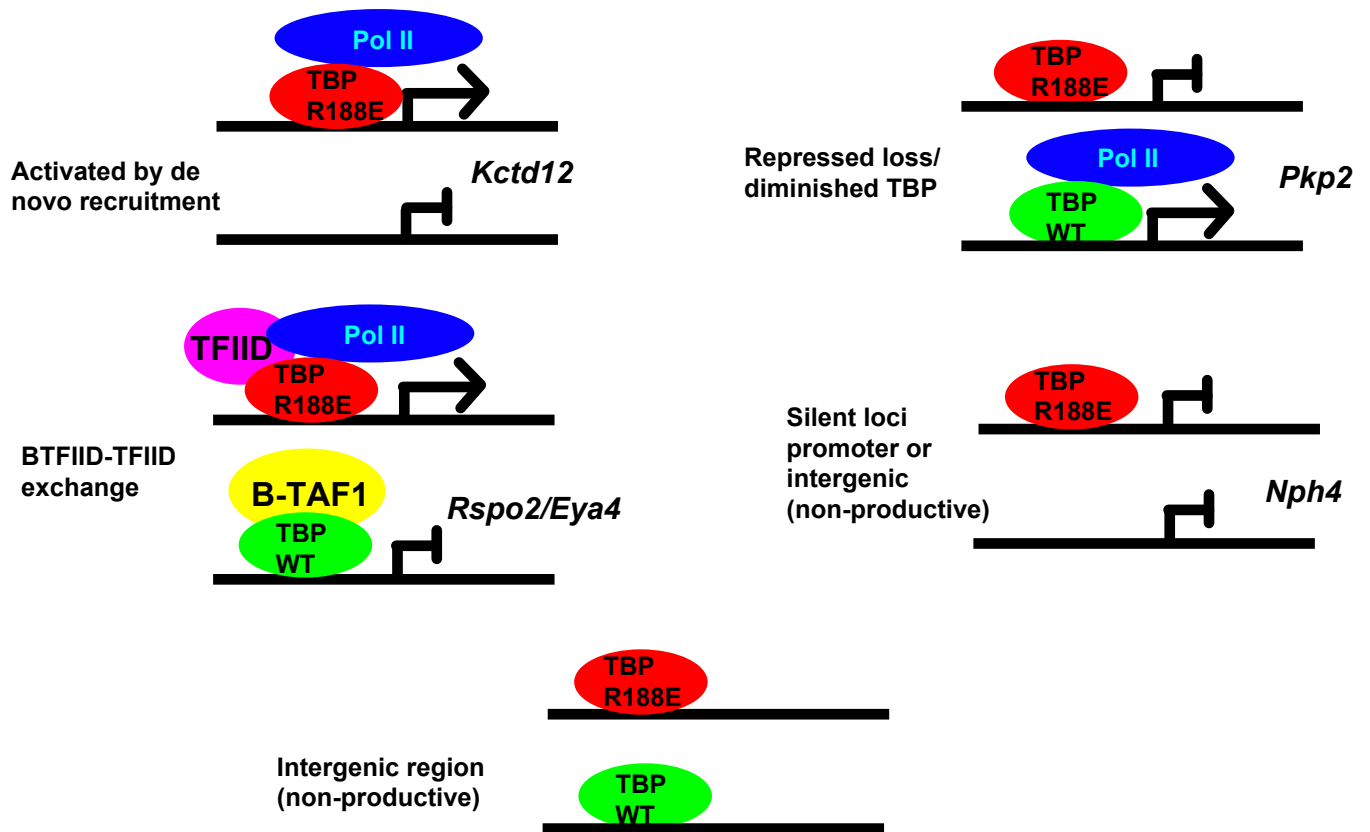
Choukrallah et al., Fig. 6





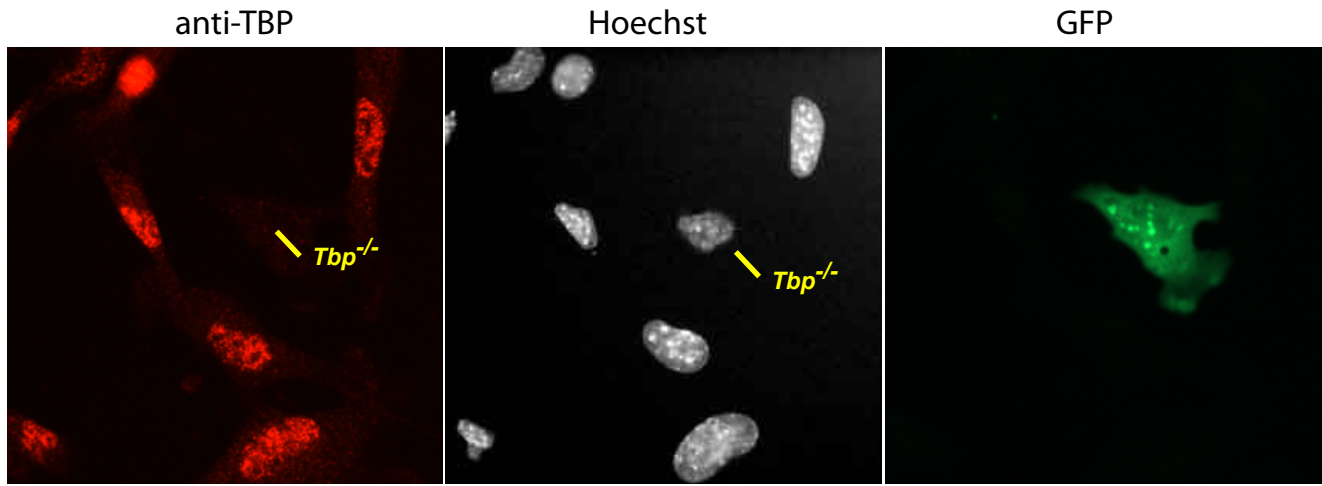


Choukrallah et al., Fig. 8

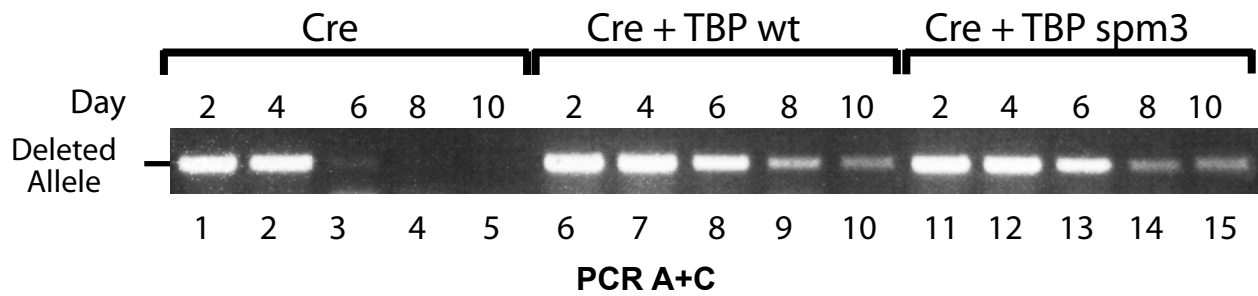


Choukrallah et al., Fig. 9

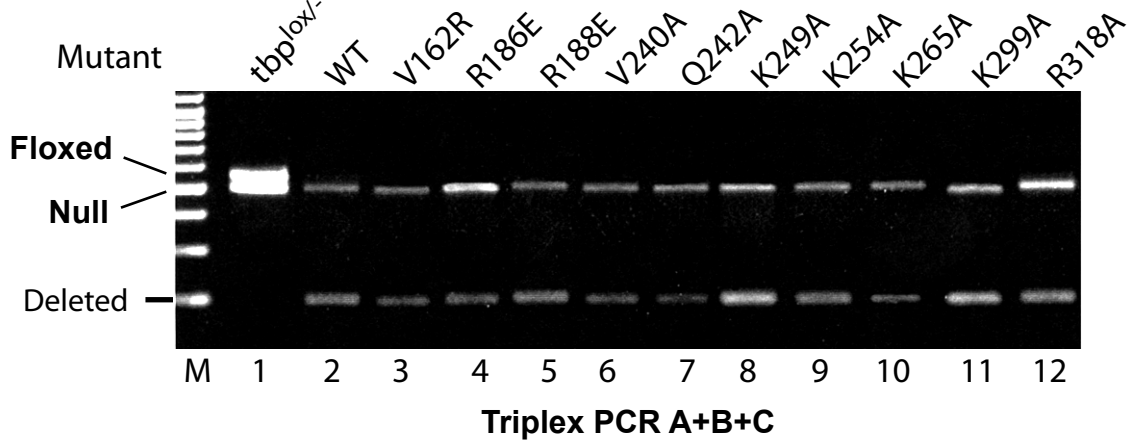
A



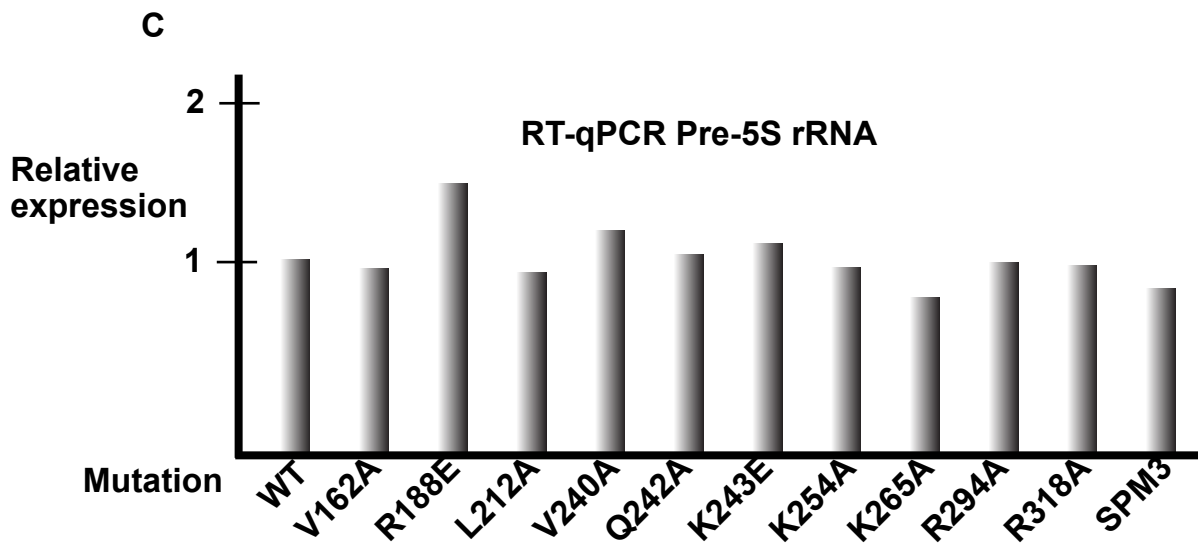
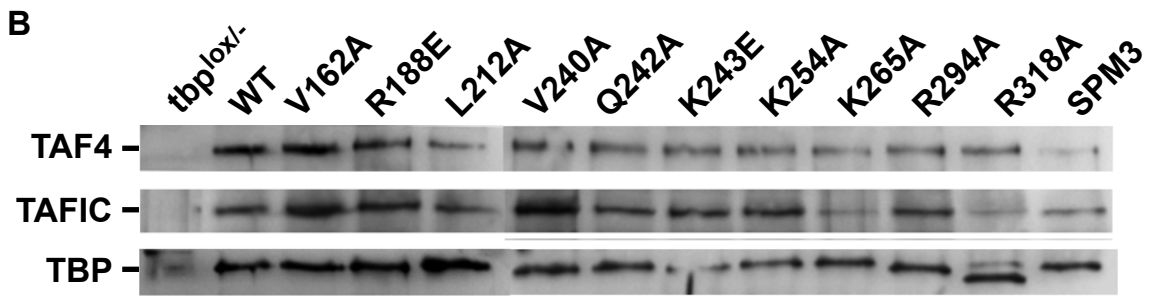
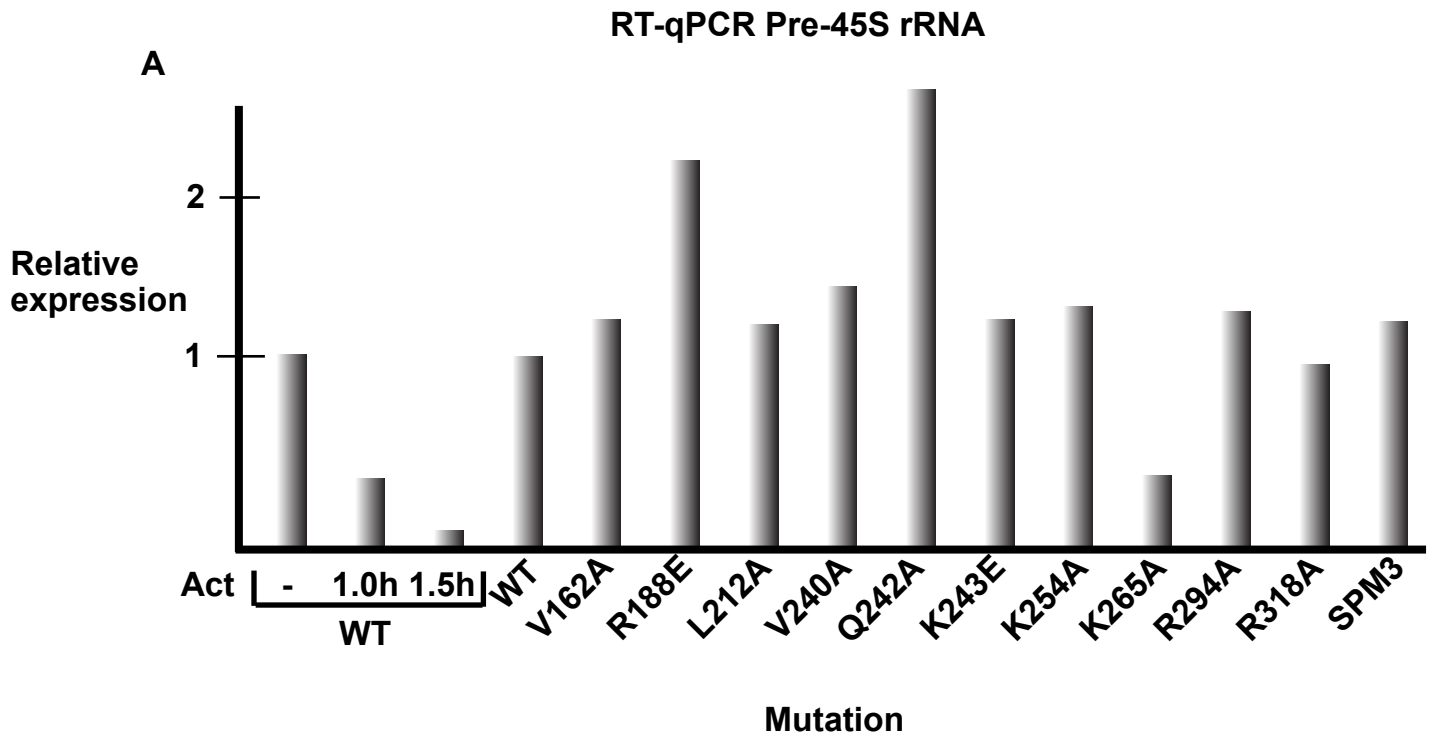
B

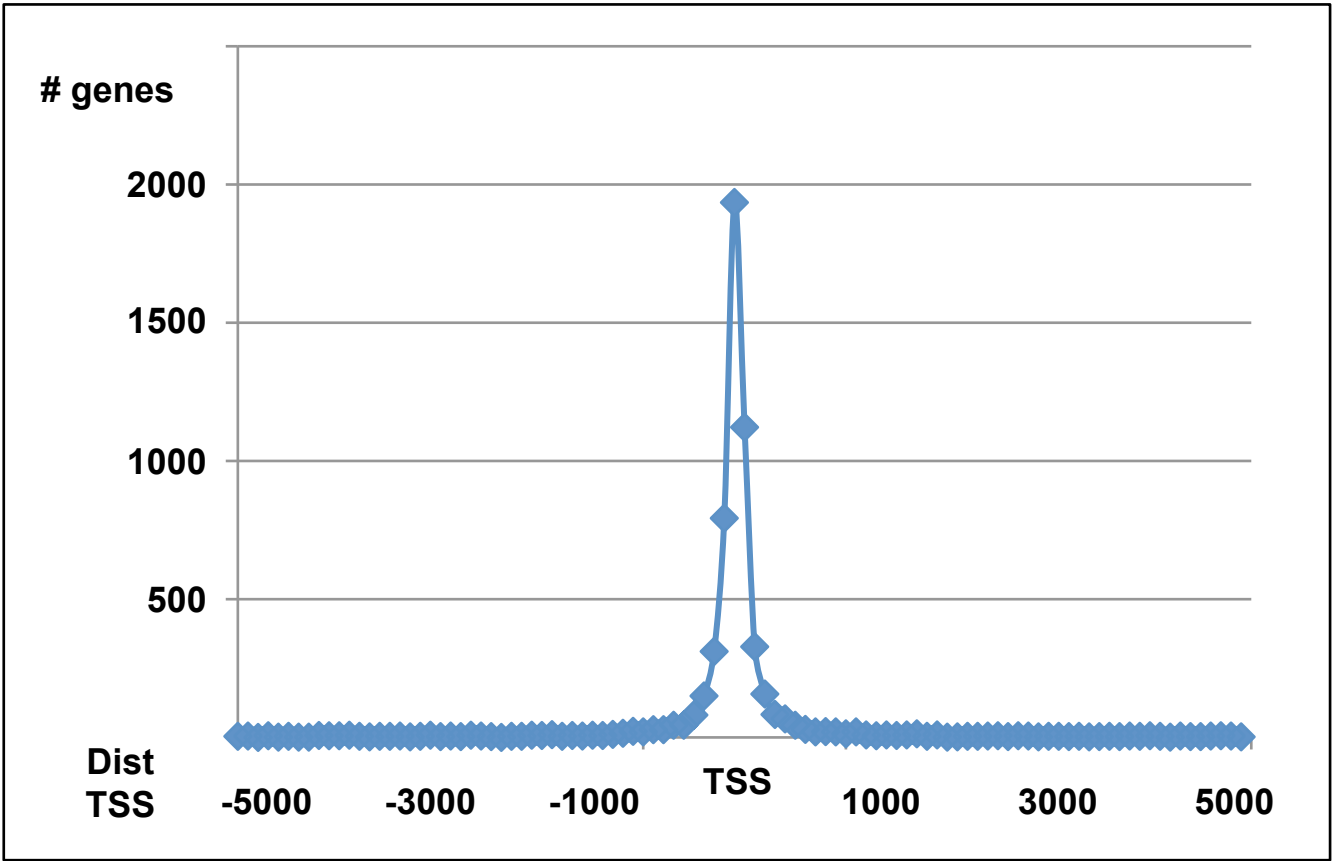


C

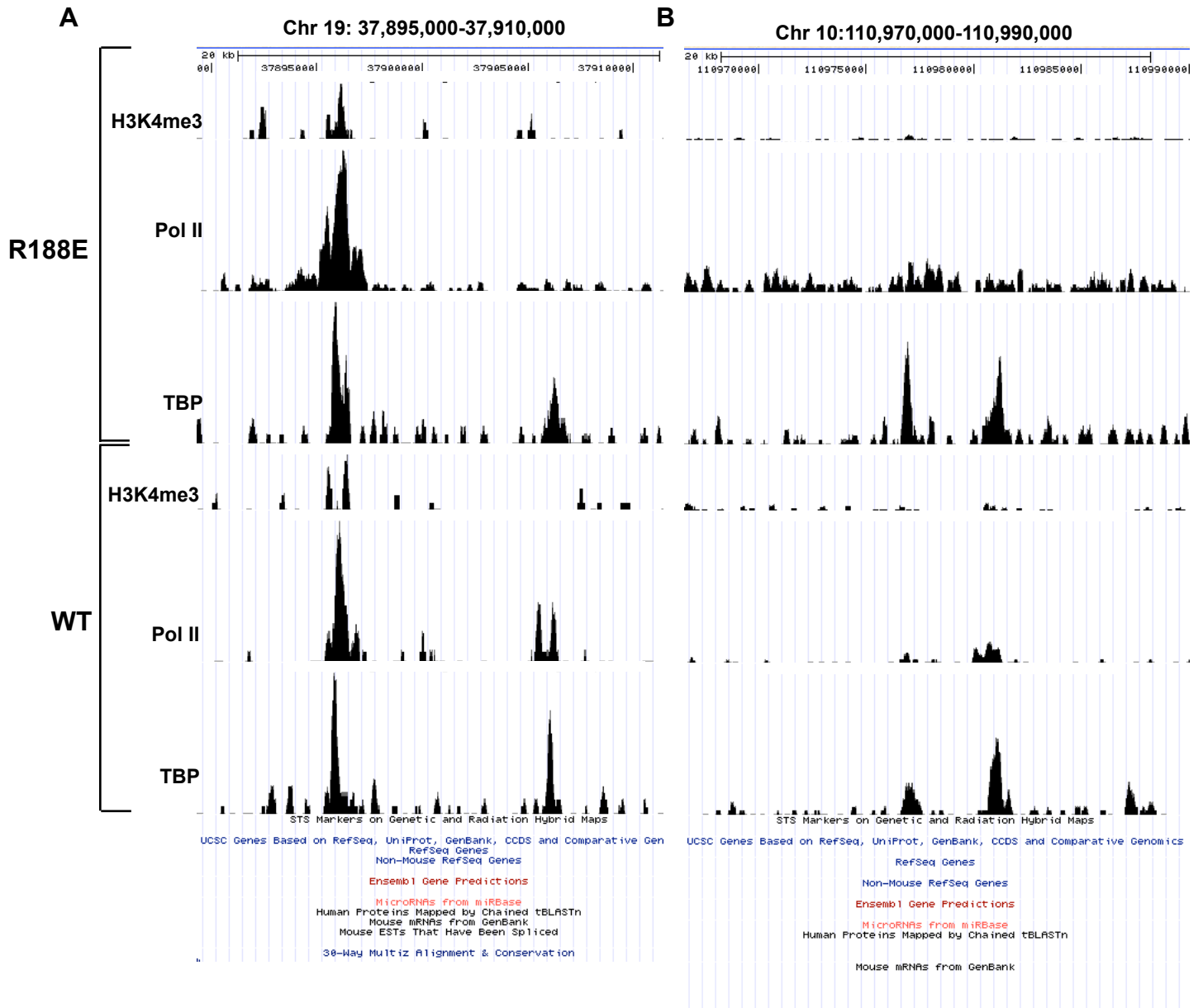


Choukrallah et al., Supplemental Fig. 1

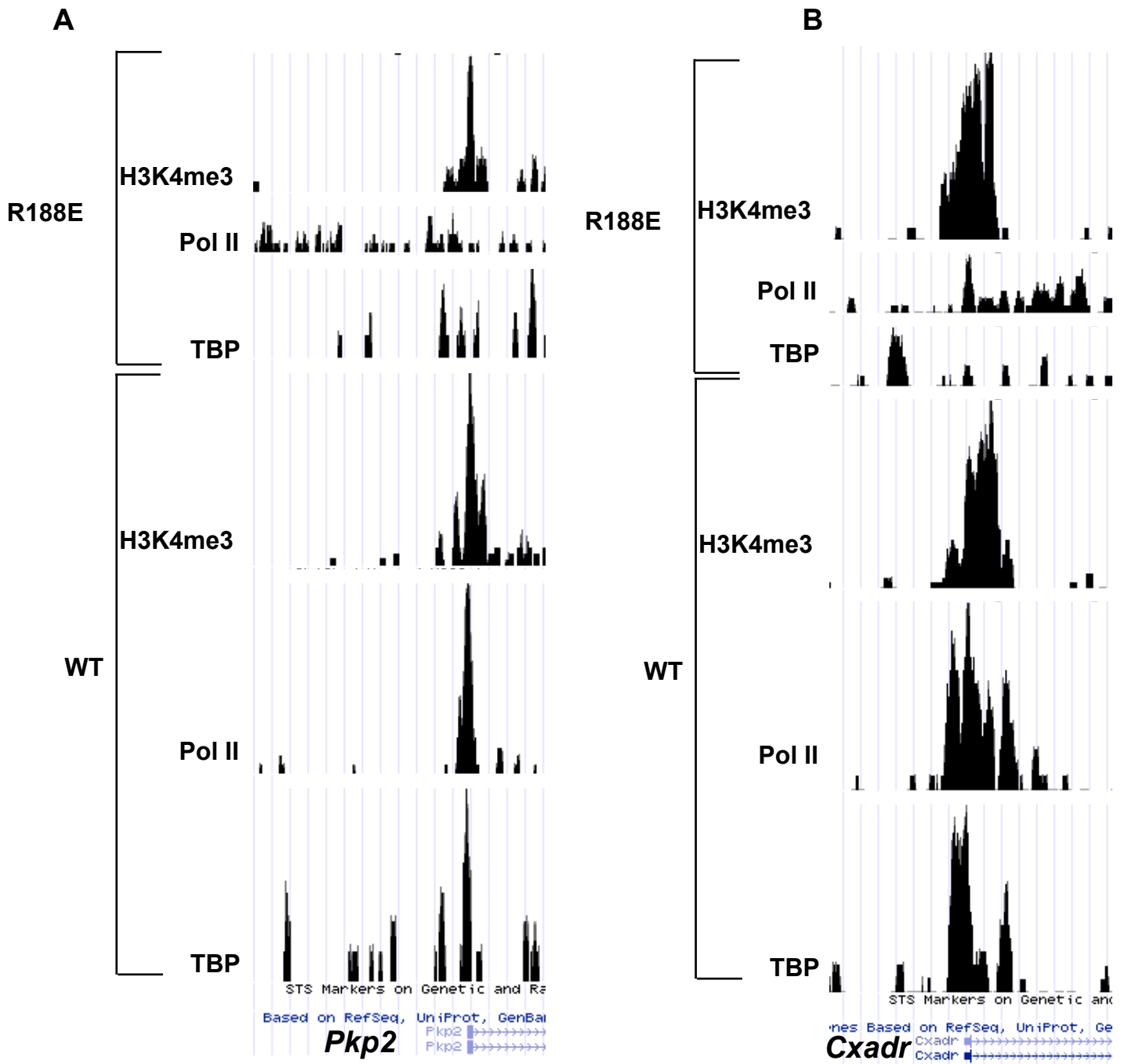




Choukrallah et al., Supplemental Figure 3

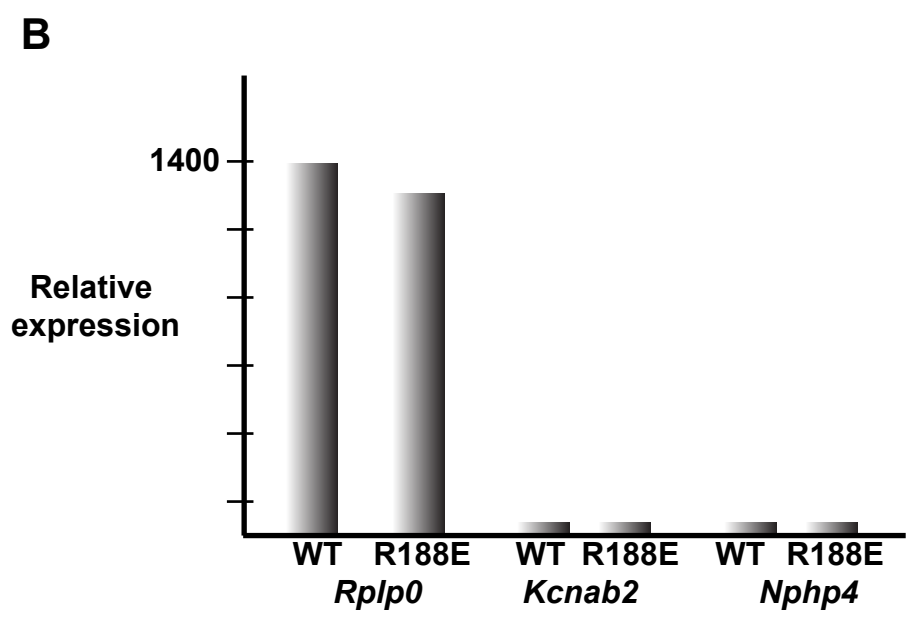
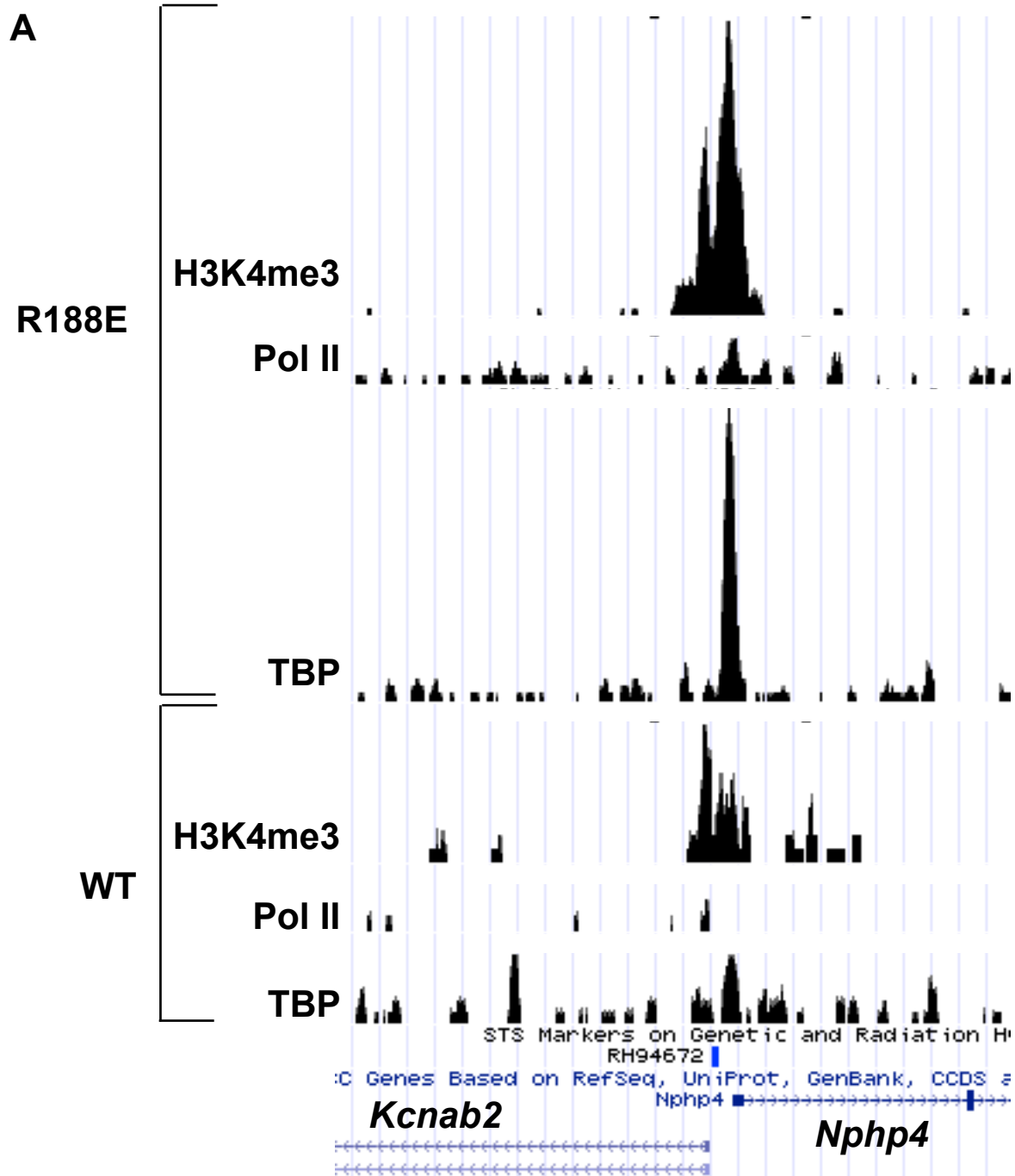


Choukrallah et al., Supplemental Fig. 4



Choukrallah et al., Supplemental Fig. 5





Choukrallah et al., Supplemental Fig. 6

## Discussion

La partie ci-dessous, vient compléter la discussion déjà présentée dans le manuscrit :  
*RNA polymerase II transcription in murine cells lacking the B-TFIID complex.*

### I. Les effets des mutations dans TBP sur sa fonction *in vivo*

Nous avons utilisé des fibroblastes modifiés génétiquement afin d'y remplacer la TBP sauvage endogène par des TBPs mutantes, où un seul résidu a été substitué par un résidu de charge opposée ou par une alanine. Nous avons établi une vingtaine de lignées cellulaires exprimant des TBP mutantes. Ces mutants assurent la viabilité cellulaire mais provoquent une baisse de la prolifération. Ce phénotype est dû à un ralentissement général du cycle cellulaire et non à l'altération de la transition d'une phase du cycle à une autre. Bien que l'expression du régulateur du cycle cellulaire p21 soit augmentée dans plusieurs mutants, nous n'avons pas détecté d'arrêt du cycle cellulaire au cours de la transition G1/S contrôlée par p21. Malgré les nombreuses analyses de cycle cellulaire que nous avons réalisées, nous n'avons pas pu déterminer les mécanismes exacts de la baisse de prolifération causée par ces mutations.

Certaines TBP mutantes sont incapables de compléter la perte de la TBP endogène. Les résidus concernés sont donc cruciaux pour la fonction de TBP *in vivo*. L'expression de certains mutants en présence de la TBP endogène provoquent la mort cellulaire, ce groupe de mutants agit probablement comme un dominant négatif. Par exemple, dans ce groupe, les mutants R203E et E286R affectent respectivement l'interaction entre TBP et l'ADN et entre TBP et TFIIB *in vitro*.

Nous étions surpris de constater que plusieurs mutations altérant l'interaction entre TBP et l'ADN *in vitro* peuvent compléter la perte de la TBP endogène et assurer la viabilité cellulaire. Dans ce cas, citons les mutants K265A, R318A et R294A. De plus, les expériences de CHIP montrent que le mutant R294A interagit avec les promoteurs des gènes non affectés par la mutation avec une efficacité équivalente à celle de la TBP sauvage. Ces résultats montrent que certaines propriétés déterminées *in vitro* ne reflètent pas toujours la situation réelle *in vivo*. Une discussion plus détaillée sur les caractéristiques des différents mutants est présentée dans la publication concernant le projet TBP.

## **II. Altération du complexe B-TFIID *in vitro* et *in vivo***

Les TBP mutantes que nous avons utilisées dans notre étude ont déjà été caractérisées dans des expériences *in vitro* par nos collaborateurs de l'équipe du Dr Timmers. Parmi les mutations que nous avons étudiées, cinq ont été décrites comme altérant l'interaction entre TBP et B-TAF1 *in vitro* : V162A, R186E, R188E, K243E et K249A. Or, les analyses réalisées dans notre système cellulaire montrent que seules les mutations R188E et K243E altèrent cette interaction *in vivo*.

## **III. Dérégulation de l'expression des gènes après l'altération de B-TFIID**

Dans la lignée R188E, un nombre très réduit de gènes est affecté par la mutation. Certains gènes sont induits alors que d'autres sont réprimés. Ce qui suggère que B-TFIID joue un rôle bivalent dans l'expression des gènes dans les fibroblastes murins. La dérégulation de l'expression de certains gènes peut toutefois être due à un effet indirect de la perte de B-TFIID. La dissection du fonctionnement des promoteurs des gènes dérégulés est nécessaire à la compréhension des mécanismes qui provoquent cette dérégulation.

## **IV. La transcription par la Pol II en absence de B-TFIID**

La perte de B-TFIID dérégule l'expression d'un petit nombre de gènes. Il était important de comprendre comment se produit cette dérégulation au niveau transcriptionnel. Il était également important de déterminer les conséquences de la perte de B-TFIID sur le recrutement de TBP et de la Pol II sur les promoteurs correspondants.

Afin de répondre à ces questions, nous avons réalisé des expériences de ChIP et de ChIP-seq en utilisant des anticorps dirigés contre TBP, la Pol II et la marque d'activation de la transcription H3K4me3. Ces expériences ont été réalisées en parallèle dans la lignée sauvage et la lignée R188E.

La comparaison des résultats obtenus dans ces deux lignées montre que la perte d'interaction entre TBP et B-TAF1 n'a pas d'effet général sur la distribution de TBP, de la Pol II et de H3K4me3. Ce qui conforte les résultats obtenus dans les analyses de transcriptome. Sur la majorité des promoteurs, le recrutement de TBP<sup>R188E</sup> et TBP<sup>wt</sup> est

similaire. Il en est de même pour la Pol II et H3K4me3 dans les deux lignées cellulaires (WT et R188E).

Nous avons tout de même réussi à isoler un groupe d'environ 400 locus, sur lesquels la TBP<sup>R188E</sup> est enrichie par rapport à la TBP<sup>wt</sup>. La plupart de ces sites sont localisés dans des régions inter- ou intra-géniques. L'enrichissement de la TBP<sup>R188E</sup> sur la majorité de ces sites ne s'accompagne pas par un enrichissement de la Pol II ni de H3K4me3 et n'affecte pas la transcription à partir de ces locus. La signification biologique du recrutement de TBP<sup>R188E</sup> sur des locus non transcrits reste à expliquer.

Il est important de noter que, de manière générale, le signal provenant de TBP et de la Pol II au niveau de ces locus est faible. La faiblesse du signal s'accompagne toujours par un coefficient de confiance bas. Pour valider l'enrichissement observé sur ces locus, il est nécessaire de réaliser des expériences de ChIP indépendantes et de vérifier ces enrichissements par qPCR. Ces validations sont en cours.

Nous nous sommes focalisés sur l'analyse des promoteurs qui sont fortement dérégulés par la mutation. Comme je l'ai mentionné plus haut, ces gènes dérégulés sont peu nombreux. Néanmoins, l'analyse détaillée de quelques promoteurs nous a permis de déterminer des mécanismes impliqués dans leur dérégulation.

Il est important de noter que la plupart des gènes dérégulés par la mutation se caractérisent par une faible expression initiale dans la lignée TBP<sup>wt</sup>. Malgré cela, la détection de cette dérégulation par RT-qPCR ou par les puces à ARN (Affymetrix) est possible et les mesures de leurs taux d'expression sont fiables, car la quantité d'ARN extraite n'est pas le facteur limitant. En revanche, l'investigation des promoteurs de ces gènes par ChIP pour évaluer le recrutement des facteurs de transcription d'intérêt est relativement difficile. En effet, pour les gènes faiblement transcrits, l'occupation de leurs promoteurs par les facteurs de transcription et par la Pol II est faible également. La quantité d'ADN obtenue lors des expériences de ChIP est un facteur limitant pour une mesure fiable de la présence de ces promoteurs par qPCR. Cela se traduit par des résultats peu reproductibles rendant toute interprétation difficile.

Pour éviter ces biais, nous avons choisi des promoteurs fortement dérégulés pour des analyses plus approfondies par ChIP. Cette analyse nous a permis d'identifier différents mécanismes impliqués dans la dérégulation des gènes induite par la perte de B-TFIID.

La répression des gènes est due principalement à l'incapacité du mutant de TBP R188E d'interagir avec les promoteurs et d'initier la transcription.

En ce qui concerne l'activation de la transcription, nous avons identifié deux mécanismes distincts :

- 1) Le premier mécanisme concerne des gènes qui ne sont pas exprimés dans les cellules exprimant la TBP<sup>WT</sup>. TBP, la pol II et la marque H3K4me3 sont absentes au niveau de ces promoteurs. En revanche, le mutant de TBP<sup>R188E</sup> est recruté à ces promoteurs, initie la formation du PIC et le recrutement de la pol II pour assurer la transcription.

- 2) Le deuxième mécanisme concerne des gènes qui ne sont pas exprimés, mais sur les promoteurs desquels se trouvent TBP et la marque H3K4me3, mais pas la pol II. A première vue, ces observations suggèrent que la TBP<sup>WT</sup> est recrutée au promoteur mais n'est pas capable de recruter la pol II, tandis que le mutant de TBP<sup>R188E</sup> recrute la pol II. Nous avons réalisé des expériences de ChIP pour évaluer la présence d'autres composants du PIC à cette catégorie de promoteurs. Ces résultats démontrent la présence de sous-unités du Médiateur et de TFIIH avec les TBP<sup>WT</sup> et TBP<sup>R188E</sup>. En revanche, la sous-unité TAF1 de TFIID est présente uniquement avec la TBP<sup>R188E</sup>. Ces résultats suggèrent que la TBP<sup>WT</sup> présente au promoteur n'appartient pas au complexe TFIID, mais probablement au complexe B-TFIID. Dans les cellules exprimant TBP<sup>R188E</sup>, le complexe B-TFIID n'étant pas formé, TFIID le remplace et permet le recrutement de la pol II et la transcription. A ces promoteurs, B-TFIID permet le recrutement d'une partie des facteurs du PIC, mais pas la pol II, réprimant ainsi le promoteur. Les résultats obtenus pour la lignée R188E ont été validés parallèlement dans la lignée exprimant le mutant K243E qui est également déficient pour l'interaction avec B-TAF1.

Dans leur ensemble, nos résultats suggèrent que B-TFIID peut agir en tant qu'activateur ou que répresseur de la transcription et ce de manière spécifique selon les gènes. La répression peut être assurée par deux mécanismes distincts : le premier consiste à inhiber l'interaction entre TBP et le promoteur, le deuxième implique un recrutement de TBP associé

probablement à B-TAF1 inhibant ainsi le recrutement de TFIID, la formation du PIC et l'initiation de la transcription.

Afin de valider ce modèle, il est nécessaire d'évaluer le recrutement de B-TAF1 sur les promoteurs des gènes dérégulés. Malheureusement, les expériences de ChIP avec des anticorps dirigés contre B-TAF1 ont toujours échoué. Ce qui rend impossible cette validation avec le protocole classique de ChIP. Notre collaborateur, le Dr Matthieu Gérard (CEA Saclay) a développé un nouveau protocole de ChIP, où l'utilisation des détergents et de la sonication est réduite (la sonication sert à fragmenter la chromatine dans l'expérience du ChIP). Ces modifications semblent améliorer l'efficacité du ChIP. Nous allons utiliser ce nouveau protocole avec les anticorps dirigés contre B-TAF1 en espérant obtenir des résultats significatifs qui nous permettraient de répondre aux questions posées.

Nos résultats sont en contradiction avec ceux obtenus dans l'équipe de notre collaborateur Marc Timmers lors de la réduction de l'expression de B-TAF1 par siRNA (de Graaf et al., 2010 ; Johannes et al., 2010). La perte d'expression de B-TAF1 augmente la fixation de TBP sur les la majorité des promoteurs. Ils interprètent ces observations par le fait que B-TAF1 peut dissocier TBP des promoteurs et ainsi baisser sa durée de résidence sur ces sites. Cela implique que dans une expérience de ChIP contre TBP, une proportion du signal représente la fixation de TBP sous forme de B-TFIID. Il est important de noter que le signal mesuré lors d'une expérience de ChIP est une moyenne sur une population de cellules et non le signal d'un promoteur unique. Il est donc possible que certains promoteurs soient occupés par TFIID dans une partie de la population et occupés par B-TFIID dans une autre partie. Cela se traduit par un équilibre sur l'ensemble de la population. Il est également possible que dans la même cellule et sur le même promoteur, TFIID et B-TFIID sont en équilibre dynamique. Lors de la perte de B-TAF1, TBP s'associe au promoteur uniquement sous forme de TFIID, cette interaction étant plus stable, elle induirait une augmentation du signal. Nous n'avons pas observé cette augmentation générale du signal de TBP dans les cellules exprimant TBP<sup>R188E</sup> ou TBP<sup>K243E</sup>. Il est donc possible que la perte d'interaction observée dans les expériences de co-immunoprécipitation est due à des conditions expérimentales stringentes, et que dans la cellule, B-TAF1 interagit toujours avec TBP. Cette hypothèse se heurte au fait que l'interaction entre TBP et B-TAF1 est forte, et que dans les mêmes conditions de purification nous avons détecté des interactions plus faibles comme par exemple l'interaction entre TBP et TFIIA.

Une explication alternative consiste à supposer que l'action de B-TAF1 sur TBP ne nécessite pas la formation d'un complexe B-TFIID stable et que même des interactions faibles et transitoires suffisent pour permettre à B-TAF1 de dissocier TBP. Il est important de noter que nos résultats adressent la fonction de l'interaction TBP-BTAF1 et la formation d'un complexe B-TFIID stable et non la fonction de B-TAF1. La confrontation de nos résultats avec ceux de nos collaborateurs, suggèrent que BTAF1 peut avoir une activité sans être stablement associé à TBP. Pour valider cette hypothèse il est nécessaire de réduire l'expression de B-TAF1 dans nos lignées cellulaires, ce qui permettrait une comparaison valable.

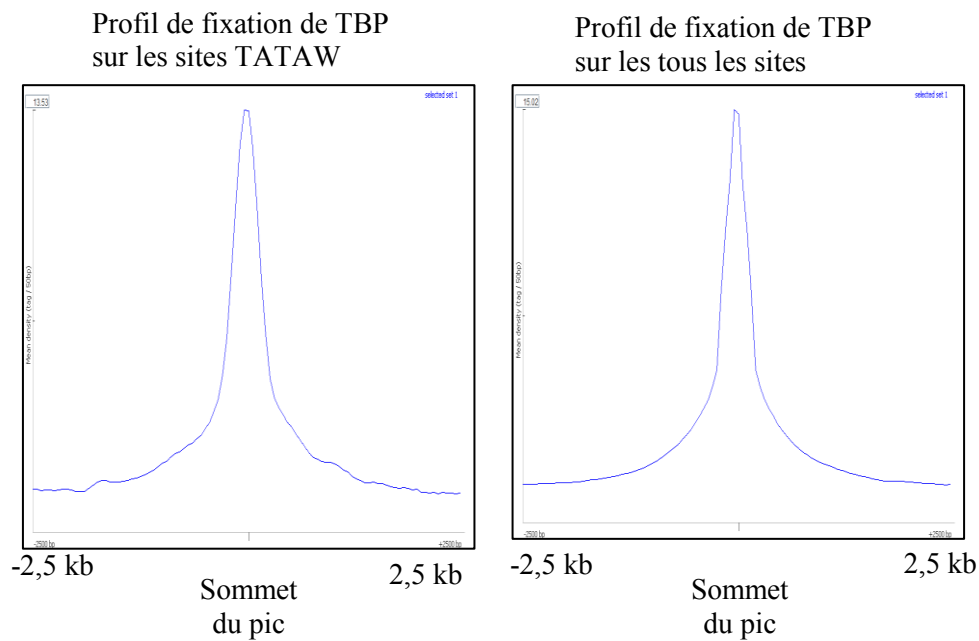
### **V. Le rôle d'élément TATA *in vivo***

L'analyse de l'architecture des promoteurs réalisée par Carninci et al. suggère que les promoteurs possédant un élément TATA et/ou Inr seraient des promoteurs à TSS unique. A l'opposé, les promoteurs dépourvus de ces éléments auraient plusieurs TSS (Carninci et al., 2006). Plusieurs auteurs suggèrent également que les promoteurs TATA-dépendants sont liés plus stablement par les GTFs (Albert et al., 2010 ; JC Andrau, communication personnelle). Dans l'étude réalisée sur TFIIB par Chip on chip, les auteurs montrent que l'élément TATA favorise la fixation de TFIIB sur les promoteurs de classe II dans les cellules B humaines (Albert et al., 2010).

Ces résultats suggèrent que les promoteurs à TSS unique possédant un élément TATA et/ou Inr seraient reconnus de façon plus précise par la machinerie de transcription de base contrairement aux promoteurs qui en sont dépourvus. Dans les expériences de CHIP-seq ou de CHIP on chip, cette propriété devrait se traduire par un pic aigu de fixation (dit « sharp peak ») des facteurs de transcription sur ces sites. A l'opposé, les promoteurs à multiples TSSs devraient avoir des pics larges et moins précis (dits « broad peaks »).

Or, nos expériences de CHIP-seq montrent que TBP est recrutée sur la majorité des promoteurs transcrits indépendamment de la présence ou de l'absence d'élément TATA dans ces promoteurs. En effet, seulement 5,4% des sites occupés par TBP possèdent un élément TATA (motif recherché TATAW). Ces sites ne présentent pas d'enrichissement particulier en TBP par rapport à des sites dépourvus de cet élément.

Le profil des pics de TBP sur les sites contenant un élément TATA est similaire à celui calculé pour tous les sites. Autrement dit, la majorité des sites de fixation de TBP ont un pic net, pointu et centré sur un sommet (la position présentant le maximum de signal) et ce indépendamment de la présence d'un élément TATA et/ou Inr. Donc, nous n'avons pas distingué les différents profils de fixation suggérés par les études précédemment citées. Cf. *Figure29*



**Figure 29 :** Comparaison des profils de fixation de TBP sur les sites possédant un élément TATA et sur la totalité des sites.

Cette comparaison montre que l'élément TATA ne change pas le profil de fixation de TBP sur les promoteurs

En résumé, nos résultats ne montrent pas de propriétés particulières caractérisant les promoteurs possédant un élément TATA, du moins dans les fibroblastes murins.



**Publication 2 :**

**Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Mol Cell Biol.* 30(1):231-44.**

Delacroix L, Moutier E, Altobelli G, Legras S, Poch O, Choukrallah MA, Bertin I, Jost B, Davidson I. (2010)

## Cell-Specific Interaction of Retinoic Acid Receptors with Target Genes in Mouse Embryonic Fibroblasts and Embryonic Stem Cells<sup>∇†</sup>

Laurence Delacroix,<sup>2</sup> Emmanuel Moutier,<sup>1</sup> Gioia Altobelli,<sup>1</sup> Stephanie Legras,<sup>1</sup> Olivier Poch,<sup>1</sup> Mohamed-Amin Choukrallah,<sup>1</sup> Isabelle Bertin,<sup>3</sup> Bernard Jost,<sup>1</sup> and Irwin Davidson<sup>1\*</sup>

*Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 1 Rue Laurent Fries, 67404 Illkirch Cédex, France<sup>1</sup>; Immunologie et Maladies Infectieuses, GIGA +2, Bat B34, 1 ave. de l'Hôpital, 4000 Liège, Belgium<sup>2</sup>; and Ecole Supérieure de Biotechnologie de Strasbourg, Pole API, 67400 Illkirch, France<sup>3</sup>*

Received 11 June 2009/Returned for modification 15 July 2009/Accepted 22 October 2009

**All-trans retinoic acid (RA) induces transforming growth factor beta (TGF- $\beta$ )-dependent autocrine growth of mouse embryonic fibroblasts (MEFs). We have used chromatin immunoprecipitation to map 354 RA receptor (RAR) binding loci in MEFs, most of which were similarly occupied by the RAR $\alpha$  and RAR $\gamma$  receptors. Only a subset of the genes associated with these loci are regulated by RA, among which are several critical components of the TGF- $\beta$  pathway. We also show RAR binding to a novel series of target genes involved in cell cycle regulation, transformation, and metastasis, suggesting new pathways by which RA may regulate proliferation and cancer. Few of the RAR binding loci contained consensus direct-repeat (DR)-type elements. The majority comprised either degenerate DRs or no identifiable DRs but anomalously spaced half sites. Furthermore, we identify 462 RAR target loci in embryonic stem (ES) cells and show that their occupancy is cell type specific. Our results also show that differences in the chromatin landscape regulate the accessibility of a subset of more than 700 identified loci to RARs, thus modulating the repertoire of target genes that can be regulated and the biological effects of RA.**

Retinoic acid (RA), the naturally active vitamin A metabolite, exerts a wide range of effects on vertebrate development and adult tissue homeostasis by regulating cell proliferation, differentiation, and apoptosis (8, 34, 42). RA activates three members of the nuclear receptor (NR) superfamily, RAR $\alpha$ , RAR $\beta$ , and RAR $\gamma$ , that function as ligand-dependent transcriptional regulators by binding, usually as heterodimers with retinoid receptors (RXRs  $\alpha$ ,  $\beta$ , and  $\gamma$ ) to RA response elements (RAREs) located in target genes (16, 17). Many RAREs are formed by a direct repeat (DR) of the consensus sequence 5'-RGKTCA-3' separated by 1, 2, or 5 nucleotides (for a review, see references 3, 4, and 6), and a large number of DR2-type elements are present within Alu repeats (27, 58).

