

Université de Montréal

**Identification *in silico* d'éléments de réponse de  
récepteurs nucléaires impliqués dans le cancer du sein**

par

David Laperrière

Département de Biochimie, Institut de Recherche en Immunologie et Cancérologie  
Faculté de Médecine

Mémoire présenté à la Faculté de Médecine  
en vue de l'obtention du grade de M.Sc.  
en bio-informatique

Avril, 2008

© David Laperrière, 2008

Université de Montréal  
Faculté de Médecine

Ce mémoire intitulé :

Identification *in silico* d'éléments de réponse de récepteurs nucléaires impliqués dans le  
cancer du sein

présenté par :

David Laperrière

a été évalué par un jury composé des personnes suivantes :

Sébastien Lemieux PhD, président-rapporteur

Sylvie Mader PhD, directeur de recherche

François Major PhD, co-directeur

David Lohnes PhD, membre du jury

## Résumé

La croissance de deux tiers des tumeurs mammaires dépend des œstrogènes. Le réseau de gènes responsable de propager les signaux prolifératifs des œstrogènes est encore mal connu. Des micropuces d'ADN de cellules de carcinome mammaire MCF7 traitées à l'œstradiol (E2) avec ou sans l'inhibiteur de synthèse protéique cycloheximide (CHX) ont permis d'identifier de nombreux gènes cibles primaires et secondaires. La séquence des promoteurs des gènes cibles a été criblée à l'aide d'une banque de 300 matrices modélisant les sites reconnus par divers facteurs de transcription. Les éléments de réponse aux œstrogènes (ERE) sont enrichis dans les promoteurs des gènes primaires. Les sites E2F sont enrichis dans les promoteurs des gènes cible secondaires. Un enrichissement similaire a été observé avec les régions liées par ER $\alpha$  et E2F1 en ChIP-on-chip pour chacune des catégories de gènes.

La croissance des cellules de carcinome mammaire est inhibée par des traitements à l'acide rétinoïque (RA). L'analyse de micropuces d'ADN de MCF7 traitées avec RA a permis d'identifier de nombreux gènes cibles potentiels. Un enrichissement d'éléments de réponse à l'acide rétinoïque (RARE) est observable dans les promoteurs de ces gènes après avoir exclus les RARE se trouvant à l'intérieur d'éléments transposables. Des RARE présents dans des éléments transposables spécifiques aux primates sont aussi fixés *in vivo* dans les promoteurs de cibles connues de RA : BTG2, CASP9 et GPRC5A. Certains gènes cibles de RA dans les MCF7 sont aussi des cibles de E2, suggérant que le contrôle que ces molécules exercent sur la prolifération est en partie attribuable à des effets opposés sur un ensemble commun de gènes.

**Mots-clés :** œstrogène, ER $\alpha$ , transcription, micropuces d'ADN, prédiction par matrice, ERE, RARE, éléments transposable

## Abstract

Two thirds of breast tumours depend on estrogens for their growth. The network of genes mediating the proliferative effect of estrogens is not fully characterized. Putative primary and secondary estrogen target genes were identified with microarray analysis of MCF7 breast cancer cells treated with estradiol (E2) in presence or absence of the protein synthesis inhibitor cycloheximide (CHX). The promoters of the target genes were screened for transcription factor binding sites with a collection of 300 matrix based DNA-binding profiles. Estrogen response elements (EREs) were enriched in the promoters of primary target genes. E2F binding sites were enriched in the promoters of secondary target genes. Similar enrichment was also observed in regions bounds by ER $\alpha$  and E2F1 in CHIP-on-chip experiments for each set of target genes.

Retinoic acid (RA) treatment of mammary carcinoma cells inhibits their growth. Putative target genes were identified through microarray analysis of MCF7 cells treated with RA. Enrichment of retinoic acid response elements (RARE) was observed in their promoters after removing the elements found within transposable elements. Although transposable elements mask the enrichment, RARE within primate specific transposable elements are bound *in vivo* by retinoic acid receptors in the promoters of known target genes BTG2, CASP9 and GPRC5A. Some of the RA target genes in MCF7 cells are also target genes of E2 suggesting that these two molecules exert their effects on cell proliferation in part by opposite action on a common set of genes.