RARs and RXRs exhibit the conserved structure of NRs, with N-terminal activation function 1 (AF-1), a central DNA binding domain, and a C-terminal ligand binding domain (LBD) that harbors ligand-dependent AF-2 (38, 40; for a review, see references 2, 41, and 51). The ability of RARs to modulate the expression of target genes results from a complex and dynamic interaction with coactivator/corepressor complexes (54). A general model proposes that unliganded RARs occupy regulatory elements at their target genes and repress their expression. Ligand binding leads to a conformational change in LBD structure, releasing the corepressor complexes

and allowing recruitment of coactivator complexes with histone acetyl- and methyltransferase activities and activation of target genes. A third scenario is ligand-dependent repression involving the recruitment of proteins such as NRIP1 (RIP140), PRAME, or TRIM24 (TIF1a) that interact with the liganded receptors to repress the transcription of target genes (11, 14, 30). An additional level of control contributing to cell specificity could occur at the DNA binding step, as RAR/RXR heterodimers may bind to distinct sets of regulatory elements in different cell types. This may be regulated by cell or tissue type differences in the epigenetic organization of the chromatin in which the regulatory elements are present (22).

We have previously shown that RA activates the transforming growth factor beta (TGF- $\beta$ ) signaling pathway to induce morphological changes and serum-independent autocrine growth of *Taf4<sup>lox/-</sup>* mouse embryonic fibroblasts (MEFs) (15). Transcriptome analysis showed that around 1,000 genes are activated or repressed by RA in these cells. Among the activated genes are TGF- $\beta$  ligands and connective tissue growth factor, which is rapidly induced by RA to induce autocrine growth (15, 28). While this study identified the genes and pathways that are activated by RA, we could not discriminate between the direct and indirect targets of the RARs.

To identify RA-regulated genes directly, we used chromatin immunoprecipitation (ChIP) coupled with array hybridization (ChIP-chip) to identify 354 target loci bound by the RARs in MEFs, only a subset of which are induced or repressed by RA. Furthermore, we also mapped 462 RAR binding loci in undifferentiated embryonic stem (ES) cells and show that RAR occupancy of target loci is cell type specific, as only a minority of the MEF loci were occupied in ES cells. We propose that cell-specific occupancy of target promoters by RARs is an

\* Corresponding author. Mailing address: Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 1 Rue Laurent Fries, 67404 Illkirch Cédex, France. Phone: 33 3 88 65 34 40 (45). Fax: 33 3 88 65 32 01. E-mail: irwin@titus.u-strasbg.fr.

† Supplemental material for this article may be found at <http://mc.manuscriptcentral.com/mcb>.

<sup>∇</sup> Published ahead of print on 2 November 2009.

important mechanism contributing to the distinct effects of RA seen in different cell types.

## MATERIALS AND METHODS

**Generation and culture of tagged RAR-expressing cells.** *Taf4<sup>lox/-</sup>* MEFs, as previously described (35), were infected with pBABE retroviruses encoding RAR $\alpha$  or RAR $\gamma$  with a hemagglutinin (HA)-3 $\times$ Flag tag at the C terminus, and populations of cells stably expressing the tagged RARs were selected with puromycin. Cells were grown at 5% CO<sub>2</sub> in Dulbecco's minimal essential medium supplemented with GlutaMAX and 10% fetal calf serum. For the generation of ES cells expressing tagged RAR, recombination targeting vectors containing HA-3 $\times$ Flag or SBP (streptavidin binding peptide)-3 $\times$ Flag (for RAR $\alpha$  and RAR $\gamma$ , respectively), followed by a neomycin resistance cassette flanked by FRT sites, were constructed. Four-kilobase homology arms corresponding to the regions immediately upstream and downstream of the stop codon were cloned on either side of these elements. The construct was electroporated into SV129 ES cells, and the neomycin-resistant clones were screened for homologous recombination by PCR and for expression of the recombinant proteins by Western blotting using the previously described anti-RAR $\alpha$  and -RAR $\gamma$  antibodies (5). All cells were treated with 1  $\mu$ M RA for the indicated times.

**ChIP, ChIP-chip, and ChIP-seq.** ChIP experiments were performed according to standard protocols (data not shown). All ChIP was performed in triplicate and analyzed by triplicate quantitative PCR (qPCR). For ChIP-chip, the total input chromatin and ChIPed material were hybridized to the extended promoter array from Agilent covering the kb -5 to +2 regions of around 17,000 cellular promoters. Data were analyzed with ChIP Analytics from Agilent (see the supplemental material). ChIP-seq was performed using an Illumina GAII sequencer, and the raw data were analyzed by the Illumina Eland pipeline. Peak detection was performed using the MACS software (63), and the peaks were annotated using GPAT (26; data not shown). The antibodies used were against HA (12CA5), pan-RAR (Sc-773), pan-RXR (Sc-774), H3K4me3 (04-745; Upstate), H3K9me3 (07-442; Upstate), H3K27me3 (07-449; Upstate), H3K9Ac (07-352; Upstate), and RNA polymerase II (Pol II; Sc-9001). Flag ChIP was performed with anti-Flag M2 affinity gel (A2220; Sigma). Real-time PCRs were performed on a Roche LightCycler using Roche SYBR green mix. The sequences of the primers used are available on request.

**Electrophoretic mobility shift assays (EMSAs).** EMSAs were performed essentially as previously described (15). After electrophoresis, the gels were dried and exposed to autoradiographic film or a PhosphorImager plate.

## RESULTS

**Identification of RAR target genes in MEFs.** To identify RAR target genes in *Taf4<sup>lox/-</sup>* MEFs, we used retroviral vectors to generate cell lines that stably express RAR $\alpha$  and RAR $\gamma$  carrying a 3 $\times$ Flag-HA tandem affinity purification (TAP) tag at the C terminus. The presence of the TAP tag does not affect RAR function in transient expression assays or activation of RAR target genes in the MEFs (data not shown). Western blotting on extracts of the infected cells showed that the exogenous TAP-tagged RARs made up only around 10 to 20% of the total RAR $\alpha$  or RAR $\gamma$  (Fig. 1A). Cells expressing exogenous RARs also displayed RA-induced morphological changes and serum-independent growth (Fig. 1B), consistent with the idea that their presence does not affect the proper RA response of the cells. ChIP assay using either Flag or HA antibodies showed that tagged RARs and endogenous RXR were efficiently recruited to the DR2-type RARE in the cellular RA binding protein 2 gene (*Crabp2*) promoter, while no enrichment was seen at the control protamine 1 gene (*Pm1*) promoter (Fig. 1C and data not shown).

We used these cells to perform ChIP-chip experiments with the Agilent extended promoter array comprising the kb -5 to +2 regions of around 17,000 cellular promoters. Duplicate tandem Flag ChIP-chip and HA ChIP-chip assays were performed on cells expressing RAR $\gamma$  that had been treated with

RA for 2 h and compared with a tandem Flag ChIP-chip on untagged cells as a control. These experiments identified 354 loci bound by RAR $\gamma$  after Flag and HA ChIP in tagged cells but not in control cells (see Table S1 and Fig. S1A in the supplemental material). For example, occupancy of the characterized DR5 element in the *Rarb* promoter was observed along with occupancy of several other known targets such as *Crabp2*, *Bhlhb2* (*Stra13* [56]), and the ubiquitin-conjugating enzyme E2C gene (*Ube2c*) (64) (Fig. 2A; data not shown). Among the identified loci are previously characterized target genes known to be RA regulated, but not as direct targets, such as the ceruloplasmin (*Cp*), matrix metalloproteinase 2 (*Mmp*), cathepsin D (*Ctsd*), or type 1 alpha collagen (*Colla1*) gene (for a review, see reference 4). Most of the occupied regions are, however, in novel target genes such as the topoisomerase II beta gene (*Top2b*), which is transcribed convergently to *Rarb*, or the interleukin-1 receptor antagonist (*Il1rn*), pyroglutamyl-peptidase I (*Pgpep1*), and tensin like C1 domain-containing phosphatase (*Tenc1*) gene promoters (Fig. 2A). Analysis of 200 of the most sharply defined RAR binding sites showed that 49% were located upstream of the transcription start site (TSS), while 42% were located downstream, principally in the first intron (Fig. 2B).

Gene ontology analysis (<http://david.abcc.ncifcrf.gov/>) showed that the largest class of target genes are involved in transcription regulation, for example, transcription factors such as *Hoxa10* and *Hoxd13* and the RAR coactivators *Hmgal* (39), *Trim16* (12), and *Top2b* (23). One of the main functions of RARs is therefore to regulate other transcriptional networks (Fig. 2D). RARs also bind a significant number of genes involved in signal transduction, protein modification, and metabolism, as well as developmental regulatory genes.

Very similar results were observed in ChIP-chip experiments with cells expressing tagged RAR $\alpha$  (Fig. 2A), as the vast majority of sites were similarly bound by RAR $\alpha$  and RAR $\gamma$  (Fig. 1A; see Table S1 and Fig. S1A in the supplemental material). ChIP-qPCR confirmed specific binding of RAR $\alpha$  and RAR $\gamma$  to promoters in the absence of RA (Fig. 3A). RA substantially increased RAR $\alpha$  binding on the *Rarb*, *Pgpep1*, *Cebpb*, and F11 receptor (*F11r*) promoters, while the strongest increase in RAR $\gamma$  binding was seen at the *Wnt10b* locus, but little or no effect was observed on the other sites. RAR and RXR occupancy was also verified using pan-RAR and pan-RXR antibodies against the endogenous proteins in MEFs that do not express the tagged RARs (Fig. 3B; see also below). Together, the above results show that most, if not all, of the sites were occupied in the absence of RA by both RAR $\alpha$  and RAR $\gamma$  but that the influence of RA is promoter specific.

**Characterization of RAR binding motifs.** To identify the RAR binding motifs, the 1,200 nucleotides flanking the highest-scoring oligonucleotide in each peak were analyzed for the presence of DR1, -2, and -5 elements that had either a perfect match to the known consensus [5'-RGKTC(A-n)<sub>1,2,5</sub>-RGKTC A-3'] or a single mismatch (see Table S2 in the supplemental material). While 149 potential DRs were identified, only 23 were fully consensus sites (Fig. 2C). This was confirmed by a more restricted analysis of the 300 nucleotides around the top 100 most sharply defined peaks that identified only 7 consensus elements (see Table S2 in the supplemental material). This analysis also revealed that 71 of these sites contained at least

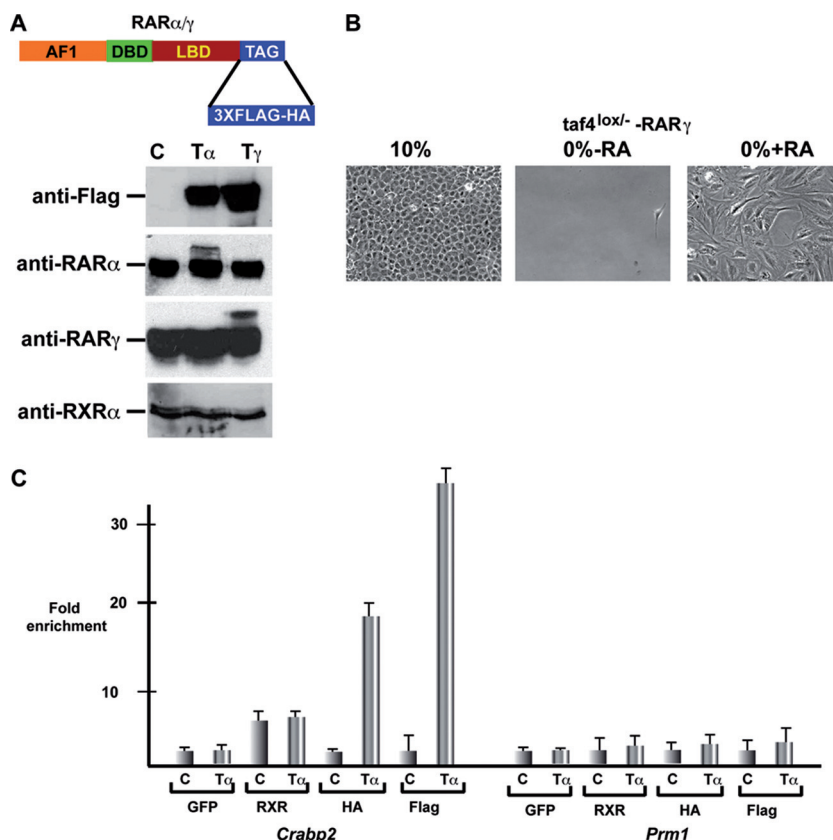


FIG. 1. Expression of tagged RARs in MEFs and ES cells. (A) Schematic representation of RAR with a C-terminal 3 $\times$ Flag-HA tag and Western blot assay showing expression of tagged RARs in total extracts from cells expressing tagged RAR $\alpha$  or RAR $\gamma$  or control cells (T $\alpha$ , T $\gamma$ , and C, respectively). In the upper panel, tagged RARs were detected with anti-Flag antibody, and in the lower panels, endogenous and tagged RARs were detected with isotype-specific antibodies and endogenous RXR was detected with a pan-RXR antibody. (B) Phase-contrast microscopy after 5 days of culture in 10% or 0% serum showing that RA treatment of *Taf4<sup>lox/-</sup>* MEFs expressing tagged RAR $\gamma$  leads to serum-independent growth. (C) ChIP-qPCR on the *Crabp2* RARE and the *Prrm1* promoter in cells expressing tagged RAR $\alpha$  and control cells treated for 2 h with RA. The antibodies used are shown below the graph. The value obtained with each antibody and the control cells was assigned a value of 1, and fold enrichment relative to this value is indicated. GFP, green fluorescent protein.

one consensus half site with a majority containing several anomalously spaced half sites but no obvious DR1, -2, or -5.

We verified RAR binding to several of the novel DRs by competition EMSAs in vitro or ChIP-qPCRs using primers centered on the DR element. In the EMSAs, the retarded RAR/RXR complex formed using an oligonucleotide comprising *Rarb* DR5 was efficiently competed by consensus DR5 elements from the *Atxn2*, *F11r*, *Top2b*, and STIP1 homology and U-box-containing protein 1 (*Stub1*) genes but not by mutated *Rarb* or *Atxn2* oligonucleotides (Fig. 4A, lanes 1 to 7). Efficient competition was also seen with DR5 elements with a single mismatch in the 5' or 3' half sites from the dehydrogenase/reductase (SDR family) member 3 (*Dhrs3*), Toll-interleukin-1 receptor domain-containing adaptor protein (*Tirap*), Rous sarcoma oncogene (*Src*), RB1-inducible coiled-coil 1 (*Rb1cc1*), and procollagen C-endopeptidase enhancer protein (*Pcolce*) genes (Fig. 4A, lanes 9, 10, and 12 to 14). In contrast, little or no competition was observed with the potential DR5 element from the fibroblast growth factor 18 (*Fgf18*) or sedoheptulokinase (*Shpk*) gene (lanes 8 and 11), both of which contain a single mismatch with respect to the consensus.

These results were confirmed by ChIP-qPCR using RAR

and RXR antibodies, where occupancy of the *Rarb*, *Atxn2*, *Dhrs3*, *Tirap*, and *F11r* loci was observed (Fig. 4F), suggesting that these elements bind the RAR/RXR both in vitro and in MEFs. In contrast, although the *Fgf18* and *Shpk* loci are clearly occupied in MEFs (see Fig. S1A in the supplemental material), no occupancy was observed using the amplicon centered on the potential DRs (Fig. 4F). The potential DRs that were identified do not coincide well with the peak detected by ChIP-chip, indicating that the RAR binds not to these DRs but rather another sequence at these loci.

Consensus or single-mismatch DR2 elements from the homeobox a10 (*Hoxa10*), periplakin (*Ppl*), *Crabp2*, growth arrest-specific 2 (*Gas2*), *Ctgf*, neuropilin 1 (*Nrp1*), and nudix (nucleoside diphosphate-linked moiety X)-type motif 6 (*Nudt6*) genes efficiently competed the formation of the retarded complex in vitro (see appropriate lanes in Fig. 4B) and were occupied in MEFs (Fig. 4G and not shown). In contrast, competition with the DR2 elements from the protein phosphatase 1-like (*Ppm1l*) and FCH domain only 2 (*Fcho2*) genes were much less efficient and no competition was seen with the cytochrome P450, family 2, subfamily d, polypeptide 22 (*Cyp2d22*) element.

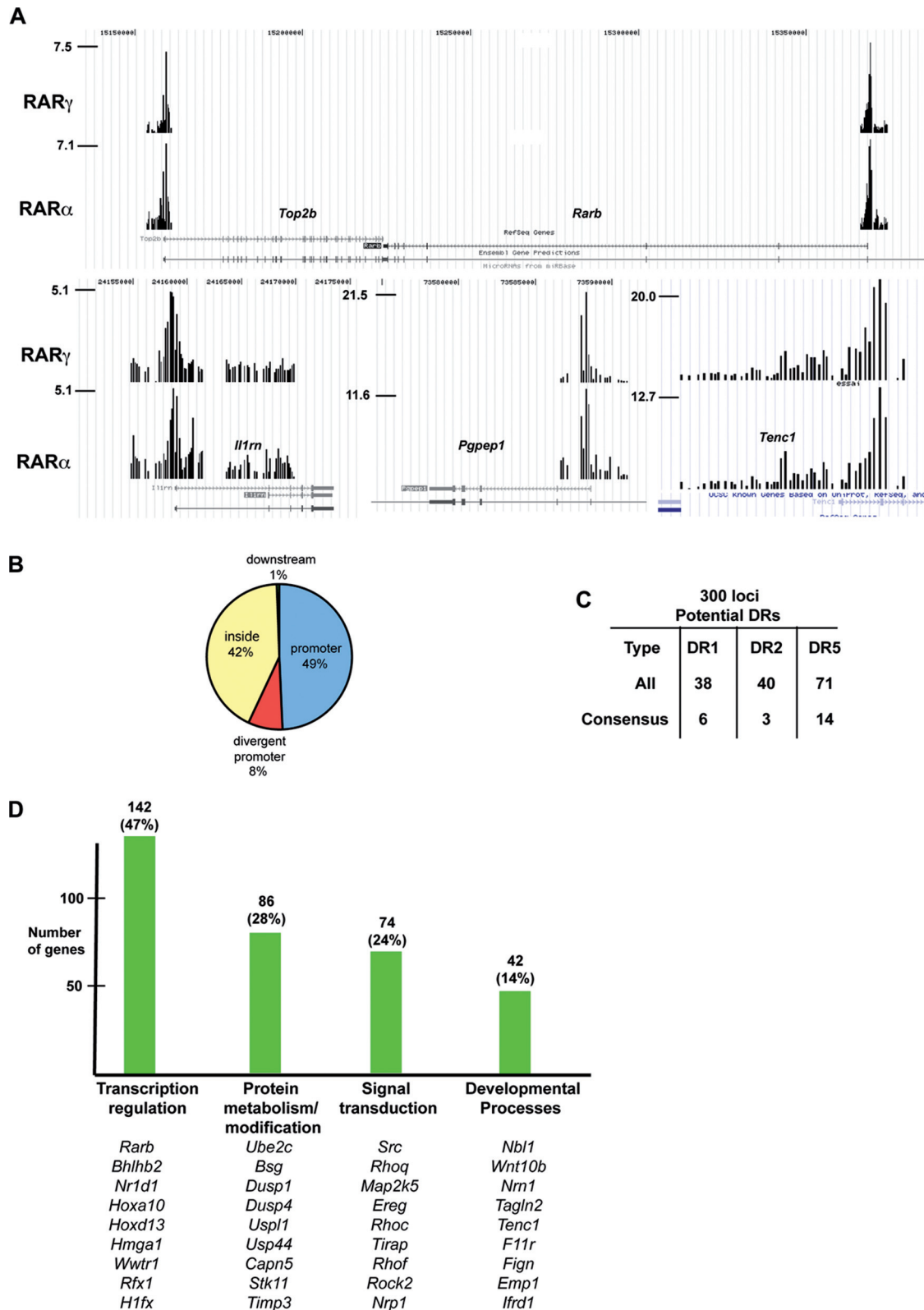


FIG. 2. Identification of RAR binding sites in MEFs. (A) Graphic representation of tandem Flag ChIP-chip results on cells expressing tagged RAR $\alpha$  and RAR $\gamma$  in the UCSC web browser (<http://genome.ucsc.edu/>) at the indicated loci. The values on the y axis show the normalized immunoprecipitate/input ratio. (B) Pie chart representation of the locations of RAR binding sites relative to the TSS. (C) Summary of the presence of classical DR-type RAR binding motifs at ChIPed loci. The total number of potential sites with one mismatch is shown along with the number of consensus sites. (D) Summary of gene ontology analysis of target genes. The number of genes in each category is shown along with representative examples.

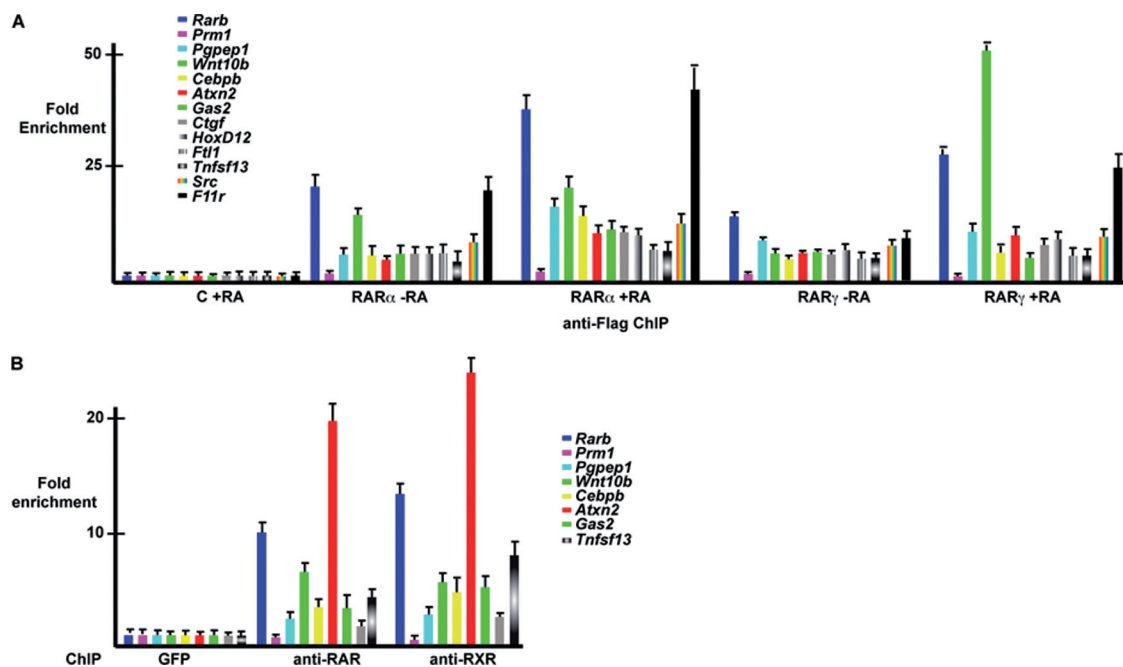


FIG. 3. Binding of RARs to target loci. (A) qPCR on the indicated loci after tandem Flag ChIP on cells expressing tagged RAR $\alpha$  or RAR $\gamma$  in the absence of RA or after 2 h of RA treatment. The value obtained for each amplicon in the control cells was set to 1, and fold enrichment relative to this value is indicated. Oligonucleotide primers are designed at the center of the ChIP peak at each locus. (B) ChIP-qPCR at the indicated loci using pan-RAR and pan-RXR antibodies. The value obtained for each amplicon after control ChIP with anti-GFP antibody was set to 1, and fold enrichment relative to this value is indicated.

For DR1 elements, competition was observed using those of the chromatin-modifying protein 1A (*Chmp1a*), calcitonin receptor activity-modifying protein 2 (*Ramp2*), NR subfamily 1, group D, member 1 (*Nr1d1*), and death effector domain-containing (*Dedd*) genes (Fig. 4C and data not shown). Surprisingly, however, no significant competition was observed with the consensus DR1 motif from the glycerophosphodiester phosphodiesterase 1 (*Gde1*) element. We noted that the spacer nucleotide in this element was C, while it was A in the *Ramp2* and *Dedd* elements that compete. Exchange of the C in *Gde1* for an A resulted in efficient competition (lane 5). Thus, half-site sequences and spacer nucleotides can play important roles in the formation of a complex with RAR/RXR. As observed in vitro, the *Nr1d1* locus, but not *Gde1*, was occupied in cells (Fig. 4G). Similar to *Fgf18*, this indicates that binding to the *Gde1* locus must be mediated by another element.

The above results indicate that many single mismatches do not abrogate RAR/RXR binding even when located at highly conserved positions such as the G in position 2 or 5'-TCA-3'. In contrast, in other elements, the equivalent mutations lead to a loss of binding. For example, in *Cyp2d22* and *Rb1cc1*, 5'-TCA-3' is mutated to 5'-TCG-3' but only *Rb1cc1* binds RAR/RXR. In *Cyp2d22*, replacing the G with an A to create a consensus leads to a strong increase in competition (*Cyp2d22C*, Fig. 4D, lanes 2 and 3). We exchanged the 3' half site of *Cyp2d22* with that of *Rb1cc1* (i.e., changing T to G at the third position). The resulting *Cyp2d22R* was a significantly better competitor than native *Cyp2d22* although less efficient than the full consensus (lanes 2 to 4). Hence, the exact sequence of one half site

plays a critical role in compensating for mismatches in the other half site.

A more striking example of this is found upon analysis of the indolethylamine *N*-methyltransferase (*Inmt*) gene binding site. At this locus, we did not find any consensus or single-mismatch DR element. However, visual inspection identified a DR5 with two mismatches that efficiently competes in an EMSA (Fig. 4E). Inverting the 5' and 3' half sites (*Inmt-RV*) led to a decrease in competition, showing that the mismatches are less well tolerated in the 5' position, where RXR normally binds (lane 4). Similarly, transforming the DR5 to a DR2 also decreases competition (lane 5). Strikingly, however, replacing 5'-GGG-3' in the 5' half site with 5'-AGG-3' or 5'-AGT-3' leads to a strong reduction in competition (lanes 6 to 7), while 5'-GGT-3' competes, but less well than native 5'-GGG-3' (lanes 3 and 8). These results show that in nonconsensus elements, mutations in one half site can be compensated for by the exact sequence of the other half site. In the case of *Inmt*, a G in the first position seems critical to compensate for mutations in the second half site.

**Association of RAR binding sites with other transcription factors.** Many RAR-bound loci did not contain identifiable DR1, -2, and -5 elements but contained consensus half sites. Multiple alignments of the regions adjacent to these half sites did not reveal the presence of an additional consensus sequence for another transcription factor that would bind together with RAR/RXR (see Fig. S2A in the supplemental material). Analysis of a larger 300 nucleotides around the peaks identified numerous consensus sites for SP1 and AP2.

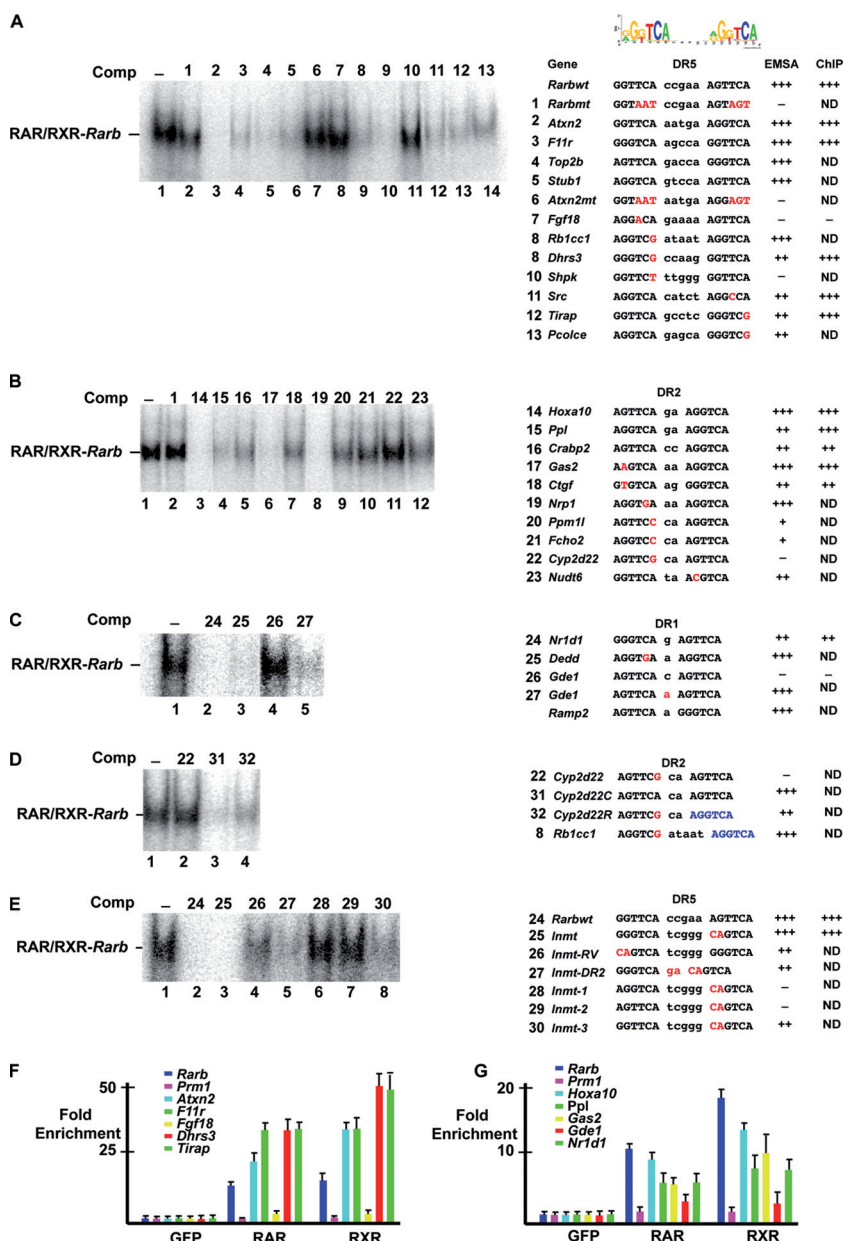


FIG. 4. Characterization of DR-type RAR binding sites. (A to E) Competition EMSAs. EMSAs were performed using *Rarbwt* DR5 as a radioactively labeled probe and a 100-fold excess of the unlabeled competitor (Comp) oligonucleotides shown above each lane. The sequences of the DRs within the oligonucleotides are indicated at the right of each panel. Mismatches with the consensus sequence are shown in red. For the chimeric DR, a color code for each half site is used. The results of competition EMSAs and ChIP assays are summarized on the right. (F and G) ChIP-qPCR on the indicated loci using amplicons centered on the DR element.

However, they were also commonly found in control promoter sets and were not enriched in the RAR-bound set.

It has recently been shown in MCF7 cells that almost 60% of the RAR-bound loci were also bound by estrogen receptor (ER) (21). We therefore examined the 354 RAR-bound loci for ER binding sites by using the consensus 5'-RGGTCWnnn WGACMY-3' (61) but identified only 3 estrogen response elements (EREs) in the MEF data set. Similarly, Hua et al. reported a significant overlap between RAR and FOXA1 binding sites. Using the consensus 5'-TRTTTRYWY-3', we observed 46 potential FOXA1 sites. Thus, assuming that they are

all *bona fide* binding sites, only 12% of the RAR-bound loci show an associated FOXA1 site. Similar results were obtained upon the analysis of 211 of the ES RAR-bound loci (see below), where we found 44 potential FOXA1 sites (20%) but no EREs.

**RA regulation of target gene expression.** We have identified 354 RAR-bound loci, but taking into account the divergent promoters and genes with close neighbors, there are 383 potential RA-regulated genes. Comparison with our previous transcriptome data shows that 267 of the 383 genes were present on the Affymetrix array, but only 48 genes were up-

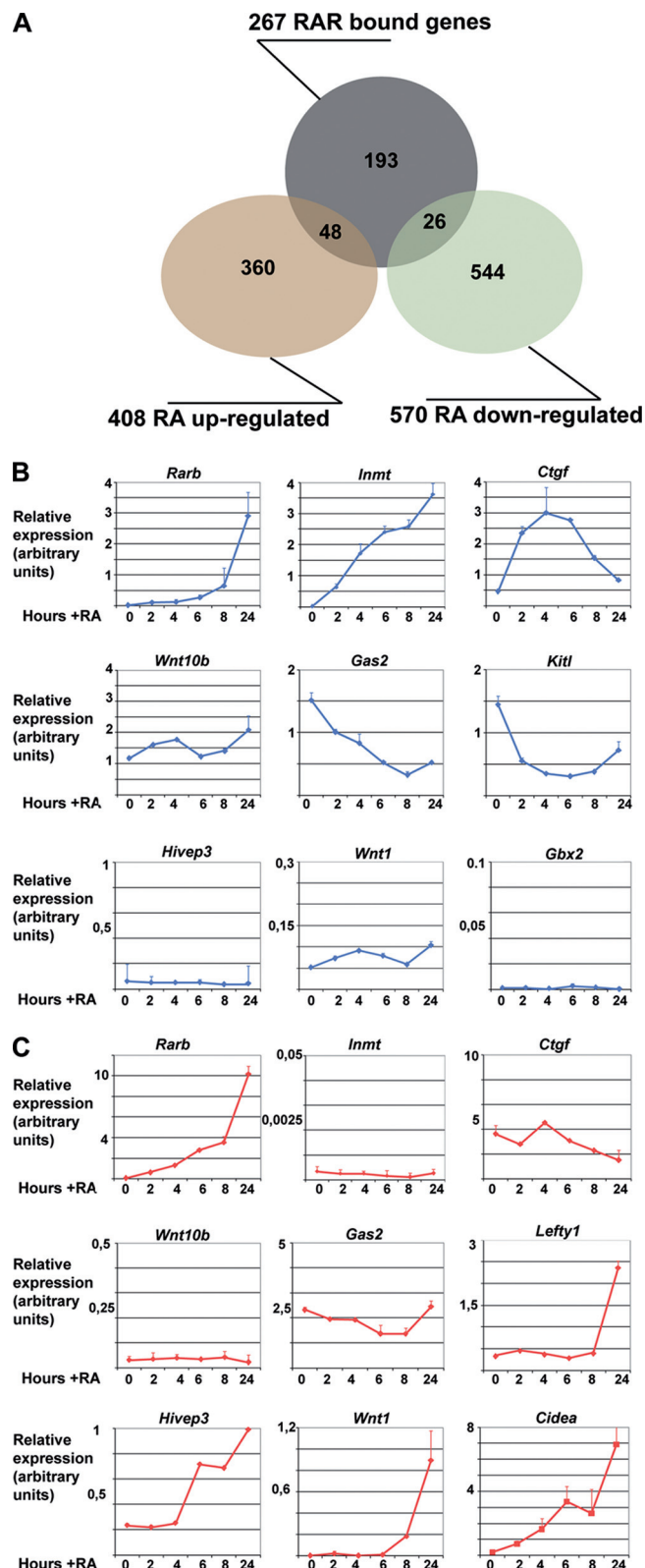


FIG. 5. Expression of RAR target genes. (A) Venn diagram of RAR-bound versus RAR-regulated genes. The numbers refer only to the ChIPed genes that are represented on the Affymetrix array as described in the text. (B) Expression of the indicated target genes in MEFs in the absence or presence of RA over a 24-h period as mea-

regulated and 26 were repressed by RA (Fig. 5A; see Table S1 in the supplemental material). As the transcriptome analysis was performed after 24 h of RA treatment, we verified gene expression to determine whether we had missed many transiently regulated genes. Reverse transcription coupled to qPCR identified several categories of genes. *Rarb* and *Inmt* are not expressed in the absence of RA and are strongly induced over a 24-h period (Fig. 5B). Others, such as *Ctgf*, show significant expression in the absence of ligand but are transiently stimulated by RA. In contrast, genes such as *Gas2* or Kit ligand (*Kitl*) are repressed by RA. However, the expression of wingless-related MMTV integration site 10b (*Wnt10b*), like that of many other genes (data not shown), is not affected by RA. These observations indicate that only around 27% of the RAR-bound genes are regulated by RA in these cells under the experimental conditions used.

We investigated several promoters more closely by using primer pairs designed to amplify the TSS and regions located at  $-2$  kb and  $+2$  kb relative to the TSS. RAR/RXR occupancy of *Rarb* DR5 was observed in the absence of RA and was strongly increased after 2 to 24 h with RA (Fig. 6A). In contrast, Pol II was observed only in the presence of ligand. This result differs from that observed previously in P19 cells, where some Pol II was present at *Rarb* in the absence of RA (31, 46). Interestingly, significant H3K4me3 and H3K9ac covalent histone modifications tightly associated with transcriptional activity (57) were observed in the absence of RA and strongly increased after RA treatment (Fig. 6B). Lower levels of both of these markers were also observed downstream of the TSS.

At the *Inmt* promoter, RAR/RXR binding to the nonconsensus DR5 located 4.2 kb upstream of the TSS was observed in the absence of RA and increased by RA (Fig. 6C). No Pol II was observed at the TSS in the absence of RA, but it was strongly recruited in its presence. Interestingly, a minor but significant RA-dependent recruitment of Pol II at the RAR binding site was also observed. H3K4me3 was seen only in the presence of RA at the TSS and in the downstream region (Fig. 6D). Thus, unlike *Rarb*, where H3K4me3 was seen in the absence of RA, this marker is fully RA dependent at *Inmt*. H3K9ac was observed at the TSS, in the downstream region, and also at the RAR binding site and was stimulated in each region by RA.

We were unable to detect such striking changes at RA-repressed promoters. At the *Gas2* promoter, RAR and RXR occupancy of the downstream DR2 element could be observed in the absence of RA and was not altered in its presence (Fig. 6E). Pol II occupancy of this promoter was observed in the absence of ligand, consistent with its constitutive expression; however, no significant changes in Pol II promoter occupancy were observed in the presence of RA. Similarly, high levels of the H3K4me3 and H3K9ac markers were observed in the absence of RA (Fig. 6F), but a transient reduction was seen in the presence of RA concomitant with the transient repression (Fig.

sured by qPCR and normalized to the ribosomal *Rplp0* transcript. (C) Normalized expression in ES cells in the absence or presence of RA for the indicated times as measured by qPCR. In panels B and C, the relative expression at each time point is shown in arbitrary units.



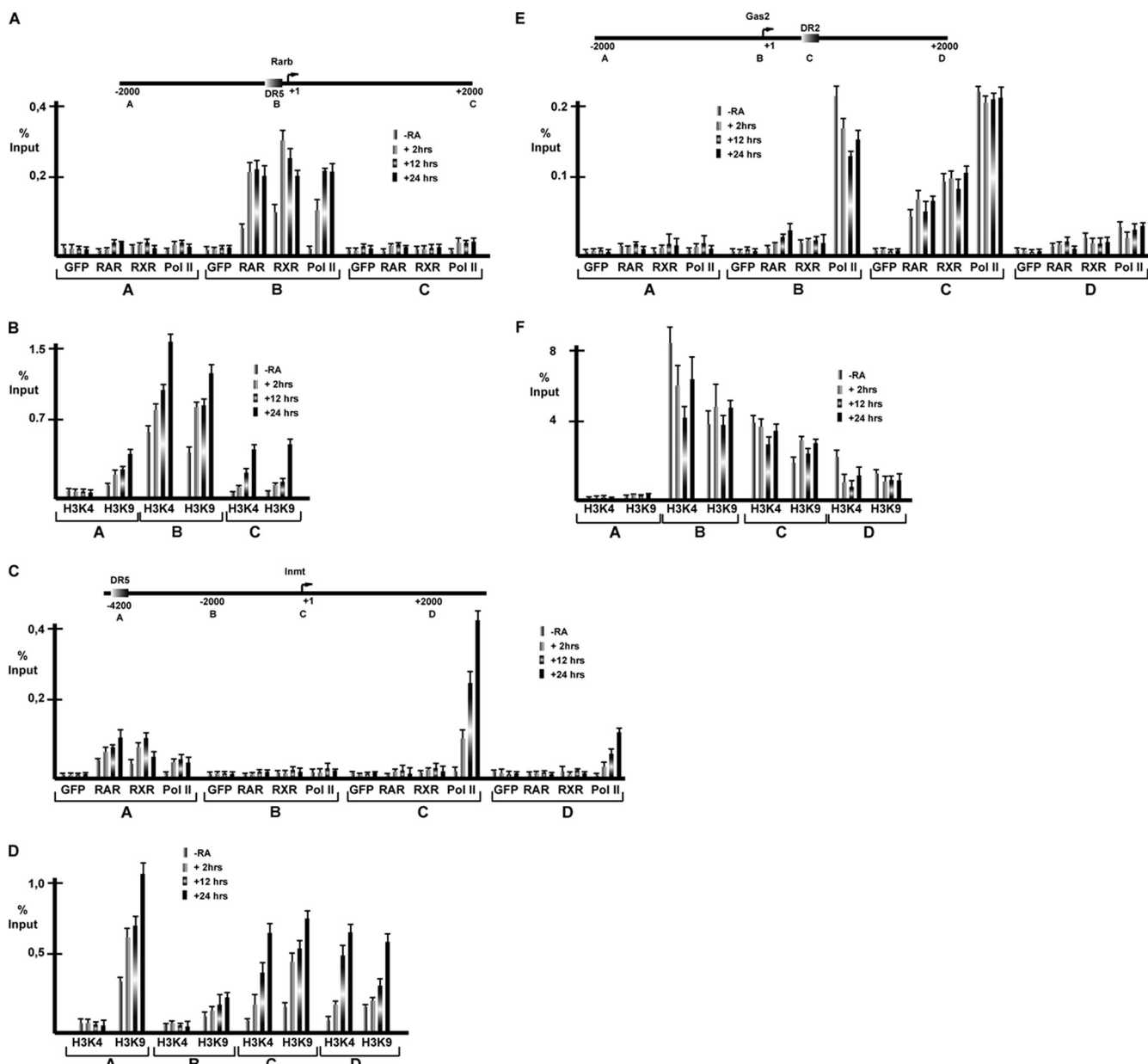


FIG. 6. ChIP at RA-regulated genes. (A and B) Schematic representation of the *Rarb* promoter with DR5 located close to the TSS. The localization of each amplicon is shown. ChIP was performed with the antibodies indicated below the graph, GFP, RAR, RXR, Pol II, H3K4me3, and H3K9ac. Results are expressed as percentages of the input. (C and D) Schematic representation of the *Inmt* promoter. The nonconsensus DR5 RAR binding site is located 4.2 kb upstream of the TSS. ChIP was performed as described above. (E and F) Schematic representation of the *Gas2* promoter with the DR2 located 300 bp downstream of the TSS.

5B). Thus, while the strong *de novo* activation is reflected by Pol II recruitment and the appearance and/or increase of histone modifications associated with transcription, at repressed genes, only minor and transient changes were seen.

**Cell-specific binding of RAR to its cognate elements.** We have previously noted that the repertoire of RA-regulated genes in MEFs is significantly different from that observed in other cell types such as F9 or ES cells (15). To examine the basis for this specificity, we generated ES cells in which the genes encoding RAR $\alpha$  and RAR $\gamma$  were modified by homologous recombination to introduce a TAP tag at the C terminus

(Fig. 7A). Western blot analysis of extracts from the tagged ES cells showed somewhat lower expression of tagged than untagged RAR, probably due to the presence of the neomycin resistance cassette in the 3' untranslated region.

We compared duplicate tandem Flag ChIP-chip on ES cells expressing tagged RAR $\gamma$  to that on untagged cells to identify 462 bound loci. Among these are many previously identified genes, such as the *Rarb*, caudal-type homeobox 1 (*Cdx1*), left-right determination factor 1 (*Lefty1*), and pluripotency factor *Pou5f1* (*Oct4*) genes, as well as a series of novel genes (see Table S3 and Fig. S1B in the supplemental material). Analysis of the 211 genes

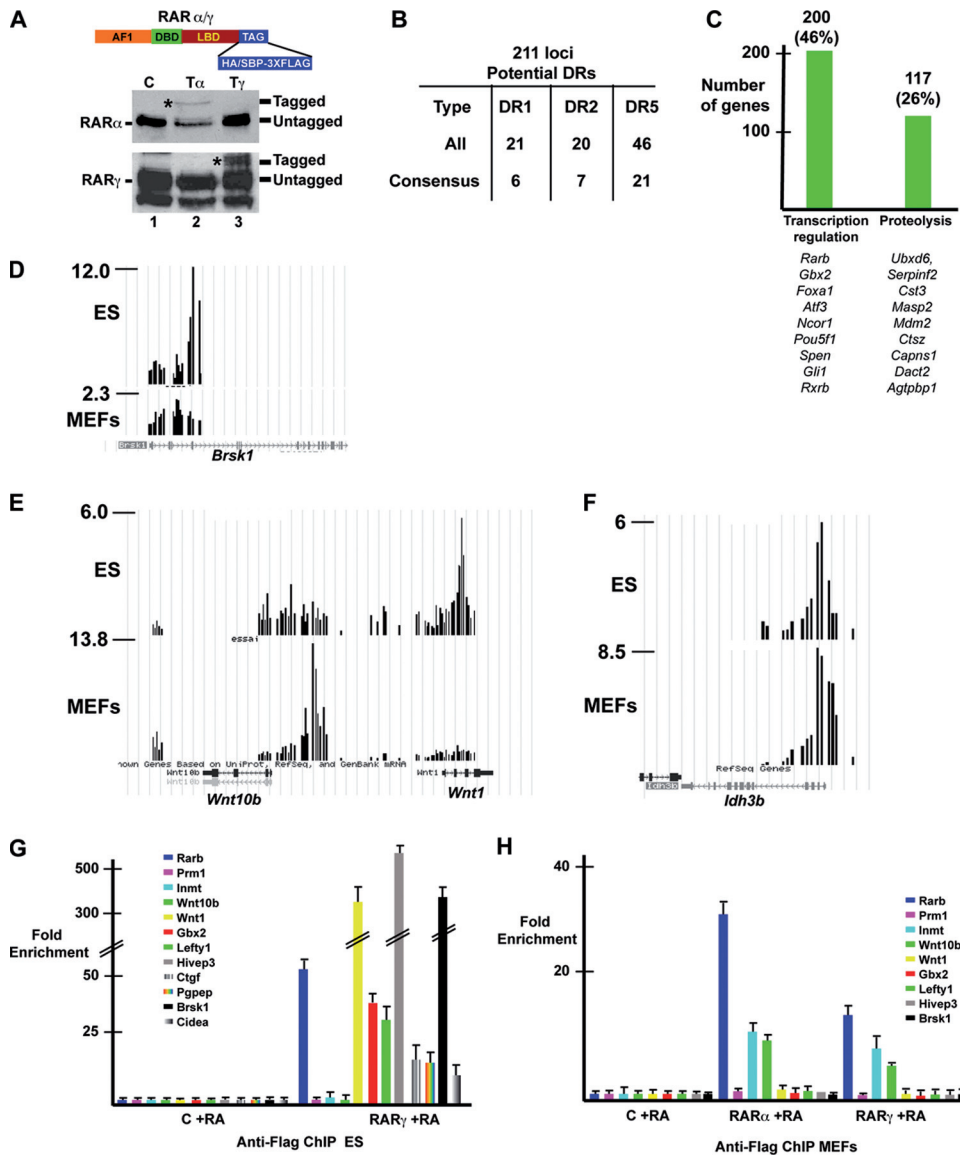


FIG. 7. Identification of RAR binding sites in ES cells. (A) Schematic representation of RARs with a C-terminal 3×Flag-SBP or 3×Flag-HA tag and Western blot assay showing the expression of tagged and endogenous RARs in extracts of recombinant or control ES cells. (B) Summary of classical DR-type RAR binding motifs at ChIPed ES cell loci as described in the legend to Fig. 1C. (C) Summary of gene ontology analysis of target genes. (D to F) Graphic representation of tandem Flag ChIP-chip results on ES cells and MEFs expressing tagged RAR $\gamma$  in the UCSC web browser at the indicated loci. The y axis shows the normalized immunoprecipitate/input ratio. (G and H) Flag ChIP-qPCR at the indicated loci in tagged and control MEFs and ES cells.

with the sharpest peaks indicated that, as observed in MEFs, the majority of the binding sites contained not consensus DR elements but either degenerate DRs or no identifiable DRs but anomalously spaced consensus half sites (Fig. 7B; see Table S3 in the supplemental material). Gene ontology analysis indicated that the largest functional classes of target genes were related to transcription regulation and proteolysis (Fig. 7C).

Comparison with MEFs showed that only 58 loci were bound in both cell types (see Table S3 in the supplemental material). For example, the brain serine/threonine kinase 1 (*Brsk1*) locus is occupied in ES cells but not in MEFs (Fig. 7D). At the *Wnt10b-Wnt1* locus, RAR binds to a site upstream of the *Wnt10b* gene in MEFs not occupied in ES cells, whereas in

ES cells, RAR occupies a site within the *Wnt1* gene (Fig. 7E). Similarly, the *HoxA* and *HoxB* loci are differentially occupied in MEFs and ES cells (see Fig. S1C in the supplemental material). In contrast, the retinitis pigmentosa isocitrate dehydrogenase 3 (NAD<sup>+</sup>) beta (*Idh3b*) locus (18) was comparably occupied in both cell types (Fig. 7F).

These results were verified by ChIP-qPCR. In ES cells, RAR occupancy of the *Lefty1*, human immunodeficiency virus type I enhancer binding protein 3 (*Hivep3*; also known as Schnurri-3 [*Shn3*]), gastrulation brain homeobox 2 (*Gbx2*), *Brsk1*, and cell death-inducing DNA fragmentation factor alpha subunit-like effector A (*Cidea*) loci, but not of the *Wnt10b* and *Inmt* loci, was observed (7G). In contrast, in MEFs, occupancy of the

*Inmt* and *Wnt10b* loci was observed, but not of the *Lefty1*, *Gbx2*, *Hivep3*, and *Brsk1* loci, while the *Rarb* and *Ctcf* loci are occupied in both cell types (Fig. 7G and H and 3A [discussed above]). These results show clearly that binding of RAR to many of its cognate elements is cell type specific.

We assessed the expression and RA regulation of several genes in the different cell types. In ES cells, *Rarb* is not expressed in the absence of RA and is progressively and strongly induced by RA (Fig. 5B and C). In contrast, *Inmt* is neither expressed nor induced by RA in ES cells. *Wnt10b* and *Gas2* were expressed at low levels in ES cells and were not regulated by RA. *Hivep3* and *Wnt1* are strongly upregulated by RA in ES cells, while they are neither expressed nor RA regulated in MEFs. *Gbx2* is also not RA regulated in MEFs.

**Chromatin modifications at target loci.** We compared the chromatin states of these loci in the two cell types. In MEFs, the highest level of H3K4me3 is observed at the constitutively active *Gas2* promoter, while it is strongly increased by RA at *Rarb* and *Inmt* (Fig. 8A). In contrast, no significant H3K4me3 was observed at *Hivep3* and *Wnt1* in the presence or absence of RA. In ES cells, a high level of H3K4me3 was seen at *Rarb* that was only mildly increased by RA (Fig. 8B). H3K4me3 was also observed at *Hivep3* and *Wnt1* and significantly increased by RA. In contrast, there was no H3K4me3 at *Inmt* in the presence or absence of RA, while at *Gas2*, levels were much lower in ES cells than in MEFs. Similar results were seen for H3K9ac (Fig. 8C and D), although it is worth noting that low but significant levels (compared to major satellite as a negative control) were seen at *Hivep3* in MEFs and at *Inmt* in ES cells, where the corresponding genes were not expressed.

In MEFs, a significantly higher level of the repressive H3K27me3 marker was observed at *Hivep3* and *Wnt1* than at the others (Fig. 8E). In ES cells, H3K27me3 at *Rarb* and *Hivep3* was decreased by RA, while it was unaffected at the other loci (Fig. 8E). Unlike MEFs, however, the low or inactive *Gas2* and *Inmt* promoters did not show H3K27me3 levels significantly higher than those of the expressed genes. In MEFs, the *Hivep3* and *Wnt1* promoters also showed higher levels of H3K9me3 than did the others, while in ES cells, none of the promoters showed a high level of this marker (Fig. 8G and H).

Together, these results show that the *Hivep3* and *Wnt1* promoters are in a repressed state in MEFs, suggesting that this chromatin conformation may restrict RAR/RXR binding to these loci. In ES cells, the *Inmt* locus shows no significant H3K4me3, in agreement with the lack of its expression, and neither this promoter nor the *Gas2* promoter showed higher levels of the repressive H3K9 and H3K27me3 markers. Thus, these promoters are not in a repressed heterochromatin state that would account for the lack of RAR binding in ES cells.

These results are similar, but not identical, to those obtained in the global mapping of H3K4me3 in ES cells and MEFs in the absence of RA (36). In this study, differential H3K4me3 at *Rarb* and *Inmt* was observed in MEFs, but *Hivep3* and *Wnt1* had H3K4me3, while we clearly do not detect this marker at these loci. This perhaps reflects differences between individual MEF lines. In contrast, in ES cells, our results are in agreement with those of Mikkelsen et al.

These observations prompted us to make a global comparison of RAR binding and H3K4me3 in MEFs and ES cells. We performed H3K4me3 ChIP-seq on our MEFs and used these

data and the Mikkelsen ES cell data for comparison (see Table S4 in the supplemental material). This analysis shows that the vast majority of genes bound by RAR in MEFs show significant H3K4me3 in the absence of RA, only a minority are in a repressed state (Fig. 8I; see Fig. S1D in the supplemental material). Similarly, although the *Wnt1* and *Hivep3* loci are in a repressed state in MEFs, the majority of the genes bound selectively in ES cells by RAR are in an active state in MEFs. A similar situation exists in ES cells, where most RAR-bound genes, whether in MEFs or in ES cells, show H3K4me3 and are in an active state or are bivalent promoters marked by both K4 and K27 trimethylation (Fig. 8J). Together, these results indicate that while the heterochromatinization in MEFs of some genes active in ES cells may explain the loss of RAR binding, this is not a general mechanism to account for the cell-specific association of RAR with its binding loci.

## DISCUSSION

**RAR binding and gene expression.** We describe the identification of 354 RAR-bound loci in the regions flanking 383 potential target genes in MEFs. These sites obviously do not correspond to the full repertoire, as the arrays cover an extended region around the TSS and many NR binding sites are known to be located far from their regulated genes (9, 43). For example, we did not find sites around the *Enpp1-Enpp3* locus, which is rapidly induced by RA in these cells (15). The RAR binding sites regulating their expression may therefore be located further upstream or downstream of the TSSs.

We initiated this study to better understand the previously described RA-TGF- $\beta$  cross talk (15). Our data confirm that RAR binds directly to a DR2 element in the *Ctcf* promoter, mediating RA induction of its expression and the resulting autocrine growth. Furthermore, we observe RAR binding to the gene encoding the TGF- $\beta$ 3 ligand, which is also strongly induced upon RA treatment. The direct activation of these two genes by RA is therefore central to the reported cross talk. Aside from these two genes, the other RAR-bound components of the TGF- $\beta$  pathway, with the exception of *Skil*, are either weakly or not RA regulated in MEFs.

In addition to the TGF- $\beta$  pathway, we observed RAR bound to several genes involved in other pathways of cell cycle regulation and transformation, such as tumor necrosis factor (ligand) superfamily member 13 (*Tnfrsf13*; also called APRIL, a proliferation-inducing ligand [47]) cyclin D1 (*Ccnd1*), metastasis marker nuclear protein 1 (*Nupr1*), growth factor *Fgf18*, and the tyrosine kinase *Src*. Furthermore, RARs bind to the gene encoding *Src*-related *Shc1(A)* and cortactin (*Cttn*), an SRC substrate involved in metastasis (60). These results suggest novel pathways by which RA may regulate cell proliferation and cancer. It is also striking that RARs can potentially regulate genes involved in degenerative pathologies such as *Atxn2* and dystrophin myotonic protein kinase (*Dmpk*) or Paget's disease through sequestosome 1 (*Sqstm1*) and tumor necrosis factor receptor superfamily member 11b (*Tnfrsf11b*, osteoprotegerin) (48). On the other hand, RARs also occupy the regulatory regions of WNK lysine-deficient protein kinase 4 (*Wnk4*) and odd-skipped-related 1 (*Osr1*), which act together to regulate salt transport and blood pressure (52). We also note that RARs bind to the Rho-GTPase-encoding genes

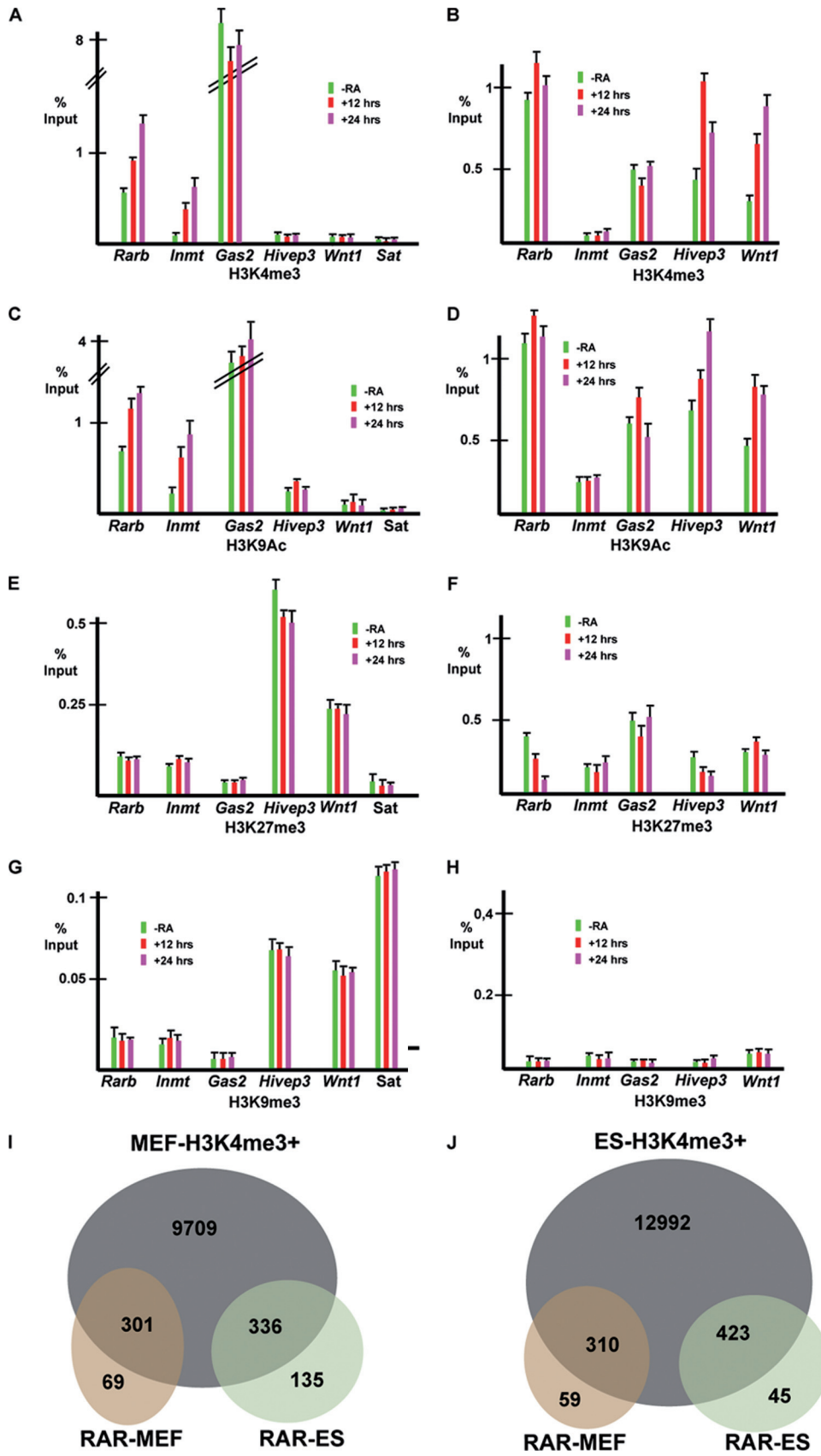


FIG. 8. Chromatin at RAR-bound loci. (A to H) ChIP-qPCRs using the indicated antibodies at loci in MEFs (A, C, E, and G) and ES cells (B, D, F, and H). Amplicons are centered on the TSS. (I and J) Venn diagrams comparing RAR-bound loci and their H3K4me3 status in MEFs (I) and ES cells (J). The total number of loci with H3K4me3 in MEFs and H3K4me3 in addition to the bivalent loci is shown as an intersection with the RAR loci bound in MEFs or ES cells.

*Rhoc*, *Rhod*, *Rhof*, *Rhoq*, *Rhobtb2*, *Rhobtb3*, and *Rnd3* (*Rhoe*), as well as the downstream effector kinase gene *Rock2* and the substrate Cofilin2 (*Cfn2*) gene. These genes act in the pathways that regulate actin dynamics and vesicle trafficking. RARs thus have the potential to control multiple cellular functions, and it is therefore important to determine if, when, and where RA may regulate these genes.

In ES cells, RAR is also found to occupy genes with a wide variety of functions. In addition to *Pou5F1*, RAR may also regulate other genes involved in pluripotency, such as *Lin28* (45) or *Utf1* (44), left-right determination through *Lefty1* and *Lefty2*, and also *Ier2* (20). RAR occupies the regulatory regions of genes encoding proteins involved in several other signaling pathways such as *Gdf3* (32) and *Nodal*, *Notch4* and its mediator *Zfp64* (55), and the *Ptch1*, *Gli1*, *Zic2*, and Tectonic (*Tect1*) components of the sonic hedgehog pathway (50), as well as *Fgf1*, *Fgf4*, and *Fgf8*. RAR regulation of one or another of these genes and their associated pathways contributes to the differentiation of ES cells in vitro and developmental processes in vivo (7, 13, 33, 42).

Although our results show RAR occupancy of genes with diverse functions, the majority of MEF RAR target genes are not RA regulated under the conditions used. The lack of RA response may, in some cases, be due to the fact that the genes are already expressed and cannot be further stimulated, but in many others, low or no basal expression is observed and yet there is no stimulation by RA. However, some of the genes not regulated in MEFs are RA regulated in other cell types or tissues. For example, *Gde1* is RA regulated in Wilms' tumor cells (64), while *Myc* is regulated in various cell types but not in MEFs (4). It remains to be determined whether all RAR-bound genes can be RA regulated in the appropriate cell type and under the appropriate conditions or whether there are instances in which RAR is always a silent partner. The observation that only a subset of RAR-bound genes are RA regulated is in keeping with what has been observed for other transcription factors. For example, only a small subset of CREB-bound genes are regulated by cyclic AMP (62) and only a subset of ER-bound genes by are regulated estrogen (61).

In contrast to what may have been expected from a model in which unliganded RAR recruits repressor complexes to silence target genes, a large majority of genes bound by unliganded RAR show significant H3K4me3 and expression. Although this may be explained by heterogeneity in the cell population where RAR is not bound in cells where the genes are active, our results rather suggest a model where unliganded RARs actively silence only a small subset of their target genes.

**Sequence diversity of RAR binding sites.** Our results show that the majority of the RAR-bound loci do not contain consensus DR1, -2, or -5 elements. At the bound MEF loci, we identified 149 potential DRs. EMSA analysis showed, however, that the only mismatch consistently found not to affect interaction was 5'-TCA-3' to 5'-TCG-3'. Taking this criterion into account, only 27 *bone fide* DR elements in MEFs and 42 in ES cells can be identified with confidence. Although elements with other mismatches can bind RAR/RXR in vitro, we cannot readily predict the binding properties of DRs.

A closer investigation of the MEF and ES cell bound loci revealed that many did not contain readily identifiable DR1, -2, or -5 elements, even considering multiple mismatches, but

rather comprised one or several anomalously spaced consensus half sites. There is no correlation between the presence and absence of recognizable DR1, -2, or -5 elements and gene regulation. Several genes with consensus DR elements are not RA regulated in MEFs, while genes with no identified DR are RA regulated. Furthermore, functional anomalously spaced half sites have previously been documented (24, 49). Multiple alignments of these half sites did not reveal the systematic presence of cryptic half sites or the presence of a consensus sequence for another transcription factor that would bind together with RAR/RXR. Our results therefore show that the majority of RAR binding sites comprise either degenerate DRs with multiple mismatches; no DR1, -2, or -5; or anomalously spaced consensus half sites. A similar observation was made for PML-RAR $\alpha$  binding sites, the majority of which lack consensus DRs (19). This was ascribed to a relaxed specificity of the PML-RAR chimera but clearly also seems to be the case with the native RAR. Alternatively, the paucity of DR elements at the identified loci could be explained if they were not, in fact, the primary RAR binding sites but were detected as a consequence of DNA loop formation due to interactions between RARs bound at more distal sites and factors bound at the proximal promoter. Formaldehyde cross linking could capture such interactions, allowing the detection of RAR at the promoter even although the authentic binding sites are in distal enhancer regions. Such interactions have been observed with other NRs (9, 59).

While this work was in progress, Hua et al. reported a significant overlap between RAR and ER binding in MCF7 breast cancer cells. In contrast, we did not find significant numbers of palindromic EREs in the RAR-bound loci in MEFs and no EREs in the 211 analyzed ES cell loci. Similarly, we found potential FOXA1 sites in only 12 to 20% of the RAR-bound loci, much less than the 50% seen in MCF7 cells. Our observations are similar to those recently reported for ER binding sites in MCF7 cells (61), where, in contrast to a previous report (10), only 6 to 10% of the ER binding loci contain FOXA1 binding sites.

**Chromatin topology contributes to cell-specific RAR occupancy.** Comparison of RAR-bound loci in MEFs and ES cells indicates that the majority are occupied in a cell-specific manner. We identified a subset of loci that are bound in ES cells but are in an inactive state in MEFs that is characterized by repressive chromatin markers and/or increased DNA methylation (our unpublished data). The compaction into a repressed heterochromatin state may account for the absence of RAR binding to these loci in MEFs. However, most ES cell-specific loci are associated with genes that are expressed and marked with H3K4me3 in MEFs. Similarly, most genes that are selectively bound in MEFs are marked by H3K4me3 and/or H3K27me3 in ES cells and are thus expressed or in a bivalent state. Thus, while chromatin condensation may explain the lack of RAR binding to a subset of sites in MEFs, there must be other mechanisms that regulate the accessibility of the majority of the loci in both MEFs and ES cells to RAR.

RA exerts distinct effects in different cell types. In ES, F9, or HL60 cells, RA induces cell cycle arrest and differentiation (1, 37, 53). In contrast, in MEFs or hepatocytes, RA promotes cell proliferation (15, 25, 29). It has been proposed that differential interaction of RAR with distinct sets of coregulatory com-

plexes may explain the observed cell-specific effects of RA (54). Our results indicate that cell-specific binding of RAR to its genomic sites is a major mechanism regulating the repertoire of target genes that can be regulated and hence the biological effects of RA.

#### ACKNOWLEDGMENTS

We thank, C. Rochette-Egly for the anti-RAR antibodies and expression vectors, J. Greenblatt and M Rudnicki for the backbone of the targeting vector, Y. Anno for help with bioinformatics, and C. Bole-Feyssot for CHIP-chip labeling and hybridization.

This work was supported by grants from the CNRS, the INSERM, the Association pour la Recherche contre le Cancer, the Ligue Nationale et Départementale Région Alsace contre le Cancer, the INCA, and the ANR Regulome project grant. I.D. is an équipe labéllisée of the Ligue Nationale contre le Cancer. L.D. and G.A. were supported by the EuTRACC program of the European Union and the French ANR.

#### REFERENCES

- Altucci, L., M. D. Leibowitz, K. M. Ogilvie, A. R. de Lera, and H. Gronemeyer. 2007. RAR and RXR modulation in cancer and metabolic disease. *Nat. Rev. Drug Discov.* **6**:793–810.
- Bain, D. L., A. F. Heneghan, K. D. Connaghan-Jones, and M. T. Miura. 2007. Nuclear receptor structure: implications for function. *Annu. Rev. Physiol.* **69**:201–220.
- Balmer, J. E., and R. Blomhoff. 2005. A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.* **96**:347–354.
- Balmer, J. E., and R. Blomhoff. 2002. Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**:1773–1808.
- Bastien, J., S. Adam-Stitah, T. Riedl, J. M. Egly, P. Chambon, and C. Rochette-Egly. 2000. TFIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J. Biol. Chem.* **275**:21896–21904.
- Bastien, J., and C. Rochette-Egly. 2004. Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **328**:1–16.
- Bibel, M., J. Richter, K. Schrenk, K. L. Tucker, V. Staiger, M. Korte, M. Goetz, and Y. A. Barde. 2004. Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**:1003–1009.
- Blomhoff, R., and H. K. Blomhoff. 2006. Overview of retinoid metabolism and function. *J. Neurobiol.* **66**:606–630.
- Carroll, J. S., X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoutte, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver, and M. Brown. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**:33–43.
- Carroll, J. S., C. A. Meyer, J. Song, W. Li, T. R. Geistlinger, J. Eeckhoutte, A. S. Brodsky, E. K. Keeton, K. C. Fertuck, G. F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E. A. Fox, P. A. Silver, T. R. Gingeras, X. S. Liu, and M. Brown. 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**:1289–1297.
- Cavailles, V., S. Dauvois, F. L'Horsset, G. Lopez, S. Hoare, P. J. Kushner, and M. G. Parker. 1995. Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *EMBO J.* **14**:3741–3751.
- Cheung, B. B., J. Bell, A. Raif, A. Bohlken, J. Yan, B. Roediger, A. Poljak, S. Smith, M. Lee, W. D. Thomas, M. Kavallaris, M. Norris, M. Haber, H. L. Liu, D. Zajchowski, and G. M. Marshall. 2006. The estrogen-responsive B box protein is a novel regulator of the retinoid signal. *J. Biol. Chem.* **281**:18246–18256.
- Dequéant, M. L., and O. Pourquie. 2008. Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.* **9**:370–382.
- Epping, M. T., L. Wang, M. J. Edel, L. Carlee, M. Hernandez, and R. Bernards. 2005. The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell* **122**:835–847.
- Fadloun, A., D. Kobi, L. Delacroix, D. Dembele, I. Michel, A. Lardinois, J. Tisserand, R. Losson, G. Mengus, and I. Davidson. 2008. Retinoic acid induces TGFbeta-dependent autocrine fibroblast growth. *Oncogene* **27**:477–489.
- Germain, P., P. Chambon, G. Eichele, R. M. Evans, M. A. Lazar, M. Leid, A. R. De Lera, R. Lotan, D. J. Mangelsdorf, and H. Gronemeyer. 2006. International Union of Pharmacology. LX. Retinoic acid receptors. *Pharmacol. Rev.* **58**:712–725.
- Germain, P., P. Chambon, G. Eichele, R. M. Evans, M. A. Lazar, M. Leid, A. R. De Lera, R. Lotan, D. J. Mangelsdorf, and H. Gronemeyer. 2006. International Union of Pharmacology. LXIII. Retinoid X receptors. *Pharmacol. Rev.* **58**:760–772.
- Hartong, D. T., M. Dange, T. L. McGee, E. L. Berson, T. P. Dryja, and R. F. Colman. 2008. Insights from retinitis pigmentosa into the roles of isocitrate dehydrogenases in the Krebs cycle. *Nat. Genet.* **40**:1230–1234.
- Hoemme, C., A. Peerzada, G. Behre, Y. Wang, M. McClelland, K. Nieselt, M. Zschunke, C. Disselhoff, S. Agrawal, F. Isken, N. Tidow, W. E. Berdel, H. Serve, and C. Muller-Tidow. 2008. Chromatin modifications induced by PML-RARalpha repress critical targets in leukemogenesis as analyzed by ChIP-Chip. *Blood* **111**:2887–2895.
- Hong, S. K., and I. B. Dawid. 2009. FGF-dependent left-right asymmetry patterning in zebrafish is mediated by *Ier2* and *Fibp1*. *Proc. Natl. Acad. Sci. U. S. A.* **106**:2230–2235.
- Hua, S., R. Kittler, and K. P. White. 2009. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137**:1259–1271.
- John, S., P. J. Sabo, T. A. Johnson, M. H. Sung, S. C. Biddie, S. L. Lightman, T. C. Voss, S. R. Davis, P. S. Meltzer, J. A. Stamatoyannopoulos, and G. L. Hager. 2008. Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol. Cell* **29**:611–624.
- Ju, B. G., V. V. Lunyak, V. Perissi, I. Garcia-Bassets, D. W. Rose, C. K. Glass, and M. G. Rosenfeld. 2006. A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science* **312**:1798–1802.
- Kato, S., H. Sasaki, M. Suzawa, S. Masushige, L. Tora, P. Chambon, and H. Gronemeyer. 1995. Widely spaced, directly repeated PuGGTCA elements act as promiscuous enhancers for different classes of nuclear receptors. *Mol. Cell. Biol.* **15**:5858–5867.
- Khetchoumian, K., M. Teletin, J. Tisserand, M. Mark, B. Herquel, M. Ignat, J. Zucman-Rossi, F. Cammas, T. Lerouge, C. Thibault, D. Metzger, P. Chambon, and R. Losson. 2007. Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha. *Nat. Genet.* **39**:1500–1506.
- Krebs, A., M. Frontini, and L. Tora. 2008. GPAT: retrieval of genomic annotation from large genomic position datasets. *BMC Bioinformatics* **9**:533.
- Laperriere, D., T. T. Wang, J. H. White, and S. Mader. 2007. Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics* **8**:23.
- Leask, A., and D. J. Abraham. 2004. TGF-beta signaling and the fibrotic response. *FASEB J.* **18**:816–827.
- Ledda-Columbano, G. M., M. Pibiri, F. Molotzu, C. Cossu, L. Sanna, G. Simbula, A. Perra, and A. Columbano. 2004. Induction of hepatocyte proliferation by retinoic acid. *Carcinogenesis* **25**:2061–2066.
- Le Douarin, B., A. L. Nielsen, J. M. Garnier, H. Ichinose, F. Jeanmougin, R. Losson, and P. Chambon. 1996. A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J.* **15**:6701–6715.
- Lefebvre, B., C. Brand, P. Lefebvre, and K. Ozato. 2002. Chromosomal integration of retinoic acid response elements prevents cooperative transcriptional activation by retinoic acid receptor and retinoid X receptor. *Mol. Cell. Biol.* **22**:1446–1459.
- Levine, A. J., and A. H. Brivanlou. 2006. GDF3 at the crossroads of TGF-beta signaling. *Cell Cycle* **5**:1069–1073.
- Maden, M. 2007. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat. Rev. Neurosci.* **8**:755–765.
- Mark, M., N. B. Ghyselinck, and P. Chambon. 2006. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu. Rev. Pharmacol. Toxicol.* **46**:451–480.
- Mengus, G., A. Fadloun, D. Kobi, C. Thibault, L. Perletti, I. Michel, and I. Davidson. 2005. TAF4 inactivation in embryonic fibroblasts activates TGF-beta signalling and autocrine growth. *EMBO J.* **24**:2753–2767.
- Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**:553–560.
- Mongan, N. P., and L. J. Gudas. 2007. Diverse actions of retinoid receptors in cancer prevention and treatment. *Differentiation* **75**:853–870.
- Nagpal, S., S. Friant, H. Nakshatri, and P. Chambon. 1993. RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization in vivo. *EMBO J.* **12**:2349–2360.
- Nagpal, S., C. Ghosh, D. DiSepio, Y. Molina, M. Sutter, E. S. Klein, and R. A. Chandraratna. 1999. Retinoid-dependent recruitment of a histone H1 displacement activity by retinoic acid receptor. *J. Biol. Chem.* **274**:22563–22568.
- Nagpal, S., M. Saunders, P. Kastner, B. Durand, H. Nakshatri, and P. Chambon. 1992. Promoter context- and response element-dependent specificity of the transcriptional activation and modulating functions of retinoic acid receptors. *Cell* **70**:1007–1100.
- Nagy, L., and J. W. Schwabe. 2004. Mechanism of the nuclear receptor molecular switch. *Trends Biochem. Sci.* **29**:317–324.
- Niederreither, K., and P. Dolle. 2008. Retinoic acid in development: towards an integrated view. *Nat. Rev. Genet.* **9**:541–553.
- Nielsen, R., T. A. Pedersen, D. Hagenbeck, P. Moulos, R. Siersbaek, E.

- Megens, S. Denissov, M. Borgesen, K. J. Francoijs, S. Mandrup, and H. G. Stunnenberg. 2008. Genome-wide profiling of PPAR $\gamma$ :RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev.* **22**:2953–2967.
44. Nishimoto, M., S. Miyagi, T. Yamagishi, T. Sakaguchi, H. Niwa, M. Muramatsu, and A. Okuda. 2005. Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol. Cell. Biol.* **25**:5084–5094.
  45. Okita, K., M. Nakagawa, H. Hyenjong, T. Ichisaka, and S. Yamanaka. 2008. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**:949–953.
  46. Pavri, R., B. Lewis, T. K. Kim, F. J. Dilworth, H. Erdjument-Bromage, P. Tempst, G. de Murcia, R. Evans, P. Chambon, and D. Reinberg. 2005. PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. *Mol. Cell* **18**:83–96.
  47. Planelles, L., J. P. Medema, M. Hahne, and G. Hardenberg. 2008. The expanding role of APRIL in cancer and immunity. *Curr. Mol. Med.* **8**:829–844.
  48. Ralston, S. H., A. L. Langston, and I. R. Reid. 2008. Pathogenesis and management of Paget's disease of bone. *Lancet* **372**:155–163.
  49. Raouf, A., V. Li, I. Kola, D. K. Watson, and A. Seth. 2000. The Ets1 proto-oncogene is upregulated by retinoic acid: characterization of a functional retinoic acid response element in the Ets1 promoter. *Oncogene* **19**:1969–1974.
  50. Reiter, J. F., and W. C. Skarnes. 2006. Tectonic, a novel regulator of the Hedgehog pathway required for both activation and inhibition. *Genes Dev.* **20**:22–27.
  51. Renaud, J. P., and D. Moras. 2000. Structural studies on nuclear receptors. *Cell. Mol. Life Sci.* **57**:1748–1769.
  52. Richardson, C., and D. R. Alessi. 2008. The regulation of salt transport and blood pressure by the WNK-SPAK/OSR1 signalling pathway. *J. Cell Sci.* **121**:3293–3304.
  53. Rochette-Egly, C., and P. Chambon. 2001. F9 embryocarcinoma cells: a cell autonomous model to study the functional selectivity of RARs and RXRs in retinoid signaling. *Histol. Histopathol.* **16**:909–922.
  54. Rosenfeld, M. G., V. V. Lunyak, and C. K. Glass. 2006. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.* **20**:1405–1428.
  55. Sakamoto, K., Y. Tamamura, K. Katsube, and A. Yamaguchi. 2008. Zfp64 participates in Notch signaling and regulates differentiation in mesenchymal cells. *J. Cell Sci.* **121**:1613–1623.
  56. Sun, H., and R. Taneja. 2000. Stra13 expression is associated with growth arrest and represses transcription through histone deacetylase (HDAC)-dependent and HDAC-independent mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **97**:4058–4063.
  57. Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**:1025–1040.
  58. Vansant, G., and W. F. Reynolds. 1995. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. U. S. A.* **92**:8229–8233.
  59. Wang, Q., J. S. Carroll, and M. Brown. 2005. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**:631–642.
  60. Weaver, A. M. 2008. Cortactin in tumor invasiveness. *Cancer Lett.* **265**:157–166.
  61. Welboren, W. J., M. A. van Driel, E. M. Janssen-Megens, S. J. van Heerijnen, F. C. Sweep, P. N. Span, and H. G. Stunnenberg. 2009. ChIP-Seq of ER $\alpha$  and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* **28**:1418–1428.
  62. Zhang, X., D. T. Odom, S. H. Koo, M. D. Conkright, G. Canetti, J. Best, H. Chen, R. Jenner, E. Herbolsheimer, E. Jacobsen, S. Kadam, J. R. Ecker, B. Emerson, J. B. Hogenesch, T. Unterman, R. A. Young, and M. Montminy. 2005. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl. Acad. Sci. U. S. A.* **102**:4459–4464.
  63. Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**:R137.
  64. Zirn, B., B. Samans, C. Spangenberg, N. Graf, M. Eilers, and M. Gessler. 2005. All-trans retinoic acid treatment of Wilms tumor cells reverses expression of genes associated with high risk and relapse in vivo. *Oncogene* **24**:5246–5251.

**Publication 3 :**

**Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. BMC Genomics, 11:530**

Igor Martianov, Mohamed-Amin Choukrallah, Arnaud Krebs, Tao Ye, Stephanie Legras, Erikjan Rijkers<sup>1</sup>, Wilfred Van Ijcken<sup>1</sup>, Bernard Jost, Paolo-Sassone-Corsi<sup>2</sup>, and Irwin Davidson (2010)



This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

**Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells**

*BMC Genomics* 2010, **11**:530 doi:10.1186/1471-2164-11-530

Igor Martianov (martiano@igbmc.fr)  
Mohamed-Amin Choukrallah (amin@igbmc.fr)  
Arnaud Krebs (arnaud.krebs@igbmc.fr)  
Tao Ye (tao.ye@igbmc.fr)  
Stephanie Legras (slegras@igbmc.fr)  
Erikjan Rijkers (h.rijkers@erasmusmc.nl)  
Wilfred Vanlijken (w.vanijcken@erasmusmc.nl)  
Bernard Jost (jost@igbmc.fr)  
Paolo Sassone-Corsi (psc@uci.edu)  
Irwin Davidson (irwin@igbmc.fr)

**ISSN** 1471-2164

**Article type** Research article

**Submission date** 22 February 2010

**Acceptance date** 29 September 2010

**Publication date** 29 September 2010

**Article URL** <http://www.biomedcentral.com/1471-2164/11/530>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

**Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells.**

Igor Martianov, Mohamed-Amin Choukrallah, Arnaud Krebs, Tao Ye, Stephanie Legras, Erikjan Rijkers<sup>1</sup>, Wilfred VanIjcken<sup>1</sup>, Bernard Jost, Paolo-Sassone-Corsi<sup>2</sup>, and Irwin Davidson#.

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 1 Rue Laurent Fries, 67404 Illkirch Cédex. , France.

1 Erasmus Medical Center, Dr Molewaterplein 50, 3015GE Rotterdam, the Netherlands.

2 Department of Pharmacology, 2115 Gillespie Neuroscience, University of California, Irvine, California 92697-4625, USA

# To whom correspondence should be addressed

E mail : [irwin@igbmc.fr](mailto:irwin@igbmc.fr). FAX: 33 3 88 65 32 01. TEL: 33 3 88 65 34 40 (45)

**Abstract.****Background.**

CREB and CREM are closely related factors that regulate transcription in response to various stress, metabolic and developmental signals. The CREM $\tau$  activator isoform is selectively expressed in haploid spermatids and plays an essential role in murine spermiogenesis.

**Results.**

We have used chromatin immunoprecipitation coupled to sequencing (ChIP-seq) to map CREM and CREB target loci in round spermatids from adult mouse testis and spermatogonia derived GC1-spg cells respectively. We identify more than 9000 genomic loci most of which are cell-specifically occupied. Despite the fact that round spermatids correspond to a highly specialised differentiated state, our results show that they have a remarkably accessible chromatin environment as CREM occupies more than 6700 target loci corresponding not only to the promoters of genes selectively expressed in spermiogenesis, but also of genes involved in functions specific to other cell types. The expression of only a small subset of these target genes are affected in the round spermatids of CREM knockout animals. We also identify a set of intergenic binding loci some of which are associated with H3K4 trimethylation and elongating RNA polymerase II suggesting the existence of novel CREB and CREM regulated transcripts.

**Conclusions.**

We demonstrate that CREM and CREB occupy a large number of promoters in highly cell specific manner. This is the first study of CREM target promoters directly in a physiologically relevant tissue *in vivo* and represents the most comprehensive experimental analysis of CREB/CREM regulatory potential to date.

## **Background.**

Cyclic AMP response element (CRE) binding protein (CREB) and cyclic AMP response element modulator (CREM) are highly related bZIP proteins that regulate transcription in response to various stress, metabolic and developmental signals [1, 2]. CREB and CREM bind to the consensus palindromic CRE 5'-TGACGTCA-3' or half-CRE 5'-TGACG-3' and 5'-CGTCA-3' elements present in the promoters of target genes. In somatic tissues, CREB activates transcription following phosphorylation of a serine residue (S133 in CREB, S117 in CREM) in the activation domain by several kinases and in response to a variety of stimuli [3]. Serine phosphorylation results in recruitment of the p300 and CREB binding protein (CBP) coactivators with histone acetyl-transferase activities [4]. Transcription regulation by CREB is also modulated by TORC (transducers of regulated CREB activity) 1 and 2 that interact with the bZIP DNA binding domain independently of serine 133 phosphorylation and modulate CREB interaction with the TAF4 subunit of TFIID, thus potentiating transcription activation [5]. TORCs play critical roles in several physiological processes such as glucose metabolism and obesity [6-9].

Multiple isoforms of CREM have been described that act either as repressors or activators [10]. Follicle stimulating hormone (FSH), by an as yet undefined mechanism, modulates the usage of alternative polyadenylation sites in mouse male germ cells, such that several destabiliser signals in the 3' untranslated region of the CREM $\tau$  activator isoform mRNA are eliminated leading to increased stability and the accumulation of the CREM $\tau$  protein to high levels in post-meiotic round spermatids [11, 12]. In these cells, CREM bypasses the requirement for phosphorylation through interaction with the LIM-only domain protein ACT (activator of CREM in testis) encoded by the *Fhl5* gene [13] specifically expressed in haploid cells. Knock out of CREM in mice leads to male sterility due to apoptosis of the round spermatids showing that it plays an essential role in spermiogenesis

[14-16]. While loss of CREM leads to apoptosis of round spermatids and a complete arrest of spermiogenesis, only a handful of CREM target genes have been described.

Here we have used chromatin immunoprecipitation coupled to massive parallel sequencing (ChIP-seq) to identify CREM target genes in round spermatids in testis and CREB target genes in spermatogonia derived GC1-spg cells. We identify more than 9000 genomic loci most of which are cell-specifically bound. Of these, more than 6700 target loci are occupied by CREM in testis, yet the expression of only a small subset of these genes are affected upon CREM knock out. We also identify intergenic CREB and CREM binding loci some of which are associated with H3K4 trimethylation suggesting the existence of novel CREB and CREM regulated transcripts.

## **Results.**

### **CREM binding sites in round spermatids of mouse testis.**

As previously reported, in testis, the CREM $\tau$  isoform is strongly and selectively expressed in round spermatids [12, 15], while CREB is expressed in spermatogonia, Sertoli, Leydig and intertubular cells, but not in round spermatids (Additional File 1, Fig. S1A and B). To identify CREM $\tau$  binding sites, we performed duplicate ChIP-seq using anti-CREM $\tau$  antibody, or anti-GFP antibody as control, and formaldehyde-fixed testis chromatin from adult mice. We reproducibly identified around 6792 CREM-bound genomic loci (Fig. 1A and B, Additional file 2, Fig. S2, Additional file 3, Table S1). Of these loci, 80% were located in the promoters of annotated RefSeq genes, but a significant number were also located in intragenic regions corresponding to potential alternate promoters, and in intergenic regions with no annotated RefSeq transcripts (see below). Amongst the CREM targets are 46 miRNAs, some of which have been described as having a brain-specific function (Additional file 4, Table S2 and see below).

We used ChIP-qPCR to validate CREM binding to sites that showed high, medium and low occupancy in the ChIP-seq experiments. We compared CREM ChIP to control ChIP with anti-GFP antibody at a series of identified binding sites and with an intragenic region from the bone morphogenic 2 inducible kinase (*Bmp2k*) gene as a negative control and the testis-specific angiotensin I converting enzyme (*Ace*) promoter [17] as a positive control (Fig. 1C). These experiments confirmed binding of CREM to the tested promoters compared to the *Bmp2k* control. Strong occupancy of the intraflagellar transport 172 homolog (*Ift172*) and small nuclear ribonucleoprotein polypeptide A (*Snrpa*) loci was observed, but no significant occupancy of the surfeit gene 6 (*Surf6*) locus was seen, when compared to the background of the anti-CREM antibody as defined by the signal seen using the *Bmp2k* control. All loci defined by more than 40 clustered reads were significantly enriched in qPCR, not all loci defined by under 40 reads showed enrichment. This number was therefore used as a cut off to obtain a high confidence list.

Bioinformatic analysis of the binding sites showed that only 13% contained a full consensus CRE element, the majority contained either a CRE with one mismatch or one or several half-CRE elements (Fig. 1A). We also identified a significant fraction of sites where no CRE (17%) or half CRE (41%) could be observed. To determine whether an additional conserved sequence motif could be identified in these sites, the 100 most highly occupied regions were analysed using the Meme programme [18]. This analysis showed that the only frequently occurring motif corresponded to the SP1 binding site. No CRE or CRE related motif could be detected in this way and no novel alternative CREM binding motif could be identified. Loci containing full CREs generally showed a higher occupancy as 90% of these sites were defined by more than 100 clustered reads, while only 5% of the no CRE class fell into this category. Analysis of the location of CREM binding sites with respect to the

transcription start site (TSS) showed a very strong preference around -150 bp irrespective of the presence of a CRE-related element (Fig. 1B).

To make a global correlation between promoter occupancy by CREM and transcription, we performed ChIP-seq using an antibody against trimethylated lysine 4 of histone H3 (H3K4me3). Trimethylation of this residue is tightly associated with transcription activity, a high density of this mark overlapping the 5' end of the gene indicates either its transcription or the presence of a poised preinitiation complex [19]. Comparison with the CREM ChIP-seq data set indicates that 95% of CREM bound promoters (ie with CREM occupied loci between -1000 and +500bp with respect to the TSS) show H3K4me3. (Additional file 3, Table S1). However, in this experiment we cannot exclude that some genes are inactive in round spermatids where CREM is bound, but are transcribed at other developmental stages or in other cell types within the testis. To make a more accurate comparison we compared CREM promoter occupancy, H3K4me3 and gene expression using a transcriptome data set of adult mouse spermatid gene expression [20]. We identified 4934 CREM-occupied promoters showing H3K4 trimethylation. Comparison with array data showed that, while the 41% of probe sets on the array showed expression values  $\geq 51$ , these accounted for 84% of the CREM and H3K4me3 associated genes (Additional file 5, Fig. S3A). We identified only 89 CREM occupied promoters (1.7% of the total CREM occupied proximal promoters) that were false negatives annotated as having no H3K4 trimethylation, but represented amongst the moderate to strongly expressed genes. These results indicate not only that the vast majority of CREM occupied promoters with H3K4 trimethylation are indeed expressed, but that they are over-represented amongst the higher expressed gene set. It should also be noted that we fail to detect any significant H3K4me3 of Leydig cell-specific genes such as *Cyp17a1* and *Star* and little or no signal was seen for the *Rhox* genes that are specific for Sertoli cells. Thus, in agreement with the above data, the contribution of these cell

types to the profile is negligible, indicating that the majority of the signal comes from transcription of genes in the germ cell lineage.

### **Differing repertoires of CREM and CREB binding sites in round spermatids and spermatogonia GC1-spg cells.**

As mentioned above CREB is not expressed in round spermatids, but is expressed in spermatogonia and in the somatic cells in testis. Due to its expression in several of the cell types that make up the testis, we cannot use total testis chromatin, even from early post-natal animals, for the ChIP-seq as we would be unable to distinguish in which cells CREB was occupying a given promoter. We therefore used GC1-spg cells that are immortalised spermatogonia [21] and express CREB, but not the CREM $\tau$  activator form (Additional file 1, Fig. S1B). ChIP-seq experiments in these cells using anti-CREB antibody allowed the identification of around 4398 binding sites (Fig. 2A, Additional file 2, Fig. S2 and Additional file 6, Table S3).

These results were confirmed by ChIP-qPCR where we observed significant, but variable occupation of all of the tested promoters. In this case even loci with small numbers of clustered reads showed significant enrichment with respect to the Bmp2k control (Fig. 2C). The majority of these promoters were also occupied by pol II (Fig. 2D).

Similar to what was observed with CREM, most CREB binding sites were observed in the promoters of annotated genes, but a significant number of intra- and intergenic sites were also observed. Bioinformatic analysis showed that only a small number comprise full canonical CREs and we identify a subset of sites with half-CRE or no identifiable CRE (Fig. 2A). Also similar to CREM, CREB binding sites are preferentially located around -150 relative to the TSS (Fig. 2B).



To assess the association of CREB loci occupancy with transcriptional activity, we further performed H3K4me3 and pol II ChIP-seq in GC1-spg cells. Comparison with CREB occupancy shows that around 60% of CREB-occupied loci are associated with high levels of pol II and H3K4me3 in these cells (Fig. 3A, clusters B and C), while a second group shows little or no pol II and H3K4me3 (Fig. 3A, cluster A and see also Additional file 6, Table S3). Comparison with publically available array data for GC1-spg cell gene expression (GSE19355) showed that similar to what was observed in round spermatids, CREB and H3K4me3 occupied promoters were strongly over-represented in the highly expressed gene set (Additional file 5, Fig. S3B).

We next compared the repertoires of CREB and CREM binding sites in round spermatids and in GC1-spg cells. To identify cell-specifically occupied sites, we took the coordinates of the CREB binding sites and determined the read density in the CREM data set in a window of +/- 5kb around these coordinates. If CREM and CREB occupy the same sites, there will be a high density of reads at these coordinates in both data sets. This comparison shows that a subset of loci show high occupancy in both cell types (Fig. 3B, cluster B), but the majority are cell-specifically occupied in either GC1-spg cells (Fig. 3B, cluster A) or in round spermatids (data not shown). In particular, CREM occupies a large set of sites in the round spermatids that are not occupied by CREB in GC1-spg cells.

For example, the genes encoding protamines 1, 2, and 3 and transition protein 2 are all selectively expressed in round spermatids and clustered on chromosome 16. Multiple binding sites are occupied by CREM at this locus in round spermatids, but not by CREB in GC1-spg cells and the locus shows H3K4 trimethylation only in testis (Fig. 4A). Similarly, the ornithine decarboxylase antizyme 3 (*Oaz3*) is occupied by CREM and active in testis, but not by CREB in GC1-spg cells, while the opposite is observed for the natriuretic peptide precursor type B (*Nppb*) locus (Fig. 4B and C). We also investigated occupancy of the well

characterised CRE in the somatostatin (*Sst*) promoter [22]. *Sst* is not expressed in testis or in GC1-spg cells as shown by the absence of H3K4me3 and pol II (data not shown), and the promoter CRE is not occupied in either cell type. However, CREB selectively occupies an alternative and conserved intragenic site in GC1-spg cells (Fig. 4D). These results show that CREM and CREB occupy distinct but overlapping repertoires of binding loci in round spermatids and GC1 cells and that genes specifically activated in round spermatids are in general not occupied in GC1 cells, but become accessible during germ cell development in vivo.

In total, 9037 distinct CREB and CREM binding loci can be identified combining the ChIP-seq in round spermatids and GC1-spg cells. Despite the wide occupancy of potential binding sites by CREM in haploid cells, not all possible sites (i.e. full CREs) are occupied as the CRE element in the *Mitf-m* promoter that is specifically active in the melanocyte lineage, or the *Hoxa5* gene are not occupied in GC1-spg or haploid cells (data not shown). These results show the existence of different classes of binding sites, those that are rather generally accessible, and a larger set that are occupied in a cell type-specific manner.

### **Genes directly regulated by CREM in haploid cells.**

CREM inactivation leads to round spermatid apoptosis at an early stage. To understand the basis of this effect, Beissbarth et al [23] used subtractive hybridisation and Affymetrix arrays to identify genes that are deregulated upon CREM inactivation. We used the results of the Affymetrix analysis to identify those genes whose promoters are bound by CREM and are direct targets. A set of 127 genes were reported to be deregulated, 102 of which are down-regulated, in CREM mutant testis [23] ([http://www.dkfz.de/tbi\\_old/crem/cremgenes.txt](http://www.dkfz.de/tbi_old/crem/cremgenes.txt)). Of these, only 58 are direct CREM targets with one or several CREM binding sites in their promoters. (Fig. 5A and Additional file 7,

Table S4). Surprisingly however, two of these *Rundc3a* and *Cnn1* do not show significant H3K4me3 and are thus likely not expressed in the testis of the animals used in our experiments.

Analysis of the CREM binding sites in the promoters of regulated target genes shows that not all have consensus CRE elements, the profile is in fact similar to that seen in the global population of target genes, and some of the directly regulated genes do not contain any identifiable CRE or half CRE elements (Fig. 5B). It has previously been suggested that the presence of a TATA element is required for a gene to respond to cAMP stimulation via CREB [24], however, of the 58 CREM regulated promoters, only 6 have a consensus TATA element. This frequency is close to that seen in the genome as a whole [25]. Thus the presence of a TATA element is not a pre-requisite for regulation by CREM and is not enriched in CREM regulated promoters. Indeed many other CREM bound promoters have TATA elements and yet are not de-regulated in the absence of CREM.

The above results show that only 1% of the total occupied promoters are actually regulated by CREM in haploid cells. The expression of the vast majority of occupied genes is not affected by loss of CREM.

### **CREB/CREM target promoters involved in other physiological processes.**

In addition to germ cell development, CREM and CREB are involved in many other physiological processes. We examined whether they occupy the promoters of target genes involved in other cell-specific functions. For example, the *Clock* and *Period (Per) 2* genes are expressed in both testis and GC1-spg cells, yet their promoters are specifically occupied only in round spermatids by CREM (Additional file 8, Fig. S4A and B). Thus, while the expression of these genes is not haploid cell-specific, the CREB/CREM binding sites of their promoters are selectively occupied in germ cells. Moreover although not regulated by CREM in testis

(see below), *Per2* expression is for example regulated by CREB in response to parathyroid hormone activation of the cAMP signalling pathway in chondrocytes [26]. Thus at least in this case, the CREM promoter occupancy is indicative of potential regulation in other cell types by physiological signalling pathways.

Mmu-mir-132 and mmu-mir-219-1 are two CREB regulated brain-enriched miRNAs that regulate circadian rhythm and in the case of mir-132, neuronal morphogenesis [27, 28]. At the mmu-mir-219-1 locus, a CRE element located downstream of the miRNA and upstream of the *Ring1* gene is occupied in both haploid and GC1-spg cells by CREM and CREB respectively (Fig. 6A, and Additional file 4, Table S2). Furthermore, an extended region of H3K4 trimethylation in testis and in GC1-spg cells is observed indicating expression of both the miRNA and the adjacent *Ring1*. A similar observation was made at the mmu-mir-132 locus (data not shown). Mmu-mir-9-3 is a miRNA involved in silencing the neuronal transcription factor REST [29]. In contrast to the mmu-mir-219-1 and mmu-mir-132 loci, the mmu-mir-9-3 locus is specifically bound by CREM and active in testis and not GC1-spg cells (Fig. 6B). Thus in testis and in GC1-spg cells, CREM and CREB occupy the promoters of genes that are involved in specific processes in other cell types, and that although expressed in testis play no known role in spermatogenesis.

As CREB/CREM are involved in many physiological and metabolic processes, we used the KEGG pathway function of DAVID annotation (<http://david.abcc.ncifcrf.gov/>) to make a more comprehensive identification of potential CREB/CREM target genes in physiological functions (see Additional file 9, Table S5). This analysis identified many potential target genes not only in cell cycle and proliferation, but also in metabolic functions such as insulin signalling.

We also noted that CREB and CREM occupy their own promoters and those of several of their essential cofactors. The complex CREM locus comprises several alternative

promoters where CREM/CREB occupy a site upstream of the major transcript that is expressed in both testis and GC1-spg cells, although as indicated above, the CREM $\tau$  activator isoform is not present in these cells (Additional file 8, Fig. S4C). Weak H3K4me3 of two of the alternate promoters is also seen specifically in testis. The genes encoding the coactivators CBP (*Crebbp*), p300 (*Ep300*) and TORC1 (*Crtc1*) are expressed in GC1-spg cells and testis, but their promoters are occupied only in haploid cells by CREM (Fig. 6C and D). These results suggest that in the appropriate cell type or in response to the appropriate signal CREB or CREM may regulate expression of their coactivators (see also [30, 31]).

### **Novel transcripts and loci potentially regulated by CREM and CREB.**

It is striking that the majority of CREB and CREM binding sites are located at around -150bp with respect to the TSS. The tight distance correlation between the CREB/CREM binding site and the transcription start site prompted us to investigate whether the CREB and CREM binding sites that are located in intergenic regions with no currently annotated RefSeq genes correspond to cryptic and as yet unidentified promoters of novel transcripts. Many of the intergenic sites are not associated with known RefSeq genes, but are associated with annotated EMBL transcripts or predicted genes most of which are not well characterised (See Additional files 3, Table S1 and Additional File 6, Table S3). However, there are also a set of loci with no nearby RefSeq annotation (Additional file 10, Table S6). Many of these loci are not associated with any predicted transcripts, but others lie close to UCSC gene predictions or Genbank transcripts corresponding to non-coding or uncharacterised RNAs. For CREM, 80% of these sites contain a CRE or half CRE, while for CREB, 55% contain CRE-related elements (Fig. 7A).