**Keywords:** estrogens, ER $\alpha$ , transcription, microarray, matrix-based detection, ERE, RARE, transposable elements

## Table des matières

Liste des tableaux .....	ix
Liste des figures .....	x
Liste des sigles et abréviations.....	xii
PREMIÈRE PARTIE .....	1
Chapitre 1 : Les récepteurs des œstrogènes et de l'acide rétinoïque.....	2
1.1 Rôles physiologiques .....	2
1.1.1 Rôles physiologiques des œstrogènes .....	2
1.1.2 Rôles physiologiques de l'acide rétinoïque .....	6
1.2 Les maladies associées aux récepteurs de l'acide rétinoïque et des œstrogènes.....	8
1.2.1 Rôles de ER et RAR dans le cancer du sein.....	9
1.3 Dualité de fonction des récepteurs nucléaires : récepteurs et facteurs de transcription .....	10
1.3.1 Organisation modulaire des récepteurs .....	10
1.3.2 La signalisation des œstrogènes et de l'acide rétinoïque .....	12
Chapitre 2 : Rôle de la chromatine dans la régulation de l'expression des gènes .....	16
2.1 Structure tridimensionnelle et dynamique de la chromatine.....	16
2.2 Les composantes des promoteurs.....	18
2.3 Régulation de la transcription à longue distance.....	20
Chapitre 3 : Impact des éléments transposables sur le génome humain .....	24
3.1 Classes d'éléments transposables.....	24
3.1.1 LTR .....	24
3.1.2 LINE.....	25
3.1.3 SINE.....	26
3.2 Distribution dans le génome humain.....	27
3.3 Impact des éléments transposables sur les gènes .....	29
3.3.1 Mécanismes limitant la transposition des TE.....	30
3.3.2 Génération de nouveaux transcrits .....	30

3.3.2 Génération de nouveaux sites de facteurs de transcription .....	32
Chapitre 4 : Méthodes biochimiques et bio-informatiques permettant l'étude de la transcription à large échelle .....	35
4.1 Identification de gènes cibles à l'aide de micropuces .....	36
4.1.1 Micropuces d'oligonucléotides Affymetrix .....	36
4.1.2 Identification des transcrits régulés.....	37
4.1.3 Études des voies de signalisation de E2 et RA à l'aide de micropuces .....	38
4.2 Identification de régions régulatrices par immunoprécipitation de la chromatine.....	40
4.2.1 Techniques de séquençage et micropuces.....	40
4.2.2 Localisation de régions fixée par ER $\alpha$ à l'échelle du génome.....	43
4.3 Reconnaissance de sites de facteurs de transcription <i>in silico</i> .....	45
4.3.1 Représentation consensus.....	46
4.3.2 Modélisation des sites de fixations par matrice .....	47
4.3.3 Matrices de sites reconnus par ER .....	51
4.3.4 Évaluation de la surreprésentation de ERE dans les promoteurs de gènes cibles .....	52
Chapitre 5 : Objectifs .....	59
DEUXIÈME PARTIE.....	61
Chapitre 6: Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution.....	62
Abstract .....	63
Background .....	64
Results .....	66
Discussion .....	72
Conclusion .....	76
Methods.....	76
List of abbreviations.....	78
Authors' contributions.....	78

Acknowledgements.....	78
Figures and Tables .....	79
Additional files.....	88
Chapitre 7: Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells.....	112
ABSTRACT.....	113
INTRODUCTION .....	113
MATERIALS AND METHODS.....	116
RESULTS .....	121
DISCUSSION.....	131
ACKNOWLEDGEMENTS.....	136
FIGURES AND TABLES .....	137
SUPPLEMENTARY DATA .....	152
REFERENCES.....	161
Chapitre 8 : Discussion et perspectives.....	169
8.1 L'incorporation de caractéristiques des sites de haute affinité de ER et RAR améliore les modèles matriciels .....	169
8.1.1 Meilleure détection des ERE à l'aide d'une matrice palindromique .....	169
8.1.2 Meilleure détection des RARE à l'aide d'une matrice dérivée de CHIP-on-chip .....	171
8.1.3 Limitations des nouvelles matrices.....	173
8.2 Les éléments transposables ont contribué des éléments de réponse de ER et RAR au cours de l'évolution.....	174
8.2.1 Amplification des RARE par les éléments transposables.....	176
8.2.2 Les éléments transposables contribuent des ERE .....	178
8.3 ER et RAR régulent leur gènes cibles sur de longue distances.....	180
8.4 Réseaux de gènes régulés par ER et RAR .....	182