A subset of these regions potentially comprise active promoters as they are marked by H3K4me3. For example, on chromosome 15, downstream of the *Slc38a2* gene, is a CREB

occupied locus that correlates with a peak of H3K4me3 and pol II (Fig. 7B). The pol II is distributed as a peak close to the H3K4me3 peak marking the 5' end of the transcript and a lower density corresponding to the potential transcribed region. A similar pol II distribution is seen over the *Slc38a2* gene where an additional peak is seen at the transcription termination site downstream of the gene, a profile typical of that previously observed for pol II [32, 33]. This second region therefore appears to comprise an additional as yet uncharacterised transcript that can be potentially regulated by CREB. A similar example is seen on chromosome 5 at the *Rnf4* locus where peaks of CREB and H3K4me3 and pol II are seen downstream of the *Rnf4* gene that are characteristic of a promoter driving the expression of an anti-sense transcript with paused and elongating pol II. (Fig. 7C).

In other cases, the CREB occupied sites lie upstream of known promoters and are associated with H3K4me3, paused pol II, but little or no elongating pol II (Additional file 11, Fig S5A and B). These may correspond to alternative promoters that are not engaged in productive transcription, or distal enhancer elements. In contrast, other conserved loci show occupation by CREB and/or CREM, but low or no H3K4me3 and pol II. Again these may correspond to remote enhancer elements. Correlation between the CREM/CREB, H3K4me3 and pol II data sets thus allows the identification of potential CREB regulated novel transcripts, promoters, or distal enhancer elements.

In the course of this analysis we also noted multiple CREB binding sites and an exceptionally high RNA pol II density and over the region covering the MEN $\epsilon/\beta$  and Malat1 (metastasis associated in lung adenocarcinoma transcript 1) loci (Additional file 11, Fig. S5C). These correspond to two non-coding RNAs. The precise function of Malat1 is not known [34], while the MEN transcripts are involved in the formation of nuclear paraspeckles [35]. These two neighbouring loci have the highest pol II density observed in the data set and are thus very highly transcribed in GC1-spg cells.

## **Discussion.**

### **An extended repertoire of CREM/CREB binding sites.**

We have used ChIP-seq to identify CREM binding sites in round spermatids and CREB binding sites in GC1-spg cells. This is the first study of CREM target promoters directly in a physiologically relevant tissue *in vivo* and represents the most comprehensive experimental analysis of murine CREB/CREM regulatory potential to date. Our data identify 9037 binding loci that are occupied in either a common or cell-specific manner. The majority of binding sites cluster to a preferential location in the nucleosome free region around 150 nucleotides upstream of the TSS. This observation differs from a bioinformatic analysis showing that the CRE element is preferentially located around -45 bp relative to the TSS [36]. This discrepancy may be explained by the fact that many of the observed binding sites are half CREs, mutated or non-CRE elements, that many not have been identified bioinformatically. Thus, while full CREs may be closer to the TSS, the overall occupancy profile is centered around -150 bp. Moreover, a bioinformatics analysis of testis-specific promoters indicated a preferential presence of CRE-like elements rather than full consensus CREs [37]. This bias is also reflected in the experimental data.

Previous studies based on ChIP-chip analysis identified smaller repertoires of sites [30, 38], while ChIP coupled to sequencing of concatenated tags in a SAGE type analysis identified more than 6300 CREB binding sites [31]. Based on this data, it has been suggested that there may be between 5000 and 19,000 binding sites in the human genome. Indeed, more than 10,000 full CREs have been predicted in the human genome of which around 4000 are

suggested to be critical for function based on their conservation in mouse or the fact that they are present as a repeated cluster in promoter regions [30]. Our results show that in round spermatids around 622 full CREs are occupied while in GC1-spg cells around 613 CREs are occupied. Thus, many CREs are occupied neither in haploid cells nor in GC1-spg cells showing that the full repertoire of CREB/CREM binding sites exceeds the 9000 that we have identified.

From comparison of ChIP-chip in three different cell lines it has previously been suggested that CREB occupies almost the same set of target genes in each cell type, and that specificity of gene regulation is conferred not by selectivity in occupation of target sites, but by cooperation of CREB with other cellular factors [30]. In contrast, by ChIP and *in vivo* footprinting in hepatoma and PC12 cells or cortical neurons, Cha-Molstad et al., [39] showed that many sites were occupied in a cell specific manner. Our data clearly favour a model in which occupation of potential CREB/CREM binding sites is largely cell-specific although this does not exclude other criteria that specify which of the occupied sites are actually regulated by CREB and CREM. We have compared our murine data sets with those of Zhang et al obtained by ChIP-chip in human HEK cells [30]. We identified murine orthologues for 2003 of the CREB-occupied human promoters. Of these, 493 are occupied by CREB in GC1-spg cells, and 922 occupied by CREM in haploid cells. This comparison further highlights the cell-specificity of CREB/CREM promoter occupancy.

Cha-Molstad et al., also suggested that CREB preferentially occupies loci that are expressed and marked with H3K4 methylation. As an example, they show that occupancy of the well studied somatostatin (*Sst*) CRE correlates with its expression. In our experiments, the *Sst* locus is expressed neither in testis nor in GC1-spg cells and the promoter CRE is not occupied by CREB or CREM. In contrast, CREB occupies a novel and conserved intronic site in GC1-spg cells that does not contain any CRE-related element. However, a more global



analysis of our CREB data set shows that a majority of CREB occupied promoters are also characterised by high pol II occupancy and H3K4 trimethylation. Thus the idea that CREB preferentially occupies promoters of expressed genes is generally valid. Similarly, CREM-occupied genes in haploid cells are over-represented amongst the highly expressed gene set. On the other hand, there are both promoter and intergenic sites that are not closely associated with H3K4 trimethylation. Thus, expression is neither an absolute pre-requisite nor outcome of CREB/CREM occupancy. Moreover as discussed above, there are promoters that are active in both testis and GC1-spg cells, but bound only by CREM in testis. There must therefore be additional factors that influence occupancy, such as DNA methylation or competition with other factors whose binding sites overlap that of CREB [30, 39].

### **Insights into the specific role of CREM in spermiogenesis.**

During spermatogenesis, the gene regulatory programme in haploid cells is characterised by the activation of a large set of genes that are uniquely expressed in this cell type and by the existence of specialised mechanisms and factors that control their expression [40-42]. Given this specialisation, one could have imagined that in contrast to somatic cells, CREM would occupy a restricted set of loci corresponding to genes that are specifically expressed and regulated in these cells. Indeed at some loci, such as the protamine and transition protein genes, there is a developmental logic to CREM promoter occupancy. In contrast however, there are numerous examples of genes which are expressed in testis and GC1-spg cells, but are occupied only by CREM in haploid cells. Similarly, CREM also binds the promoters of genes that have no known or obvious function in spermatogenesis. Thus, contrary to what one may have expected for a highly specialised cell type with a unique gene expression programme, haploid cells clearly provide a much more permissive environment for accessibility than GC1-spg cells.

It has previously been shown that TBP, TFIIB, RNA polymerase II, and TFIIA are all strongly up-regulated in haploid spermatids [43, 44]. It has been proposed that overexpression of the general transcription machinery at this stage is required to produce the large amounts of some mRNAs that have to be synthesised and stored for translation during the elongation and remodeling phase [43]. There is also evidence that overall mRNA levels are generally higher in haploid cells than in somatic cells due to ‘promiscuous’ transcription where in addition to transcription of highly expressed testis-specific genes and of normal ‘housekeeping’ genes, many other genes are also transcribed albeit at lower levels [43]. In agreement with this idea, our present ChIP-seq experiments show widespread occupancy by CREM of target genes that have no specific role in spermatogenesis, but whose promoters are active in haploid cells. The euchromatin of round spermatids is therefore highly accessible for transcription factor binding which, together with high levels of the basal transcription machinery, may lead to the expression of a large set of genes and generally higher mRNA levels .

Strikingly, only a very restricted set of CREM-occupied genes are affected by CREM inactivation. Comparison of the 127 genes whose expression is affected with those that are bound by CREM identifies 58 genes as direct targets. For this analysis, we used only the Affymetrix data of Beissbarth et al [23] that may underestimate the real number of affected genes. At the time of this study, many haploid expressed genes were not well characterised and were identified only as ESTs. For example, we have previously characterised H1T2 (*H1fnt*) a histone H1 variant specifically expressed in haploid cells [45]. The *H1fnt* promoter is occupied by CREM, selectively activated in round spermatids and is therefore a direct CREM target. The full data set of Beissbarth et al actually comprises more than 367 regulated transcripts many of which are only identified as ESTs. If the proportion of direct CREM targets from the Affymetrix data can be extended to the full set, this would suggest that CREM directly regulates between 150-200 genes in male germ cells. This number still

represents only a small fraction of the CREM bound promoters. CREM clearly occupies many more promoters than it actually regulates. Neither the presence of a full CRE nor a TATA element are pre-requisites to predict whether a gene will be regulated by CREM. The criteria that determine whether a promoter will be regulated or not by CREM remain to be determined.

Inactivation of CREM leads to an early apoptosis of haploid cells around step 4 of their development [14, 15]. Amongst the direct CREM targets are genes that encode protamines, transition proteins and outer dense fiber protein whose expression initiates around this stage, but whose gene products are required only later during the elongation stage. Reduced expression of these genes may account for the reduced sperm count and abnormalities seen in heterozygous CREM mutant mice, but do not explain the early round spermatid apoptosis. Previously, Beissbarth et al [23] suggested that down regulation of the anti-apoptotic genes *Bcl2* and *Bcl6b* (BAZF) may contribute to germ cell apoptosis. *Bcl2* is not a direct CREM occupied promoter (although the promoters of the related *Bcl2l1*, *Bcl2l11* and *Bcl2l13* genes are occupied by CREM), but *Bcl6b* is a direct target. Apoptosis may potentially result from direct deregulation of this gene or of one or several of the other direct CREM target genes involved in signal transduction and metabolic processes.

## **Conclusions.**

Together our results have identified a large repertoire of CREM bound loci in haploid cells of which a small subset involved in signalling and metabolic processes are deregulated upon CREM inactivation and may explain the haploid cell apoptosis. In contrast to previous studies on smaller data sets, we have shown that promoter occupancy by CREB and CREM is highly cell specific in haploid and GC1 cells. This is the most comprehensive evaluation of CREB/CREM regulatory potential to date and we have identified a large set of promoters that may be regulated by these factors in different tissues and physiological processes.

Correlation between the CREM/CREB, H3K4me3 and pol II data sets has identified many novel potential CREB regulated transcripts, and revealed the presence of CREB at novel start sites and distal regulatory elements.

## **Methods.**

### **ChIP-seq.**

Chromatin was prepared from the testes of 12 week old adult 129/SV mice. Animal breeding and manipulation were performed according to French national ethical standards using approved protocols. The animals were sacrificed and the testes removed and opened directly in 20 ml fixing solution (0.4 % formaldehyde in 1x PBS) in a Petri dish. Seminiferous tubules were fixed for 10 minutes with agitation. The cross-linking reaction was stopped by adding 2 ml 2M glycine solution. Seminiferous tubules were then centrifuged at 4000 rpm and the pellet was lysed with 2 ml cell lysis buffer (50mM Tris pH8.0, 10 mM EDTA, 1% SDS) for 15 minutes. GC1-spg cells were obtained from the ATCC collection and cultured in DMEM with 10% foetal calf serum. Cells were grown to 90% confluency and crosslinked with formaldehyde 0.4% for 10 min at room temperature and the reaction was stopped by adding glycine to final concentration 0.2M for 10 minutes at room temperature. Fixed cells were rinsed twice with PBS and resuspended in lysis buffer. Lysate from cells or testis was sonicated 30 min (10 sec on / 30 sec off) in Diagenode water bath-sonicator and centrifuged at 14000 rpm for 10 min. The cleared supernatant was used immediately in ChIP experiments or stored at -80°C. ChIP was performed by standard procedures.

ChIP-seq was performed using an Illumina GAII sequencer and the raw data analysed by the Illumina Eland pipeline. Peak detection was performed using the MACS software (<http://liulab.dfci.harvard.edu/MACS/>) [46], and the peaks annotated using GPAT ([47], [http://bips.u-strasbg.fr/GPAT/Gpat\\_home.html](http://bips.u-strasbg.fr/GPAT/Gpat_home.html)). Peak detection with MACS was performed

using the GFP-ChIP as negative control. Clustering was performed by first generating density (.wig format files) counting the number of tags in a 25pb sliding window for each ChIP-seq data set. The coordinates of the CREB binding sites were used as reference, and the tag density for the Pol II and H3K4me3 data sets were extracted from the .wig files in the 10 kb around the CREB site. A matrix of binding sites and densities was generated and subjected to K-means clustering using the Cluster 3.0 software (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv>). The clustered matrix was visualized using Java TreeView. (<http://jtreeview.sourceforge.net/>).

Comparison of the ChIP-seq data with the array expression data was carried out by first performing an RMA normalisation of the mouse spermatid expression .Cel files from [20] and from the mock silenced GC1-spg data sets in the GEO data base (GSE19355). Excel files with the normalised data were compared to those comprising a list of CREM (or CREB) and H3K4me3-occupied promoters, thus assigning an expression value to each of the corresponding genes.

The following antibodies were used : the anti-CREMtau rabbit polyclonal antibody has been previously described [14, 48], CREB (06-863, Upstate), H3K4me3 (MC315 04-745, Millipore), Pol II (H-224 Sc-9001 Santa Cruz). Real-time PCR was performed on a Roche Lightcycler using Roche SYBR Green mix. The following primer sequences were used.

*Ift172*: 5'-TTTTGCCTCTCGTAGCACCT-3', 5'-ATGCAAACCTAACGCAAACC-3'

*Gosr2*: 5'-TAAAACTGGGAGGGATGCAG-3', 5'-CTAGGCCCATCTCATTTC-3'

*SnrpA*: 5'-CACAACGCTCTTGAAGGTCA-3', 5'-GGAACGGACCAATCAAAGAA-3'

*Cdc6*: 5'-AGACCTGGGGCTGTCCTATT-3', 5'-CCAAAGCCGCTCTACTTCAG-3'

*Camk2d*: 5'-AAGCACGAGCACATAAGCAA-3', 5'-GCAGATTTAGAGGCCTACGG-3'

*Cd63*: 5'-ACCTGGTTTTGCCATCTCTG-3', 5'-AAGCCTTAGAGCTCCCTTGC-3'

*Ace*: 5'-GCTGGCACATTGCTCTATGA-3', 5'-AGCAGAGCAGCAGAAAGAGG-3'

*Ankrd52*: 5'-TGACGTCAGGGAGAGAGCTT-3', 5'-CCCCATCACAAGGAGAGAAA-3'

*SURF6*: 5'-GCCTTTCCCTTCATTCCTC-3', 5'-GACCTGAGTATGGCGTGGAT-3'

Bmp2k intragenic: 5'-TCCCTGCACGTACTTTAGGC-3', 5'-  
TTGGAGCACTCAAAGGTGTG-3'  
*Rtbdn*: 5'-TCCACGGTGCTGAAATATGA-3', 5'-CCAGGAACCGAGTTACAGGA-3'  
*Stat3*: 5'-CTCCCCACGCAATCTAGTA-3', 5'-AAGCATTTGGGTTTGTGGAG-3'  
*Pcna*: 5'-GGGTTGGTAGTTGTCTGCTGT-3', 5'-AGCACCTTCTTCAGGATGGA-3'  
*Cfos*: 5'-TTAGGACATCTGCGTCAGCA-3', 5'-CCCGTCTTGGCATAACATCTT-3'  
*Fn1*: 5'-CCTCCCTTTCCTTCGAGTCT-3', 5'-GCCAGGTCTGGGACTAGAGA-3'  
*B2m*: 5'-CTGACCGGCCTGTATGCTAT-3', 5'-TCCACCCTGTAGCCTCAAAG-3'  
*Crebzf*: 5'-CCGGGTAGTACCGATGAGTG-3', 5'-CTTCGCTCCAGCCAGAGTAT-3'  
*Klf7*: 5'-CCAGCGTGTACAGTGCAGAT-3', 5'-TGTCAGTGAGTTTGC GTTCC-3'

### **Bioinformatics analysis**

Detection of CRE and half CRE elements in the CREB and CREM bound loci was performed by first extracting a region of 500 nucleotides centered on the peak summit. These sequences were then analysed using the motif search programme of Genomatix software. Additional analysis of the bound regions was performed using the Meme programme ([http://meme.nbcr.net/meme4\\_1/cgi-bin/meme.cgi](http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi)). [18]

### **Authors' contributions.**

IM designed and performed all of the ChIP assays, the immunostaining and immunoblot experiments, MAC, AK and TY developed and applied the algorithm for read density matrix clustering and AK also developed the ChIP-seq peak annotation programme, SL and ER performed the bioinformatics analyses of the ChIP-seq data, peak detection, motif detection, comparison with transcriptome data, WVI and BJ performed ChIP-seq library preparation and Illumina sequencing, PSC and ID conceived the project designed the experiments and ID wrote the manuscript. All authors have read and approved the final manuscript.

### **Acknowledgements**

We thank, G. Altobelli and Y. Anno for help with bioinformatics and C. Bole-Feysot for technical assistance. This work was supported by grants from the CNRS, the INSERM, the Association pour la Recherche contre le Cancer, the Ligue Nationale et Départementale Région Alsace contre le Cancer, the INCA and the EuTRACC programme of the European Union. ID is an 'équipe labéllisée' of the Ligue Nationale contre le Cancer.

## References

1. Sassone-Corsi P: **Transcription factors responsive to cAMP**. *Annu Rev Cell Dev Biol* 1995, **11**:355-377.
2. Hummler E, Cole TJ, Blendy JA, Ganss R, Aguzzi A, Schmid W, Beermann F, Schutz G: **Targeted mutation of the CREB gene: compensation within the CREB/ATF family of transcription factors**. *Proc Natl Acad Sci U S A* 1994, **91**(12):5647-5651.
3. Johannessen M, Delghandi MP, Moens U: **What turns CREB on?** *Cell Signal* 2004, **16**(11):1211-1227.
4. Bannister AJ, Kouzarides T: **The CBP co-activator is a histone acetyltransferase**. *Nature* 1996, **384**:641-643.
5. Conkright MD, Canettieri G, Sreaton R, Guzman E, Miraglia L, Hogenesch JB, Montminy M: **TORCs: transducers of regulated CREB activity**. *Mol Cell* 2003, **12**(2):413-423.
6. Sreaton RA, Conkright MD, Katoh Y, Best JL, Canettieri G, Jeffries S, Guzman E, Niessen S, Yates JR, 3rd, Takemori H *et al*: **The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector**. *Cell* 2004, **119**(1):61-74.
7. Koo SH, Flechner L, Qi L, Zhang X, Sreaton RA, Jeffries S, Hedrick S, Xu W, Boussouar F, Brindle P *et al*: **The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism**. *Nature* 2005, **437**(7062):1109-1111.
8. Altarejos JY, Goebel N, Conkright MD, Inoue H, Xie J, Arias CM, Sawchenko PE, Montminy M: **The Creb1 coactivator Crtc1 is required for energy balance and fertility**. *Nat Med* 2008, **14**(10):1112-1117.
9. Wang B, Goode J, Best J, Meltzer J, Schilman PE, Chen J, Garza D, Thomas JB, Montminy M: **The insulin-regulated CREB coactivator TORC promotes stress resistance in Drosophila**. *Cell Metab* 2008, **7**(5):434-444.
10. Foulkes NS, Sassone-Corsi P: **More is better: activators and repressors from the same gene**. *Cell* 1992, **68**(3):411-414.
11. Foulkes NS, Mellstrom B, Benusiglio E, Sassone-Corsi P: **Developmental switch of CREM function during spermatogenesis: from antagonist to activator**. *Nature* 1992, **355**(6355):80-84.
12. Foulkes NS, Schlotter F, Pevet P, Sassone-Corsi P: **Pituitary hormone FSH directs the CREM functional switch during spermatogenesis**. *Nature* 1993, **362**(6417):264-267.
13. Fimia GM, De Cesare D, Sassone-Corsi P: **CBP-independent activation of CREM and CREB by the LIM-only protein ACT**. *Nature* 1999, **398**(6723):165-169.
14. Nantel F, Monaco L, Foulkes NS, Masquillier D, LeMeur M, Henriksen K, Dierich A, Parvinen M, Sassone-Corsi P: **Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice**. *Nature* 1996, **380**(6570):159-162.
15. Blendy JA, Kaestner KH, Weinbauer GF, Nieschlag E, Schutz G: **Severe impairment of spermatogenesis in mice lacking the CREM gene**. *Nature* 1996, **380**(6570):162-165.
16. Nantel F, Sassone-Corsi P: **CREM: A transcriptional master switch during the spermatogenesis differentiation program**. *Front Biosci* 1996, **1**:d266-269.
17. Zhou Y, Sun Z, Means AR, Sassone-Corsi P, Bernstein KE: **cAMP-response element modulator tau is a positive regulator of testis angiotensin converting enzyme transcription**. *Proc Natl Acad Sci U S A* 1996, **93**(22):12262-12266.
18. Bailey TLaE, C.: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers**. *Proceedings of the Second International Conference on*



- Intelligent Systems for Molecular Biology* 1994, **AAAI Press, Menlo Park, California**:28-36.
19. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ: **How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.** *Nat Struct Mol Biol* 2007, **14**(11):1025-1040.
  20. Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SS, Demougin P, Gattiker A, Moore J, Patard JJ, Wolgemuth DJ, Jegou B *et al*: **The conserved transcriptome in human and rodent male gametogenesis.** *Proc Natl Acad Sci U S A* 2007, **104**(20):8346-8351.
  21. Hofmann MC, Narisawa S, Hess RA, Millan JL: **Immortalization of germ cells and somatic testicular cells using the SV40 large T antigen.** *Exp Cell Res* 1992, **201**(2):417-435.
  22. Montminy MR, Sevarino KA, Wagner JA, Mandel G, Goodman RH: **Identification of a cyclic-AMP-responsive element within the rat somatostatin gene.** *Proc Natl Acad Sci U S A* 1986, **83**(18):6682-6686.
  23. Beissbarth T, Borisevich I, Horlein A, Kenzelmann M, Hergenbahn M, Klewe-Nebenius A, Klaren R, Korn B, Schmid W, Vingron M *et al*: **Analysis of CREM-dependent gene expression during mouse spermatogenesis.** *Mol Cell Endocrinol* 2003, **212**(1-2):29-39.
  24. Conkright MD, Guzman E, Flechner L, Su AI, Hogenesch JB, Montminy M: **Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness.** *Mol Cell* 2003, **11**(4):1101-1108.
  25. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC *et al*: **Genome-wide analysis of mammalian promoter architecture and evolution.** *Nat Genet* 2006, **38**(6):626-635.
  26. Hinoi E, Ueshima T, Hojo H, Iemata M, Takarada T, Yoneda Y: **Up-regulation of per mRNA expression by parathyroid hormone through a protein kinase A-CREB-dependent mechanism in chondrocytes.** *J Biol Chem* 2006, **281**(33):23632-23642.
  27. Cheng HY, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, Nakazawa T, Shimizu K, Okamura H, Impey S *et al*: **microRNA modulation of circadian-clock period and entrainment.** *Neuron* 2007, **54**(5):813-829.
  28. Vo N, Klein ME, Varlamova O, Keller DM, Yamamoto T, Goodman RH, Impey S: **A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis.** *Proc Natl Acad Sci U S A* 2005, **102**(45):16426-16431.
  29. Wu J, Xie X: **Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression.** *Genome Biol* 2006, **7**(9):R85.
  30. Zhang X, Odom DT, Koo SH, Conkright MD, Canettieri G, Best J, Chen H, Jenner R, Herbolsheimer E, Jacobsen E *et al*: **Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues.** *Proc Natl Acad Sci U S A* 2005, **102**(12):4459-4464.
  31. Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH: **Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions.** *Cell* 2004, **119**(7):1041-1054.
  32. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells.** *Cell* 2007, **130**(1):77-88.

33. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322**(5909):1845-1848.
34. Wilusz JE, Freier SM, Spector DL: **3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA.** *Cell* 2008, **135**(5):919-932.
35. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL: **MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles.** *Genome Res* 2009, **19**(3):347-359.
36. Bellora N, Farre D, Alba MM: **Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters.** *BMC Genomics* 2007, **8**:459.
37. Yamashita A, Goto N, Nishiguchi S, Shimada K, Yamanishi H, Yasunaga T: **Computational search for over-represented 8-mers within the 5'-regulatory regions of 634 mouse testis-specific genes.** *Gene* 2008, **427**(1-2):93-98.
38. Euskirchen G, Royce TE, Bertone P, Martone R, Rinn JL, Nelson FK, Sayward F, Luscombe NM, Miller P, Gerstein M *et al*: **CREB binds to multiple loci on human chromosome 22.** *Mol Cell Biol* 2004, **24**(9):3804-3814.
39. Cha-Molstad H, Keller DM, Yochum GS, Impey S, Goodman RH: **Cell-type-specific binding of the transcription factor CREB to the cAMP-response element.** *Proc Natl Acad Sci U S A* 2004, **101**(37):13572-13577.
40. Kimmins S, Kotaja N, Davidson I, Sassone-Corsi P: **Testis-specific transcription mechanisms promoting male germ-cell differentiation.** *Reproduction* 2004, **128**(1):5-12.
41. Kimmins S, Sassone-Corsi P: **Chromatin remodelling and epigenetic features of germ cells.** *Nature* 2005, **434**(7033):583-589.
42. Kolthur-Seetharam U, Martianov I, Davidson I: **Specialization of the general transcriptional machinery in male germ cells.** *Cell Cycle* 2008, **7**(22):3493-3498.
43. Schmidt EE, Schibler U: **High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids.** *Development* 1995, **121**:2373-2383.
44. Schmidt EE, Ohbayashi T, Makino Y, Tamura T, Schibler U: **Spermatid-specific overexpression of the TATA-binding protein gene involves recruitment of two potent testis-specific promoters.** *JBiolChem* 1997, **272**:5326-5334.
45. Martianov I, Brancorsini S, Catena R, Gansmuller A, Kotaja N, Parvinen M, Sassone-Corsi P, Davidson I: **Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis.** *Proc Natl Acad Sci U S A* 2005, **102**(8):2808-2813.
46. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W *et al*: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**(9):R137.
47. Krebs A, Frontini M, Tora L: **GPAT: retrieval of genomic annotation from large genomic position datasets.** *BMC Bioinformatics* 2008, **9**:533.
48. Delmas V, van der Hoorn F, Mellstrom B, Jegou B, Sassone-Corsi P: **Induction of CREM activator proteins in spermatids: down-stream targets and implications for haploid germ cell differentiation.** *Mol Endocrinol* 1993, **7**(11):1502-1514.

## Figure Legends

### **Fig. 1. Summary of CREM Chip-seq data in haploid cells.**

**A.** Left hand pie chart shows genomic annotation of CREM binding sites with respect to known RefSeq genes. Promoter associated sites are arbitrarily confined to the regions between -5000 and +2000bp with respect to the transcription start sites (TSS). The right hand pie chart indicates the frequency of occurrence of consensus or mutated CRE and half CREs at the CREM binding sites. **B.** Frequency of binding site localisation with respect to TSS. **C.** CREM ChIP-qPCR on the indicated loci. The GFP ChIP-qPCR is shown as control. Results are from three independent ChIP experiments. In panel C the numbers above each column represent the corresponding number of clustered reads in the ChIP seq experiments.

### **Fig. 2. Summary of CREB Chip-seq data in GC1-spg cells.**

**A-D.** The layout is as described in Fig. 1.

### **Fig. 3. Correlative analysis of CREB and CREM binding site occupancy.**

**A.** Association of CREB, pol II and H3K4me3 in GC1-spg cells. Comparison of pol II and H3K4me3 tag density in the region of +/- 5kb around the CREB-occupied loci. Clustering analysis identifies 3 classes, A; CREB loci with low or no pol II and H3K4me3, B-C; CREB loci with high pol II and H3K4me3 corresponding to transcription on the sense and anti-sense strands, respectively. **B.** Comparison of CREM tag density in the region of +/- 5kb around the CREB-occupied loci. Clustering identifies 2 groups A; CREB loci with low or no CREM and B; CREB loci with high CREM. **C.** Venn diagramme comparing common and cell-specific CREB/CREM occupancy of binding sites using the CREB data in panel B and the equivalent analysis of the CREM occupied loci (data not shown).

**Fig. 4. Specific occupancy of target promoters by CREM and CREB.**

**A-D.** Graphic representation of ChIP-seq results as .wig format files in UCSC web browser at the indicated loci. The same format has been used for Figs. 6 and 7 as well as Additional file 1, Figs. S4 and 5. The ChIP-seq results for CREB and CREM in GC1-spg and testis haploid cells respectively are shown along with results for H3K4me3 in GC1-spg cells (K4-GC1) and testis (K4-T). In addition, the mammalian conservation track from the UCSC browser is also included. CREM and CREB binding peaks are indicated by arrows.

**Fig. 5. CREM bound genes that are deregulated upon CREM inactivation in testis.**

**A.** Venn diagramme comparing the total number of genes with CREM binding in their promoters with the genes deregulated in the Affymetrix analysis of Beisbarth et al. **B** Pie chart indicating the frequency of occurrence of consensus or mutated CRE and half CREs at the CREM binding sites in the promoters of the 58 deregulated genes.

**Fig. 6. Binding of CREB and CREM to the promoters of brain enriched miRNAs in haploid and GC1-spg cells or to the promoters of the coactivator CBP (*Crebbp*) and TORC1 (*Crtc1*) genes.**

**A-D.** Graphic representation of ChIP-seq results in UCSC web browser at the indicated loci using the nomenclature described in Fig. 4.

**Fig. 7. Identification of novel CREB occupied regulatory elements and transcripts.**

**A.** Pie chart indicating the frequency of occurrence of consensus or mutated CRE and half CREs at the CREM and CREB intergenic binding sites. **B-C.** Graphic representation of ChIP-seq results in UCSC web browser at the indicated loci using the nomenclature described in Fig. 4, with the addition of the GC1-spg pol II ChIP-seq data set. In panels B-C, the

presence and direction of transcription of potential novel transcripts in the regions covered by elongating pol II are indicated by arrows.

## **Additional files**

### **Additional File 1, Figure S1.**

Expression of the CREB and CREM $\tau$  proteins in adult mouse testis. **A.** Immunofluorescence on sections from adult mouse testis with the anti-CREB and CREM $\tau$  antibodies. The signal for the antibodies in red, the Hoechst stained nuclei and the merged views are shown as indicated. Representative examples of the different cell types are indicated by arrows. L; Leydig cells, S; Sertoli cells; Sp; spermatogonia; P; pachytene spermatocytes, I; intertubular cells, R; round spermatids, E; elongated spermatids. These results reveal a distinct expression of CREB and CREM $\tau$  in the testis, where CREB is present in the spermatogonia, the Sertoli, Leydig and intertubular cells, while CREM $\tau$  is present in haploid round spermatids. 20 fold magnification. **B.** Western blots with the CREM $\tau$  and CREB antibodies on 20ug of extracts from the indicated cells or tissues. ES is undifferentiated mouse E14 embryonic stem cells, COS is COS-1, and 3T3 is NIH3T3. The CREB antibody detects a single polypeptide corresponding to CREB in each of the extracts, while the CREM $\tau$  antibody detects the CREM $\tau$  proteins only in the testis extract.

### **Additional File 2, Figure S2.**

Example of ChIP-seq data over an 11,191,448 bp region of mouse chromosome 3qA3-3qB. The results for CREB and CREM in GC1 and haploid cells respectively are shown along with the ChIP-seq for H3K4me3 in GC1 cells (K4-GC1) and testis (K4-T). Representative CREB, CREM binding sites are indicated. Active promoters marked by H3K4me3 are also indicated. The \* in the CREB and CREM channels show examples of loci that are specifically occupied

GC1 cells or haploid cells respectively. The \* in the K4-GC1 and K4-T show examples promoters that are specifically or preferentially active in GC1 cells or testis, respectively.

**Additional File 3, Table S1.**

CREM bound loci in mouse male haploid germ cells. Excel table of annotated loci bound by CREM. Page 1 shows annotation of bound loci to RefSeq genes. N is number, Chr chromosome, start and end shows the coordinates of the binding sites and summit, the coordinates of the center of the site. The number of clustered reads and the pvalue and the local fold enrichment are indicated. For annotation, the RefSeq transcript ID, the gene symbol, and TSS coordinates are indicated along with the distance between the summit of the peak and the TSS. No match signifies that there was no RefSeq annotation within 5kb from the peak summit position. The presence of a perfect CRE, a CRE with 0 or 1 mismatch or a half CRE are then indicated. The H3K4me3 status of the loci is also indicated. Page 2 shows annotation of bound loci to Ensembl transcripts. Layout is as in page 1 except that the RefSeq transcripts are replaced by Ensembl annotation.

**Additional File 4, Table S2.**

MiRNAs that are targets of CREM in mouse male haploid germ cells. Excel table of CREM bound miRNAs in haploid cells. Layout is essentially as in Table 1.

**Additional File 5, Figure S3.**

Comparison of CREM and H3K4me3 ChIP-seq with adult mouse round spermatid gene expression. **A.** The probe set expression values were divided into classes of  $\leq 10$ , 11-50, 51-500 and  $\geq 501$  and the % of the total probe sets in each category are represented by the black bars as a % of the total. The % of the total number of CREM and H3K4me3 occupied

promoters in each expression category is shown with the shaded bars. **B.** A similar representation is shown for the probe set expression values and CREB and H3K4me3 occupied promoters in GC1-spg cells.

**Additional File 6, Table S3.**

CREB bound loci in GC1 cells. Excel table of binding sites. Page 1 shows annotation of bound loci to RefSeq genes. Page 2 shows annotation of bound loci to Ensembl transcripts. Layout is as in Table S1.

**Additional File 7, Table S4.**

Comparison of genes whose expression is affected by CREM inactivation in testis and CREM binding. Page 1 of the Excel table reproduces the Affymetrix data of Beissbarth et al., (1) ([http://www.dkfz.de/tbi\\_old/crem/cremgenes.txt](http://www.dkfz.de/tbi_old/crem/cremgenes.txt)) concerning genes whose expression is deregulated in the testis of CREM knock-out mice. Page 2 shows the data for CREM occupancy of the subset of loci that map to the promoters of the genes in page 1. Note that some promoters contain several CREM occupied sites; Page 3 shows the list of the genes whose expression is deregulated from page 1 and whose promoters are occupied by CREM.

**Additional File 8, Figure S4.**

UCSC web browser graphic view of CREB and CREM binding to the indicated loci. **A-B.** Selective binding of CREM to the *Clock* and *Per 2* promoters. **C.** CREB, CREM occupancy and H3K4me3 over the *Crem* locus.

**Additional File 9, Table S5.**



Functional annotation genes whose promoters are occupied by CREM (page 1) and CREB (page 2). Functional annotation was performed using the KEGG pathway of David (<http://david.abcc.ncifcrf.gov/>). In addition to the list of genes in each identified functional category, are shown the number of genes (gene count) the % of the total number of genes, and pvalue.

**Additional File 10, Table S6.**

Excel table of CREM and CREB bound sites with no annotation to RefSeq genes or Ensembl transcripts. Pages 1 and 2 show the data concerning CREM and CREB binding sites for which no RefSeq annotation was found within 5Kb of the peak summit and which were designated 'no match' in Tables S1 and S3. Layout is as in Table S1.

**Additional File 11, Figure S5.**

**A-B.** UCSC web browser graphic view of CREB, H3K4me3 and pol II occupancy of intergenic sites upstream of annotated genes. In panel A, no elongating pol II is seen between the upstream site and the annotated promoter, while in panel B, a low level of elongating pol II is observed, suggesting the existence of an alternative promoter. **C.** Multiple CREB binding sites and high pol II occupancy of the locus encoding the Malat1 and MEN $\epsilon/\beta$  non-coding RNAs.

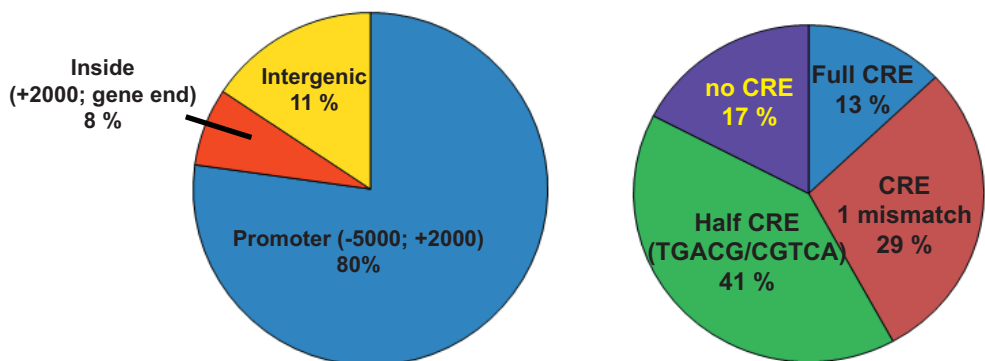
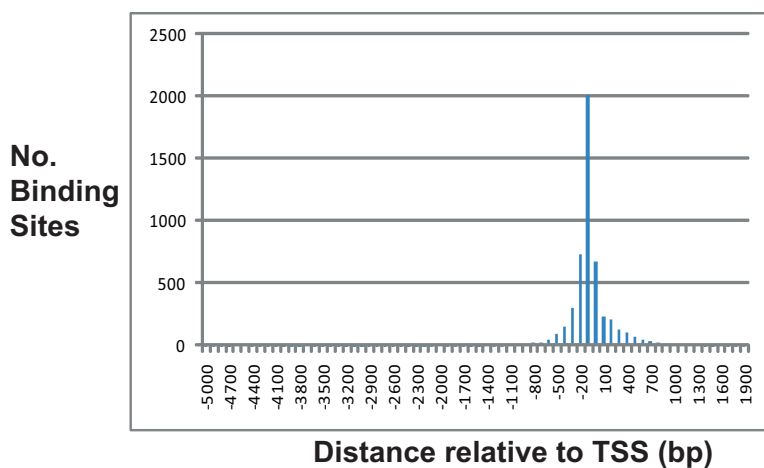
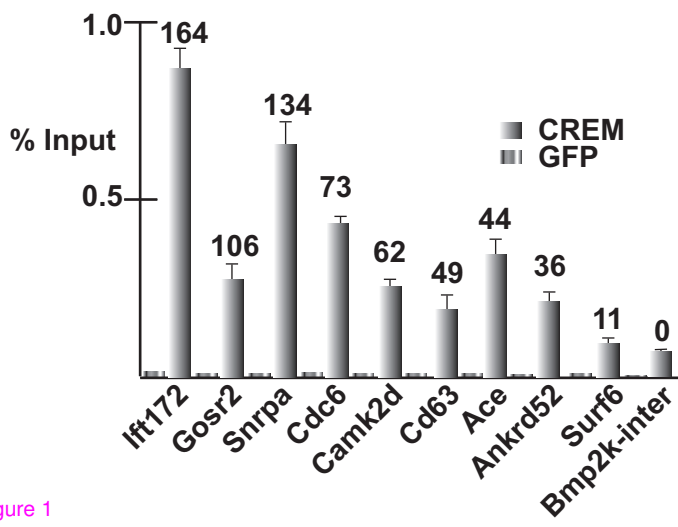
**A****6792 binding sites****B****C**

Figure 1

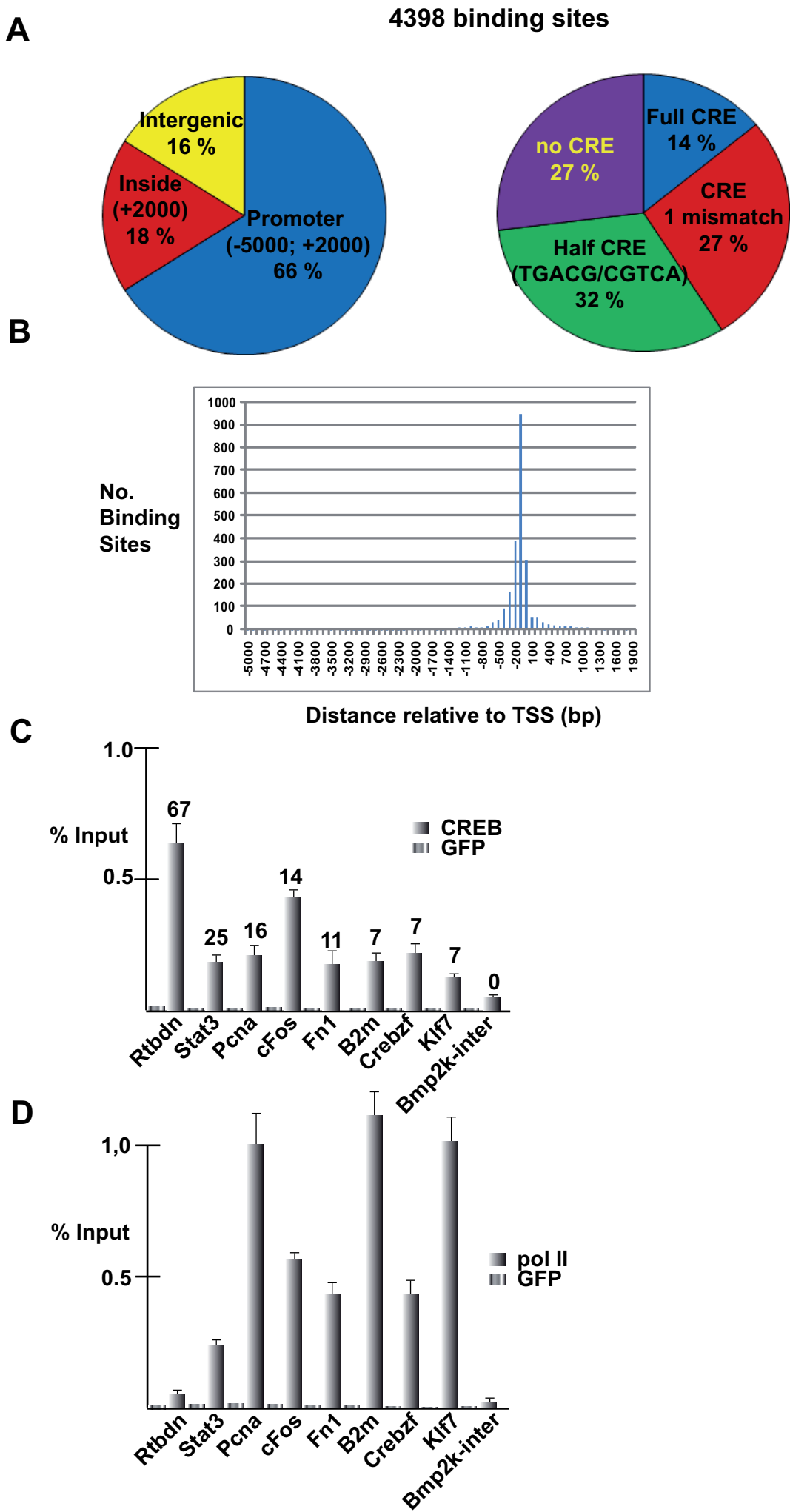


Figure 2

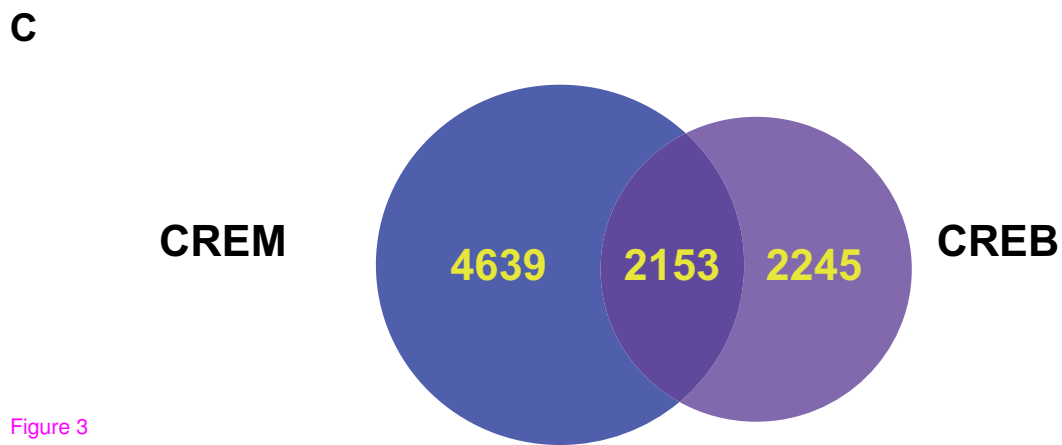
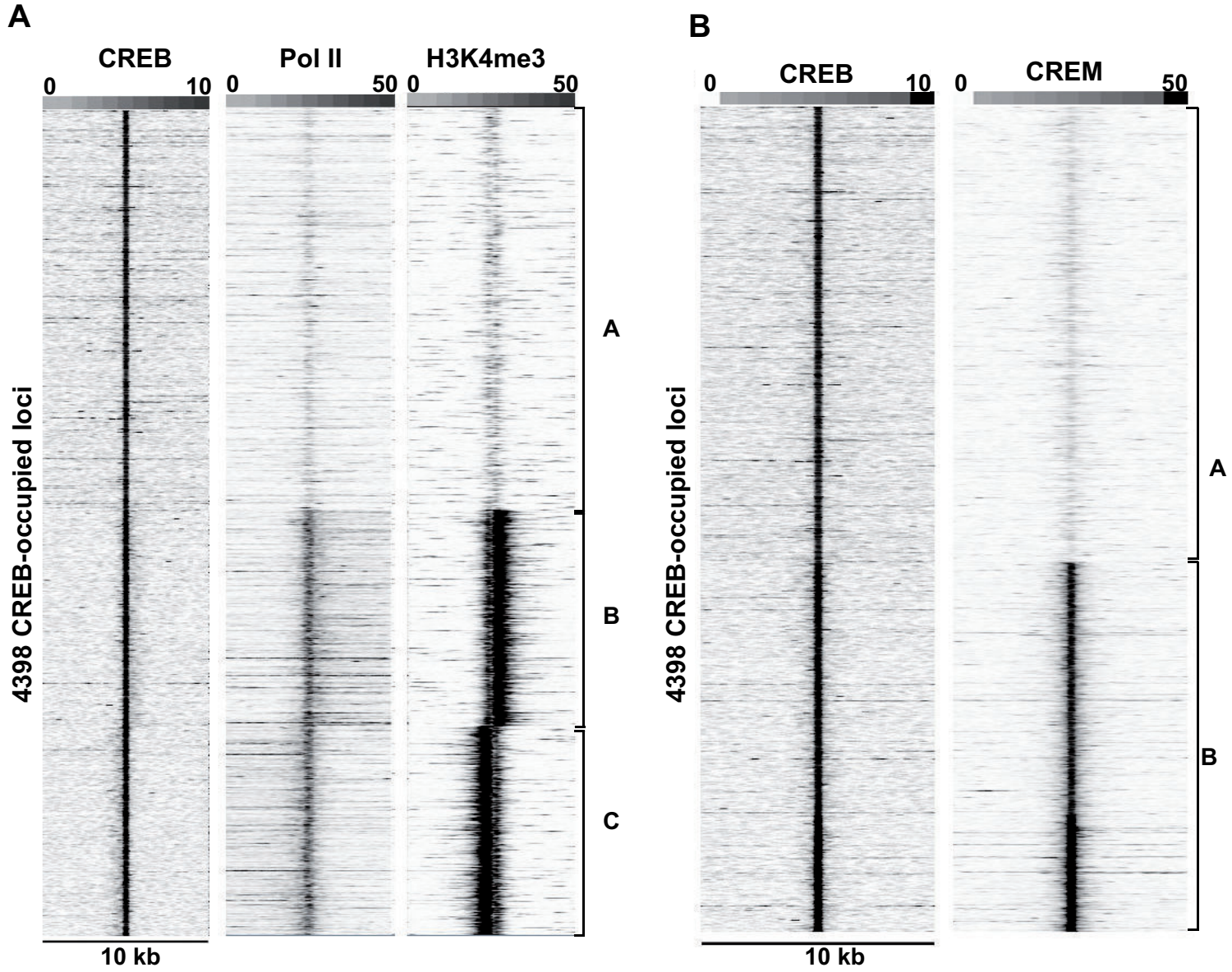


Figure 3

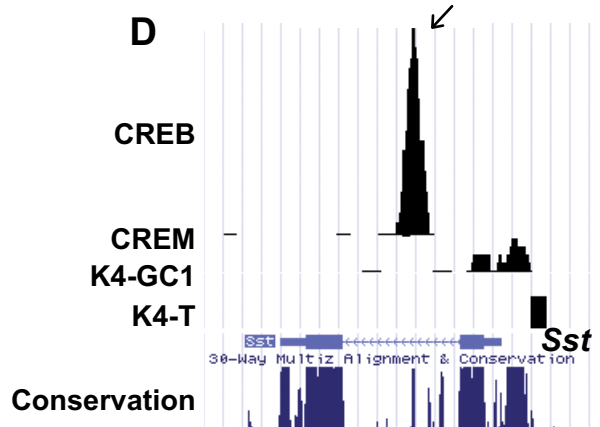
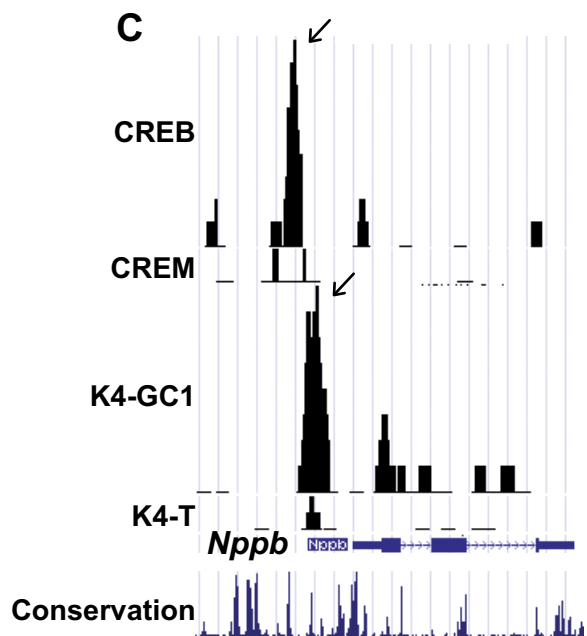
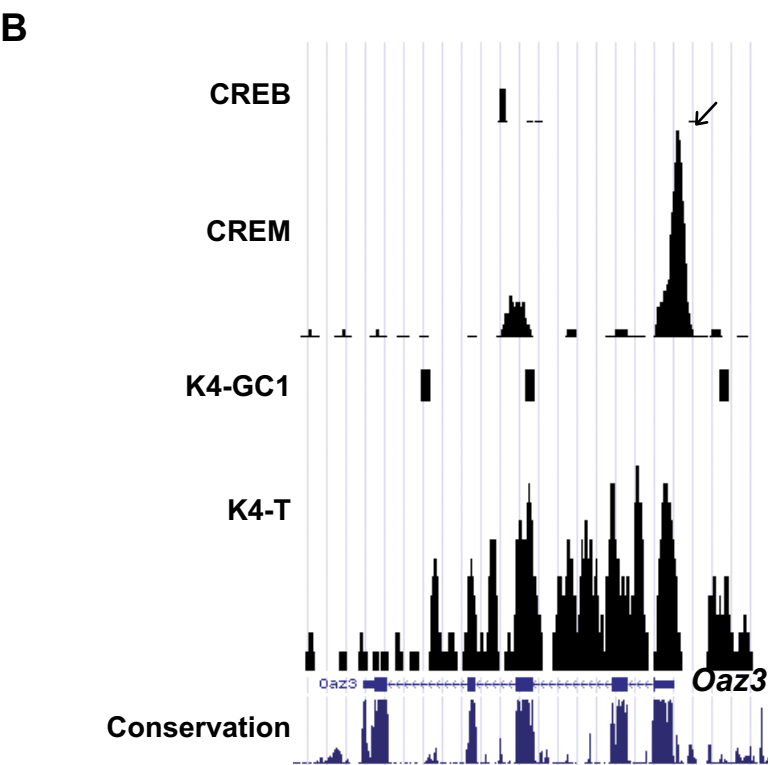
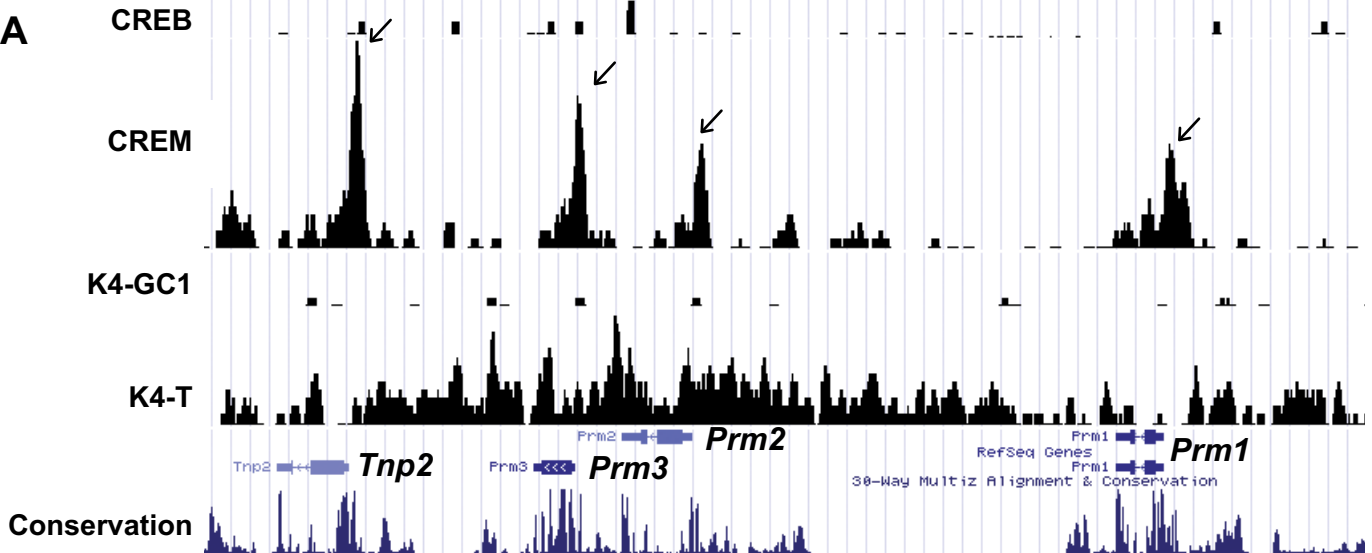


Figure 4

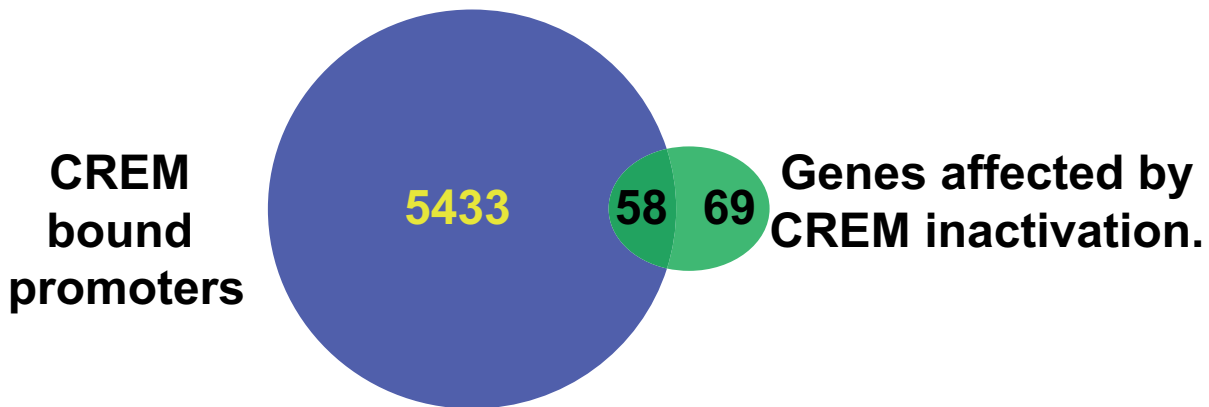
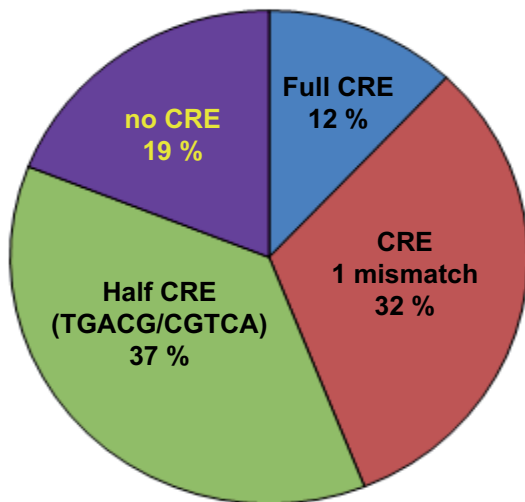
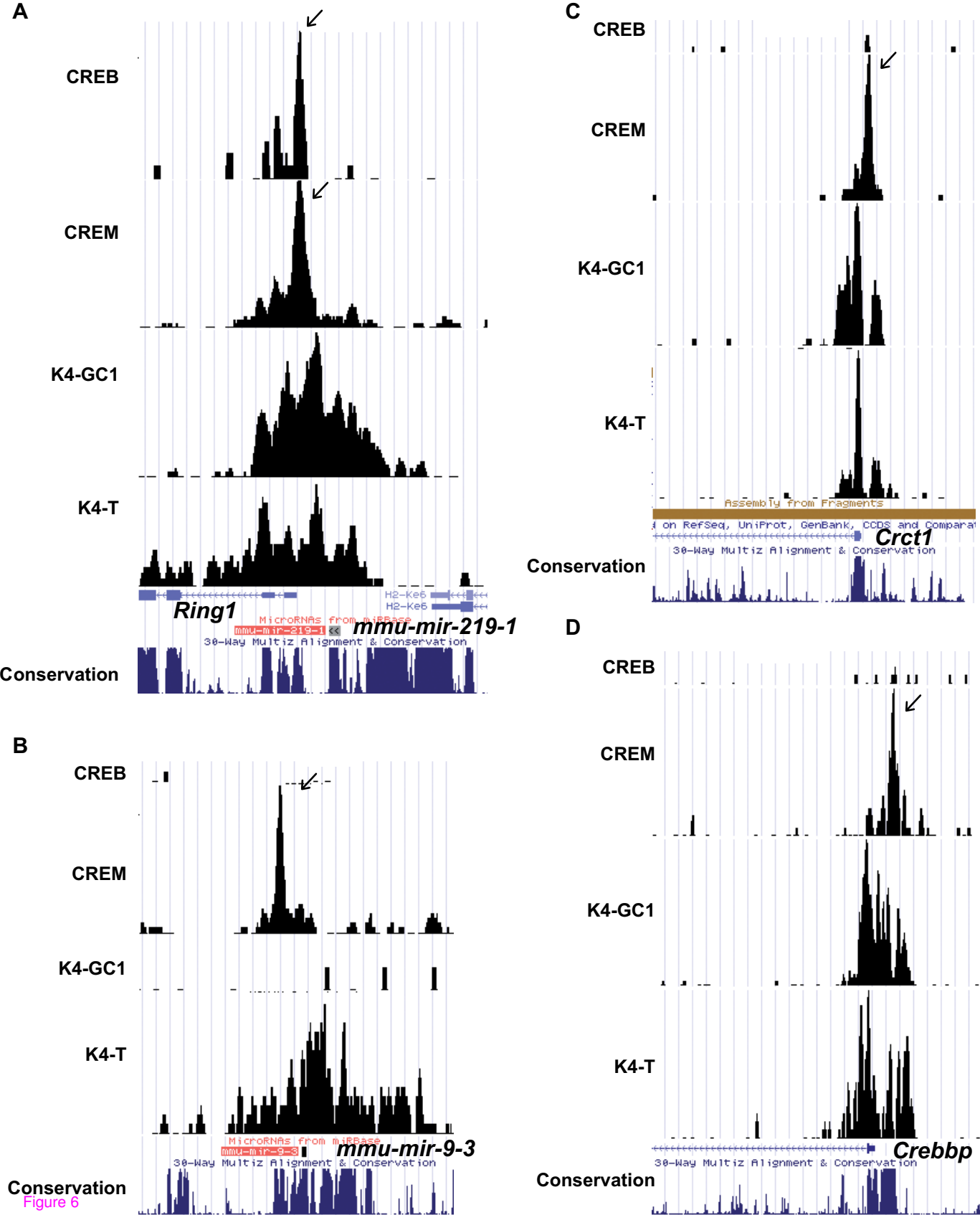
**A****B**

Figure 5



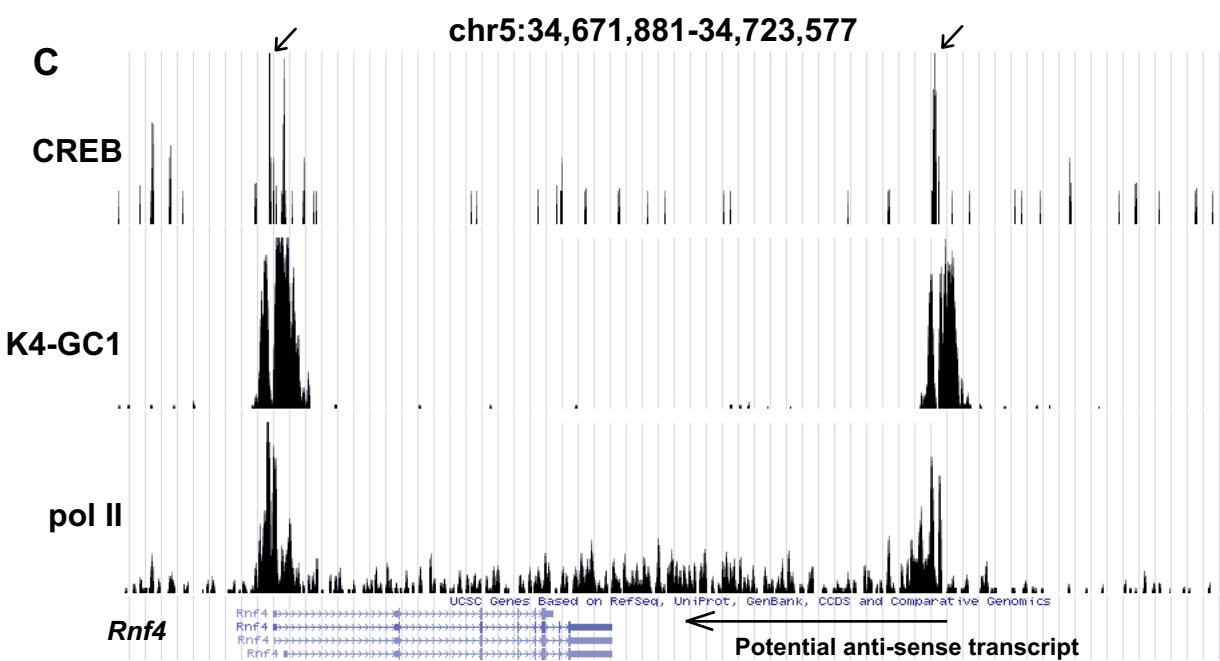
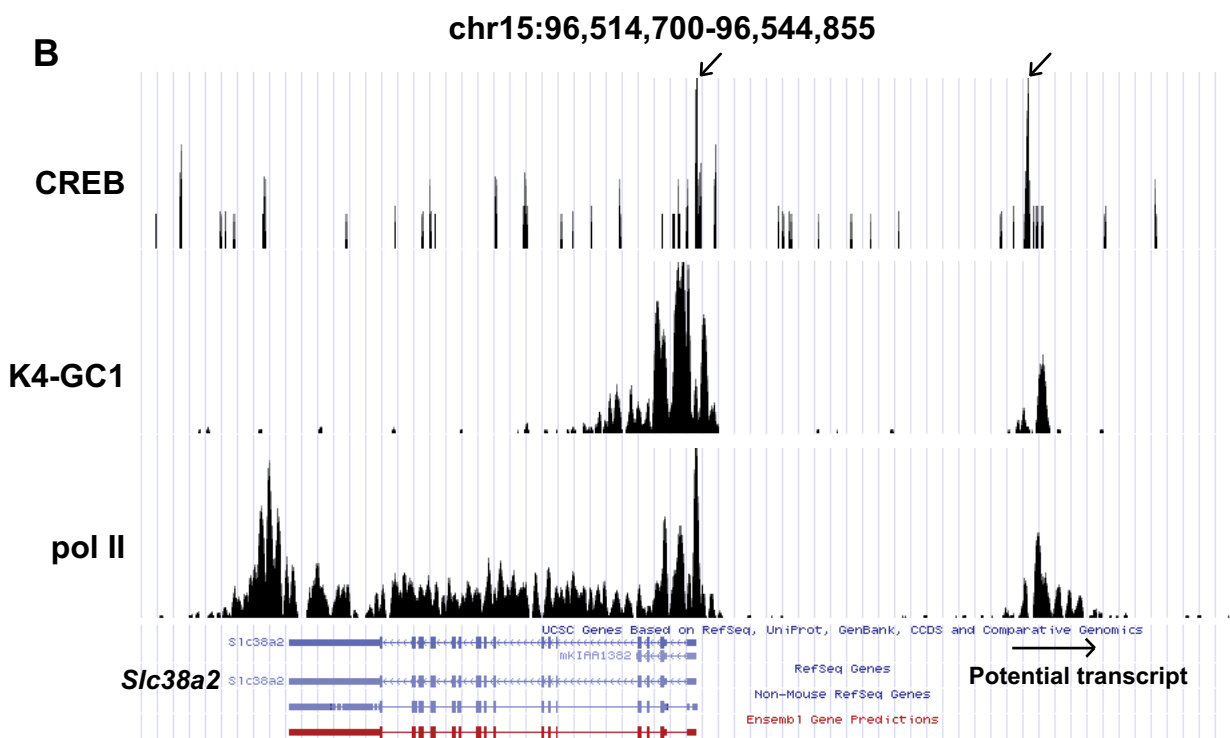
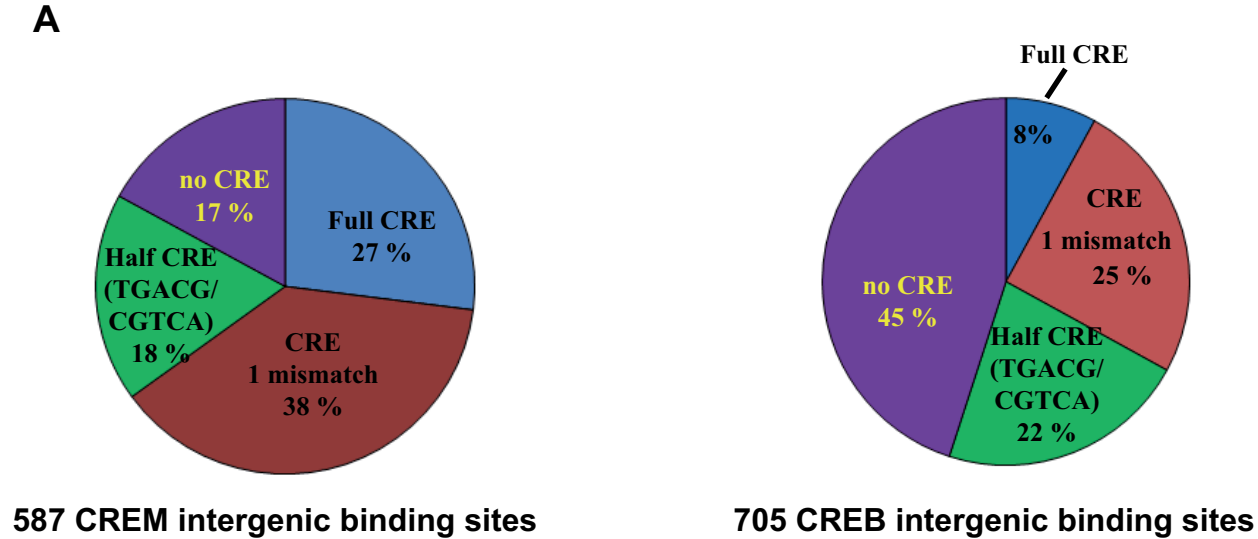


Figure 7



**Additional files provided with this submission:**

Additional file 1: Supplemental Figure 1.jpg, 1821K

<http://www.biomedcentral.com/imedia/2534992864580882/supp1.jpeg>

Additional file 2: Supplemental Figure 2.jpg, 2107K

<http://www.biomedcentral.com/imedia/1599658392458088/supp2.jpeg>

Additional file 3: Supplemental Table-1.xls, 2929K

<http://www.biomedcentral.com/imedia/1517035075458088/supp3.xls>

Additional file 4: Supplemental Table-2.xls, 30K

<http://www.biomedcentral.com/imedia/1324131912458088/supp4.xls>

Additional file 5: Supplemental Fig-3.pdf, 251K

<http://www.biomedcentral.com/imedia/1834785054580882/supp5.pdf>

Additional file 6: Supplemental Table-3.xls, 1832K

<http://www.biomedcentral.com/imedia/1347512161458088/supp6.xls>

Additional file 7: Supplemental Table-4.xls, 69K

<http://www.biomedcentral.com/imedia/1035412679458088/supp7.xls>

Additional file 8: Supplemental Fig-4.jpg, 1745K

<http://www.biomedcentral.com/imedia/2054941070458088/supp8.jpeg>

Additional file 9: Supplemental Table-5.xls, 40K

<http://www.biomedcentral.com/imedia/6127019494580882/supp9.xls>

Additional file 10: Supplemental Table-6.xls, 213K

<http://www.biomedcentral.com/imedia/4926335045808824/supp10.xls>

Additional file 11: Supplemental Fig-5.jpg, 1556K

<http://www.biomedcentral.com/imedia/1113698318458088/supp11.jpeg>

**Publication 4 :**

**seqMINER: An integrated ChIP-seq data interpretation platform**

**Tao Ye\*<sup>1</sup>, Arnaud R. Krebs\*<sup>§1</sup>, Amin M. Choukrallah<sup>1</sup>, Celine Keime<sup>2</sup>, Frederic Plewniak<sup>3</sup>, Irwin Davidson<sup>1</sup> and Laszlo Tora**

Publication en révision, soumise à Nucleic Acids Research

**seqMINER: An integrated ChIP-seq data interpretation platform**

|                  |   |
|------------------|---|
| Journal:         | <i>Nucleic Acids Research</i>                           |
| Manuscript ID:   | NAR-01571-Met-G-2010                                    |
| Manuscript Type: | 3 Methods Manuscript (Online Publication)               |
| Key Words:       | ChIP-seq, data analysis, tool, bioinformatics, genomics |
|                  |   |

SCHOLARONE™  
Manuscripts

Review

# seqMINER: An integrated ChIP-seq data interpretation platform

**Manuscript Type: Resource**

**Tao Ye\*<sup>1</sup>, Arnaud R. Krebs\*<sup>§1</sup>, Amin M. Choukrallah<sup>1</sup>, Celine Keime<sup>2</sup>, Frederic Plewniak<sup>3</sup>,**

**Irwin Davidson<sup>1</sup> and Laszlo Tora<sup>§1</sup>**

Address: <sup>1</sup>Department of Functional Genomics, <sup>2</sup> Microarray and Sequencing Platform, <sup>3</sup> Bioinformatics Platform. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 596, Université de Strasbourg, BP 10142-67404 ILLKIRCH Cedex, CU de Strasbourg, France.

<sup>§</sup>Corresponding authors

\*These authors contributed equally to the work

Email addresses:

[laszlo@igbmc.fr](mailto:laszlo@igbmc.fr)

[krebs@igbmc.fr](mailto:krebs@igbmc.fr)

## ABSTRACT

In a single experiment, Chromatin Immunoprecipitation combined with high throughput sequencing (ChIP-seq) provides genomewide information about a given covalent histone modification or transcription factor occupancy. However, time efficient bioinformatics resources for extracting biological meaning out of these gigabyte-scale datasets are often a limiting factor for data interpretation by biologists. We created seqMINER, an integrated portable ChIP-seq data interpretation platform with optimized performances for efficient handling of multiple genomewide datasets. seqMINER allows comparison and integration of multiple ChIP-seq datasets and extraction of qualitative as well as quantitative information. seqMINER can handle the biological complexity of most experimental situations and proposes supervised methods to the user in data categorization according to the analysed features. In addition, through multiple graphical representations, seqMINER allows visualisation and modelling of general as well as specific patterns in a given dataset. Moreover, seqMINER proposes a module to quantitatively analyse correlations and differences between datasets. To demonstrate the efficiency of seqMINER, we have carried out a comprehensive analysis of genome wide chromatin modification data in mouse embryonic stem cells to understand the global epigenetic landscape and its change through cellular differentiation.

1  
2 INTRODUCTION  
3  
4  
5  
6

7 Chromatin immunoprecipitation (ChIP) allows the quantitative measurement of protein (i.e.  
8 transcription factor) occupancy or the presence of post-translational epigenetic histone  
9 modifications at defined genomic loci. Recent technological developments combining ChIP with  
10 direct high throughput sequencing of the immunoprecipitated DNA fragments (ChIP-seq) allowed the  
11 analysis of transcription factor genomic occupancy or epigenetic chromatin marks at the genome  
12 wide (GW) scale in an relatively unbiased manner. Over the last years, numerous studies have used  
13 ChIP-seq as a central method to create GW binding maps for a particular genomic feature (Ku et al.  
14 2008; Mikkelsen et al. 2007), installing ChIP-seq as the gold standard in the functional genomics  
15 toolbox.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

27  
28 Each ChIP-seq run generates data at the gigabyte scale that requires successive steps of  
29 bioinformatics treatment prior to biological interpretation (reviewed in (Johnson et al. 2007)). In a  
30 standard analysis pipeline, two major steps can currently be distinguished. First, genomic locations  
31 presenting relevant enrichment in the ChIP-seq signal are identified and annotated with respect to  
32 known genomic sequence features (genes, transcripts, repeat elements etc). Second, by performing  
33 multiple rounds of analyses using various methods, the biological meaning of the dataset has to be  
34 extracted and often compared with other datasets. Many methods and softwares have been released  
35 over the past months in order to easily perform the initial analysis stage with good accuracy  
36 (reviewed in (Johnson et al. 2007; Laajala et al. 2009)). However, unlike the first step that  
37 distinguish relevant signal from noise to provide information on a given factor or chromatin mark,  
38 the second analysis stage requires the combined use of various methodologies to answer multiple  
39 complex biological questions. Thus, development of integrated complementary approaches is a  
40 prerequisite to make ChIP-seq analysis as easy and routine as possible.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

58 Several initiatives devoted to particular biological questions have already contributed to  
59 enrich the ChIP-seq analysis toolbox. For example, numerous tools, integrated or not in larger  
60

1  
2 analysis pipelines, have been proposed that allow: annotation of genomic features present in the  
3  
4 neighbourhood of the relevant enrichment signals (Ji et al. 2008; Krebs et al. 2008; Shin et al.  
5  
6 2009), detection and *de novo* definition of consensus DNA motifs (Ji et al. 2008), cross and  
7  
8 compare information from distinct datasets (Blankenberg et al. 2007), and comprehensive  
9  
10 visualization of the obtained GW results (Kent et al. 2002; Manske and Kwiatkowski 2009; Skinner  
11  
12 et al. 2009). More recently, sophisticated statistical approaches were proposed to predict association  
13  
14 between sets of genomic locations and numerous genomic features (Bock et al. 2009; Hon et al.  
15  
16 2008). Nevertheless, due to the multiplicity of biological questions that may be asked by the ChIP-  
17  
18 seq method, many analysis issues remain un-addressed.  
19  
20  
21  
22

23 In particular, the definition of the genomic rules governing the function of a particular factor  
24  
25 is a complex, but central question. The most common method for addressing this question is to  
26  
27 investigate the frequency of co-occurrence of the analysed factor/mark with other genomic features  
28  
29 by (i) extracting relevant signals (peaks) in the two datasets to be compared and (ii) calculating the  
30  
31 number of events that are close enough between the datasets to be considered as overlapping. This  
32  
33 approach, that has been previously used (Blahnik et al.; Blankenberg et al. 2007), presents several  
34  
35 limitations. First, the analysis is biased towards the peak-searching step and the empirical cut-off  
36  
37 values used to discriminate between relevant and background signals. The second drawback comes  
38  
39 from comparisons of datasets containing either sharp (i.e. a transcription factor) or non-discrete  
40  
41 binding sites (i.e. spread histone marks like H3K27me3), where peak detection is more difficult.  
42  
43 Third, an increasing number of studies show that particular factors have more than a single function  
44  
45 in the genome (i.e.(Bilodeau et al. 2009)), challenging the direct interpretation of one to one  
46  
47 comparisons. Finally, it becomes clear that definition of genomic regulatory elements cannot rely on  
48  
49 a single feature, but need integration of multiple sources of information to be properly identified  
50  
51 (Heintzman et al. 2009; Wang et al. 2008). Moreover, tools able to analyse multiple information  
52  
53 sources are required to allow comprehensive qualitative and quantitative ChIP-seq analysis.  
54  
55  
56  
57  
58  
59  
60

To overcome the above described limitations, we have developed seqMINER, an integrated

1 user friendly platform that addresses central questions in the ChIP-seq analysis workflow.  
2  
3 seqMINER was designed in order to make it as easy as possible for the biologist to carry out in  
4  
5 depth interpretation of the analysed datasets to answer their biological question. The purpose of  
6  
7 seqMINER is to allow qualitative and quantitative comparisons between a reference set of genomic  
8  
9 positions and multiple ChIP-seq datasets. Different analysis modules have been implemented to  
10  
11 allow users to search for full characteristics of a particular feature genomewide. Starting from a set  
12  
13 of reference coordinates that can be a list of ChIP-seq enrichment clusters (peaks) for a particular  
14  
15 target (i.e. a transcription factor), seqMINER proposes two complementary methods to analyse the  
16  
17 signal enrichment status in multiple other tracks: i) a qualitative method that computes a density  
18  
19 array over a defined window around the reference coordinate; ii) a quantitative method that  
20  
21 computes enrichment value over a defined widow around the reference coordinate. Following these  
22  
23 steps, automated as well as manual reordering methods of the analysed loci are implemented to  
24  
25 assist users in defining functional subgroups in their data. Finally graphical representations of the  
26  
27 data are proposed through heatmaps and dotplots to illustrate particular, as well as general  
28  
29 properties of the data. We show that, the different resources of seqMINER taken together allow a  
30  
31 comprehensive analysis of genome-wide chromatin modification data to understand the global  
32  
33 epigenetic landscape and its change through cellular differentiation.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

## 44 MATERIAL AND METHODS

### 45 seqMINER algorithms

46  
47 seqMINER uses two methods to quantify ChIP-seq signal depending on the type of analysis  
48  
49 performed (Figure 2). For both methods, prior to analysis, all reads were extended from a user  
50  
51 defined size (default 200bp). For the calculation of densities over a defined window, methods were  
52  
53 derived from the one generally used to generate density files (i.e. (Zhang et al. 2008) except that tag  
54  
55 extension was performed only on the direction of the tag (not in both strand orientations). Then a  
56  
57  
58  
59  
60



1  
2 user defined number of bins are created around the reference coordinate and for each bin the  
3  
4 maximal number of overlapping tags is computed (Figure 2).  
5

6  
7 To calculate a single quantitative value for a binding site, tag density was defined as the number of  
8  
9 tags present or overlapping in user defined window (default 2kb) around the reference site (Figure  
10  
11 2). ChIP-seq enrichments (e) were defined as  $e = \log_2 \left( \frac{\text{foreground tags} + q}{\text{background tags} + q} \right)$   
12  
13 where q is an empirical constant in the range of 10 that normalises the enrichment value; foreground  
14  
15 tags are the density value computed in the data track ; background tags, the density value in the  
16  
17 control track. Use of the constant q reduces the influence of the signal variation in the noise  
18  
19 measurement of the background sample on the ratio calculation.  
20  
21  
22

### 23 **Data source and treatment**

24  
25  
26 ChIP-seq datasets were downloaded from the public data bank Gene Expression Omnibus  
27  
28 (<http://www.ncbi.nlm.nih.gov/gds>) under the accession number: GSM307618 (H3K4me3-ESC);  
29  
30 GSM281696 (H3K4me3-brain); GSM307625 (input-ES); GSM307620 (H3K36me3-ES) and  
31  
32 GSM307619 (H3K27me3-ES). Reference coordinates list was established using MACS (Wang et  
33  
34 al. 2008).  
35  
36

### 37 **Code repository**

38  
39  
40 The source code, tutorial and wiki for seqMINER are available at [http://bips.u-](http://bips.u-strasbg.fr/seqminer/)  
41  
42 [strasbg.fr/seqminer/](http://bips.u-strasbg.fr/seqminer/) under General Public License (GPL3).  
43  
44  
45  
46  
47  
48

## 49 **RESULTS**

### 50 **Correlative integration of multiple datasets.**

51  
52  
53  
54 Most genomic studies aim to define the function of a particular regulatory factor by  
55  
56 understanding globally how it affects other co-occurring events in a regulatory circuit (i.e. chromatin  
57  
58 modifications, binding of other factors) and the consequences on outputs of the regulated system  
59  
60 (i.e gene expression). Thus we designed seqMINER to allow the integration of multiple ChIP-seq

1  
2 datasets in a quick and user friendly manner.

3  
4 seqMINER uses a list of BED formatted genomic coordinates (i.e. binding sites of a  
5 particular factor, set of genes, set of promoters etc.) as a reference for investigating information in  
6 other genomic datasets. Three stages can be distinguished in the analysis process (Figure 1). First,  
7 in the data collection module, seqMINER collects the read density from a reference dataset over a  
8 user-defined window around a set of coordinates and then calculates the read density in the same  
9 window in one or multiple other data sets. Second, in the clustering module, seqMINER, uses a  
10 supervised clustering procedure (k-means) to organise the identified loci presenting similar read  
11 densities within the specified window. Third, in the visualization module, seqMINER allows at  
12 glance visualisation of the entire output dataset through various graphical representations.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 seqMINER proposes two related complementary tag (read) density based methods to  
26 analyse the signal enrichment status in multiple tracks (Figure 2). The first is a qualitative method  
27 that defines general patterns and functional sub-groups in the dataset: densities over a window  
28 around the reference coordinates can be calculated at different resolutions in multiple tracks (Figure  
29 2A). The created matrix can be organized by supervised k-means clustering to isolate groups of loci  
30 having similar features as developed by (Heintzman et al. 2009; Heintzman et al. 2007). At this  
31 stage, clusters can optionally be reorganized manually according to the biological significance.  
32 Visualisation of the whole dataset is in this case achieved through heatmaps. This method allows  
33 easy visualisation of signal distribution over multiple loci and identifies general patterns over the  
34 dataset, which can be plotted as average profiles. Information on signal distribution, that can be an  
35 important biological feature, is conserved (i.e. broad, sharp enrichment peaks). However,  
36 visualisation of quantitative phenomenon is more complex. The second method allows extraction of  
37 quantitative information. Raw tag counts as well as normalised enrichment over a control track can  
38 be calculated (Figure 2B). The created matrix allows easy plotting of quantitative information and  
39 interpretation for one to one comparisons. Moreover, the method produces numerical data that are  
40 necessary to integrate sequencing data in larger mathematical models of a particular system.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 All visual features as well as the list of loci can be exported for further analysis with other  
3 methodologies (i.e. gene annotation (Krebs et al. 2008), ontology (Dennis et al. 2003)). Moreover,  
4 output data can be used in new clustering rounds using different sets of data or analysis parameters.  
5  
6 This possibility facilitates multiple iterative steps of analysis inherent to the genomic data analysis.  
7  
8  
9  
10  
11  
12

### 13 **Performance optimisation**

14  
15 seqMINER takes advantage of the portability of the Java Virtual Machine (JVM) allowing  
16 an installation free multiplatform usage. seqMINER was designed to optimise all time limiting steps  
17 in the manipulation of the large raw sequencing datasets. For all the computational tasks, the  
18 implementation combines an optimal use of the Random Access Memory (RAM), avoiding  
19 repetitive access to the hard drive, and Simultaneous Multi Threading (SMT) to fully use available  
20 computational resources and lower the time of analysis. SMT permits the program to execute  
21 multiple independent threads to better utilize the resources provided by modern processor  
22 architectures. Particular attention was paid to the design of the genomic data storing objects to  
23 optimize space occupancy and efficiency of data retrieval algorithms. seqMINER performances  
24 were assessed by training the analysis on three standard size CHIP-seq datasets (from 191-554MB)  
25 using as an input reference set different subsets of the ENSEMBL transcript database (33881 input  
26 coordinates) (Figure 3). All software trainings were performed on a PC running under windows  
27 with standard performances (CPU-3.0GHz core-duo; RAM – 4Go). For all the tested sets, the  
28 limiting step appears to be the loading of the dataset in the memory (Figure 3), which is usually  
29 performed only once and followed by multiple cycles of analysis. Nevertheless, the time required  
30 for a complete analysis by seqMINER for all tested datasets, regardless of the number of reference  
31 coordinates used, and including data loading is less than 30 seconds (Figure 3).  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

56 To bring the analysis time by seqMINER to acceptable levels we had to optimize the  
57 clustering process for the density arrays. The density array based method generates a large array of  
58 values (i.e. for a 10kb window at a 25bp resolution, an array of 400 values is created per dataset  
59  
60

1 analysed) for each reference position. The use of existing implementation (de Hoon et al. 2004) of  
2 clustering algorithms to organize these large datasets appeared to considerably slow down the  
3 analysis workflow (over an hour for the 8470 reference sites and two datasets). Thus, we decided to  
4 create a novel implementation of this algorithm using recent technology developments in  
5 programming resources. We implemented the k-means algorithm, taking advantage of SMT  
6 technology and using Java Machine Learning libraries (Abeel et al. 2009) for the data structure  
7 implementation. This new algorithm dramatically improves the clustering speed compared to  
8 existing implementations (Figure 3) since less than 10 seconds are needed for clustering all the  
9 tested datasets, while more than an hour was needed for the smallest dataset tested with the former  
10 algorithm. Thus this new implementation used in seqMINER brings the efficiency of clustering  
11 routines to the requirements of ChIP-seq data analysis.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

28 These combined efforts allow the analysis of multiple raw sequencing datasets in a few  
29 seconds, considerably speeding up the ChIP-seq analysis workflow and facilitating the testing of  
30 multiple biological assumptions in a time efficient manner.  
31  
32  
33  
34  
35  
36

### 37 **Genome wide analysis of the chromatin landscape of genes in embryonic stem cells**

38 Comparative analysis of epigenetic profiles in several cell types allows a global view and  
39 better understanding of chromatin dynamics and its role in gene regulation. In the last few years,  
40 numerous genomic studies focussed on embryonic stem cells (ESC) that can both self-renew  
41 indefinitely or differentiate in cell types that from the three primary germ layers. These interesting  
42 properties were broadly studied and were shown to be highly dependent on transcriptional and  
43 epigenetic regulatory networks (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006). In order  
44 to train seqMINER and demonstrate its efficiency in addressing complex biological questions, we  
45 decided to (i) make a comprehensive description of the chromatin states at all annotated promoters  
46 of mouse ESCs, and (ii) quantitatively compare the active chromatin marking in ESCs with that  
47 observed in the differentiated brain tissue.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 First, we used seqMINER to characterize the epigenetic profile of mouse genes in ESCs  
3  
4 using three representative histone modifications known to mark active or inactive genes. The active  
5  
6 histone marks H3K4me3 and H3K36me3, are enriched in the promoter regions and transcribed  
7  
8 gene bodies, respectively. In contrast, the repressive mark, H3K27me3, is known to give a broader  
9  
10 distribution over genes. As reference coordinates, we used transcription start sites (TSS; assumed to  
11  
12 be at the 5' end of annotated genes) of the 33881 mouse genes referenced in ENSEMBL (v58)  
13  
14 database. Tags densities from each dataset were collected in a window of 10 kb around the reference  
15  
16 coordinates and the collected values were subjected to k-means clustering.  
17  
18  
19

20  
21 Out of this initial clustering, five groups of loci could be distinguished (Figure 4A) amongst  
22  
23 which we can identify three clusters of active genes transcribed on the negative strand, on the  
24  
25 positive strand and in both directions, as determined by the H3K36me3 mark (Figure 4A middle  
26  
27 panel). Additionally, two clusters of inactive genes could be identified, one marked by both  
28  
29 H3K4me3 and H3K27me3 marks (known as bivalent promoters) and one showing no significant  
30  
31 enrichment of the three studied marks. Interestingly, while the bivalent loci are known to  
32  
33 correspond to transiently repressed genes required for later differentiation of ESCs, the unmarked  
34  
35 loci likely represent strongly repressed genes that could harbor constitutive heterochromatin marks.  
36  
37 Thus, from heterogeneous starting datasets and based only on three chromatin features, we could  
38  
39 identify consistent gene categories for which average profiles can be automatically drawn (Figure  
40  
41 4B), illustrating general regulatory features of these loci.  
42  
43  
44  
45  
46

47  
48 Second, we aimed to test seqMINER in a study to compare quantitative changes in the  
49  
50 histone H3K4me3 mark in ESCs and brain tissue. In order to compare these two datasets, we  
51  
52 generated a list of reference coordinates enriched in H3K4me3 in ESCs. We used both the  
53  
54 quantitative and the qualitative methods proposed in seqMINER to compare the signal in the two  
55  
56 selected datasets over the reference loci. First, we used the qualitative method to organize groups in  
57  
58 the dataset (Figure 5A). Two major clusters could be identified, the first cluster contains loci equally  
59  
60 enriched in H3K4me3 mark in ES and brain cells, the second cluster contains loci with much higher

1  
2 H3K4me3 signal in ESCs relative to brain (cluster 2 in Figure 5A). This analysis suggests that  
3  
4 cluster 2 comprises only ES specific genes. However, by performing an additional round of  
5  
6 clustering on cluster 2, three sub-clusters were identified (2.1, 2.2 and 2.3), illustrating the  
7  
8 importance of the possibility given by seqMINER to perform iterative rounds of analysis to  
9  
10 organize the data with precision. Analysis of the sub-clusters after this second clustering step  
11  
12 indicates that the seemingly ES specific loci can now be reclassified as weakly, moderately or  
13  
14 strongly enriched in the brain (Figure 5B). The results now suggest for example that cluster 2.1 is  
15  
16 also transcribed in the brain cells analyzed.  
17  
18  
19

20  
21 Second, in order to quantitatively determine the differences between ESC and brain, we  
22  
23 collected enrichments over the input loci (H3K4me3 bound regions) using the quantitative method  
24  
25 implemented in seqMINER (Figure 5B). Then using the visualization module of seqMINER, we  
26  
27 plotted enrichment in ESC versus brain. Using this method, we again observe a heterogeneous  
28  
29 pattern when using the total set of loci (Figure 5C) with a large proportion showing similar  
30  
31 enrichments in ESC and brain tissue (Pearson=0.71), but also a subset of loci that are enriched in  
32  
33 ESCs, but not in brain. Interestingly, when we collect the same data on subsets isolated with the  
34  
35 qualitative method (subset 2 and 2.3) and overlay this information on the plot (Figure 5D, subset 2  
36  
37 in blue and 2.3 in green), we can confirm that these subsets correspond to the loci that change their  
38  
39 status between the two conditions. Additional information is gained by this method since we  
40  
41 observe that most of the loci that are differentially enriched in H3K4me3 between the two cell types  
42  
43 display low enrichment in ESC. Consequently, loci that are highly enriched in the H3K4me3 mark  
44  
45 are invariant in both cells types and probably correspond to housekeeping genes (Figure 5C).  
46  
47  
48  
49

50  
51 These analyses showed how the quantitative and qualitative methods implemented in  
52  
53 seqMINER can be combined to select populations of loci having similar features and quantitatively  
54  
55 follow their behavior in different conditions. Importantly, we show how clusters can be used as  
56  
57 reference coordinates for iterative rounds of analysis to detect potential sub-populations. Finally  
58  
59 through some biological examples, we illustrate the possibilities given by seqMINER to create  
60

1  
2 visual representations of the isolated genomic features.  
3  
4  
5

## 6 **DISCUSSION**

7  
8  
9 Extracting the biological meaning from genomewide studies requires the use of various  
10 methodologies to answer complex biological questions. With seqMINER, we attempted to develop  
11 a standalone analysis platform that allows users to make biological interpretation of their high  
12 throughput sequencing data. We implemented two methods that allow high level correlative  
13 integration of multiple sequencing datasets to identify general as well as specific genomic  
14 characteristics that emerge from the analysed features. We provided a number of visualisation  
15 methods, in order to provide the possibility for rapid assessment of the datasets from different  
16 analysis perspectives (i.e. general patterns, sub-groups, quantitative data comparisons). Thus, we  
17 provide a novel broad bioinformatics resource that should provide analysis solutions in many ChIP-  
18 seq based biological applications. Additionally, we significantly improved the efficiency of broadly  
19 used clustering algorithms by re-implementing them using recent technology developments (i.e.  
20 SMT). seqMINER should prove to be a valuable resource for a broad range of applications in the  
21 bioinformatics community.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 By using Java Virtual Machine (JVM) for code interpretation, we aimed to avoid both  
40 operating system (OS) incompatibility and to limit user efforts for installation making seqMINER  
41 accessible to a broad range of users. However, on most OS, by default, only a small RAM memory  
42 space is attributed to the application, limiting the number of genomic features that can be analysed  
43 simultaneously. In order to overcome this limitation, the user has to manually set the desired RAM  
44 space dedicated for seqMINER analysis. Another implementation choice regarding memory usage  
45 was to load all the necessary data in the RAM prior to analysis to avoid repetitive hard disk access  
46 and speed up the computational tasks. This can be a limitation for seqMINER usage on local  
47 computers that are commonly equipped with 2-4 GB RAM and a 32bit operation system that limits  
48 the maximum RAM of Java virtual machine at about 1.5GB. Thus, to perform multi-file  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 comparison (~10<sup>6</sup> sequencing raw files), the use of seqMINER on a server is recommended.  
3

4 A current limitation in ChIP-seq data inter-comparison comes from technical issues in the  
5 ChIP experiment itself. It appears to be relatively difficult to obtain quantitatively comparable data  
6 due to high variation in immunoprecipitation efficiency that is affected by a multitude of  
7 imponderable factors (antibody specificity and efficiency, natural abundance of the particular  
8 ChIPed feature etc). In its present version, seqMINER generally assumes that all datasets present a  
9 similar enrichment ranges. Particularly, for the clustering procedure, seqMINER equally weighs the  
10 density arrays of each data set. However, if a data set presents significantly higher enrichment  
11 compared to the others, it will likely over-contribute to the clustering and eventually lead to sub-  
12 optimal locus organisation in the final cluster.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 A simple way to overcome this problem is to apply normalisation filters upstream of the  
26 seqMINER analysis to normalise the signal between datasets. Several methods have already been  
27 proposed for normalising high throughput sequencing datasets prior to comparison (i.e. (Rozowsky  
28 et al. 2009; Taslim et al. 2009; Welboren et al. 2009)). However, no real consensus has appeared in  
29 the field to date on whether methods that are suitable for a particular experimental design will be  
30 widely applicable. Thus, it is difficult to propose a robust and broadly accepted normalisation  
31 method to circumvent this problem. Another approach could be to interfere directly at the clustering  
32 stage. One can imagine developing a user supervised method allowing imposition of different  
33 weights to the clusters to correct for experimental variations between data sets. Methods for  
34 correcting experimental bias will likely be developed in the future that will facilitate seqMINER  
35 usage.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51 In the present version of seqMINER, we aimed to address numerous biological questions  
52 commonly raised in classical ChIP-seq analysis pipelines. As users may have specific questions  
53 they wish to ask and as the arrival of future technologies extends the scope of the questions that can  
54 be asked, we designed seqMINER in a flexible and open mode. We implemented seqMINER in an  
55 organized architecture (following Model-View-Controller guidelines) and making the source code  
56  
57  
58  
59  
60



1  
2 available open source (GPL3). For example, support for new rich data formats (ie: SAM-BAM (Li  
3  
4 et al. 2009)) or dedicated methods for integrating ChIP-seq with RNA-seq will be the next  
5  
6  
7 developments needed to improve seqMINER functionality.  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## FUNDING

ARK is recipient of a fellowship from INSERM-Région Alsace and the Association pour la Recherche sur le Cancer (ARC). AMC is a recipient of a fellowship from the French Ministry of Education, Research and Technology (MNERT) and ARC. This work was funded by grants from ANR (GenomATAC; ANR-09-BLAN-0266), the EU (EUTRACC, LSHG-CT-2007-037445), the Institut National du Cancer, the Ligue National contre le cancer. ID is an 'équipe labélisée' of the Ligue National contre le cancer.

## AKNOWLEDGMENTS

We are grateful to Stadler M for technical suggestions, De Dieuleveult M, Legras S, Strub T for useful comments on the software.

## **Figure Legends**

**Figure 1.** Schematic representation of the general workflow of seqMINER. seqMINER takes as an input a single set of reference loci (i.e. gene promoters or binding sites) and multiple raw sequencing datasets. seqMINER can collect tag densities or calculate enrichment values around the set of reference coordinates. Using a combination of automated clustering and manual reordering methods, seqMINER helps the user to create functional groups within the reference set. Each specific sub-group in the dataset can be visualised as heatmaps, dotplots or average plots and the groups of loci can be exported for further analysis.

**Figure 2.** Schematic representation of data collection methods implemented in seqMINER. In both methods, prior to quantification, reads are extended from a user defined value (default 200bp). (A) Maximal number of overlapping reads method: a user defined number of bins are created around the reference coordinate and for each bin the maximal number of overlapping tags is computed. (B) Method summing the total number of tags: the number of tags presents or overlapping a user-defined window (default 2kb) around the reference site are counted.

**Figure 3:** Time to complete a typical analysis with seqMINER. Time required for the different stages of the analysis (namely data loading, generation of distributions and clustering by seqMINER, using raw ChIP-seq datasets of various sizes (191, 554 and 397 MBs) and various number of reference coordinates (8470-33881). The analysis was performed on a PC running under windows with standard performances (CPU-3.0 GHz core-duo; RAM – 4Go).