8.4.1 Convergence de résultats de CHIP-on-chip et des prédictions de ERE pour les cibles primaires des œstrogènes .....	184
8.4.2 Amplification du signal des œstrogènes par la famille E2F .....	185
8.4.3 Interférence entre les gènes régulés par E2 et RA .....	187
8.5 Conclusion .....	190
Bibliographie.....	192
ANNEXE : Figures et tableaux supplémentaires.....	I



## Liste des tableaux

Table 1 Sites de fixation de récepteurs nucléaires provenant de TE validées par CHIP .....	33
Table 2 Micropuces avec traitement E2.....	39
Table 3 Micropuces avec traitement RA.....	39
Table 4 Fréquence des ERE à proximité de régions liées par ERa CHIP-on-chip .....	45
Table 5 Symboles représentant les acides nucléiques.....	47
Table 6 Enrichissement de RARE DR5 dans les promoteurs de gènes cibles observé avec les deux matrices.....	V
Table 7 Facteurs de transcription régulés par E2 ayant des matrices disponibles dans TRANSFAC.....	VIII
Table 8 Gènes Up E2+CHX avec des ERE prédits impliqués dans la prolifération et la régulation de la croissance .....	IX
Table 9 Gènes Up E2-CHX avec des sites E2F prédits impliqués dans la prolifération et le cycle cellulaire .....	IX
Table 10 Gènes ayant des réponses différentes suite à des traitements avec E2 et RA.....	XIII
Table 11 Gènes ayant des réponses différentes suite à des traitements avec E2 dans les MCF7 et RA dans les SKBR3.....	XIV
Table 12 Facteurs de transcription régulés par RA dans les MCF7.....	XVI

## Liste des figures

Figure 1 Structure de l'estradiol (E2).....	3
Figure 2 Structure de l'acide rétinoïque (RA).....	6
Figure 3 Domaines fonctionnels des récepteurs nucléaires .....	11
Figure 4 Voies de signalisation des récepteurs nucléaires .....	12
Figure 5 Éléments de réponse des récepteurs nucléaires .....	14
Figure 6 Empaquetage de la chromatine.....	17
Figure 7 Classes de promoteurs .....	19
Figure 8 Régions impliquées dans la régulation de la transcription des gènes.....	21
Figure 9 Territoires chromatiniens et régions frontières.....	23
Figure 10 Structure des LTR.....	25
Figure 11 Structure des LINE .....	26
Figure 12 Structure des SINE .....	27
Figure 13 Proportion du génome humain correspondant à des éléments transposés.....	28
Figure 14 Distribution des TE à proximité des gènes.....	29
Figure 15 Exemples d'exons dérivés de TE.....	32
Figure 16 Fréquence des TE près des groupes de gènes HOX et KLK .....	34
Figure 17 Gènes sur les puces Affymetrix.....	37
Figure 18 Régions identifiées à l'aide de micropuces et séquençage.....	41
Figure 19 Distribution des ERE dans les régions liées par ER $\alpha$ ChIP-PET .....	42
Figure 20 Contribution des TE aux régions liées par ER $\alpha$ .....	43
Figure 21 Exemple de matrice représentant un demi ERE .....	48
Figure 22 Exemple de matrice pondérée.....	49
Figure 23 Matrice ERE de TRANSFAC.....	52
Figure 24 Simulation Monte Carlo des ERE 70% pour les gènes induits par les œstrogènes seulement en présence de cycloheximide .....	54