**Figure 4:** Characterisation of epigenetic profiles of mouse promoters in ESCs using seqMINER.

(A) Read densities of regions surrounding the whole set of TSS (assumed to be the 5'-end of the annotated transcript) of mouse genes from ENSEMBL (v58). TSSs were used as reference coordinates to collect data in publicly available H3K4me3 , H3K36me3 and H3K27me3 datasets. Tag densities from each ChIP-seq dataset were collected within a window of 10 kb around the reference coordinates, the collected data were subjected to k-means clustering. The five major clusters are indicated. (B) Using seqMINER, the average profile for selected clusters was

1  
2 automatically calculated and plotted. The H3K4me3 mean profile for transcripts actively  
3 transcribed on the negative strand (pink), positive (blue) strand and on both strands (green) was  
4 calculated and represented.  
5  
6  
7  
8  
9

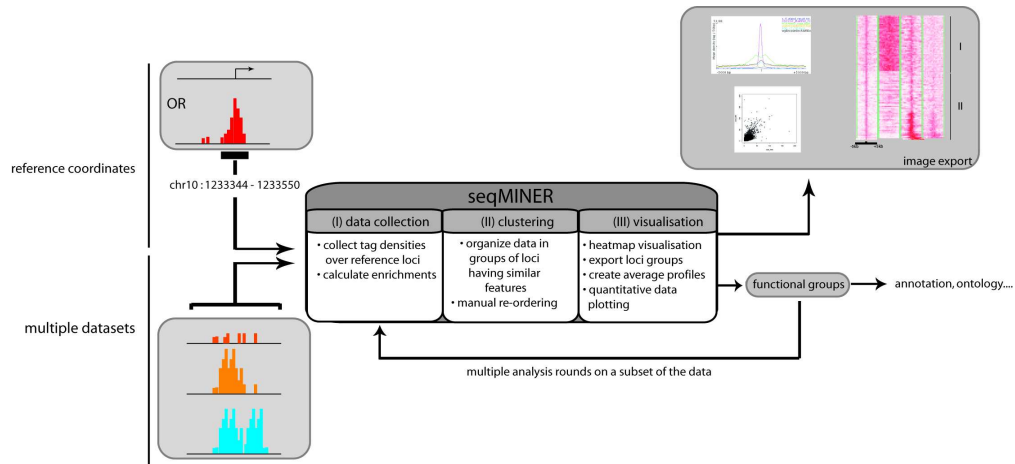
10  
11 **Figure 5:** Quantitative changes of H3K4me3 mark in mouse brain cells relative to mESCs

12 Tag densities of regions surrounding the H3K4me3 enriched loci in ESCs. Publicly available ChIP-  
13 seq datasets for H3K4me3 in ESCs and in brain cells were used in this comparative analysis. (A)  
14 H3K4me3 enriched loci in ESC were detected using MACS software, these loci were used as  
15 reference coordinates. Tag densities from H3K4me3-ESC and H3K4me3-brain datasets were  
16 collected within a window of 10 kb around the reference coordinates, and then the density files  
17 were subjected to k-means clustering. Two major clusters can be isolated: cluster 1 contains loci  
18 with significant and equal enrichment of H3K4me3 in both ESC and brain; cluster 2 contains loci  
19 with higher enrichment of H3K4me3 in ESC relative to brain tissue. (B) As a second step of  
20 analysis, the loci in cluster 2 were used as reference. The densities around these loci were re-  
21 collected and a second round of clustering was performed. After the second round of clustering  
22 three sub-clusters can be isolated; sub cluster 2.1, 2.2 and 2.3 corresponding to loci weakly,  
23 moderately and strongly enriched in H3K4me3 mark in ESC relative to brain respectively. Note that  
24 the bottom of the sub cluster 2.3 that has H3K4me3 enrichment distant from the cluster center was  
25 not considered as a separate entity since we focused our analysis on the differential signal in the  
26 cluster center. (C) Quantification of the changes observed between the two conditions. Dotplot  
27 representing the H3K4me3 enrichment in ESC versus brain. Enrichments were calculated for  
28 H3K4me3-ESC and H3K4me3-brain datasets within a window of 2 kb around the complete set of  
29 reference coordinates (black dots), (D) and against previously isolated subsets of the reference  
30 coordinates (subset 2 in blue and subset 2.3 in green).  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**References:**

- 1  
2  
3  
4  
5  
6  
7 Abeel, T., Y.V. de Peer, and Y. Saeys. 2009. Java-ML: A Machine Learning Library. *Journal of*  
8 *Machine Learning Research* **2009**: 931-934.
- 9 Bernstein, B.E., T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M.  
10 Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, and E.S. Lander. 2006.  
11 A bivalent chromatin structure marks key developmental genes in embryonic stem cells.  
12 *Cell* **125**: 315-326.
- 13 Bilodeau, S., M.H. Kagey, G.M. Frampton, P.B. Rahl, and R.A. Young. 2009. SetDB1 contributes to  
14 repression of genes encoding developmental regulators and maintenance of ES cell state.  
15 *Genes Dev* **23**: 2484-2489.
- 16 Blahnik, K.R., L. Dou, H. O'Geen, T. McPhillips, X. Xu, A.R. Cao, S. Iyengar, C.M. Nicolet, B.  
17 Ludascher, I. Korf, and P.J. Farnham. Sole-Search: an integrated analysis program for peak  
18 detection and functional annotation using ChIP-seq data. *Nucleic Acids Res* **38**: e13.
- 19 Blankenberg, D., J. Taylor, I. Schenck, J. He, Y. Zhang, M. Ghent, N. Veeraraghavan, I. Albert, W.  
20 Miller, K.D. Makova, R.C. Hardison, and A. Nekrutenko. 2007. A framework for  
21 collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly.  
22 *Genome Res* **17**: 960-964.
- 23 Bock, C., K. Halachev, J. Buch, and T. Lengauer. 2009. EpiGRAPH: user-friendly software for  
24 statistical analysis and prediction of (epi)genomic data. *Genome Biol* **10**: R14.
- 25 Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig,  
26 A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, and R.  
27 Jaenisch. 2006. Polycomb complexes repress developmental regulators in murine embryonic  
28 stem cells. *Nature* **441**: 349-353.
- 29 de Hoon, M.J., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software.  
30 *Bioinformatics* **20**: 1453-1454.
- 31 Dennis, G., Jr., B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, and R.A. Lempicki. 2003.  
32 DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**:  
33 P3.
- 34 Heintzman, N.D., G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee,  
35 R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D.  
36 Green, V.V. Lobanenko, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, and B. Ren.  
37 2009. Histone modifications at human enhancers reflect global cell-type-specific gene  
38 expression. *Nature* **459**: 108-112.
- 39 Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van  
40 Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, and B. Ren.  
41 2007. Distinct and predictive chromatin signatures of transcriptional promoters and  
42 enhancers in the human genome. *Nat Genet* **39**: 311-318.
- 43 Hon, G., B. Ren, and W. Wang. 2008. ChromaSig: a probabilistic approach to finding common  
44 chromatin signatures in the human genome. *PLoS Comput Biol* **4**: e1000201.
- 45 Ji, H., H. Jiang, W. Ma, D.S. Johnson, R.M. Myers, and W.H. Wong. 2008. An integrated software  
46 system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**: 1293-1300.
- 47 Johnson, D.S., A. Mortazavi, R.M. Myers, and B. Wold. 2007. Genome-wide mapping of in vivo  
48 protein-DNA interactions. *Science* **316**: 1497-1502.
- 49 Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler.  
50 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.
- 51 Krebs, A., M. Frontini, and L. Tora. 2008. GPAT: retrieval of genomic annotation from large  
52 genomic position datasets. *BMC Bioinformatics* **9**: 533.
- 53 Ku, M., R.P. Koche, E. Rheinbay, E.M. Mendenhall, M. Endoh, T.S. Mikkelsen, A. Presser, C.  
54 Nusbaum, X. Xie, A.S. Chi, M. Adli, S. Kasif, L.M. Ptaszek, C.A. Cowan, E.S. Lander, H.  
55  
56  
57  
58  
59  
60

- 1  
2 Koseki, and B.E. Bernstein. 2008. Genomewide analysis of PRC1 and PRC2 occupancy  
3 identifies two classes of bivalent domains. *PLoS Genet* **4**: e1000242.  
4  
5 Laajala, T.D., S. Raghav, S. Tuomela, R. Lahesmaa, T. Aittokallio, and L.L. Elo. 2009. A practical  
6 comparison of methods for detecting transcription factor binding sites in ChIP-seq  
7 experiments. *BMC Genomics* **10**: 618.  
8  
9 Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E.  
10 Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker,  
11 B. Yuan, G.W. Bell, E. Herbolsheimer, N.M. Hannett, K. Sun, D.T. Odom, A.P. Otte, T.L.  
12 Volkert, D.P. Bartel, D.A. Melton, D.K. Gifford, R. Jaenisch, and R.A. Young. 2006. Control  
13 of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-  
14 313.  
15  
16 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R.  
17 Durbin. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:  
18 2078-2079.  
19  
20 Manske, H.M. and D.P. Kwiatkowski. 2009. LookSeq: a browser-based viewer for deep sequencing  
21 data. *Genome Res* **19**: 2125-2132.  
22  
23 Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W.  
24 Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C.  
25 Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, and B.E.  
26 Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-  
27 committed cells. *Nature* **448**: 553-560.  
28  
29 Rozowsky, J., G. Euskirchen, R.K. Auerbach, Z.D. Zhang, T. Gibson, R. Bjornson, N. Carriero, M.  
30 Snyder, and M.B. Gerstein. 2009. PeakSeq enables systematic scoring of ChIP-seq  
31 experiments relative to controls. *Nat Biotechnol* **27**: 66-75.  
32  
33 Shin, H., T. Liu, A.K. Manrai, and X.S. Liu. 2009. CEAS: cis-regulatory element annotation system.  
34 *Bioinformatics* **25**: 2605-2606.  
35  
36 Skinner, M.E., A.V. Uzilov, L.D. Stein, C.J. Mungall, and I.H. Holmes. 2009. JBrowse: a next-  
37 generation genome browser. *Genome Res* **19**: 1630-1638.  
38  
39 Taslim, C., J. Wu, P. Yan, G. Singer, J. Parvin, T. Huang, S. Lin, and K. Huang. 2009. Comparative  
40 study on ChIP-seq data: normalization and binding pattern characterization. *Bioinformatics*  
41 **25**: 2334-2340.  
42  
43 Wang, Z., C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, W.  
44 Peng, M.Q. Zhang, and K. Zhao. 2008. Combinatorial patterns of histone acetylations and  
45 methylations in the human genome. *Nat Genet* **40**: 897-903.  
46  
47 Welboren, W.J., M.A. van Driel, E.M. Janssen-Megens, S.J. van Heeringen, F.C. Sweep, P.N. Span,  
48 and H.G. Stunnenberg. 2009. ChIP-Seq of ERalpha and RNA polymerase II defines genes  
49 differentially responding to ligands. *Embo J* **28**: 1418-1428.  
50  
51 Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoutte, D.S. Johnson, B.E. Bernstein, C. Nussbaum, R.M.  
52 Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS).  
53 *Genome Biol* **9**: R137.  
54  
55  
56  
57  
58  
59  
60



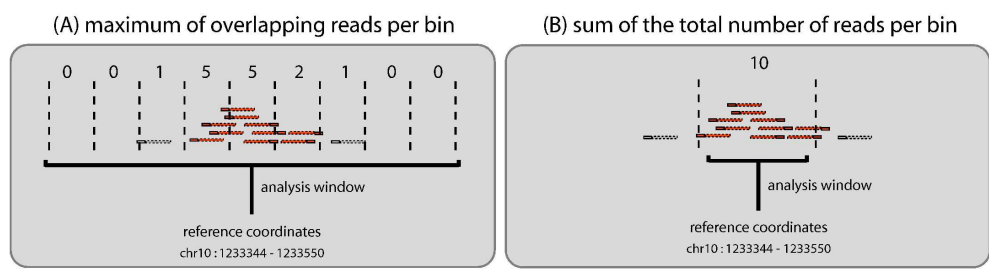
Ye T, et al Figure 1

Figure 1  
188x99mm (300 x 300 DPI)

Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

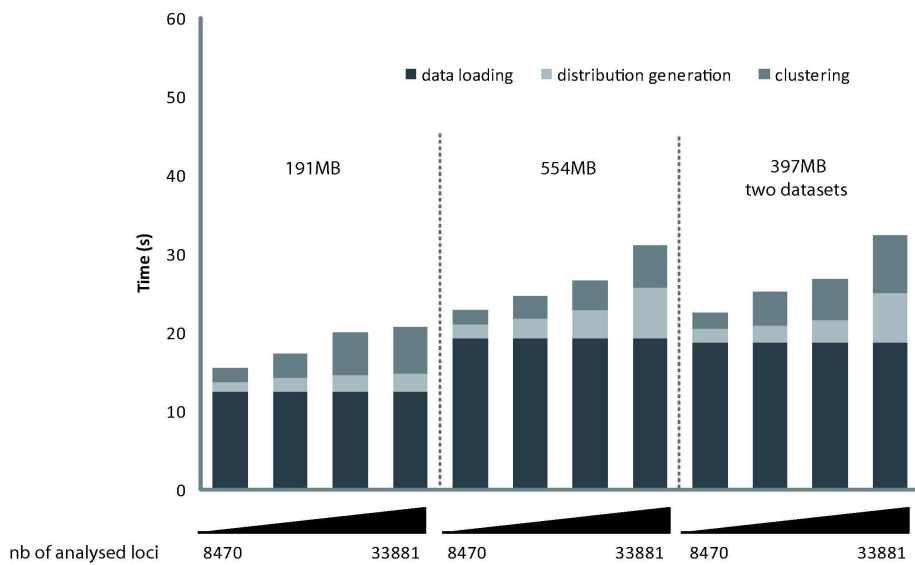


Ye T, et al Figure 2

Figure 2  
185x62mm (600 x 600 DPI)

Peer Review



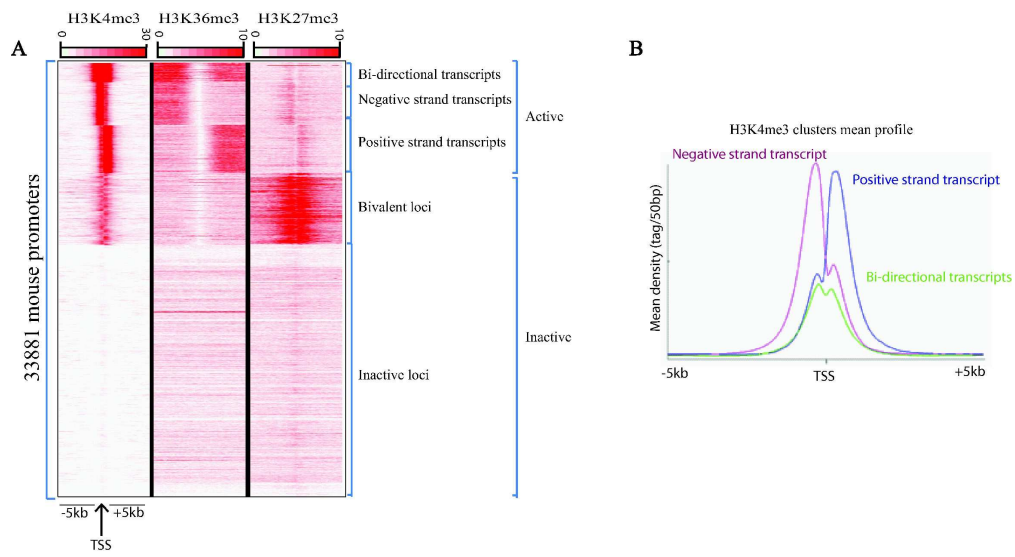


Ye T, et al Figure 3

Figure 3  
109x89mm (600 x 600 DPI)

view

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

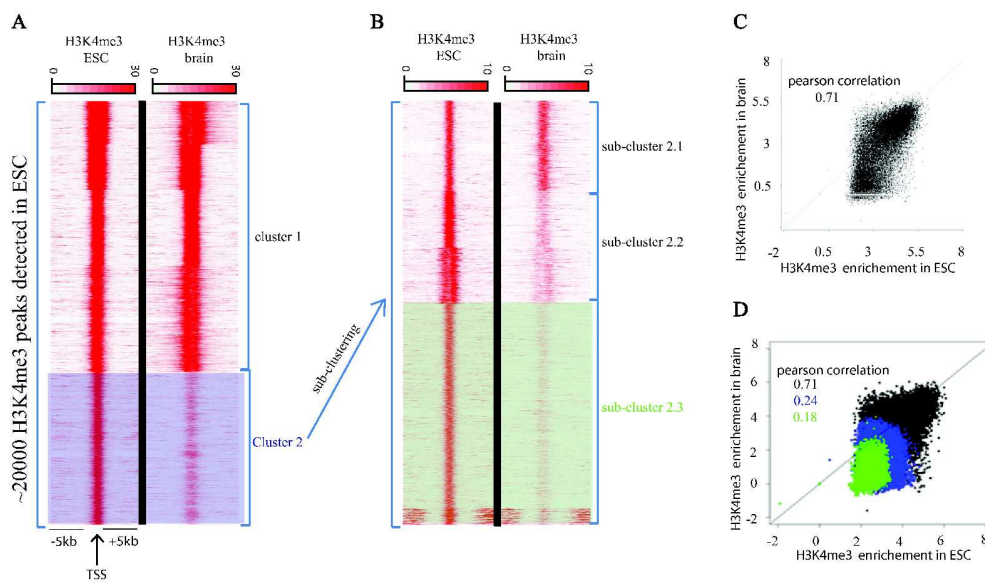


Ye T, et al Figure 4

Figure 4  
197x122mm (600 x 600 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Ye T, et al Figure 5

Figure 5  
203x146mm (600 x 600 DPI)

review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**PARTIE II : ETUDE DE L'ACTION DE  
L'ACIDE RÉTINOÏQUE DANS LA  
DIFFÉRENCIATION NEURONALE DES  
CELLULES SOUCHES EMBRYONNAIRES**

## AVANT-PROPOS

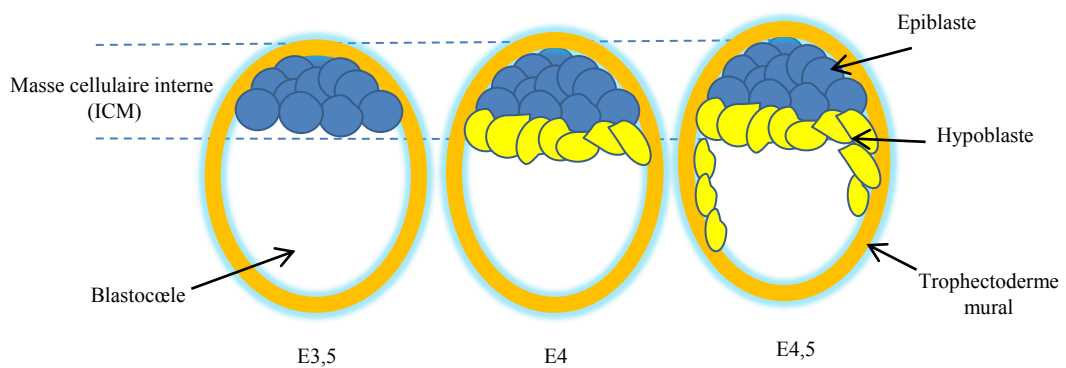
Parallèlement aux études concernant les sous-unités de TFIID et leurs paralogues, l'équipe du Dr. Davidson s'intéresse également au rôle de l'acide rétinoïque (AR) et de ses récepteurs (RARs) dans différents processus physiologiques. Ces deux axes de recherches se rejoignent puisque TAF4 est un cofacteur de RAR (Mengus et al., 1996). Le projet AR/RARs avait pour objectif premier de caractériser le rôle des RARs dans les cellules souches embryonnaires (ES) murines. La cartographie des sites de fixation des RARs dans les ES et les MEFs a permis de démontrer que l'occupation des promoteurs par ces récepteurs était spécifique à chaque lignée cellulaire et que la voie de l'AR exerce des fonctions différentes selon le contexte physiologique. Ces travaux ont fait l'objet d'une publication récente (Delacroix et al., 2010).

Connaissant le rôle de l'AR dans le développement embryonnaire et son implication dans la différenciation neuronale *in vivo*, nous avons orienté nos travaux de recherches sur la différenciation neuronale des cellules ES *in vitro*, où l'AR joue un rôle clé (Bibel et al., 2004, 2007).

Lors de ma dernière année de thèse, j'ai repris ce projet de recherche que Laurence Delacroix avait initié. Avant de quitter le laboratoire, Laurence a optimisé et adapté le protocole de différenciation neuronale dépendante de l'AR des cellules ES *in vitro*. J'ai utilisé la technique de séquençage à haut débit des ARNs (RNA-seq) pour identifier les gènes dont l'expression est modifiée par le traitement par l'AR durant la différenciation neuronale. Nous nous sommes particulièrement intéressés aux facteurs de transcription dont l'expression est induite par l'AR lors de ce processus, notamment HES3 (*Hairy and Enhancer of Split 3*). Afin de caractériser le rôle de ce facteur dans la différenciation neuronale, nous avons utilisé la technique de shRNA (*short hairpin RNA*) pour réduire son expression. Les résultats de cette expérience montrent que HES3 est nécessaire à l'expression des marqueurs neuronaux tels que BRN2 (POU3F2) et MAPt (*Microtubule Associated Protein tau*).

Cette partie du manuscrit comportera une introduction succincte de la biologie des cellules ES, de la différenciation neuronale, des propriétés des RARs et des facteurs HES. Les résultats préliminaires seront ensuite discutés pour clore ce manuscrit.

## INTRODUCTION



**Figure 1:** Les lignages cellulaires du blastocyste

Au 4<sup>ème</sup> jour du développement, la masse cellulaire interne contient 20 à 40 cellules. Les cellules qui sont en contact avec le liquide du blastocœle acquièrent un phénotype migratoire. Le jour suivant, elles se déplacent en utilisant la face interne du trophoctoderme mural comme support. Ces cellules vont former l'endoderme primitif. Les cellules situées au niveau du trophoctoderme polaire forment l'épiblaste. Les cellules ES sont isolées en culture à partir de l'ICM au jour 3,5.

Adapté du "livre Le développement d'un mammifère : la souris" de Thierry Darribère

## I. Les cellules souches embryonnaires pluripotentes

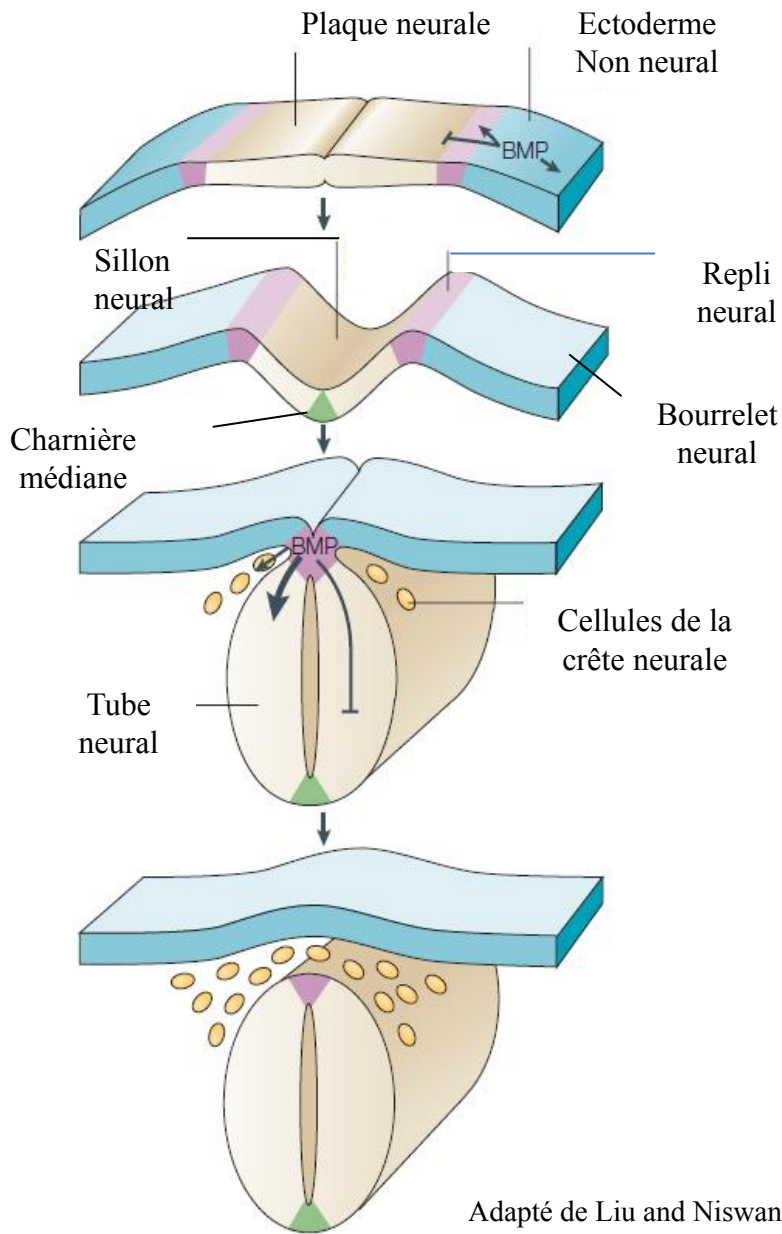
Les cellules souches embryonnaires murines sont isolées directement en culture à partir de la masse cellulaire interne (ICM : *Inner Cell Mass*) du blastocyste âgé de 3,5 jours. Au moment de l'implantation du blastocyste, les cellules de l'ICM se divisent en deux groupes pour former deux feuillets primitifs, l'épiblaste (ectoderme primitif) qui va donner naissance au futur embryon et l'hypoblaste qui contribuera à la formation des structures extra-embryonnaires. Les cellules de l'ICM sont donc à l'origine des tous les organes et tissus de l'embryon, elles sont dites pluripotentes (*Cf. Figure 1*).

Les cellules ES sont capables de s'autorenouveler à l'infini sans perdre leur pluripotence. *In vitro*, elles peuvent être différenciées en plusieurs types cellulaires spécialisés tels que les myoblastes, les ostéoblastes ou encore les neurones. Les protocoles de différenciation *in vitro* présentent un outil précieux pour étudier les mécanismes moléculaires impliqués dans la différenciation et la spécialisation des différents types cellulaires lors du développement embryonnaire.

La pluripotence des cellules ES dépend de l'activation de la voie de signalisation STAT et des facteurs de transcription NANOG et OCT3/4 (POU5F1). La voie de signalisation STAT, activée par l'interaction de différents signaux extracellulaires tels les cytokines IL6 (*Interleukine 6*) et LIF (*Leukemia Inhibitory Factor*) avec un récepteur commun GP130, conduit à l'activation du cycle cellulaire et au maintien de l'état indifférencié. L'inactivation de *Stat3* ou de *Gp130* est létale au stade précoce du développement embryonnaire (Takeda et al., 1997). *In vitro*, l'action du LIF est suffisante pour maintenir les cellules ES dans un état indifférencié. Les cellules ES cultivées sans LIF se différencient spontanément (pour revue, Cazillis et al., 2005).

Les facteurs de transcription OCT4 et NANOG jouent un rôle central dans le contrôle de la différenciation des cellules ES. Ces deux facteurs sont exprimés dans les cellules de l'ICM et de l'épiblaste, puis disparaissent au cours du développement précoce. *In vitro*, la réduction du taux d'expression de NANOG induit la différenciation des cellules ES, alors que sa surexpression permet leur maintien dans un état indifférencié même en absence du LIF (Hatano et al, 2005 ; Niwa et al., 2001). Chez la souris, l'inactivation du gène *Pou5f1*





**Figure 2 :** La formation du tube neural dans la région dorsale de l’embryon

Après 7,5 jours de gestation, l’épiblaste s’épaissit pour former la plaque neurale. Il se forme un sillon puis une charnière : la charnière médiane. La plaque neurale prend alors la forme d’un V. La charnière se referme, les bords des bourrelets neuraux se rejoignent dans la région médiane. A 8,5 jours de gestation, les bourrelets fusionnent le long de la ligne médiane de l’embryon. Le tube neural s’isole progressivement. Les cellules de la crête neurale migrent latéralement le long du tube.

empêche l'établissement d'une population cellulaire pluripotente et provoque l'arrêt du développement (Nichols et al., 1998).

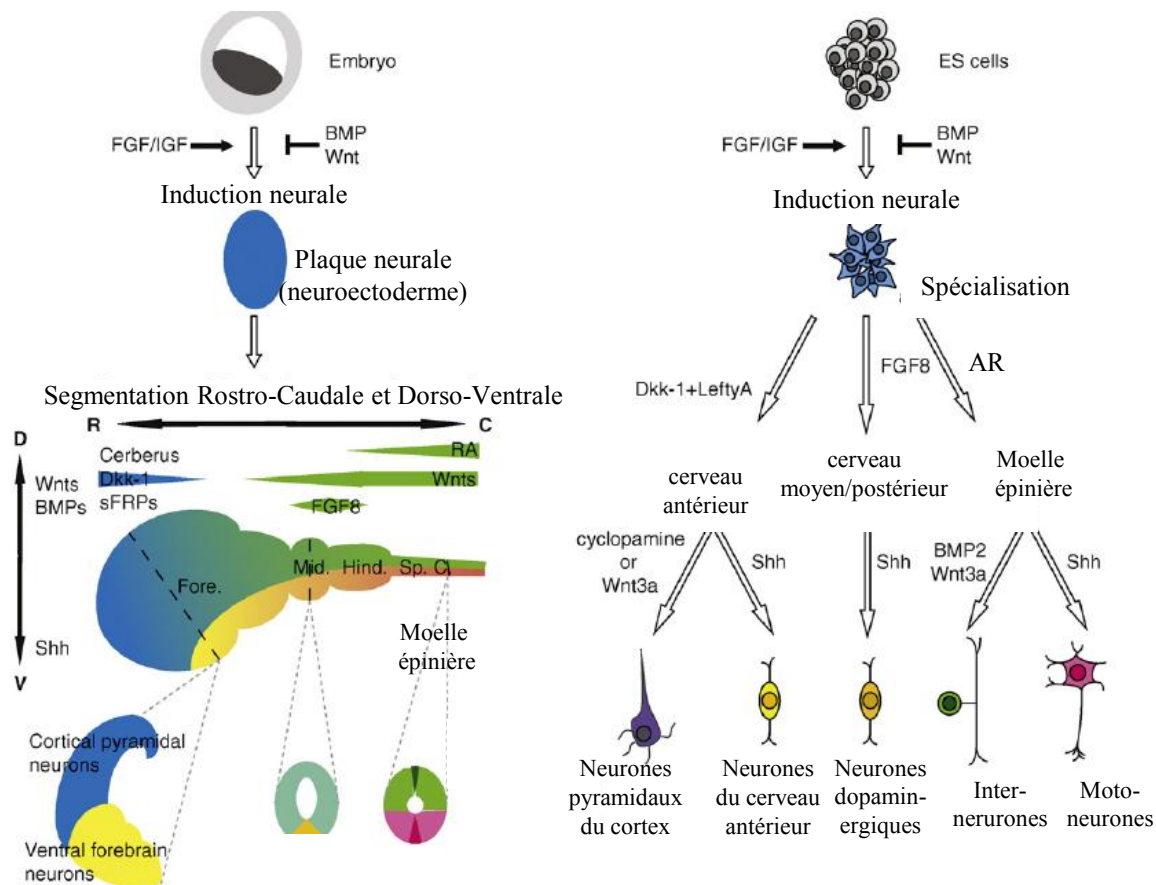
### **II. La neurulation chez l'embryon de souris**

L'induction neuronale chez l'embryon commence par la différenciation de l'épiblaste en neuroectoderme. Ce tissu prend la forme d'un tube creux : le tube neural, il constitue le progéniteur du système nerveux central (CNS). Parallèlement, dans la région dorsale du tube neural, des cellules se multiplient et migrent pour former les crêtes neurales, à l'origine du système nerveux périphérique (pour revue, Liu and Lee, 2005). (Cf. Figure 2)

Les facteurs BMP (*Bone Morphogenic Proteins*) jouent un rôle central dans la détermination du devenir des cellules épiblastiques lors du développement embryonnaire. BMP4 induit la différenciation épidermique de l'épiblaste, alors que tout événement inhibant l'action de ce facteur, induit la différenciation neuroectodermique. BMP4 contrôle donc la différenciation neuronale précoce selon un modèle par défaut. Chez les amphibiens, la voie de signalisation BMP est inhibée par les protéines Noggin, Follistatin et Chordin qui agissent comme des antagonistes des facteurs BMP (pour revues, Gaspar et al, 2010 ; Liu and Niswander, 2005). L'expression ectopique de ces trois inhibiteurs dans l'embryon provoque la formation d'un tube neural secondaire. La suite du processus de différenciation neuronale qui va conduire à la ségrégation des différents types neuronaux est contrôlée par d'autres voies de signalisation, telles que les voies FGF (*Fibroblast Growth Factor*), Wnt, SHH (*Sonic HedgeHog*) et celle de l'AR. La figure 3 illustre les différents destins des cellules proneuronales ainsi que les voies de signalisations impliquées.

### **III. La voie de signalisation de l'acide rétinoïque**

L'acide rétinoïque est un dérivé du rétinol (Vitamine A). Il joue des rôles multiples lors du développement embryonnaire et dans l'homéostasie de plusieurs tissus adultes. En effet, l'AR est impliqué dans la définition des axes embryonnaires, la régionalisation du système nerveux central ou encore le contrôle de la neurogenèse (pour revue, Niederreither and Dollé, 2008). L'AR exerce ses fonctions en modulant l'expression de gènes cibles via ses récepteurs RARs (*Retinoic Acid Receptors*) et RXRs (*Retinoid X Receptors*), deux facteurs de transcription de la famille des récepteurs nucléaires (NR).



Adapté de Gaspar et al., 2010

**Figure 3 :** Différenciation des cellules proneurales issues des cellules ES en différents types de cellules neuronales

L'action opposée de SHH et BMP sur les cellules proneurales conduit à la production de cellules précurseurs du plancher et de cellules de la crête neurale. Les gradients de concentration des différents morphogènes définissent l'identité neuronale des précurseurs. La localisation embryonnaire correspondant à l'identité acquise par les cellules ainsi que la voie de signalisation impliquée sont indiquées.

### III.1. Le cycle de l'acide rétinoïque

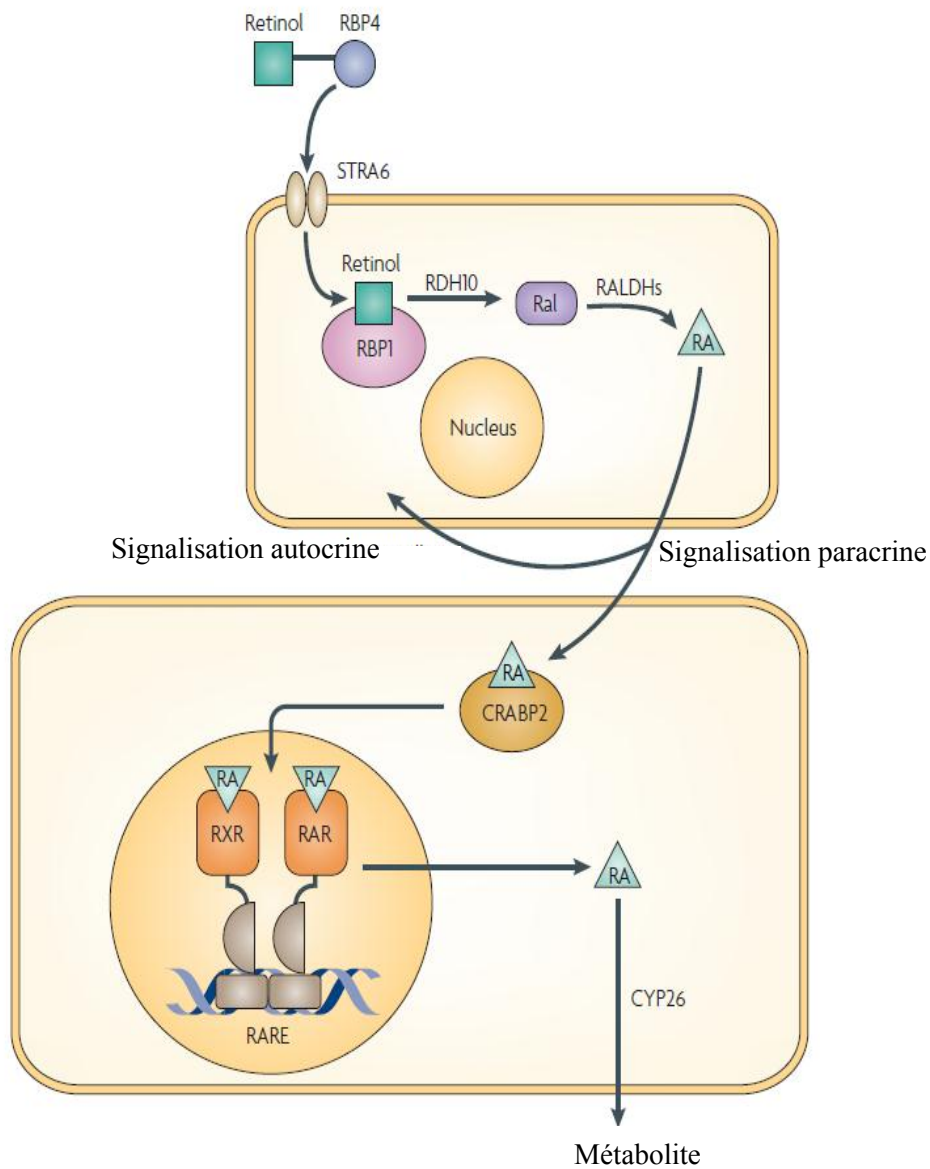
Le rétinol est importé dans la cellule grâce à l'action conjointe de la protéine RBP4 (*Retinol-Binding Protein 4*) et du récepteur membranaire STRA6 (*Stimulated by Retinoic Acid gene 6*). Une fois dans le cytoplasme, deux étapes majeures ont lieu : dans un premier temps, le rétinol est transformé en rétinaldéhyde (Ral) par RDH10 (*Rétinol Déshydrogénase 10*) ou par l'alcool déshydrogénase 7 (ADH7). Ensuite, le rétinaldéhyde est transformé en AR par une rétinaldéhyde déshydrogénase (RALDH<sub>1-3</sub>). L'import de l'AR dans le noyau se fait par l'intermédiaire de la protéine CRABP2 (*Cellular Retinoic-Acid Binding Protein*) (Maden, 2007). Cf. Figure 4

La vitamine A, et par conséquent son dérivé actif l'AR, sont indispensables au développement pré- et post-natal ainsi qu'à la vie adulte, en assurant des fonctions dans la croissance, la survie ou encore la vision (Mark et al., 2006). Au cours de l'embryogenèse, l'AR régule la mise en place de l'axe antéro-postérieur, agit sur la croissance et la mise en place du cerveau postérieur, est impliqué dans le développement des membres, mais peut également réguler la morphogenèse et la différenciation d'organes tels que le cœur, les reins ou le pancréas (Niederreither and Dollé, 2008). L'importance de la voie de l'AR au cours du développement a été mise en évidence par différentes expériences de génétique chez la souris. La délétion des gènes codant pour les récepteurs de l'acide rétinoïque par recombinaison homologue ou encore la délétion de certains enzymes de la voie de synthèse de l'AR affectent les structures tissulaires ainsi que les organes de l'embryon.

### III.2. Les récepteurs de l'acide rétinoïque

Les RARs font partie de la grande famille des récepteurs nucléaires (NRs) qui comprend également les récepteurs aux hormones stéroïdiennes, à la vitamine D3 et aux hormones thyroïdiennes. Les NRs peuvent être divisés en deux sous-types selon qu'ils homodimerisent et fixent des éléments palindromiques (récepteurs stéroïdiens), ou qu'ils hétérodimerisent avec le RXR pour fixer des éléments en répétition directes. La famille comprend également des récepteurs dits 'orphelins' pour lequel aucun ligand n'est décrit.

Les RARs appartiennent au second sous-type et s'hétérodimerisent avec les RXRs. Il existe deux familles de récepteurs de l'AR, les RARs et les RXRs. La famille des RARs comporte RAR $\alpha$  (NR1B1), RAR $\beta$  (NR1B2) et RAR $\gamma$  (NR1B3) alors que la famille des



**Figure 4 :** Le cycle de l'acide rétinoïque

Le rétinol est lié par RBP4 qui, en interagissant avec le récepteur membranaire STRA6 permet l'import du rétinol dans le cytoplasme. Une cascade enzymatique permet la synthèse de l'AR, qui peut agir de manière autocrine ou paracrine. L'AR est ensuite importé dans le noyau par CRABP2. En se liant à ses récepteurs, l'AR module l'expression des gènes cibles. L'AR est ensuite catabolisé par l'enzyme CYP26.

RXR $\alpha$  (NR2B1), RXR $\beta$  (NR2B2) et RXR $\gamma$  (NR2B3). Les RARs se caractérisent par une forte affinité pour l'AR tout-trans et son isomère 9-cis alors que les RXRs ont une forte affinité pour l'isomère 9-cis et une faible affinité pour l'AR tout-trans. Pour chaque récepteur, il existe plusieurs isoformes qui résultent de l'utilisation de promoteurs différents et/ou d'un épissage alternatif (Germain et al., 2006a et b)

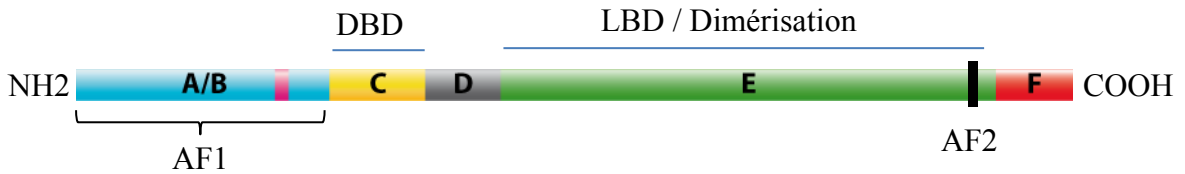
L'expression des RARs et RXRs est régulée de façon spatio-temporelle dans l'organisme au cours du développement (pour revue, Dolle, 2009). En fonction de l'étape du développement ou du tissu cellulaire considéré, un hétérodimère RAR/RXR spécifique va être principalement actif et responsable de la régulation d'un ensemble de gènes. Alors que les RARs n'ont que les RXRs pour partenaire d'hétérodimérisation, les RXRs peuvent soit s'homodimériser, soit s'hétérodimériser avec de nombreux autres récepteurs nucléaires tels que VDR, TR, PPAR, LXR, FXR, CAR, PXR, NGFIB et NURR1 (Germain et al., 2006b).

### III.2.1 Organisation structurale des récepteurs de l'acide rétinoïque

Les récepteurs de l'AR adoptent une organisation structurale commune à la majorité des récepteurs nucléaires. Cette organisation consiste en six régions, nommées de A à F. Les régions C et E contiennent respectivement le DBD (*DNA Binding Domain*) et le LBD (*Ligand Binding Domain*) (Germain et al., 2006a). Cf. *Figure 5*

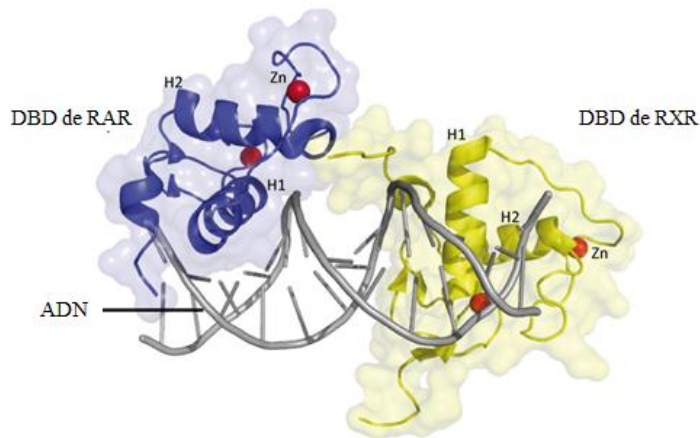
### III.2.2. Le domaine de liaison à l'ADN : DBD

Le DBD des RARs est un domaine globulaire composé de deux motifs en doigt de zinc reliés par une boucle et suivi d'une extension C-terminale (Zechel et al., 1994a; Zechel et al., 1994b) (Cf. *figure 6*). Le DBD permet l'interaction avec l'ADN au niveau des éléments de réponse à l'AR (RARE : *Retinoic Acid Response Element*). Les RAREs sont constitués par deux répétitions directes d'un hexamère (hémisites) séparées par 1,2 ou 5 nucléotides (DR1, 2 et 5). Les RAREs sont reconnus par un hétérodimère RAR/RXR (Kliwer et al., 1992). Chaque partenaire établit des contacts avec un des deux hémisites composant le RARE (Rastinejad et al., 2000). *In vitro*, RAR et RXR se lient avec la plus grande affinité au niveau de RAREs composés de deux répétitions directes (DR) d'un motif consensus 5'-(A/G)G(G/T)TCA-3' séparés par 5 nucléotides (DR5) (Leid et al., 1992 ; Umesono et al., 1991). Cependant, la cartographie des sites de liaison de RAR $\alpha$  et RAR $\gamma$  dans les MEFs et les ES par CHIP on chip, montre que la majorité de ces sites sont dépourvus de DR canoniques



**Figure 5 :** Organisation structurale des récepteurs nucléaires.

Les différentes régions (A/B, C, D, E et F) sont représentées. La région C correspond au DBD. La région D/E contient le LBD et le domaine de dimérisation. Ces deux régions sont conservées dans les NRs. Les régions A/B, D et F, sont variables en séquence et en taille. La plupart des NRs possèdent deux fonctions de transactivation : AF-1 dans la région N-terminale qui est ligand-indépendante ; AF-2, dans la partie C-terminale du domaine E (en noir), qui est régulée par le ligand.



Adapté Rastinejad et al., 2000

**Figure 6 :** Structure tridimensionnelle du DBD

Organisation modulaire et structure tridimensionnelle des DBD de RAR et de RXR liés à un élément de réponse de type DR1. Zn : Atome de Zinc.

(Delacroix et al., 2010). Cela suggère l'existence d'autres mécanismes d'interaction entre les récepteurs à l'AR et les promoteurs des gènes cibles de ce dernier.

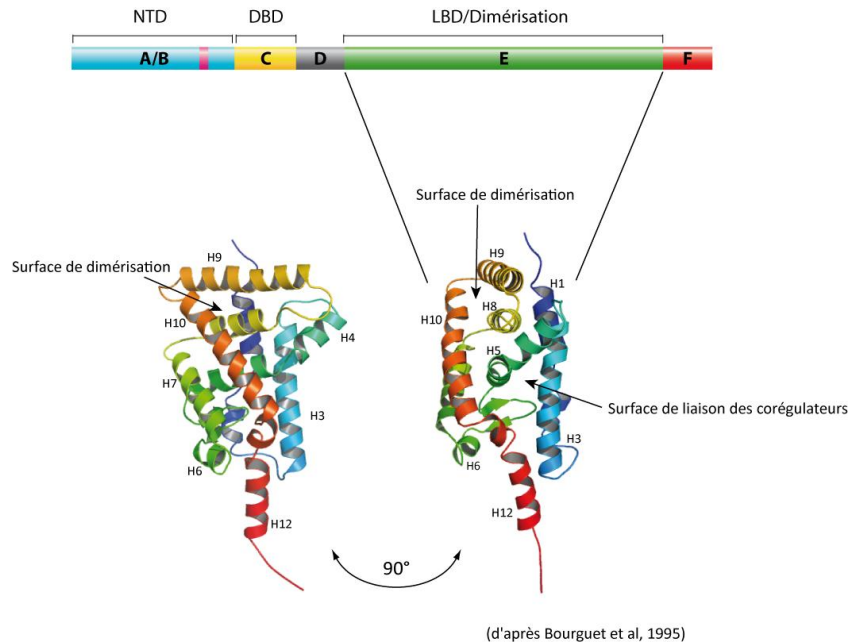
### III.2.3. Le domaine de liaison du ligand : LBD

Le LBD comporte quatre sites fonctionnels nécessaires à l'activité des RARs : 1) la surface de dimérisation avec le LBD du partenaire ; 2) les sites de liaison pour les corépresseurs ; 3) la poche de liaison du ligand ou LBP (*Ligand Binding Pocket*) et 4) l'hélice H12 qui renferme la fonction de transactivation dépendante du ligand appelée (AF2). Une fois le ligand fixé, l'hélice H12 vient refermer la poche du ligand et former une nouvelle surface d'interaction pour les coactivateurs.

Présent chez tous les NRs, le LBD est localisé dans la partie C-terminale. Il est fortement conservé tant au niveau de la séquence (Laudet et al., 1992) qu'au niveau de la structure tridimensionnelle (Wurtz et al., 1996). Il est constitué de 12 hélices  $\alpha$ , notées de H1 à H12 (sauf dans le cas de RAR $\gamma$ , où l'hélice H2 est absente) et d'un feuillet  $\beta$  court à deux brins  $\beta$ , le tout formant un domaine globulaire. L'ensemble des hélices adopte une forme de "sandwich" composé de trois feuillets d'hélices antiparallèles. (*Cf. Figure 7*)

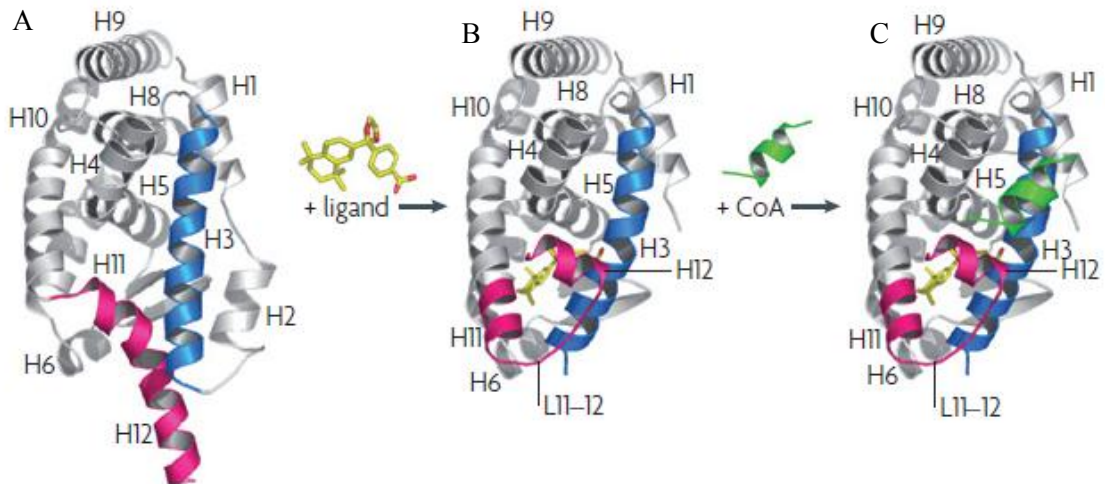
L'hélice H12 est un composant crucial du LBD puisqu'elle contrôle l'activité transcriptionnelle dépendante du ligand des RARs (AF2). La comparaison des structures cristallographiques de RXR $\alpha$  avec et sans ligand a permis d'établir la mécanique du fonctionnement de l'hélice H12 (Bourguet et al., 1995 ; Renaud et al., 1995). En absence de ligand, l'hélice H12 est exposée au solvant, en position étendue vers le bas du LBD (le haut du DBD étant au niveau de H9). Cette surface sert d'ancrage aux complexes corépresseurs en absence de ligand. Après la liaison du ligand, une série d'interactions intramoléculaires entraîne le repositionnement de l'hélice H11 en continuité de l'hélice H10 ainsi que le repliement de l'hélice H12, refermant la LBP pour piéger le ligand (Moras and Gronemeyer, 1998). Une fois l'hélice H12 positionnée, des changements structuraux au niveau des hélices H3 et H4 et de la boucle H3-H4 induisent la formation d'une nouvelle surface d'interaction pour le recrutement de coactivateurs transcriptionnels (pour revue, Lera et al., 2007). *Cf. Figure 8*





**Figure 7 :** Structure tridimensionnelle du LBD

Le LBD de RXR est composé de 12 hélices  $\alpha$  formant un sandwich de trois feuillets d'hélices antiparallèles. Les deux vues montrent les surfaces d'interaction avec le partenaire d'hétérodimérisation et avec les corégulateurs, ainsi que la position exposée de l'hélice H12.



Adapté de Lera et al., 2007

**Figure 8 :** Changements conformationnels du LBD du RXR induits par la liaison du ligand

A : RXR libre, l'hélice H12 est en position étendue

B : RXR lié par le ligand (HoloRXR). L'hélice H12 bascule vers le corps de RXR et ferme la poche qui contient le ligand.

C : Dans la conformation adoptée par l'holoRXR, les hélices H3, H4, H5 et H12 constituent une surface hydrophobe reconnue par le motif LXXLL des coactivateurs (schématisé en vert).

### III.2.4. La fonction d'activation AF1

Elle est située dans la région A/B et est peu conservée entre les membres de la superfamille des NR, voire absente dans certains cas. Sa structure tridimensionnelle est intrinsèquement désordonnée ce qui rend son élucidation quasiment impossible (Combet et al., 2000). Elle est capable d'activer la transcription indépendamment du ligand dans le contexte d'un récepteur dont le LBD est invalidé (Bocquel et al., 1989 ; Tora et al., 1989). Dans le contexte d'un récepteur complet, elle peut néanmoins être dépendante de la fixation du ligand dans le LBD (Kumar et Litwack, 2009).

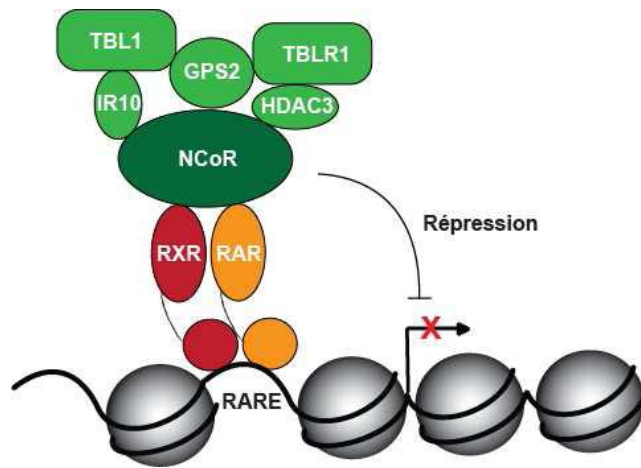
### **III.3. Modèle classique de l'activation des gènes cibles de l'AR**

Dans le modèle classique, en absence de ligand, les NR non-stéroïdiens comme les RARs sont localisés dans le noyau, liés à l'ADN au niveau de leurs éléments de réponse. Ils interagissent avec des complexes corépresseurs contenant des HDACs. La liaison du ligand entraîne un changement conformationnel du LBD qui entraîne le relargage des complexes répresseurs et la liaison de complexes coactivateurs contenant des HATs. Ce modèle est remis en question par des études récentes qui montrent que les RARs peuvent être trouvés dissociés de l'ADN (Bruck et al., 2009) voire dans le cytoplasme ou en contact avec la membrane cellulaire où ils auraient différents effets non-génomiques (pour revue Rochette-Egly et Germain, 2009).

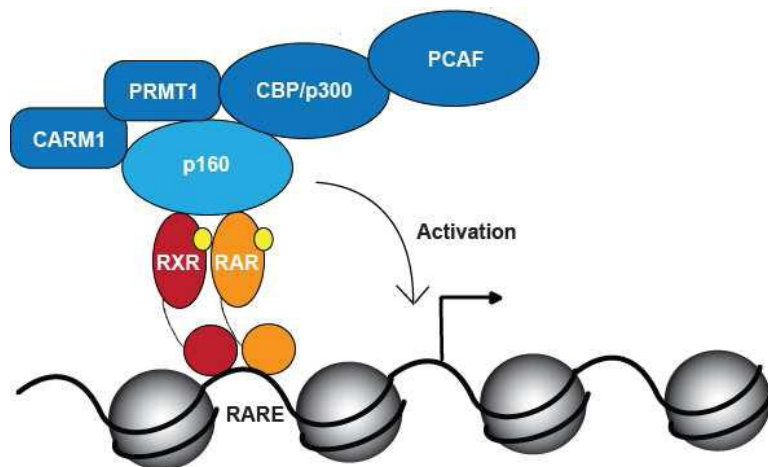
#### III.3.1. Les corépresseurs des récepteurs à l'AR

Les corépresseurs des NR se caractérisent par la présence d'un motif spécifique nommé *CoRNRbox* (*CoRepressor Nuclear Receptor box*). Il est constitué de la séquence d'acides aminés LXX I/H IXXX I/L (L : leucine ; I : isoleucine, H : histidine et X un acide aminé quelconque) (Perissi et al., 1999). Ce motif est reconnu par les NR en absence de ligand.

Les corépresseurs historiques des récepteurs nucléaires sont les protéines NCoR (*Nuclear receptor CoRepressor*) et SMRT (*Silencing Mediator of Retinoic and Thyroid receptors*) (Horlein et al., 1995 ; Chen and Evans, 1995). Ces deux protéines présentent une forte homologie de séquence (Privalsky, 2004) et des fonctions similaires de répression de la transcription.



**Figure 9 :** Complexe répresseur associé à un dimère RAR/RXR en absence de ligand.



**Figure 10 :** Complexes activateur associés à un dimère RAR/RXR en présence de ligand

SMRT et NCoR ne possèdent pas de fonctions enzymatiques propres, mais peuvent, par l'intermédiaire de leurs domaines SANT (Swi3/Ada2/NCoR/TFIIID) et RD (Repression Domain) recruter des complexes multiprotéiques contenant des activités enzymatiques. Parmi les partenaires de SMRT et NCoR se trouvent HDAC3 (Histone DeAcetylase -3), TBL1 (Transducin Beta-Like protein 1), TBLR1 (Transducin Beta-Like Related protein 1) et GPS2 (G protein Pathway Suppressor 2) (Guenther et al., 2000 ; Li et al., 2000 ; pour revue Perissi et al., 2010). *Cf. Figure 9*

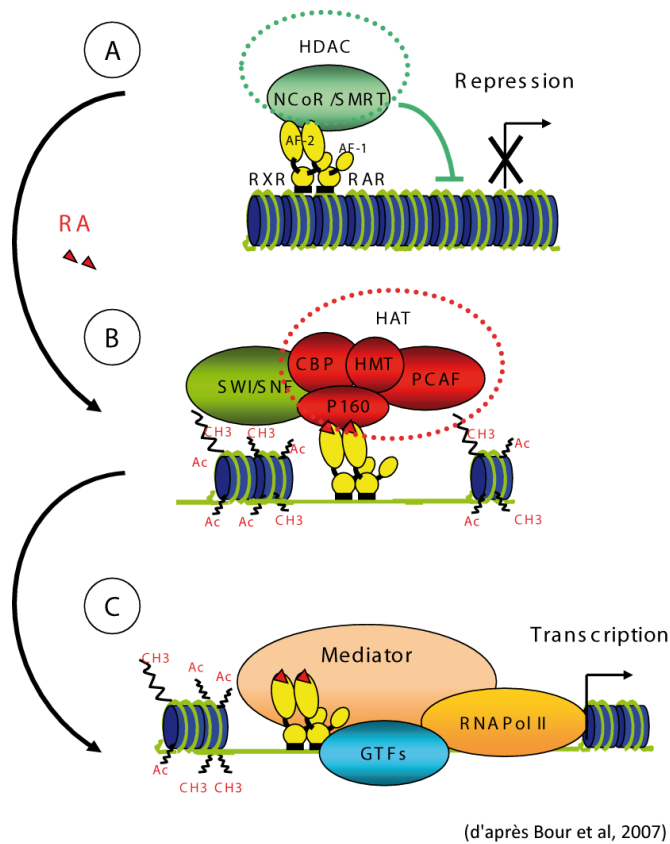
### III.3.2. Les Coactivateurs des récepteurs à l'AR

SRC1 fut l'un des premiers coactivateurs des NR à être découvert (Onate et al., 1995). Il s'est avéré par la suite que ce facteur fait partie d'une famille de protéines d'environ 160 kDa (la famille p160) comprenant plusieurs membres de séquences proches : NCOA1 (SRC1), NCOA2 (TIF2 GRIP1, SRC2) et NCOA3 (SRC3, pCIP, ACTR, AIB1, TRAM1, RAC3) (Hong et al., 1997 ; Torchia et al., 1997 ; Voegel et al., 1996 ; Yao et al., 1996). Les protéines de la famille p160 interagissent avec les NR uniquement lorsqu'ils sont liés à leur ligand. La conformation du LBD associé au ligand permet de recruter ces protéines par l'intermédiaire de leur domaine NRbox (Nuclear Receptor box) de séquence LXXLL (Le Douarin et al., 1996; Heery et al., 1997) qui constitue une hélice amphipathique capable de s'arrimer au niveau du sillon hydrophobe formé par les hélices 3, 4, 5 et 12 du LBD (Darimont et al., 1998) *Cf. Figure 10*

Les coactivateurs de la famille p160 fonctionnent en partie grâce à leur domaine à activité HAT (Chen et al., 1997 ; Spencer et al., 1997). Ils sont aussi capables d'interagir avec d'autres protéines, d'une part les coactivateurs de la famille p300 qui possèdent eux aussi une activité HAT et sont capables d'interagir avec la machinerie de transcription de base (Chan et Lathangue, 2001), d'autre part des coactivateurs à activité HMT tels que CARM1 et PRMT1 (Chen et al., 1999), mais aussi des protéines à activité kinase, ubiquitine-ligase ou ATPase (Perissi et Rosenfeld, 2005 ; Rosenfeld et al., 2006).

### III.3.3. Modèle séquentiel d'activation des gènes cibles de l'AR

L'implication de ces différentes classes de protéines suggère un recrutement interdépendant et ordonné en plusieurs étapes dont la dissection a été réalisée par l'analyse d'expériences de transcription *in vitro* (Dillworth et Chambon, 2001) et de CHIP (Shang et al., 2000 ; Reid et al., 2003 ; Métivier et al., 2003 ; pour revue Perissi et Rosenfeld, 2005). Des



(d'après Bour et al, 2007)

**Figure 11** : Modèle séquentiel de l'activation de la transcription des gènes cibles de l'AR

A) En absence de ligand, l'hétérodimère RAR/RXR est présent au niveau de l'ADN associé au complexe corépresseur, réprimant ainsi l'expression des gènes cibles.

B) La liaison du ligand (AR) induit la dissociation des corépresseurs et le recrutement des complexes coactivateurs.

C) Une fois la chromatine décompactée, le médiateur, les GTFs et la Pol II sont recrutés permettant l'initiation de la transcription.

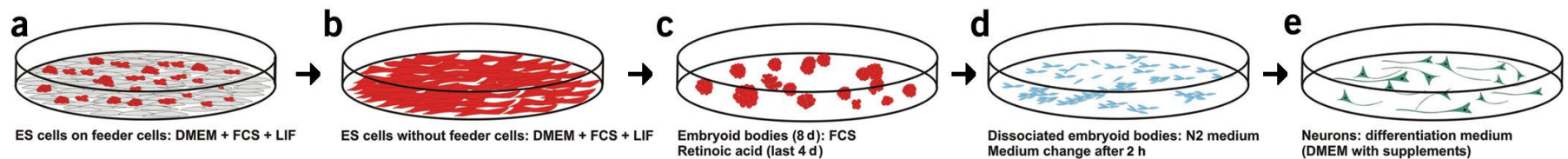
cycles de recrutement et dégradation des différents acteurs ont lieu durant l'activation de l'expression des gènes cibles, cycles dont le rythme est régulé de manière prépondérante par l'ubiquitination des différentes protéines impliquées. L'activation par des NRs est également accompagnée par des cycles de méthylation et déméthylation des dinucléotides CpGs présents dans les promoteurs des gènes cibles. Metivier (2008) et X et Ried 2008. *Cf. Figure 11*

#### **IV. La différenciation neuronale dépendante de l'AR *in vitro***

Avant de parler de la différenciation neuronale dépendante de l'AR des cellules ES, il est important de noter que ces cellules cultivées dans un milieu minimum sans influence extérieure (sans LIF et sans cellules nourricières), expriment spontanément des marqueurs neuronaux tels Sox1 et Nestine (Smukler et al., 2006). Cela conforte l'idée selon laquelle, l'induction neurale se fait par défaut. Cependant, cette différenciation spontanée n'aboutit pas à une population de neurones mais donne plusieurs types cellulaires. Cette différenciation anarchique peut être basculée vers une différenciation neuronale en bloquant les voies BMP, TGF- $\beta$  ou WNT (Chambers et al., 2009 ; Pera et al., 2004 ; Vallier et al., 2004; Aubert et al., 2002 ; Fuentealba et al., 2007). En revanche, l'activation des voies de signalisation FGF/IGF et celle de l'AR stimule la différenciation neuronale (Ying et al., 2003; Bibel et al., 2004). Ces différentes études permettent d'établir un modèle expliquant l'action des différentes voies de signalisation dans le processus de différenciation neuronale lors du développement embryonnaires et/ou *in vitro* (*Cf. Figure 3*).

De nombreux protocoles existent pour la différenciation neuronale des cellules ES *in vitro*. Tous ces protocoles sont basés sur la manipulation des voies de signalisation citées plus haut. Le protocole basé sur l'induction neuronale par l'AR, développé par Bibel et al (Bibel et al., 2004 et 2007), permet l'obtention d'une population pure et homogène de neurones. De plus, ce protocole permet d'étudier l'action de l'AR durant la différenciation neuronale.

Ce protocole comprend cinq étapes majeures illustrées dans la figure 12. Le maintien des cellules ES dans un état indifférencié nécessite leur culture en présence de LIF ajouté dans le milieu de culture et de cellules nourricières (feeders). Les feeders sont des fibroblastes inactivés qui ne se divisent plus. Il faut dans un premier temps éliminer les feeders de la culture par plusieurs passages successifs pour obtenir une population pure de cellules ES. Le LIF étant maintenu dans le milieu de culture, les cellules ES restent dans un état indifférencié pendant plusieurs jours. Les cellules ES sont ensuite cultivées pendant huit jours sous forme



**Figure 12** : Protocole de différenciation neuronale dépendante de l'AR

a) En présence de cellules nourricières (feeders), les cellules ES se multiplient sous forme de colonies. b) Plusieurs passages sans feeders mais en présence de LIF permettent d'éliminer ces dernières et d'enrichir la population en ES. c) Les cellules ES sont ensuite cultivées dans des boîtes non adhérentes pendant 4 jours, ce qui favorise la formation de corps embryoides (EB). Durant les 4 jours suivants, les EB sont traités par l'AR. d) Au 8ème jour, les EB sont dissociés et cultivés dans le milieu N2. Quelques heures après leurs dissociation, ces cellules projettent des prolongements et adoptent une forme allongée. e) 48h plus tard, ces cellules établissent des contacts entre cellule via leurs prolongements. L'ajout d'un milieu adapté à la culture neuronale permet d'obtenir une population homogène de neurones glutaminergiques.

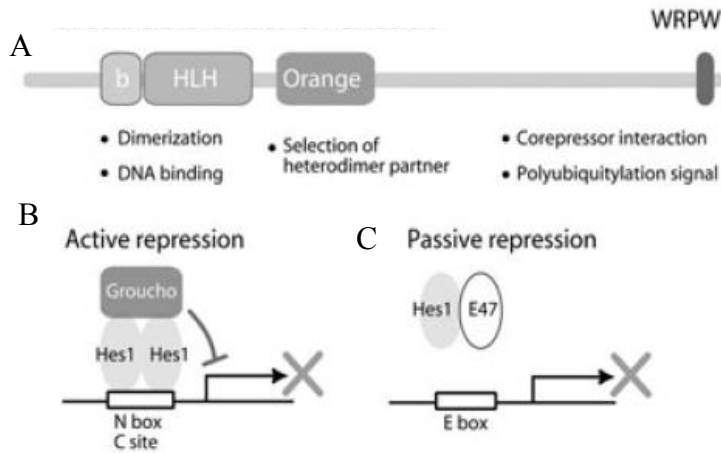
de corps embryoïdes (grappes non adhérentes ou EB : *Embryoid Bodies*) et traitées avec l'AR les quatre derniers jours. Les EB sont ensuite dissociés et les cellules réparties dans des boîtes de culture tapissées de poly-D-lysine/laminine en présence de milieu N2. Ce milieu contient de nombreuses hormones, telles que l'insuline, la transferrine et la progestérone (Bibel et al., 2007), favorisant la neurogenèse et le maintien des neurones en culture. A cette étape, les cellules présentent les caractéristiques de cellules radiales gliales neurogéniques (*Cf. figure 12*). Le remplacement du milieu N2 par le milieu de culture approprié permet de générer des neurones glutaminergiques (Bibel et al., 2004). Les neurones obtenus par ce protocole de différenciation présentent les caractéristiques des neurones pyramidaux, qui se trouvent dans le cortex cérébral et l'hippocampe *in vivo*.

## V. Les facteurs HES dans la différenciation neuronale

Parmi les gènes dont l'expression est fortement induite après le traitement des EB par l'AR se trouvent ceux qui codent pour les facteurs de transcription HES3 et HES5. Les gènes *Hes* des mammifères sont les orthologues des gènes "hairy et enhancer of split" de la drosophile, qui codent pour des répresseurs de la transcription. Cette famille comprend sept membres, parmi lesquels HES1, HES3 et HES5 sont exprimés dans les cellules souches neurales (Kageyama et al., 2006 et 2008). Les facteurs de transcription de la famille HES possèdent à leur extrémité N-terminale un domaine conservé formé d'une région basique, suivi d'un motif bHLH. Le domaine basique contient un résidu proline, impliqué dans la reconnaissance d'une séquence d'ADN particulière appelée N-box (5'-CACNAG-3'). Le domaine C-terminal des facteurs HES contient la séquence Trp-Arg-Pro-Trp (WRPW domain) nécessaire au recrutement des corépresseurs et à leur fonction transcriptionnelle (*Cf. Figure 13*). Ainsi, le corépresseur TLE/Grg, homologue de Groucho de la drosophile, interagit avec ce domaine WRPW et permet de recruter des HDACs (Kageyama et al., 2005).

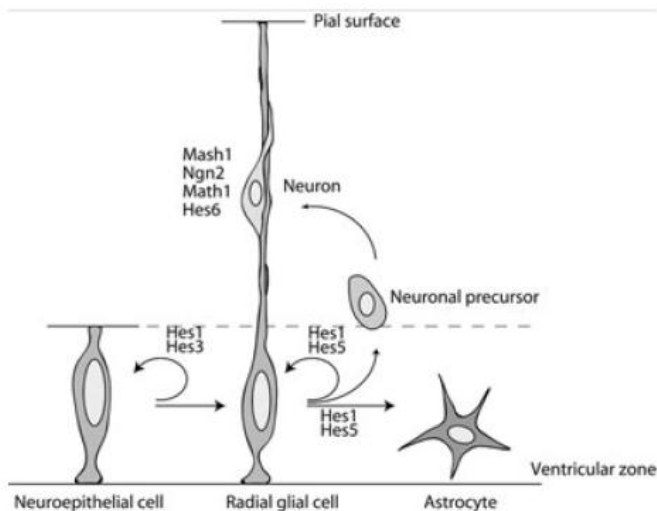
La fonction des facteurs HES dans la prolifération, la différenciation et la spécification des cellules souches neurales a été étudié *in vivo* (Kageyama et al., 2008). Au cours du développement du système nerveux central, les cellules neuroépithéliales subissent des divisions symétriques. Après la formation du tube neural, elles se transforment progressivement en cellules radiales gliales, dont le corps cellulaire se trouve dans la zone ventriculaire mais dont les extensions vont de la surface interne à la surface externe du tube neural. Ces cellules radiales gliales subissent des divisions asymétriques, chaque division permettant de générer une cellule radiale gliale et un précurseur neuronal. Ces précurseurs





**Figure 13** : Organisation structurale des facteurs HES et mécanismes d'action

A : Organisation structurale des protéines HES, les différents motifs et leurs rôles sont indiqués  
 B: L'homodimère de HES1 est fixé sur l'ADN au niveau de l'élément N-box, il recrute des corépresseurs comme Groucho pour réprimer la transcription.  
 C: HES1 peut également séquestrer des activateurs comme E47 et ainsi inhiber la transcription de manière passive.



**Figure 14** : Rôle des facteurs HES dans la différenciation neuronale *in vivo*

Des expériences d'inactivation (knock out) chez la souris, ont permis de mettre en évidence l'importance des facteurs HES1, HES3 et Hes5 lors de la neurogénèse. HES1 et HES3 sont impliqués dans le maintien des cellules souches neuroépithéliales. Le couple Hes1/Hes5 est impliqué dans le maintien des cellules radiales gliales ou la différenciation de ces cellules en astrocytes, suivant le stade du développement.

neuronaux donnent naissance à des neurones, qui vont migrer le long des fibres gliales afin d'atteindre les couches extérieures. Une fois la neurogenèse achevée, les cellules radiales gliales se différencient en astrocytes (Kageyama et al., 2006).

Les facteurs de transcription HES1, HES3, et HES5 sont impliqués dans ce processus de neurogenèse. Des expériences de double invalidation *in vivo* ont permis d'associer les gènes *Hes* aux différentes étapes de la neurogenèse décrites ci-dessus. Ainsi, HES3 et HES1 sont nécessaires pour le maintien des cellules neuroépithéliales, alors que HES5 et HES1 régulent le maintien des cellules radiales gliales. Après la neurogenèse, les cellules radiales gliales se différencient en astrocytes, et ce processus dépend également de l'action des facteurs HES1 et HES5. L'étude de souris transgéniques montre que la surexpression des gènes *Hes* lors des stades précoces du développement neuronal maintient les cellules souches neurales et empêche leur différenciation. En revanche, à des stades ultérieurs, cette surexpression promeut la formation des astrocytes, ceci indiquant qu'en fonction du stade de développement, les gènes *Hes* peuvent avoir des activités opposées (Kageyama et al., 2008) (*Cf. Figure-14*).

## OBJECTIFS

La différenciation neuronale est un processus qui nécessite une cascade d'événements impliquant plusieurs voies de signalisation. Parmi les effecteurs des voies de signalisation se trouvent plusieurs facteurs de transcription, tels que les RARs pour la voie de l'AR et les SMADs pour la voie BMP. La régulation de la transcription est au centre du contrôle du processus de différenciation. Ce processus implique l'expression différentielle dans le temps et l'espace des régulateurs et des facteurs de transcription. Il implique également des modifications épigénétiques. L'ensemble de ces actions définit le transcriptome d'une cellule non différenciée, d'un précurseur ou d'une cellule différenciée.

Durant la différenciation neuronale induite par l'AR, les facteurs de transcription responsables de la pluripotence tels OCT4 et NANOG sont réprimés, tandis que des facteurs de transcription spécifiques à chaque stade de la différenciation neuronale sont induits suivant une cinétique précise. Afin de mieux comprendre les mécanismes qui contrôlent cette cinétique, il est important de caractériser les réseaux transcriptionnels impliqués dans la différenciation neuronale et de définir la chronologie des actions des différents facteurs de transcription impliqués dans ce processus.

Le projet que j'ai intégré durant ma dernière année de thèse, s'inscrit dans cette optique. La stratégie que nous avons adoptée se base sur trois axes majeurs : 1) utiliser la technique de RNA-seq pour définir le transcriptome des cellules ES aux différents stades de la différenciation neuronale induite par l'AR en utilisant le protocole de Bibel et al ; 2) déterminer la cinétique d'expression des facteurs de transcription durant ces stades ; 3) sélectionner un ensemble de facteurs de transcription afin de mieux les caractériser en utilisant les techniques de shRNA et de ChIP-seq. La technique de shRNA permet de réduire l'expression de ces facteurs et de déterminer l'effet de cette réduction sur la différenciation neuronale, alors que le ChIP-seq permet d'identifier les gènes cibles directes de ces facteurs.

# MATÉRIEL ET MÉTHODES

## I. Technique de RNA-seq (3' end sequencing)

Comme son nom l'indique, la technique "3' end sequencing" consiste à séquencer l'extrémité 3' des ARNm. Cette technique permet de quantifier l'abondance d'un ARNm donné. Cette technique consiste en 8 étapes illustrées dans la figure 15 : 1) purification des ARNm poly-adénylés en utilisant des billes couplées à une amorce oligo dT (oligo deoxythymidylic) ; 2) synthèse des ADNc à partir de ces ARN ; 3) clivage de l'ADNc en 5' avec l'enzyme DpnII ; 4) purification des fragments 3' ; 5) ligation d'un bras (linker) en 5' de l'ADNc contenant le site de restriction MME ; 6) clivage de l'ADNc avec l'enzyme MME qui coupe l'ADN 16 nucléotides en aval de son site de reconnaissance ; 7) ligation en 3' d'un second bras qui servira de matrice pour séquencer l'ADNc et 8) séquençage de l'étiquette de 16 nt générée.

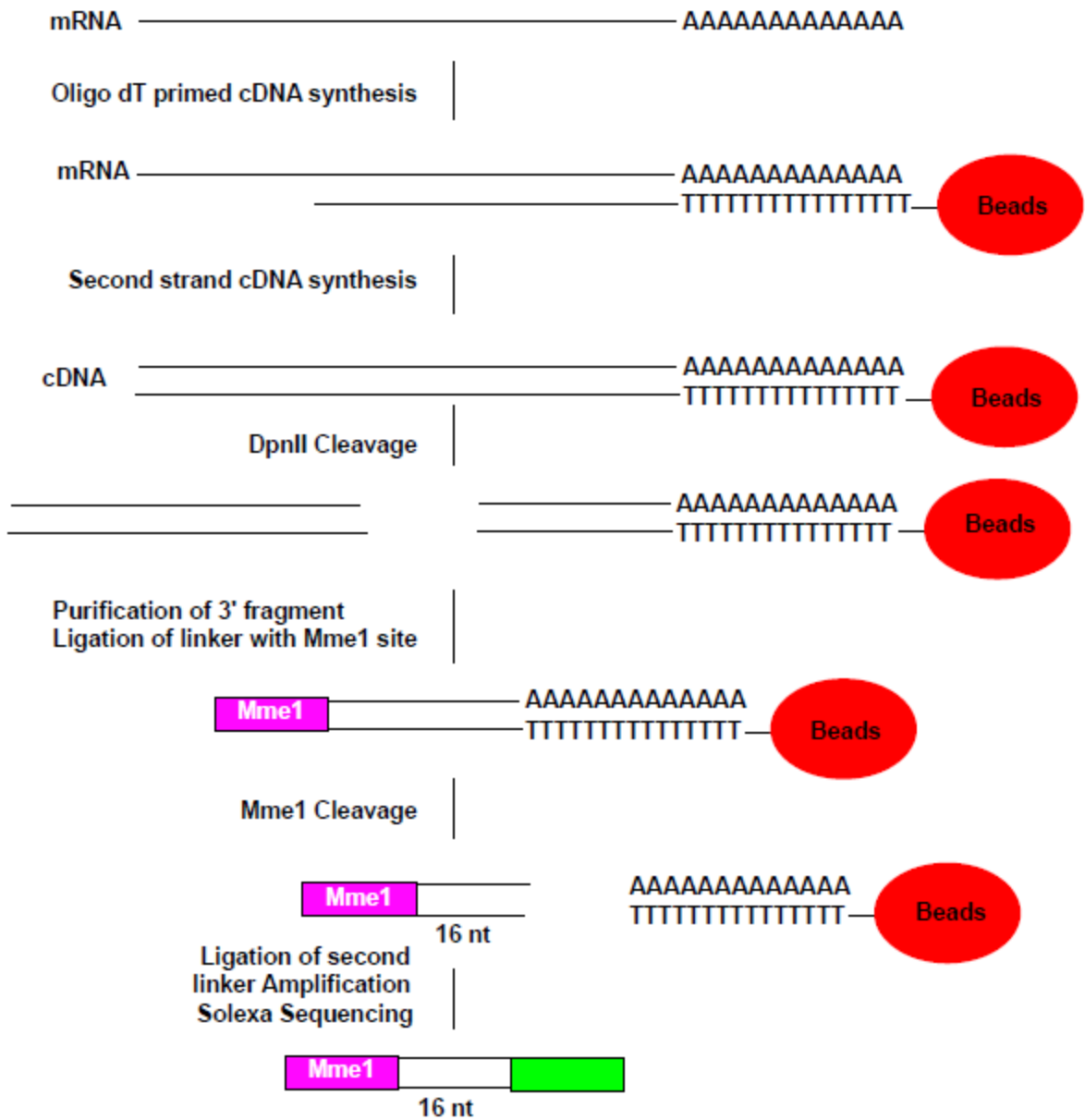
Les étiquettes générées vont être séquencées, annotées et quantifiées. Le nombre d'étiquettes appartenant à un ARNm donné reflète son abondance et donc le taux d'expression du gène correspondant.

## II. Génération de lignées cellulaires stables exprimant les shRNA

Nous avons choisi la méthode basée sur l'infection virale pour exprimer les shRNA d'intérêts dans les cellules ES. Cette méthode comporte trois étapes majeures : 1) produire des particules lentivirales contenant les séquences du shRNA et du gène de résistance à la puromycine ; 2) utiliser ces particules virales pour infecter les cellules ES et ainsi y exprimer le shRNA et le gène de résistance à la puromycine et 3) sélectionner les cellules exprimant le shRNA grâce à leur résistance à la puromycine. Les cellules ES ainsi sélectionnées forment une lignée stable exprimant le shRNA d'intérêt (*Cf. Figure 16*).

La production des particules lentivirales se fait dans les cellules HEK (cellules embryonnaires humaines de rein). La lignée que nous utilisons porte le nom H293T. Ces cellules ont la propriété d'exprimer l'antigène T du virus SV40 (*large T antigen*), ce qui

## Digital Gene Expression Pipeline



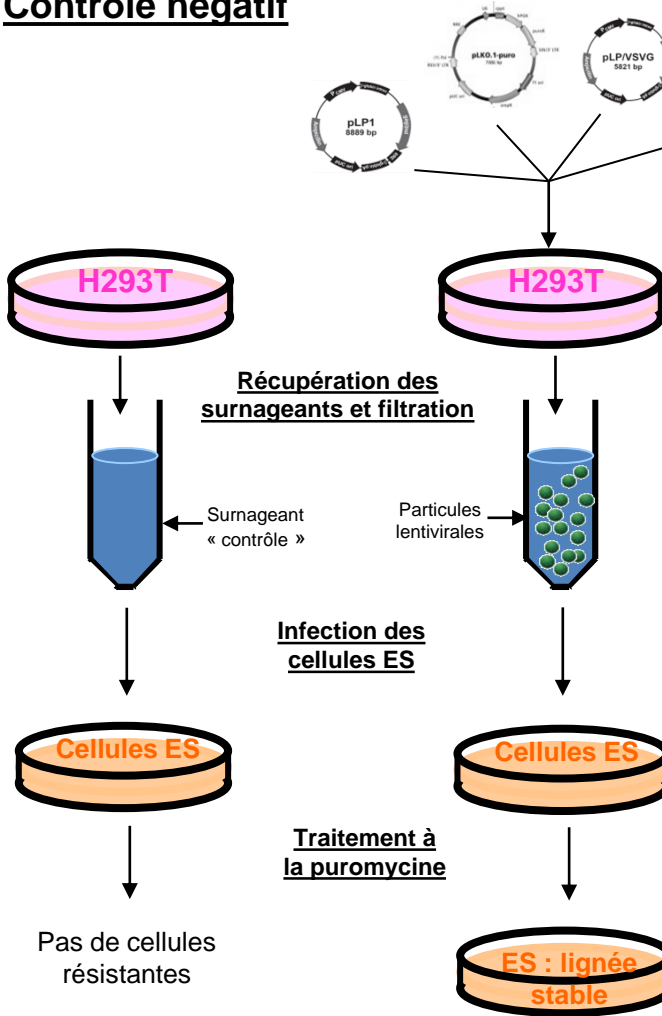
Comparison of tag sequence with bioinformatic library of all predicted tags downstream of DpnII sites in Ensembl annotated transcripts.  
 Ensembl Tag Table software  
 (A. Paulson Stowers Institute)

Figure 15 : Schématisation de la technique de RNA-seq

permet d'obtenir une forte expression des gènes présents sur des plasmides contenant l'origine de répllication du SV40.

Les cellules H293T sont transfectées par plusieurs plasmides codant pour les protéines de la capsid virale et des enzymes nécessaires à la répllication et l'intégration du génome viral. Le génome viral est remplacé par un plasmide codant pour le shRNA et le gène de résistance à la puromycine. Ce plasmide comporte également le signal d'encapsidation qui lui permet d'être empaqueté dans la capsid virale. La particule virale ainsi formée peut infecter des cellules ES et y introduire la séquence d'ADN correspondant au shRNA, qui grâce à l'intégrase virale va être intégrée dans le génome dans la cellule hôte et peut y être transcrit.

## Contrôle négatif



## Contrôle positif

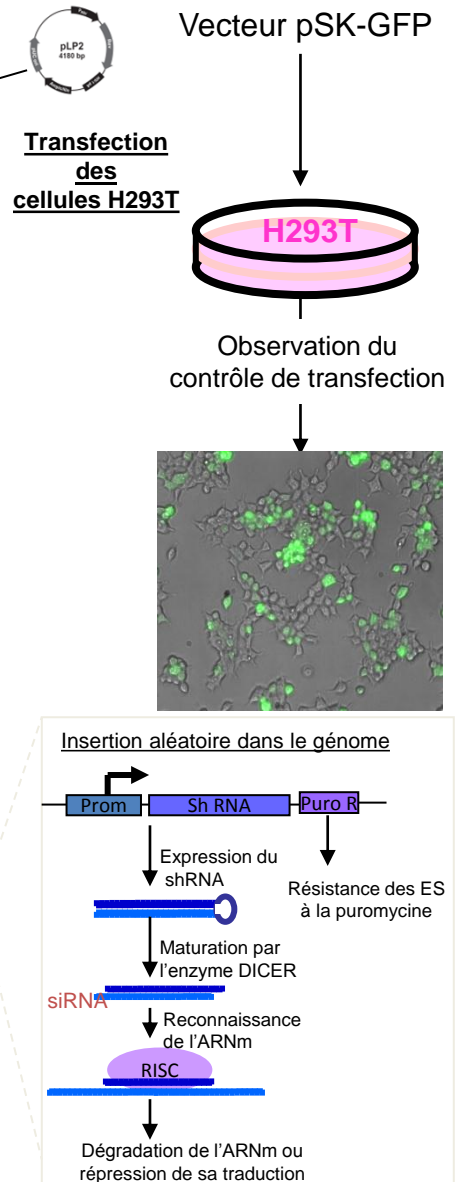


Figure 16: Génération de lignées stables exprimant des shRNA.

# RÉSULTATS

Afin de déterminer les variations du transcriptome des cellules ES durant leur différenciation en neurones, nous avons réalisé des préparations d'ARN à différents temps après le traitement des corps embryoïdes par l'AR (Figure). Ces ARNs ont servi à réaliser des expériences de RNA-seq (DGE).

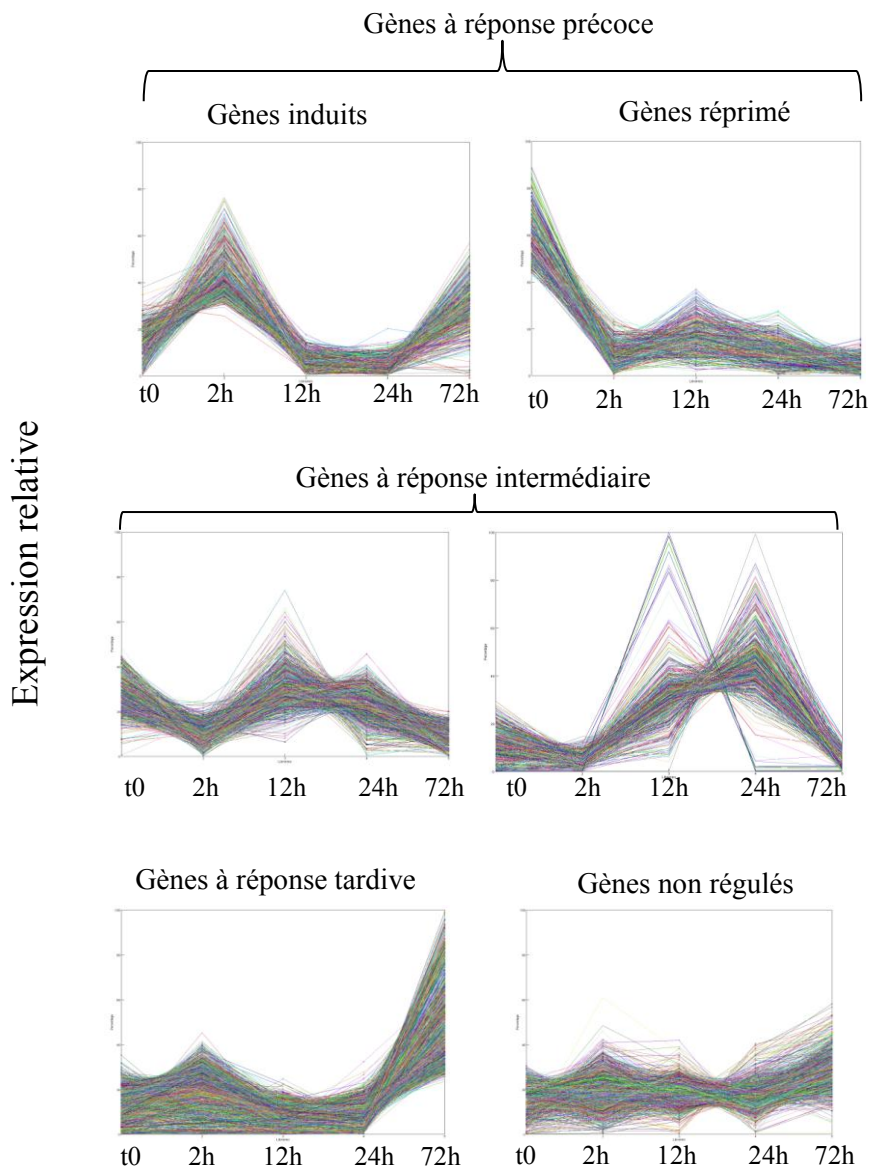
## **I. Définition de groupes de gènes ayant un profil d'expression similaire ou "clustering"**

Le terme "clustering" désigne une opération de répartition des données où les variables ayant des valeurs proches ou adoptant un comportement similaire sont groupées dans le même cluster (groupe). Le clustering des gènes exprimés dans les cellules des corps embryoïdes a été basé sur leur profil de variation d'expression après le traitement par l'AR.

L'analyse de clustering nous a permis de distinguer quatre groupes majeurs (*Cf. Figure 17*). Le premier groupe contient des gènes à réponse précoce (régulés après 2 h de traitement), le deuxième groupe contient des gènes à réponse intermédiaire (régulés après 12 ou 24 h de traitement), le troisième groupe contient des gènes à réponse tardive (régulés après 3 jours de traitement) et le quatrième et dernier groupe correspond aux gènes dont l'expression n'est pas modifiée par l'AR. L'ensemble des groupes et des sous-groupes est illustré dans la figure 15. Cette analyse montre que l'induction des gènes par l'AR est séquentielle et précise dans le temps. En effet, un grand nombre de gènes induits, atteignent un pic d'expression après 2, 12 ou 24h de traitement puis reviennent à leur état initial d'expression. D'autres gènes sont induits tardivement après 72h de traitement. En revanche, la plupart des gènes réprimés après le traitement par l'AR restent dans cet état réprimé tout au long du processus de différenciation. Cette observation suggère que l'induction par l'AR serait transitoire pour la majorité des gènes durant la différenciation neuronale alors que la répression dépendante de ce ligand serait stable dans le temps. Parmi les gènes réprimés, se trouvent notamment les gènes codant pour les facteurs de transcription responsables de la pluripotence, NANOG, OCT4 et SOX2.

Parmi les gènes induits par l'AR se trouvent plusieurs cibles connues de ce dernier, tel que RAR $\beta$ 2 et l'enzyme Cyp26 qui ne sont pas impliqués dans la neurogenèse. Mais nous





**Figure 17:** Les différents profils d'expression des gènes après traitement des EB par l'AR

Le clustering permet d'identifier plusieurs groupes avec des profils d'expression distincts. Les temps de traitement par l'AR sont indiqués.

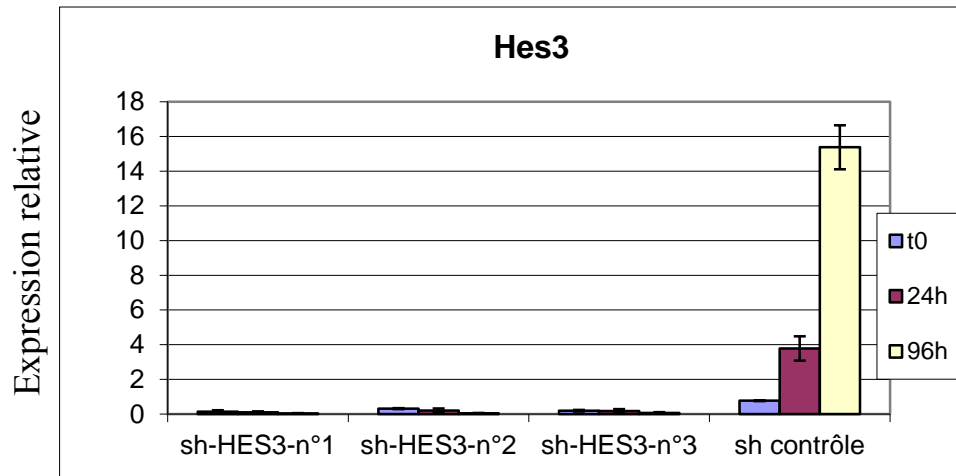
trouvons également plusieurs marqueurs neuronaux, tel que Nestine et PAX6. Nous nous sommes particulièrement intéressés aux facteurs de transcription qui sont induits rapidement par l'AR. Cette catégorie comporte notamment GBX2 (*Gastrulation Brain Homeobox 2*) et HES3 (*Hairy and Enhancer of split*) qui sont fortement induits par l'AR après 12 h de traitement. Nous avons choisi ces deux facteurs pour caractériser leur rôle dans la différenciation neuronale dépendante de l'AR. La première étape de cette étude consistait à réduire l'expression de ces deux facteurs en utilisant la technique de shRNA. Nous avons réussi à établir des lignées cellulaires ES exprimant stablement les shRNAs dirigés contre ces deux facteurs. Malheureusement les molécules de shRNA dirigés contre GBX2 sont peu efficaces et ne réduisent pas significativement son expression. En revanche, celles dirigées contre HES3 réduisent considérablement l'expression de ce dernier. Seuls les résultats obtenus pour HES3 seront présentés ici.

## **II. Génération des lignées stables dans lesquelles l'expression de HES3 est réduite**

Nous avons utilisé indépendamment trois shRNAs (*shHes3-1*, -2, -3) de séquences différentes pour réduire l'expression de HES3. Un shRNA ne reconnaissant aucun ARN murin a été utilisé comme contrôle (shC). Chaque shRNA a été utilisé pour générer une lignée cellulaire stable selon le protocole détaillé plus haut.

Afin de vérifier la capacité de chaque séquence de shRNA (sh-HES3-1, -2, -3) à induire une baisse significative de l'expression de HES3, l'expression de ce gène a été mesurée par RT-qPCR dans les quatre lignées pendant le processus de différenciation neuronale induite par l'AR. Les EB issus de ces lignées ont été traités par l'AR, des ARN ont été extraits après 24 et 96 h de traitement. Ces ARNs ont servi à réaliser les expériences de RT-qPCR.

Dans la lignée contrôle (shC), l'expression de *Hes3* est induite dès 24 h de traitement par l'AR. Cette induction est stable dans le temps car elle augmente jusqu'à 96 h. En revanche, dans les lignées exprimant les trois shRNAs dirigés contre HES3, cette induction est presque totalement abolie (*Cf. Figure 18*). Ces observations démontrent une très bonne efficacité des 3 shRNAs. Nous avons donc réussi à générer trois lignées stables dans lesquelles l'expression de HES3 est fortement réduite.



**Figure 18** : Expression de HES3 dans les lignées sh-HES3 et dans la lignée contrôle

Après traitement par l'AR, l'expression de HES3 a été mesurée par RT-qPCR dans les lignées exprimant les shRNAs dirigés contre HES3 (sh-HES3 n°1-3) et dans la lignée exprimant le shRNA contrôle. HES 3 est fortement induit dans la lignée exprimant le shRNA contrôle alors que cette induction est abolie dans les lignées sh-HES3. Les durées d'induction par l'AR sont indiquées par le code couleur.

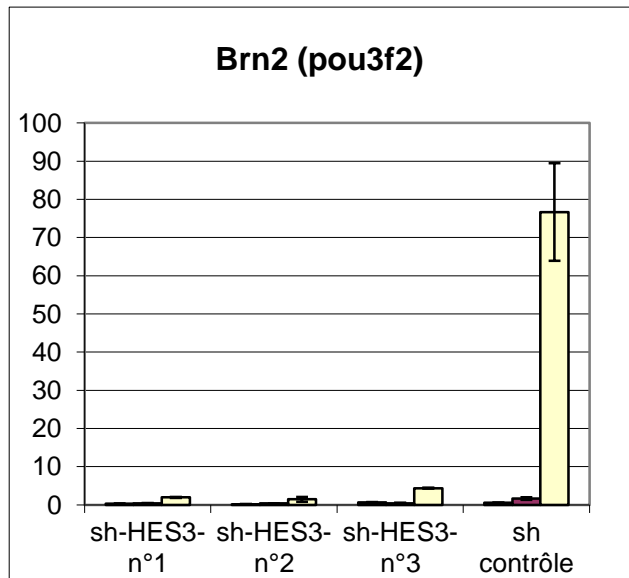
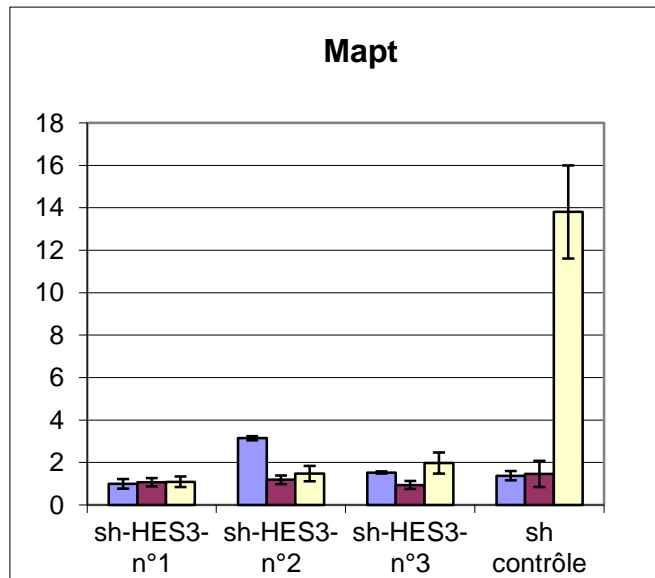
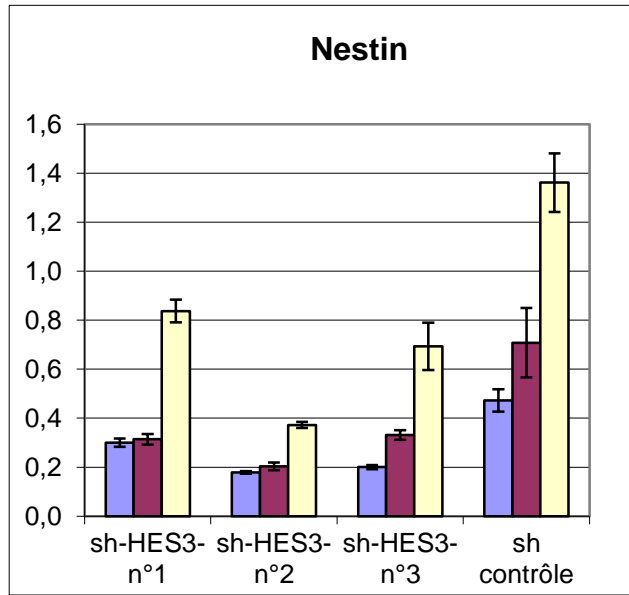
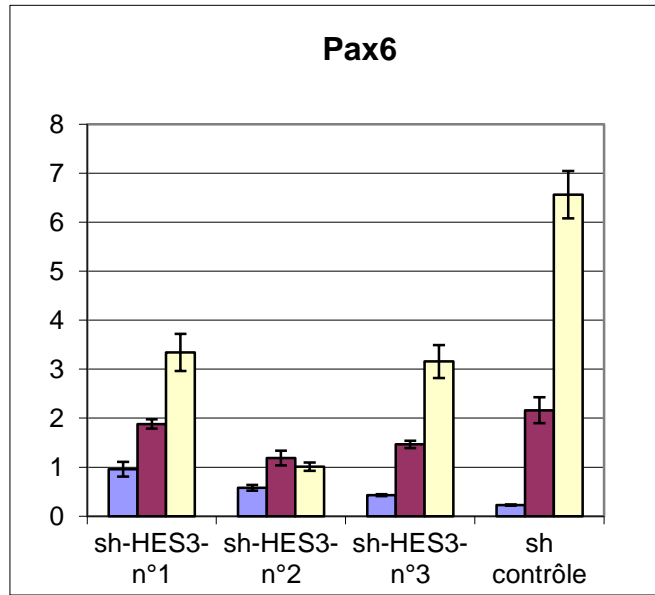
Afin de vérifier si l'expression des shRNAs a un effet global sur la réponse à l'AR, nous avons également vérifié l'expression des gènes *Cyp26a1*, *Hoxb1*, *Oct4* et *Nanog* dans les lignées sh-HES3 après traitement par l'AR. Dans les cellules ES, *Cyp26a1* et *Hoxb1* sont fortement induits par traitement par l'AR alors que *Oct4* et *Nanog* sont réprimés. Les résultats montrent que l'expression des sh-HES3 n'affecte pas le profil d'activation ou de répression de ces gènes. Ces observations démontrent que l'effet des sh-HES3 est spécifique pour ce messager et que la perte de HES3 n'a pas un effet global sur la réponse transcriptionnelle à l'AR. Le fait que les gènes de pluripotence *Oct4* et *Nanog* sont normalement réprimés dans les lignées sh-HES3 suggère que la perte de HES3 n'affecte pas l'initiation de la différenciation.

### **III. L'expression de HES3 est nécessaire à l'expression des marqueurs de la différenciation neuronale.**

Les résultats ci-dessus montrent que la perte d'expression de *Hes3* n'affecte pas les premières étapes de la différenciation. Nous avons donc choisi d'analyser les conséquences de la perte d'expression de HES3 sur l'activation des marqueurs de la différenciation neuronale. Parmi les marqueurs proneuronaux dont l'expression augmente suite au traitement par l'AR dans les EB, se trouve *Pax6*, un marqueur des cellules radiales gliales dont l'expression augmente dès 12 h de traitement par l'AR et qui reste fortement exprimé après 96 h. De façon similaire, le gène codant pour la *Nestine* est également fortement induit par l'AR. L'expression de ces deux marqueurs est affectée dans les trois lignées sh-HES3. Cela suggère que HES3 joue un rôle dans l'expression de ces deux gènes (Cf. Figure 19).

Nous avons également déterminé si l'expression de HES3 est nécessaire à l'activation de l'expression du facteur de transcription neurogénique BRN2 et du marqueur neuronal MAPT. *Brn2* et *Mapt* ne sont exprimés qu'après 96 heures de différenciation et représentent des gènes exprimés uniquement au stade de précurseurs neuronaux. L'activation de ces deux gènes est presque totalement abolie dans les 3 lignées sh-HES3. Cf. Figure 19

L'ensemble de ces observations démontre que la perte d'expression de *Hes3* réduit celle de *Pax6* nécessaire à la spécification des cellules radiales gliales, mais de façon spectaculaire abolit l'expression des marqueurs neuronaux *Brn2* et *Mapt*. HES3 semble donc nécessaire à la différenciation neuronale des cellules ES *in vitro*.



**Figure 19:**

Expression des marqueurs Pax6, Nestin, Brn2 et Mapt dans les lignées sh-HES3 et dans la lignée contrôle

La perte de HES3 affecte l'expression de Pax6 et Nestin et abolit complètement l'expression de Brn2 et de Mapt.

Les temps d'induction sont indiqués par le code couleur



## Discussion et perspectives

La différenciation neuronale des cellules ES *in vitro* est induite par le traitement des EBs par l'AR. L'AR induit l'expression d'un ensemble de gènes dont certains codent pour des facteurs de transcription qui agissent en aval pour réguler la suite du programme de différenciation.

Grace à la technique de RNA-seq, nous avons pu établir la cinétique d'expression des gènes régulés par l'AR durant la différenciation neuronale. Parmi ces gènes, se trouvent plusieurs facteurs de transcription. La vision globale de ce projet consiste à choisir certains de ces facteurs pour mieux caractériser leurs rôles dans la différenciation neuronale. Ce qui permettra de mieux comprendre les réseaux transcriptionnels qui contrôlent ce processus.

La stratégie choisie consiste dans un premier temps à réduire l'expression de ces facteurs par la technique de shRNA. L'avantage majeur de cette technique est qu'elle permet une réduction permanente des ARNs cibles. Cette propriété est importante pour étudier les facteurs impliqués dans la différenciation neuronale, car le protocole de cette différenciation dure en moyenne trois semaines.

Afin d'initier ce projet ambitieux, j'ai commencé la mise au point de cette stratégie en étudiant les facteurs GBX2 et HES3. Comme je l'ai mentionné plus haut, les shRNAs dirigés contre GBX2 ne permettaient pas de réduire son expression. Nous avons alors focalisé notre attention sur les facteurs HES.

Les résultats de RNA-seq et de RT-qPCR montrent que HES1 est fortement exprimé dans les EB et que son expression n'est pas affectée par l'AR, tandis que l'expression de HES3 et de HES5 est fortement induite. L'expression des trois gènes est maintenue tout au long des 96 heures nécessaires à la spécification des cellules radiales gliales. Ce profil d'expression suggère que l'activation et le maintien de l'expression de HES3 et HES5 sont importants pour cette spécification. Afin de tester cette possibilité, nous avons réprimé l'expression de HES3 afin de comprendre son rôle dans la différenciation. Nos résultats montrent que l'expression stable de trois shARNs différents ciblant HES3 permet de réprimer fortement et spécifiquement son expression. La perte d'expression de HES3 n'affecte pas

l'activation d'autres gènes par l'AR, ni la répression des gènes de pluripotence. Par conséquent, la perte de HES3 n'affecte pas la réponse à l'AR et l'initiation du processus de différenciation.