Figure 25 Évaluation de l'enrichissement des ERE à l'aide de simulations et tests statistiques.....	56
Figure 26 Enrichissement des ERE et des sites E2F dans les promoteurs de gènes cibles..	57
Figure 27 Absence d'enrichissement des ERE dans les promoteurs de gènes cibles E2F et SP1 .....	58
Figure 28 Comparaison des scores des matrices ERE .....	I
Figure 29 courbe ROC des matrices ERE.....	II
Figure 30 Matrice RARE DR5 provenant d'une compilation de la littérature .....	III
Figure 31 Matrice RARE DR5 dérivée de données de <i>ChIP-on-chip</i> .....	III
Figure 32 Comparaison des scores assignés par les matrices RARE DR5 .....	IV
Figure 33 Courbes ROC des matrices RARE DR5.....	VI
Figure 34 Distribution des RARE DR5 à +/- 25 kpb du TSS de gènes régulés par RA.....	VII
Figure 35 Contribution des TE aux ERE et RARE prédits dans le génome humain .....	VII
Figure 36 Enrichissement DR5 gènes induits par RA dans les MCF7 .....	X
Figure 37 Enrichissement DR5 gènes induits par RA en présence de cycloheximide dans les SKBR3 .....	XI
Figure 38 Gènes régulés par E2 et RA.....	XII

## Liste des sigles et abréviations

ADN	:	acide désoxyribonucléique
ADNc	:	ADN complémentaire (ARN retranscrit en ADN)
AF	:	fonction activatrice 1 ou 2
AP1	:	<i>activator protein 1</i>
APL	:	leucémie aiguë promyélocytaire
AR	:	récepteur des androgènes
ARE	:	éléments de réponse aux androgènes
ARN	:	acide ribonucléique
BRE	:	<i>Transcription Factor II B Recognition Element</i>
ChIP	:	immunoprécipitation de la chromatine
ChIP-on-chip	:	ChIP suivit d'hybridation sur des micropuces d'ADN
CHX	:	cycloheximide (inhibiteur de synthèse protéique)
DBD	:	domaine de liaison à l'ADN
DPE	:	<i>downstream promoter element</i>
E2	:	17 $\beta$ -estradiol
ER	:	récepteurs des oestrogènes ( $\alpha$ ou $\beta$ )
ERE	:	éléments de réponse aux estrogènes
GR	:	récepteur des glucocorticoïdes
LBD	:	domaine de liaison du ligand
LINE	:	<i>Long Interspersed Nucleotide Elements</i>
LTR	:	<i>Long Terminal Repeat</i>
MCF7	:	lignée cellulaire de carcinome mammaire exprimant ER
pb	:	paire de base (nucléotide)
PR	:	récepteur de la progestérone
RA	:	acide rétinoïque
RAR	:	récepteurs de l'acide rétinoïque ( $\alpha$ ou $\beta$ ou $\gamma$ )

RARE :	éléments de réponse à l'acide rétinoïque
RXR :	récepteur rétinoïde X ( $\alpha$ ou $\beta$ ou $\gamma$ )
SKBR3:	lignée cellulaire de carcinome mammaire exprimant RAR mais pas ER
SINE :	<i>Short Interspersed Nucleotide Elements</i>
SP1 :	<i>specificity protein 1</i>
SRC :	<i>steroid receptor coactivator</i>
SWI/SNF :	famille de protéine reliée à SWI/SNF ( <i>matrix-associated, actin-dependant regulator of chromatin</i> )
TE :	éléments transposables (SINE, LINE, LTR)
TSS :	site d'initiation de la transcription
VDR :	récepteur de la vitamine D
VDRE :	éléments de réponse à la vitamine D

# **PREMIÈRE PARTIE**

Chapitre 1 : Les récepteurs des œstrogènes et de l'acide rétinoïque

Chapitre 2 : Rôle de la chromatine dans la régulation de l'expression des gènes

Chapitre 3 : Impact des éléments transposables sur le génome humain

Chapitre 4 : Méthodes biochimiques et bio-informatiques permettant l'étude de la transcription à large échelle

Chapitre 5 : Objectifs

# Chapitre 1 : Les récepteurs des œstrogènes et de l'acide rétinoïque

Les récepteurs des œstrogènes (ER  $\alpha$  et  $\beta$ ) et de l'acide rétinoïque (RAR  $\alpha$ ,  $\beta$  et  $\gamma$ ) sont des facteurs de transcription appartenant à la superfamille des récepteurs nucléaires. Comme dans le cas de plusieurs membres de cette famille, l'impact des RAR et ER sur la transcription des gènes est modulé par la liaison de leurs ligands respectifs. Les œstrogènes et l'acide rétinoïque jouent des rôles cruciaux dans de nombreux processus biologiques incluant le développement, la croissance et l'homéostasie. En raison de leur importance dans ces processus, des dérèglements de la signalisation par les récepteurs sont associés à plusieurs pathologies. Les ER et RAR contrôlent la transcription de leurs gènes cibles par plusieurs voies : ils peuvent à la fois lier des séquences spécifiques d'ADN et interagir avec d'autres facteurs impliqués dans la transcription ou dans la signalisation intracellulaire.

## 1.1 Rôles physiologiques

### 1.1.1 Rôles physiologiques des œstrogènes

Les œstrogènes sont des hormones stéroïdiennes dérivées du cholestérol. La principale forme d'œstrogène est l'estradiol (E2, Figure 1). Plusieurs enzymes de la famille cytochrome P450 interviennent dans la conversion du cholestérol en androgènes, puis ces derniers sont convertis par aromatisation en œstrogènes. Chez la femme, avant la ménopause, la majeure partie des œstrogènes sont synthétisés par les ovaires et ensuite relâchés dans la circulation sanguine avant d'atteindre les tissus cibles. Chez l'homme et chez la femme après la ménopause, les œstrogènes sont synthétisés localement dans les tissus cibles par aromatisation des androgènes [1, 2].

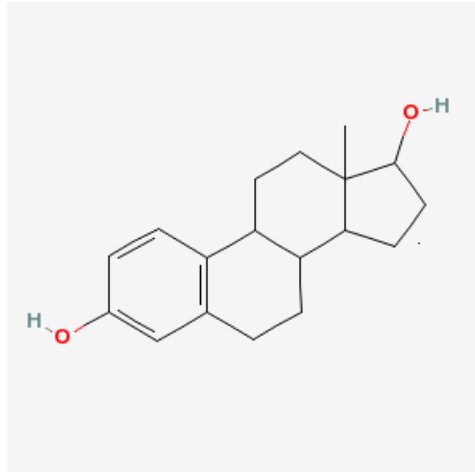


Figure 1 Structure de l'estradiol (E2)

La figure provient de <http://pubchem.ncbi.nlm.nih.gov>.

### 1.1.1.1 Le tissu osseux

Les œstrogènes régulent l'homéostasie du squelette tant chez les femmes que chez les hommes [3, 4]. Les os sont composés de cellules, appelées ostéocytes, entourées d'une matrice extracellulaire minéralisée. Deux types cellulaires régénèrent constamment la matrice osseuse : les ostéoclastes résorbent l'ancienne matrice osseuse rapidement et les ostéoblastes synthétisent plus lentement la nouvelle. Les œstrogènes sont impliqués dans le maintien de l'équilibre entre ostéoblastes et ostéoclastes en modulant la durée de vie de ces deux types cellulaires. Les œstrogènes protègent les ostéoblastes de l'apoptose et ils induisent celle-ci chez les ostéoclastes. Les risques d'ostéoporose liés à l'âge sont dus à la baisse de synthèse d'œstrogènes chez la femme après la ménopause et à la baisse de synthèse d'androgènes chez l'homme.