Nos résultats indiquent que le facteur HES3 n'agit pas à l'étape d'initiation de la différenciation mais à une étape plus tardive où son absence se traduit par la diminution de l'expression de *Pax6* et de *Nestin*. Son expression est encore plus importante lors du début de la transition des cellules radiales gliales en neurones caractérisée par l'expression des marqueurs neurogéniques comme *Brn2* et *Mapt*. L'activation de ces deux gènes est totalement inhibée en absence de HES3. Nous n'avons pas encore vérifié si l'expression de HES1 et de HES5 est affectée en absence de HES3. Par conséquent, nous ne pouvons pas affirmer que HES3 possède une fonction spécifique qui ne peut pas être remplie par les autres membres de sa famille. De même, nous n'avons pas encore poursuivi la différenciation jusqu'au stade terminal des neurones glutaminergiques. Il sera intéressant de déterminer si la répression de l'expression de HES3 provoque une absence totale de neurones dans la culture comme le suggère la perte d'expression des marqueurs *Brn2* et *Mapt*. Je n'ai pas pu réaliser cette expérience importante, car malheureusement, pendant les derniers mois de ma thèse nous avons rencontré des difficultés à obtenir une population homogène de neurones. Le protocole de différenciation neuronale comporte plusieurs étapes et fait intervenir plusieurs composants. Il est difficile de trouver rapidement le grain de sable qui fait gripper la mécanique de différenciation.

Toutefois, les résultats obtenus peuvent être interprétés de différentes façons. Il est possible que la diminution de l'expression de *Pax6* reflète un défaut de spécification des cellules radiales gliales de sorte que ces cellules ne soient plus capables de se différencier correctement par la suite en neurones. De façon alternative, HES3 pourrait agir directement dans les cellules radiales gliales pour assurer l'activation des marqueurs neuronaux. Cette interprétation se heurte aux observations précédentes qui montrent que les facteurs HES sont des répresseurs de la transcription et non pas des activateurs. Il est donc probable que l'absence d'expression de *Brn2* et de *Mapt* ne résulte pas d'un effet direct de HES3, mais reflète un défaut plus global de la spécification des cellules radiales gliales.

Afin de mieux comprendre la fonction de HES3 dans les cellules ES, nous proposons d'étudier l'expression d'autres marqueurs de la différenciation. Cela peut être fait par

RT-qPCR en ciblant certains gènes spécifiquement. En revanche, une analyse globale d'expression génique par RNA-seq dans les cellules sh-HES3 à des temps différents permettra d'identifier l'ensemble des changements d'expression génique induit par la perte de l'expression de HES3. Une telle analyse permettra d'étudier plus précisément la fonction de ce facteur, notamment de déterminer si la perte de son expression entraîne un changement d'identité cellulaire, soit en altérant l'identité des neurones, soit en générant un type cellulaire différent (cellules gliales ou astrocytes). De plus, nous pouvons coupler les résultats d'expression génique avec des résultats de ChIP-seq contre HES3 afin de définir les gènes cibles directes de ce facteur.

Nos résultats obtenus *in vitro* qui montrent que HES3 est nécessaire à la différenciation neuronale sont en partie contradictoires avec les fonctions définies pour les facteurs HES dans la différenciation neuronale *in vivo*. Il a été proposé que les facteurs HES sont des antagonistes des facteurs pro-neuronaux à domaine bHLH comme ASCL1, ATOH1 et les NEUROGs. L'expression des facteurs HES est importante pour maintenir les progéniteurs neuronaux et la perte de leur expression induit une différenciation prématurée en neurones. En revanche, nos résultats semblent montrer que la perte de son expression bloque la différenciation en neurones plutôt que de la faciliter. Les expériences de RNA-seq et de ChIP-seq proposées ci-dessus permettront de résoudre ce paradoxe.



## Bibliographie

- Adamkewicz JI, Hansen KE, Prud'homme WA, Davis JL, Thorner J.** (2001) High affinity interaction of yeast transcriptional regulator, Mot1, with TATA box-binding protein (TBP). *J Biol Chem.* 13;276(15):11883-94.
- Agresti A, Bianchi ME.** (2003) HMGB proteins and gene expression. *Curr. Opin. Genet. Dev.* 13:170-8.
- Ahn SH, Diaz RL, Grunstein M, Allis CD.** (2006). Histone H2B deacetylation at lysine 11 is required for yeast apoptosis induced by phosphorylation of H2B at serine 10. *Mol. Cell.* 24:211-220.
- Albert TK, Grote K, Boeing S, Meisterernst M.** (2010) Basal core promoters control the equilibrium between negative cofactor 2 and preinitiation complexes in human cells. *Genome Biol.* 11:R33.
- Allison LA, Moyle M, Shales M, Ingles CJ.** (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell.* 42:599-610.
- Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P.** (2001) Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* 309:1-18.
- Andrau JC, Van Oevelen CJ, Van Teeffelen HA, Weil PA, Holstege FC, Timmers HT.** (2002) Mot1p is essential for TBP recruitment to selected promoters during in vivo gene activation. *EMBO J.* 21:5173-83.
- Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, Kornberg RD.** (1999) Conserved structures of mediator and RNA polymerase II holoenzyme. *Science.* 283:985-7.
- Aubert J, Dunstan H, Chambers I, Smith A.** (2002) Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat. Biotechnol.* 20:1240-5.
- Auble DT, Wang D, Post KW, Hahn S.** (1997) Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol. Cell. Biol.* 17:4842-51.
- Aubert J, Dunstan H, Chambers I, Smith A.** (2002) Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat. Biotechnol.* 20:1240-5.
- Auty R, Steen H, Myers LC, Persinger J, Bartholomew B, Gygi SP, Buratowski S.** (2004) Purification of active TFIID from *Saccharomyces cerevisiae*. Extensive promoter contacts and co-activator function. *J. Biol. Chem.* 279:49973-81.
- Bajic VB, Tan SL, Christoffels A, Schönbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, Kawai J, Hume DA, Carninci P, Hayashizaki Y.** (2006) Mice and men: their promoter properties. *PLoS Genet.* 2:e54.
- Barboric M, Nissen RM, Kanazawa S, Jabrane-Ferrat N, Peterlin BM.** (2001) NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell.* 8:327-37.
- Bártfai R, Balduf C, Hilton T, Rathmann Y, Hadzhiev Y, Tora L, Orbán L, Müller F.** (2004) TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol.* 14:593-8.

- Bartolomei MS, Halden NF, Cullen CR, Corden JL.** (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol. Cell Biol.* **8**:330-9.
- Bartholomew B, Dahmus ME, Meares CF.** (1986) RNA contacts subunits Ilo and Iic in HeLa RNA polymerase II transcription complexes. *J. Biol. Chem.* **261**:14226-31.
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES.** (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**:169-181.
- Bibel M, Richter J, Schrenk K, Tucker KL, Staiger V, Korte M, Goetz M, Barde YA.** (2004) Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7**:1003-9.
- Bibel M, Richter J, Lacroix E, Barde YA.** (2007) Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* **2**:1034-43.
- Birck C, Poch O, Romier C, Ruff M, Mengus G, Lavigne AC, Davidson I, Moras D.** (1998) Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell.* **94**:239-49.
- Bird A.** (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6-21.
- Bocquel MT, Kumar V, Stricker C, Chambon P, Gronemeyer H.** (1989) The contribution of the N- and C-terminal regions of steroid receptors to activation of transcription is both receptor and cell-specific. *Nucleic Acids Res.* **17**:2581-95.
- Bogdanović O, Veenstra GJ.** (2009) DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma.* **118**:549-65.
- Bourbon HM, Aguilera A, Ansari AZ, Asturias FJ, Berk AJ, Bjorklund S, Blackwell TK, Borggreffe T, Carey M, Carlson M, Conaway JW, Conaway RC, Emmons SW, Fondell JD, Freedman LP, Fukasawa T, Gustafsson CM, Han M, He X, Herman PK, Hinnebusch AG, Holmberg S, Holstege FC, Jaehning JA, Kim YJ, Kuras L, Leutz A, Lis JT, Meisterernest M, Naar AM, Nasmyth K, Parvin JD, Ptashne M, Reinberg D, Ronne H, Sadowski I, Sakurai H, Sipiczki M, Sternberg PW, Stillman DJ, Strich R, Struhl K, Svejstrup JQ, Tuck S, Winston F, Roeder RG, Kornberg RD.** (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell.* **14**:553-7.
- Bourguet W, Ruff M, Chambon P, Gronemeyer H, Moras D.** (1995) Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature.* **375**:377-82.
- Bradbury EM, Inglis RJ, Matthews HR, Sarnier N.** (1973) Phosphorylation of very-lysine-rich histone in Physarum polycephalum. Correlation with chromosome condensation. *Eur. J. Biochem.* **33**:131-9.
- Bruck N, Vitoux D, Ferry C, Duong V, Bauer A, de Thé H, Rochette-Egly C.** (2009) A coordinated phosphorylation cascade initiated by p38MAPK/MSK1 directs RARalpha to target promoters. *EMBO J.* **28**:34-47.
- Bryant GO, Martel LS, Burley SK, Berk AJ.** (1996) Radical mutations reveal TATA-box binding protein surfaces required for activated transcription in vivo. *Genes Dev.* **10**:2491-504.
- Bulger M, Groudine M.** (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**:2465-77.
- Buratowski S, Hahn S, Sharp PA, Guarente L.** (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature.* **334**:37-42.

- Burke TW, Kadonaga JT.** (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* **11**, 3020-3031.
- Bushnell DA, Westover KD, Davis RE, Kornberg RD.** (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science.* **303**:983-8.
- Cadena DL, Dahmus ME.** (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**:12468-74.
- Carles C, Treich I, Bouet F, Riva M, Sentenac A.** (1991) Two additional common subunits, ABC10 alpha and ABC10 beta, are shared by yeast RNA polymerases. *J. Biol. Chem.* **266**:24092-6.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y.** (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**:626-35.
- Cavallini B, Huet J, Plassat JL, Sentenac A, Egly JM, Chambon P.** (1988) A yeast activity can substitute for the HeLa cell TATA box factor. *Nature.* **334**:77-80.
- Cazillis M, Lelièvre V, Gressens P.** (2005) Neural differentiation of murine embryonic stem cells ES. *Med. Sci. (Paris).* **21**:484-90.
- Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L.** (2009) Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**:275-80
- Chan HM, and La Thangue NB.** (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* **114**:2363-2373.
- Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D.** (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science.* **18**:1780-2.
- Chasman DI, Flaherty KM, Sharp PA, Kornberg RD.** (1993) Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**:8174-8.
- Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR.** (1999) Regulation of transcription by a protein methyltransferase. *Science* **284**:2174-2177.
- Combet C, Blanchet C, Geourjon C, Deléage G.** (2000) NPS@: Network protein sequence analysis. *TIBS* **25**:147-50.
- Chen H, Lin RJ, Schiltz RL, Chakravarti D, Nash A, Nagy L, Privalsky ML, Nakatani Y, Evans RM.** (1997) Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**:569-580.
- Chen JD, and Evans RM.** (1995) A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature.* **377**:454-457.
- Chen Z, Manley JL.** (2000) Robust mRNA transcription in chicken DT40 cells depleted of TAF(II)31 suggests both functional degeneracy and evolutionary divergence. *Mol. Cell. Biol.* **20**:5064-76.
- Cheung WL, Ajiro K, Samejima K, Kloc M, Cheung P, Mizzen CA, Beeser A, Etkin LD, Chernoff J, Earnshaw WC, Allis CD.** (2003). Apoptotic phosphorylation of histone

H2B is mediated by mammalian sterile twenty kinase. *Cell*. **113**:507-517.

**Chiang CM, Roeder RG.** (1995) Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science*. **267**:531-6.

**Chicca JJ 2nd, Auble DT, Pugh BF.** (1998) Cloning and biochemical characterization of TAF-172, a human homolog of yeast Mot1. *Mol. Cell Biol*. **18**:1701-10.

**Chou S, Chatterjee S, Lee M, Struhl K.** (1999) Transcriptional activation in yeast cells lacking transcription factor IIA. *Genetics*. **153**:1573-81.

**Cler E, Papai G, Schultz P, Davidson I.** (2009) Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell. Mol. Life Sci*. **66**:2123-34.

**Coin F, Egly JM.** (1998) Ten years of TFIID. *Cold Spring Harb. Symp. Quant. Biol*. **63**:105-10.

**Coin F, Proietti De Santis L, Nardo T, Zlobinskaya O, Stefanini M, Egly JM.** (2006) p8/TTD-A as a repair-specific TFIID subunit. *Mol. Cell*. **21**:215-26.

**Collart MA.** (1996) The NOT, SPT3, and MOT1 genes functionally interact to regulate transcription at core promoters. *Mol. Cell Biol*. **16**:6668-76.

**Corden JL, Cadena DL, Ahearn JM Jr, Dahmus ME.** (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **82**:7934-8.

**Core LJ, Lis JT.** (2008) Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*. **319**:1791-2.

**Core LJ, Waterfall JJ, Lis JT.** (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. **322**:1845-8.

**Couture JF, Hauk G, Thompson MJ, Blackburn GM, Trievel RC.** (2006). Catalytic roles for carbon-oxygen hydrogen bonding in SET domain lysine methyltransferases. *J. Biol. Chem*. **281**:19280-19287.

**Cramer P, Bushnell DA, Kornberg RD.** (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*. **292**:1863-76.

**Darimont BD, Wagner RL, Apriletti JW, Stallcup MR, Kushner PJ, Baxter JD, Fletterick RJ, Yamamoto KR.** (1998) Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev*. **12**:3343-3356.

**Darst RP, Dasgupta A, Zhu C, Hsu JY, Vroom A, Muldrow T, Auble DT.** (2003) Mot1 regulates the DNA binding activity of free TATA-binding protein in an ATP-dependent manner. *J. Biol. Chem*. **278**:13216-26.

**Davidson I, Martianov I, Viville S.** (2004) TBP, a universal transcription factor? *Med. Sci. (Paris)*. **20**:575-9.

**Davis JL, Kunisawa R, Thorner J.** (1992) A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell Biol*. **12**:1879-92.

**Davis JA, Takagi Y, Kornberg RD, Asturias FA.** (2002) Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell*. **10**:409-15.

**de Graaf P, Mousson F, Geverts B, Scheer E, Tora L, Houtsmuller AB, Timmers HT.** (2010) Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *J. Cell Sci*.

**Delacroix L, Moutier E, Altobelli G, Legras S, Poch O, Choukrallah MA, Bertin I, Jost B, Davidson I.** (2010) Cell-specific interaction of retinoic acid receptors with target genes in mouse embryonic fibroblasts and embryonic stem cells. *Mol. Cell Biol*. **30**:231-44.

- Delaval K, Govin J, Cerqueira F, Rousseaux S, Khochbin S, Feil R.** (2007). Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* **26**:720-729.
- de Lera AR, Bourguet W, Altucci L, Gronemeyer H.** (2007) Design of selective nuclear receptor modulators: RAR and RXR as a case study. *Nat. Rev. Drug. Discov.* **6**:811-20.
- Deng W, Roberts SG.** (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**:2418-23.
- Dikstein R, Ruppert S, Tjian R.** (1996) TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. *Cell.* **84**:781-90.
- Dilworth FJ, Chambon P.** (2001) Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. *Oncogene* **20**:3047-3054.
- Ding N, Zhou H, Esteve PO, Chin HG, Kim S, Xu X, Joseph SM, Friez MJ, Schwartz CE, Pradhan S, Boyer TG.** (2008) Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol. Cell.* **31**:347-59.
- Dollé P.** (2009) Developmental expression of retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7**:e006.
- Dotson MR, Yuan CX, Roeder RG, Myers LC, Gustafsson CM, Jiang YW, Li Y, Kornberg RD, Asturias FJ.** (2000) Structural organization of yeast and mammalian mediator complexes. *Proc. Natl. Acad. Sci. U.S.A.* **97**:14307-10.
- Doyon Y, Cayrou C, Ullah M, Landry AJ, Côté V, Selleck W, Lane WS, Tan S, Yang XJ, Côté J.** (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**:51-64.
- Dynlacht BD, Hoey T, Tjian R.** (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell.* **66**:563-76.
- Eberhardy SR, Farnham PJ.** (2001) c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. *J. Biol. Chem.* **276**:48562-71.
- Eberhardy SR, Farnham PJ.** (2002) Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J. Biol. Chem.* **277**:40156-62.
- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M.** (2008) Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell.* **31**:337-46.
- Fang SM, Burton ZF.** (1996) RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J. Biol. Chem.* **271**:11703-9.
- Fang J, Chen T, Chadwick B, Li E, Zhang Y.** (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *J. Biol. Chem.* **279**:52812-52815.
- Featherstone M.** (2002). Coactivators in transcription initiation: here are your orders. *Curr. Opin. Genet. Dev.* **12**:149-155.
- Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD.** (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature.* **438**:1116-1122.
- Flores O, Maldonado E, Reinberg D.** (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.* **264**:8913-21.

- Flores O, Lu H, Reinberg D.** (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIIH. *J. Biol. Chem.* **267**:2786-93.
- Fondell JD, Ge H, Roeder RG.** (1996) Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl. Acad. Sci. U.S.A.* **93**:8329-33.
- Forget D, Langelier MF, Thérien C, Trinh V, Coulombe B.** (2004) Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell Biol.* **24**:1122-31.
- Fournier C, Goto Y, Ballestar E, Delaval K, Hever AM, Esteller M, Feil R.** (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J.* **21**:6560-6570.
- Frith et al., 2010 (pas trouvé)
- Fuda NJ, Ardehali MB, Lis JT.** (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature.* **461**:186-92.
- Fuentealba LC, Eivers E, Ikeda A, Hurtado C, Kuroda H, Pera EM, De Robertis EM.** (2007) Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell.* **131**:980-93.
- Gadbois EL, Chao DM, Reese JC, Green MR, Young RA.** (1997) Functional antagonism between RNA polymerase II holoenzyme and global negative regulator NC2 in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **94**:3145-50.
- Garbett KA, Tripathi MK, Cencki B, Layer JH, Weil PA.** (2007) Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. *Mol. Cell Biol.* **27**:297-311.
- Gazdag E, Santenard A, Ziegler-Birling C, Altobelli G, Poch O, Tora L, Torres-Padilla ME.** (2009) TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev.* **23**:2210-23.
- Ge H, Martinez E, Chiang CM, Roeder RG.** (1996) Activator-dependent transcription by mammalian RNA polymerase II: in vitro reconstitution with general transcription factors and cofactors. *Methods Enzymol.* **274**:57-71.
- Ge H, Roeder RG.** (1994a) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell.* **78**:513-23.
- Ge H, Roeder RG.** (1994b) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J. Biol. Chem.* **269**:17136-40.
- Geiger JH, Hahn S, Lee S, Sigler PB.** (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science.* **272**:830-6.
- Geisberg JV, Holstege FC, Young RA, Struhl K.** (2001) Yeast NC2 associates with the RNA polymerase II preinitiation complex and selectively affects transcription in vivo. *Mol. Cell Biol.* **21**:2736-42.
- Germain P, Chambon P, Eichele G, Evans RM, Lazar MA, Leid M, De Lera AR, Lotan R, Mangelsdorf DJ, Gronemeyer H.** (2006a) International Union of Pharmacology. LX. Retinoic acid receptors. *Pharmacol. Rev.* **58**:712-25.
- Ghavi-Helm Y, Michaut M, Acker J, Aude JC, Thuriaux P, Werner M, Soutourina J.** (2008) Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes Dev.* **22**:1934-47.
- Giglia-Mari G, Coin F, Ranish JA, Hoogstraten D, Theil A, Wijgers N, Jaspers NG, Raams A, Argentini M, van der Spek PJ, Botta E, Stefanini M, Egly JM, Aebersold R,**

- Hoeijmakers JH, Vermeulen W.** (2004) A new, tenth subunit of TFIID is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat. Genet.* **36**:714-9.
- Gilfillan S, Stelzer G, Piaia E, Hofmann MG, Meisterernst M.** (2005) Efficient binding of NC2.TATA-binding protein to DNA in the absence of TATA. *J. Biol. Chem.* **280**:6222-30.
- Gill G, Pascal E, Tseng ZH, Tjian R.** (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* **91**:192-6.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD.** (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science.* **292**:1876-82.
- Grob P, Cruse MJ, Inouye C, Peris M, Penczek PA, Tjian R, Nogales E.** (2006) Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. *Structure.* **14**:511-20.
- Groft CM, Uljon SN, Wang R, Werner MH.** (1998) Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc. Natl. Acad. Sci. U.S.A.* **95**:9117-22.
- Gu W, Malik S, Ito M, Yuan CX, Fondell JD, Zhang X, Martinez E, Qin J, Roeder RG.** (1999) A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell.* **3**:97-108.
- Guglielmi B, Soutourina J, Esnault C, Werner M.** (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **104**:16062-7.
- Guenther MG, Lane WS, Fischle W, Verdin E, Lazar MA, Shiekhhattar R.** (2000) A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness. *Genes Dev.* **14**:1048-57.
- Ha I, Roberts S, Maldonado E, Sun X, Kim LU, Green M, Reinberg D.** (1993) Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* **7**:1021-32.
- Hampsey M.** (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**:465-503.
- Han J, Zhou H, Horazdovsky B, Zhang K, Xu RM, Zhang Z.** (2007). Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. *Science* **315**:653-655.
- Hassa PO, Buerki C, Lombardi C, Imhof R, Hottiger MO.** (2003) Transcriptional coactivation of nuclear factor- $\kappa$ B-dependent gene expression by p300 is regulated by poly(ADP)-ribose polymerase-1. *J. Biol. Chem.* **278**:45145-53.
- Hatano SY, Tada M, Kimura H, Yamaguchi S, Kono T, Nakano T, Suemori H, Nakatsuji N, Tada T.** (2005) Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122**:67-79.
- Hayashida T, Sekiguchi T, Noguchi E, Sunamoto H, Ohba T, Nishimoto T.** (1994) The CCG1/TAFII250 gene is mutated in thermosensitive G1 mutants of the BHK21 cell line derived from golden hamster. *Gene.* **141**:267-70.
- Heery DM, Kalkhoven E, Hoare S, Parker MG.** (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**:733-736.
- Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, Young RA.** (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell.* **2**:43-53.
- Hermann A, Schmitt S, Jeltsch A.** (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* **278**:31717-31721.

- Hermann A, Gowher H, Jeltsch A.** (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.* **61**:2571-87.
- Hernandez N.** (1993) TBP, a universal eukaryotic transcription factor? *Genes Dev.* **7**:1291-308.
- Hilton TL, Li Y, Dunphy EL, Wang EH.** (2005) TAF1 histone acetyltransferase activity in Sp1 activation of the cyclin D1 promoter. *Mol Cell Biol.*;25(10):4321-32.
- Ho CK, Shuman S.** (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell.* **3**:405-11.
- Hobbs NK, Bondareva AA, Barnett S, Capecchi MR, Schmidt EE.** (2002) Removing the vertebrate-specific TBP N terminus disrupts placental beta2m-dependent interactions with the maternal immune system. *Cell.* **110**:43-54.
- Hoey T, Weinzierl RO, Gill G, Chen JL, Dynlacht BD, Tjian R.** (1993) Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell.* **72**:247-60.
- Høiby T, Zhou H, Mitsiou DJ, Stunnenberg HG.** (2007) A facelift for the general transcription factor TFIIA. *Biochim. Biophys. Acta.* **1769**:429-36.
- Holmes MC, Tjian R.** (2000) Promoter-selective properties of the TBP-related factor TRF1. *Science.* **288**:867-70.
- Hong H, Kohli K, Garabedian MJ, Stallcup MR.** (1997) GRIP1, a transcriptional coactivator for the AF-2 transactivation domain of steroid, thyroid, retinoid, and vitamin D receptors. *Mol. Cell. Biol.* **17**:2735-44.
- Horlein AJ, Naar AM, Heinzl T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamei Y, Soderstrom M, Glass CK, and et al.** (1995) Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**:397-404.
- Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR.** (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **2104**:829-838.
- Hyland EM, Cosgrove MS, Molina H, Wang D, Pandey A, Cottee RJ, Boeke JD.** (2005). Insights into the role of histone H3 and histone H4 core modifiable residues in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **25**:10060-10070.
- Imhof A, Yang XJ, Ogryzko VV, Nakatani Y, Wolffe AP, Ge H.** (1997) Acetylation of general transcription factors by histone acetyltransferases. *Curr. Biol.* **7**:689-92.
- Ito M, Yuan CX, Malik S, Gu W, Fondell JD, Yamamura S, Fu ZY, Zhang X, Qin J, Roeder RG.** (1999) Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol. Cell.* **3**:361-70.
- Jacobson RH, Ladurner AG, King DS, Tjian R.** (2000) Structure and function of a human TAFII250 double bromodomain module. *Science.* **288**:1422-5.
- Jawhari A, Uhring M, De Carlo S, Crucifix C, Tocchini-Valentini G, Moras D, Schultz P, Poterszman A.** (2006) Structure and oligomeric state of human transcription factor TFIIE. *EMBO Rep.* **7**:500-5.
- Johannes F, Wardenaar R, Colomé-Tatché M, Mousson F, de Graaf P, Mokry M, Guryev V, Timmers HT, Cuppen E, Jansen RC.** (2010) Comparing genome-wide chromatin profiles using ChIP-chip or ChIP-seq. *Bioinformatics.* **26**:1000-6.



- Kageyama R, Ohtsuka T, Hatakeyama J, Ohsawa R.** (2005) Roles of bHLH genes in neural stem cell differentiation. *Exp. Cell Res.* **306**:343-8.
- Kageyama R, Ohtsuka T, Kobayashi T.** (2008) Roles of Hes genes in neural development. *Dev. Growth Differ.* **50** Suppl 1:S97-103.
- Kamada K, Shu F, Chen H, Malik S, Stelzer G, Roeder RG, Meisterernst M, Burley SK.** (2001) Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell.* **106**:71-81.
- Kanazawa S, Soucek L, Evan G, Okamoto T, Peterlin BM.** (2003) c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene.* **22**:5707-11.
- Kannan P, Yu Y, Wankhade S, Tainsky MA.** (1999) PolyADP-ribose polymerase is a coactivator for AP-2-mediated transcriptional activation. *Nucleic Acids Res.* **27**:866-74.
- Keriel A, Stary A, Sarasin A, Rochette-Egly C, Egly JM.** (2002) XPD mutations prevent TFIIH-dependent transactivation by nuclear receptors and phosphorylation of RARalpha. *Cell.* **109**:125-35.
- Kettenberger H, Armache KJ, Cramer P.** (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell.* **114**:347-57.
- Keveaney et al., 1993
- Kim YJ, Björklund S, Li Y, Sayre MH, Kornberg RD.** (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell.* **77**:599-608.
- Kim JL, Nikolov DB, Burley SK.** (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature.* **365**:520-7.
- Kim Y, Geiger JH, Hahn S, Sigler PB.** (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature.* **365**:512-20.
- Kim JB, Sharp PA.** (2001) Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J. Biol. Chem.* **276**:12317-23.
- Kimura M, Suzuki H, Ishihama A.** (2002) Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Mol. Cell Biol.* **22**:1577-88.
- Klejman MP, Zhao X, van Schaik FM, Herr W, Timmers HT.** (2005) Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. *Nucleic Acids Res.* **33**:5426-36.
- Kliewer SA, Umesono K, Mangelsdorf DJ, Evans RM.** (1992) Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D3 signalling. *Nature.* **355**:446-9.
- Kobor MS, Simon LD, Omichinski J, Zhong G, Archambault J, Greenblatt J.** (2000) A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **20**:7438-49.
- Koike H, Kawashima-Ohya Y, Yamasaki T, Clowney L, Katsuya Y, Suzuki M.** (2004) Origins of protein stability revealed by comparing crystal structures of TATA binding proteins. *Structure.* **12**:157-68.
- Koleske AJ, Young RA.** (1994) An RNA polymerase II holoenzyme responsive to activators. *Nature.* **368**:466-9.

- Komarnitsky P, Cho EJ, Buratowski S.** (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**:2452-60.
- Kouzarides T.** (2007) Chromatin modifications and their function. *Cell* **128**:693-705.
- Kostrewa D, Zeller ME, Armache KJ, Seizl M, Leike K, Thomm M, Cramer P.** (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature.* **462**:323-30.
- Kumar R, Litwack G.** (2009) Structural and functional relationships of the steroid hormone receptors' N-terminal transactivation domain. *Steroids* **74**:877-83.
- Kuras L, Kosa P, Mencia M, Struhl K.** (2000) TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. *Science.* **288**:1244-8.
- Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH.** (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**:34-44.
- Langelier MF, Forget D, Rojas A, Porlier Y, Burton ZF, Coulombe B.** (2001) Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIIF in transcription initiation by RNA polymerase II. *J. Biol. Chem.* **276**:38652-7.
- Laudet V, Hänni C, Coll J, Catzeflis F, Stéhelin D.** (1992) Evolution of the nuclear receptor gene superfamily. *EMBO J.* **11**:1003-13.
- Lavigne AC, Gangloff YG, Carré L, Mengus G, Birck C, Poch O, Romier C, Moras D, Davidson I.** (1999) Synergistic transcriptional activation by TATA-binding protein and hTAFII28 requires specific amino acids of the hTAFII28 histone fold. *Mol. Cell Biol.* **19**:5050-60.
- Laybourn PJ, Dahmus ME.** (1990) Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**:13165-73.
- Le Douarin B, vom Baur E, Zechel C, Heery D, Heine M, Vivat V, Gronemeyer H, Losson R, Chambon P.** (1996) Ligand-dependent interaction of nuclear receptors with potential transcriptional intermediary factors (mediators). *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **351**:569-578.
- Lee M, Struhl K.** (2001) Multiple functions of the nonconserved N-terminal domain of yeast TATA-binding protein. *Genetics.* **158**:87-93.
- Lee SK, Kim JH, Lee YC, Cheong J, Lee JW.** (2000) Silencing mediator of retinoic acid and thyroid hormone receptors, as a novel transcriptional corepressor molecule of activating protein-1, nuclear factor-kappaB, and serum response factor. *J. Biol. Chem.* **275**:12470-4.
- Leurent C, Sanders SL, Demény MA, Garbett KA, Ruhlmann C, Weil PA, Tora L, Schultz P.** (2004) Mapping key functional sites within yeast TFIID. *EMBO J.* **23**:719-27.
- Leid M, Kastner P, Lyons R, Nakshatri H, Saunders M, Zacharewski T, Chen JY, Staub A, Garnier JM, Mader S, et al.** (1992) Purification, cloning, and RXR identity of the HeLa cell factor with which RAR or TR heterodimerizes to bind target sequences efficiently. *Cell* **68**:377-395.
- Leonhardt H, Page AW, Weier HU, Bestor TH.** (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**:865-873.
- Lescure A, Lutz Y, Eberhard D, Jacq X, Krol A, Grummt I, Davidson I, Chambon P, Tora L.** (1994) The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *EMBO J.* **13**:1166-75.

- Li, X.Y., Bhaumik, S.R., Green, M.R.** (2000). Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* 288, 1242–1244.
- Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA.** (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell Biol.* **22**:480-491
- Liu A, Niswander LA.**(2005) Bone morphogenetic protein signalling and vertebrate nervous system development. *Nat Rev Neurosci.* 6(12):945-54. Review.
- Lu H, Flores O, Weinmann R, Reinberg D.** (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **88**:10004-8.
- Maas NL, Miller KM, DeFazio LG, Toczyski DP.** (2006). Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol. Cell* **23**:109-119.
- Maden M.** (2007) Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat. Rev. Neurosci.* **8**:755-65.
- Madison JM, Winston F.** (1997) Evidence that Spt3 functionally interacts with Mot1, TFIIA, and TATA-binding protein to confer promoter-specific transcriptional control in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **17**:287-95.
- Maile T, Kwoczynski S, Katzenberger RJ, Wassarman DA, Sauer F.** (2004) TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science.* **304**:1010-4.
- Malik S, Gu W, Wu W, Qin J, Roeder RG.** (2000) The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. *Mol. Cell.* **5**:753-60.
- Mantovani R.** (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene.* **239**:15-27.
- Marco E, García-Nieto R, Gago F.** (2003) Assessment by molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1. *J. Mol. Biol.* **328**:9-32.
- Mark M, Ghyselinck NB, Chambon P.** (2006) Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu. Rev. Pharmacol. Toxicol.* **46**:451-80.
- Marsden MP, Laemmli UK.** (1979) Metaphase chromosome structure: evidence for a radial loop model. *Cell* **17**:849-58.
- Marshall NF, Price DH.** (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**:12335-8.
- Marshall NF, Peng J, Xie Z, Price DH.** (1996) Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**:27176-83.
- Martianov I, Brancorsini S, Gansmuller A, Parvinen M, Davidson I, Sassone-Corsi P.** (2002) Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. *Development.* **129**:945-55.
- Martin C, Zhang Y.** (2005). The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**:838-849.
- Matsui T, Segall J, Weil PA, Roeder RG.** (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem.* **255**:11992-6.
- Meinhart A, Kamenski T, Hoepfner S, Baumli S, Cramer P.** (2005) A structural perspective of CTD function. *Genes Dev.* **19**:1401-15.

- Meisterernst, M., Roy, A. L., Lieu, H.M., and Roeder, R.G.** 1991 a. Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* 66:981–993.
- Meisterernst M, Roeder RG** 1991 b. Family of proteins that interact with TFIID and regulate promoter activity. *Cell*. 1;67(3):557-67.
- Mengus G, May M, Carré L, Chambon P, Davidson I.** (1997) Human TAF(II)135 potentiates transcriptional activation by the AF-2s of the retinoic acid, vitamin D3, and thyroid hormone receptors in mammalian cells. *Genes Dev.* **11**:1381-95.
- Merino A, Madden KR, Lane WS, Champoux JJ, Reinberg D.** (1993) DNA topoisomerase I is involved in both repression and activation of transcription. *Nature.* **365**:227-32.
- Metivier R, Penot G, Hubner MR, Reid G, Brand H, Kos M, Gannon F.** (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**:751-763.
- Metzger E, Yin N, Wissmann M, Kunowska N, Fischer K, Friedrichs N, Patnaik D, Higgins JM, Potier N, Scheidtmann KH, Buettner R, Schüle R.** (2008). Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nat. Cell Biol.* **10**:53-60.
- Meyer KD, Donner AJ, Knuesel MT, York AG, Espinosa JM, Taatjes DJ.** (2008) Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *EMBO J.* **27**:1447-57.
- Mialon A, Sankinen M, Söderström H, Junttila TT, Holmström T, Koivusalo R, Papageorgiou AC, Johnson RS, Hietanen S, Elenius K, Westermarck J.** (2005) DNA topoisomerase I is a cofactor for c-Jun in the regulation of epidermal growth factor receptor expression and cancer cell proliferation. *Mol. Cell Biol.* **25**:5040-51.
- Mittal V, Hernandez N.** (1997) Role for the amino-terminal region of human TBP in U6 snRNA transcription. *Science.* **275**:1136-40.
- Mitchell PJ, Tjian R.** (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science.* **245**:371-8.
- Mizzen CA, Yang XJ, Kokubo T, Brownell JE, Bannister AJ, Owen-Hughes T, Workman J, Wang L, Berger SL, Kouzarides T, Nakatani Y, Allis CD.** (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell.* **87**:1261-70.
- Mohrmann L, Verrijzer CP.** (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim. Biophys. Acta.* **1681**:59-73.
- Moqtaderi, Z., Bai, Y., Poon, D., Weil, P.A., and Struhl, K.** 1996. TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* 383:188–191.
- Moras D, Gronemeyer H.** (1998) The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**:384-91.
- Mortusewicz O, Roth W, Li N, Cardoso MC, Meisterernst M, Leonhardt H.** (2008) Recruitment of RNA polymerase II cofactor PC4 to DNA damage sites. *J. Cell Biol.* **183**:769-76.
- Näär AM, Beurang PA, Zhou S, Abraham S, Solomon W, Tjian R.** (1999) Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature.* **398**:828-32.
- Neuwald AF, Hirano T.** (2000) HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res.* **10**:1445-52.

**Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H, Smith A.** (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*. **95**:379-91.

**Niederreither K, Dollé P.** (2008) Retinoic acid in development: towards an integrated view. *Nat. Rev. Genet.* **9**:541-53.

**Nikolov DB, Hu SH, Lin J, Gasch A, Hoffmann A, Horikoshi M, Chua NH, Roeder RG, Burley SK.** (1992) Crystal structure of TFIID TATA-box binding protein. *Nature*. **360**:40-6.

**Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK.** (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*. **377**:119-28.

**Nikolov DB, Chen H, Halay ED, Hoffman A, Roeder RG, Burley SK.** (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. U.S.A.* **93**:4862-7.

**Niwa H.** (2001) Molecular mechanism to maintain stem cell renewal of ES cells. *Cell Struct. Funct.* **26**:137-48.

**Nolte F, Hofmann WK.** (2010) Molecular mechanisms involved in the progression of myelodysplastic syndrome. *Future Oncol.* **6**:445-55.

**Oelgeschläger T, Tao Y, Kang YK, Roeder RG.** (1998) Transcription activation via enhanced preinitiation complex assembly in a human cell-free system lacking TAFII. *Mol. Cell*. **1**:925-31.

**Okano M, Bell DW, Haber DA, Li E.** (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**:247-257.

**Okuda M, Tanaka A, Arai Y, Satoh M, Okamura H, Nagadoi A, Hanaoka F, Ohkuma Y, Nishimura Y.** (2004) A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. *J. Biol. Chem.* **279**:51395-403.

**Onate SA, Tsai SY, Tsai MJ, O'Malley BW.** (1995) Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**:1354-1357.

**Olins AL, Olins DE.** (1974) Spheroid chromatin units (v bodies). *Science*. **183**:330-2.

**Orphanides G, Lagrange T, Reinberg D.** (1996) The general transcription factors of RNA polymerase II. *Genes Dev.* **10**:2657-83.

**Pan G, Greenblatt J.** (1994) Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J. Biol. Chem.* **269**:30101-4.

**Papai G, Tripathi MK, Ruhlmann C, Layer JH, Weil PA, Schultz P.** (2010) TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature*. **465**:956-60.

**Parvin JD, Sharp PA.** (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*. **73**:533-40.

**Pavri R, Lewis B, Kim TK, Dilworth FJ, Erdjument-Bromage H, Tempst P, de Murcia G, Evans R, Chambon P, Reinberg D.** (2005) PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. *Mol. Cell*. **18**:83-96.

- Payne JM, Laybourn PJ, Dahmus ME.** (1989) The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit I $\alpha$ . *J. Biol. Chem.* **264**:19621-9.
- Pera MF, Andrade J, Houssami S, Reubinoff B, Trounson A, Stanley EG, Ward-van Oostwaard D, Mummery C.** (2004) Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. *J. Cell Sci.* **117**:1269-80.
- Pereira LA, van der Knaap JA, van den Boom V, van den Heuvel FA, Timmers HT.** (2001) TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol. Cell. Biol.* **21**:7523-34.
- Pereira LA, Klejman MP, Timmers HT.** (2003) Roles for BTAf1 and Mot1p in dynamics of TATA-binding protein and regulation of RNA polymerase II transcription. *Gene.* **315**:1-13.
- Pereira LA, Klejman MP, Ruhlmann C, Kavelaars F, Oulad-Abdelghani M, Timmers HT, Schultz P.** (2004) Molecular architecture of the basal transcription factor B-TFIID. *J. Biol. Chem.* **279**:21802-7.
- Perissi V, Staszewski LM, McInerney EM, Kurokawa R, Kronos A, Rose DW, Lambert MH, Milburn MV, Glass CK, Rosenfeld MG.** (1999) Molecular determinants of nuclear receptor-corepressor interaction. *Genes Dev.* **13**:3198-3208.
- Perissi V, Rosenfeld MG.** (2005) Controlling nuclear receptors: the circular logic of cofactor cycles. *Nat. Rev. Mol. Cell. Biol.* **6**:542-554.
- Perissi V, Jepsen K, Glass CK, Rosenfeld MG.** (2010) Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* **11**:109-23.
- Perry J, Kleckner N.** (2003) The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell.* **112**:151-5.
- Peterlin BM, Price DH.** (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell.* **23**:297-305.
- Peterson MG, Tanese N, Pugh BF, Tjian R.** (1990) Functional domains and upstream activation properties of cloned human TATA binding protein. *Science.* **248**:1625-30.
- Pham AD, Sauer F.** (2000) Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science.* **289**:2357-60.
- Prather DM, Larschan E, Winston F.** (2005) Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **25**:2650-9.
- Prelich G.** (1997) *Saccharomyces cerevisiae* BUR6 encodes a DRAP1/NC2 $\alpha$  homolog that has both positive and negative roles in transcription in vivo. *Mol. Cell Biol.* **17**:2057-65.
- Privalsky ML.** (2004) The role of corepressors in transcriptional regulation by nuclear hormone receptors. *Annu. Rev. Physiol.* **66**:315-60.
- Pugh BF, Tjian R.** (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell.* **61**:1187-97.
- Pugh BF, Tjian R.** (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**:1935-45.
- Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA.** (2010) c-Myc regulates transcriptional pause release. *Cell.* **141**:432-45.
- Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, Aebersold R.** (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIIH. *Nat. Genet.* **36**:707-13.
- Rastinejad F, Wagner T, Zhao Q, Khorasanizadeh S.** (2000) Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.* **19**:1045-54.

- Reid G, Hubner MR, Metivier R, Brand H, Denger S, Manu D, Beaudouin J, Ellenberg J, Gannon F.** (2003) Cyclic, proteasome-mediated turnover of unliganded and liganded ERalpha on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell* **11**:695-707.
- Reinberg D, Roeder RG.** (1987) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. *J. Biol. Chem.* **262**:3310-21.
- Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, Moras D.** (1995) Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature.* **378**:681-9.
- Rochette-Egly C, Germain P.** (2009) Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **8**;7:e005.
- Roeder RG.** (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**:327-35.
- Rosenfeld MG, Lunyak VV, Glass CK.** (2006) Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.* **20**:1405-28.
- Roth SY, Denu JM, Allis CD.** (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* **70**:81-120.
- Ruppert S, Tjian R.** (1995) Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* **9**:2747-55.
- Ryu S, Zhou S, Ladurner AG, Tjian R.** (1999) The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature.* **397**:446-50.
- Saha A, Wittmeyer J, Cairns BR.** (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **7**:437-447.
- Samuels M, Fire A, Sharp PA.** (1982) Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J. Biol. Chem.* **257**:14419-27.
- Sawadogo M, Roeder RG.** (1985) Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell.* **43**:165-75.
- Sawadogo M, Sentenac A.** (1990) RNA polymerase B (II) and general transcription factors. *Annu. Rev. Biochem.* **59**:711-54.
- Saxonov S, Berg P, Brutlag DL.** (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* **103**:1412-7.
- Schaeffer L, Moncollin V, Roy R, Staub A, Mezzina M, Sarasin A, Weeda G, Hoeijmakers JH, Egly JM.** (1994) The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *EMBO J.* **13**:2388-92.
- Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M.** (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell.* **103**:843-852.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL.** (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**:844-847.
- Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M.** (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**:1263-1271.

- Serizawa H, Conaway JW, Conaway RC.** (1993) Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. *Nature*. **363**:371-4.
- Sikorski et al., 2007
- Shimada M, Niida H, Zineldeen DH, Tagami H, Tanaka M, Saito H, Nakanishi M.** (2008). Chk1 is a histone H3 threonine 11 kinase that regulates DNA damage-induced transcriptional repression. *Cell* **132**:221-232.
- Smale ST, Kadonaga JT.** (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449-479.
- Smallwood A, Esteve PO, Pradhan S, Carey M.** (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* **21**:1169-1178.
- Smukler SR, Runciman SB, Xu S, van der Kooy D.** (2006) Embryonic stem cells assume a primitive neural stem cell fate in the absence of extrinsic influences. *J. Cell Biol.* **172**:79-90.
- Solow S, Salunek M, Ryan R, Lieberman PM.** (2001) Taf(II) 250 phosphorylates human transcription factor IIA on serine residues important for TBP binding and transcription activity. *J. Biol. Chem.* **276**:15886-92.
- Spencer TE, Jenster G, Burcin MM, Allis CD, Zhou J, Mizzen CA, McKenna NJ, Onate SA, Tsai SY, Tsai MJ, O'Malley BW.** (1997) Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature* **389**:194-8.
- Stargell LA, Struhl K.** (1995) The TBP-TFIIA interaction in the response to acidic activators in vivo. *Science*. **269**:75-8.
- Strubin M, Struhl K.** (1992) Yeast and human TFIID with altered DNA-binding specificity for TATA elements. *Cell*. **68**:721-30.
- Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S.** (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* **279**:27816-27823.
- Svejstrup JQ, Vichi P, Egly JM.** (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biochem. Sci.* **21**:346-50.
- Taatjes DJ.** (2010) The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem. Sci.* **35**:315-22.
- Takeda K, Noguchi K, Shi W, Tanaka T, Matsumoto M, Yoshida N, Kishimoto T, 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.
- Tan S, Hunziker Y, Sargent DF, Richmond TJ.** (1996) Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*. **381**:127-51.
- Tang H, Sun X, Reinberg D, Ebright RH.** (1996) Protein-protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex. *Proc. Natl. Acad. Sci. U.S.A.* **93**:1119-24.
- Tansey WP, Ruppert S, Tjian R, Herr W.** (1994) Multiple regions of TBP participate in the response to transcriptional activators in vivo. *Genes Dev.* **8**:2756-69.
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ.** (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**:1025-1040.
- Thomas JO.** (2001) HMG1 and 2: architectural DNA-binding proteins. *Biochem. Soc. Trans.* **29**:395-401.
- Thomas JO, Travers AA.** (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem. Sci.* **26**:167-74.
- Thomas MC, Chiang CM.** (2006) The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**, 105-178.



- Timmers HT, Sharp PA.** (1991) The mammalian TFIID protein is present in two functionally distinct complexes. *Genes Dev.* **5**:1946-56.
- Timmers HT, Meyers RE, Sharp PA.** (1992) Composition of transcription factor B-TFIID. *Proc. Natl. Acad. Sci. U.S.A.* **89**:8140-4.
- Tora L, White J, Brou C, Tasset D, Webster N, Scheer E, Chambon P.** (1989) The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* **59**:477-487.
- Tora L.** (2002) A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* **16**:673-5.
- Torchia J, Rose DW, Inostroza J, Kamei Y, Westin S, Glass CK, Rosenfeld MG.** (1997) The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**:677-684.
- Tsukiyama T.** (2002). The in vivo functions of ATP-dependent chromatin-remodelling factors. *Nat. Rev. Mol. Cell Biol.* **3**:422-429.
- Umesono K, Murakami KK, Thompson CC, Evans RM.** (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell.* **65**:1255-66.
- Usheva A, Maldonado E, Goldring A, Lu H, Houbavi C, Reinberg D, Aloni Y.** (1991) Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell.* **69**:871-81.
- Vakoc CR, Mandat SA, Olenchock BA, Blobel GA.** (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19**:381-391.
- Vallier L, Reynolds D, Pedersen RA.** (2004) Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**:403-21.
- Van Dyke MW, Roeder RG, Sawadogo M.** (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science.* **241**:1335-8.
- van Werven FJ, van Bakel H, van Teeffelen HA, Altelaar AF, Koerkamp MG, Heck AJ, Holstege FC, Timmers HT.** (2008) Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. *Genes Dev.* **22**:2359-69.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT.** (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell.* **131**:58-69.
- Voegel JJ, Heine MJ, Zechel C, Chambon P, Gronemeyer H.** (1996) TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.* **15**:3667-3675.
- Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S, Handa H.** (1998a) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**:343-56.
- Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H.** (1998b) Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* **17**:7395-403.

- Wassarman DA, Sauer F.** (2001) TAF(II)250: a transcription toolbox. *J. Cell Sci.* **114**:2895-902.
- Watanabe T, Hayashi K, Tanaka A, Furumoto T, Hanaoka F, Ohkuma Y.** (2003) The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol. Cell Biol.* **23**:2914-26.
- Watt F, Molloy PL.** (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2**:1136-1143.
- Wei Y, Yu L, Bowen J, Gorovsky MA, Allis CD.** (1999). Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**:99-109.
- Wei W, Dorjsuren D, Lin Y, Qin W, Nomura T, Hayashi N, Murakami S.** (2001) Direct interaction between the subunit RAP30 of transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. *J. Biol. Chem.* **276**:12266-73.
- Weil PA, Luse DS, Segall J, Roeder RG.** (1979) Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. *Cell.* **18**:469-84.
- Werthen S, Mitschler A, Romier C, Gangloff YG, Thuault S, Davidson I, Moras D.** (2002) Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J. Biol. Chem.* **277**:45502-9.
- West ML, Corden JL.** (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics.* **140**:1223-33.
- Willy PJ, Kobayashi R, Kadonaga JT.** (2000) A basal transcription factor that activates or represses transcription. *Science.* **290**:982-5.
- Wolffe AP, Pruss D.** (1996) Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell* **84**:817-819.
- Woychik NA, Liao SM, Kolodziej PA, Young RA.** (1990) Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev.* **4**:313-23.
- Wu SY, Chiang CM.** (1998) Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. *J. Biol. Chem.* **273**:12492-8.
- Wu SY, Zhou T, Chiang CM.** (2003) Human mediator enhances activator-facilitated recruitment of RNA polymerase II and promoter recognition by TATA-binding protein (TBP) independently of TBP-associated factors. *Mol. Cell Biol.* **23**:6229-42.
- Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, Gronemeyer H.** (1996) A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* **3**:206.
- Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD.** (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell.* **121**:859-872.
- Xie X, Kokubo T, Cohen SL, Mirza UA, Hoffmann A, Chait BT, Roeder RG, Nakatani Y, Burley SK.** (1996) Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature.* **380**:316-22.
- Yao TP, Ku G, Zhou N, Scully R, Livingston DM.** (1996) The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc. Natl. Acad. Sci. U.S.A.* **93**:10626-31.

- Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H.** (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell.* **21**:227-37.
- Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H.** (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell.* **97**:41-51.
- Yankulov KY, Bentley DL.** (1997) Regulation of CDK7 substrate specificity by MAT1 and TFIIH. *EMBO J.* **16**:1638-46.
- Yokomori K, Verrijzer CP, Tjian R.** (1998) An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. *Proc. Natl. Acad. Sci. U.S.A.* **95**:6722-7.
- Young RA.** (1991) RNA polymerase II. *Annu. Rev. Biochem.* **60**:689-715.
- Ying QL, Stavridis M, Griffiths D, Li M, Smith A.** (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**:183-6.
- 
- Zechel C, Shen XQ, Chambon P, Gronemeyer H.** (1994a) Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *EMBO J.* **13**:1414-24.
- Zechel C, Shen XQ, Chen JY, Chen ZP, Chambon P, Gronemeyer H.** (1994b) The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *EMBO J.* **13**:1425-33.
- Zehring WA, Lee JM, Weeks JR, Jokerst RS, Greenleaf AL.** (1988) The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **85**:3698-702.
- Zhang Y, Reinberg D.** (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**:2343-2360.
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE.** (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* 2009;10(6):R62. Epub 2009 Jun 9.
- Zhou Q, Boyer TG, Berk AJ.** (1994) Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. *Genes Dev.* **7**:180-7.
- Zurita M, Merino C.** (2003) The transcriptional complexity of the TFIIH complex. *Trends Genet.* **19**:578-84.