### 1.1.1.2 Le système cardiovasculaire

L'incidence de maladies cardiovasculaires est beaucoup moins élevée chez les femmes avant la ménopause que chez les hommes du même âge. Après la ménopause, les risques de maladies chez les femmes rejoignent celui des hommes. Plusieurs études épidémiologiques ont suggéré un effet protecteur des œstrogènes sur le système cardiovasculaire [5-7]. Une partie des effets bénéfiques est attribuable à l'impact des œstrogènes sur le taux de lipoprotéines associées au cholestérol dans le sang. Les œstrogènes augmentent le taux de HDL (« High Density Lipoproteins ») et diminuent la concentration des LDL (« Low Density Lipoproteins ») qui peuvent contribuer à l'apparition de graves problèmes vasculaires comme l'artériosclérose. Les œstrogènes agissent aussi plus directement sur les cellules musculaires lisses et l'endothélium qui composent les vaisseaux sanguins. Ces deux types cellulaires expriment les récepteurs des œstrogènes. La prolifération des cellules musculaires lisses est une des conséquences de lésions des vaisseaux sanguins. Plusieurs études utilisant des modèles animaux démontrent que des traitements aux œstrogènes atténuent la prolifération des cellules musculaires lisses [8]. Les œstrogènes modulent aussi la production d'oxyde nitrique (NO, un vasodilatateur) et d'endothéline (un vasoconstricteur) par les cellules de l'endothélium des vaisseaux sanguins.

### 1.1.1.3 L'appareil reproducteur et les glandes mammaires

Les différentes composantes de l'appareil reproducteur sont de cibles majeures des hormones stéroïdiennes. Ces hormones jouent des rôles prépondérants dans la différenciation de ces tissus au cours du développement fœtal ainsi que dans leur maturation à la puberté [9, 10]. Même si le sexe de l'embryon est déterminé dès la fécondation, toutes les structures du tractus reproducteur ont *a priori* un programme de différenciation de type féminin. Chez les embryons de sexe génétique masculin, l'expression du facteur de transcription SRY (« sex determining region Y »), situé sur le

chromosome Y, initie la différenciation des gonades indifférenciées en testicules. Par la suite, les structures du tractus reproducteur se masculinisent sous l'action des androgènes et des hormones antimüllériennes (AMH). Les œstrogènes ne semblent pas jouer de rôle dans la différenciation et le développement initial du tractus reproducteur féminin. Des études menées avec des souris mutantes ont observé des problèmes de maturation de l'appareil reproducteur suite à l'inactivation des récepteurs des œstrogènes [8]. Malgré un développement embryonnaire normal du tractus reproducteur, l'absence de ER a un impact sur la fertilité des souris mâles et femelles. Les souris mâles présentent une atrophie testiculaire et une spermatogenèse déficiente tant au niveau du nombre de spermatozoïdes produits que de la motilité de ces derniers malgré des niveaux d'androgènes circulant normaux. Chez les souris femelles, l'absence de ER a des conséquences sur la maturation de l'ensemble du tractus reproducteur. Les ovaires de ces souris sont anovulatoires et élargis. Les cellules épithéliales des trompes de Fallope et l'utérus sont insensibles aux œstrogènes. Normalement, l'augmentation de la concentration d'hormones durant le cycle ovarien entraîne la prolifération et la différenciation de ces cellules.

À l'instar du tractus reproducteur féminin, le développement embryonnaire et foetal des glandes mammaires ne dépend pas des oestrogènes. Elles sont constituées de tissus épithéliaux tubulo-alvéolaires et d'un stroma de tissus annexes (adipocytes, vaisseaux sanguins, etc...) [10]. Les glandes mammaires sont rudimentaires à la naissance, les stéroïdes ovariens sont essentiels aux étapes subséquentes de développement à la puberté, durant la vie adulte et la grossesse [8, 11]. Jusqu'à la ménopause, les oestrogènes stimulent la prolifération et la différenciation de l'épithélium des canaux lactophores et la formation de lobes alvéolaires nécessaire à la lactation durant les cycles ovariens. Une partie des effets des oestrogènes sur les cellules épithéliales provient de l'induction par ER du récepteur de la progestérone (PR) et de facteurs de croissance.

## 1.1.2 Rôles physiologiques de l'acide rétinoïque

L'acide rétinoïque (RA, Figure 2) est un métabolite actif de la vitamine A. La seule source de cette vitamine, également appelée rétinol, est l'alimentation étant donné qu'aucune espèce animale peut la synthétiser *de novo*. La majeure partie de la vitamine A provenant de l'alimentation est mise en réserve dans le foie. Chez les mammifères, de 50 à 80% du rétinol total de l'organisme se trouve dans le foie [12]. Le rétinol stocké dans le foie est libéré dans la circulation sanguine en fonction des besoins de l'organisme. Dans les tissus cibles, deux familles d'enzymes interviennent dans la conversion de la vitamine A en acide rétinoïque [13]. Des alcools déshydrogénases oxydent le rétinol de façon réversible en un métabolite intermédiaire appelé rétinal. Le rétinal peut ensuite être oxydé de façon irréversible en acide rétinoïque par des aldéhydes déshydrogénases.

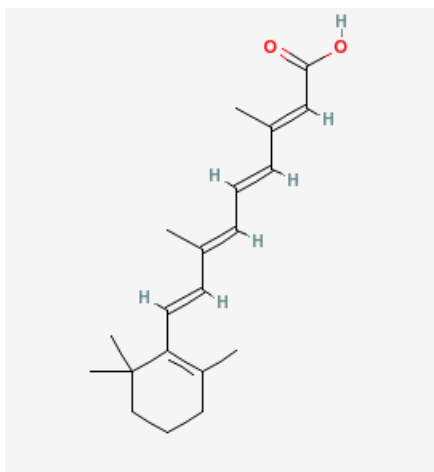


Figure 2 Structure de l'acide rétinoïque (RA)

La forme tout-trans de RA est représentée. La figure provient de <http://pubchem.ncbi.nlm.nih.gov>.

### 1.1.2.1 Développement embryonnaire

Les rétinoïdes sont essentiels pour le développement de l'embryon. Une carence ou un excès d'acide rétinoïque durant l'embryogénèse est tératogène [14]. Les déformations du

fœtus affectent le cœur, le squelette, les membres, les yeux, le cerveau et le système nerveux central. Plusieurs études ont été réalisées avec des souris mutantes pour un des RAR ou une combinaison de ces récepteurs. Les mutants nuls pour un seul des RAR présentent des anomalies limitées dans la plupart des cas suggérant une redondance fonctionnelle des récepteurs lors du développement. Les souris mutantes pour deux des trois RAR meurent *in utero* ou peu après la naissance. L'anatomie de ces doubles mutants reproduit la plupart des anomalies associées au syndrome de carence fœtale en vitamine A. La synthèse d'acide rétinoïque à partir du rétinol par l'embryon est également nécessaire au début de la morphogenèse et à sa survie. L'inactivation chez des embryons de souris du gène de la rétinaldéhyde déhydrogénase-2 (Raldh2), une des enzymes responsable de la conversion du rétinol en acide rétinoïque, cause plusieurs malformations et leur mort avant la fin de la gestation [15]. L'administration d'acide rétinoïque à la mère permet de compenser la plupart des défauts observés. Ces embryons mutants ont plusieurs anomalies du système nerveux central, leur région frontonasale est tronquée et leur cœur comporte une seule cavité dilatée. Ils présentent aussi un raccourcissement de l'axe antéro-postérieur et leurs membres ne sont pas formés. La détermination de l'axe antéro-postérieur de l'embryon nécessite l'expression des gènes homéotiques HOX. Les gènes HOX sont induits de façon séquentielle en présence d'acide rétinoïque [16]. L'expression de certains d'entre eux est due à la fixation à l'ADN de RAR sur des éléments de réponse à l'acide rétinoïque (RARE) [17-21].

### **1.1.2.2 La vision**

Chez les vertébrés, la vitamine A est essentielle au développement embryonnaire des yeux et à la vision [12, 22]. L'œil est l'organe où l'on retrouve la plus grande concentration de rétinol après le foie. Une mauvaise vision sous faible éclairage est un des premiers signes d'une carence alimentaire en vitamine A. Dans les cellules en bâtonnet de la rétine, le rétinol dérivé du rétinol est lié de façon covalente à des protéines appelées opsines pour former un complexe hautement sensible à la lumière. Les photons de la

lumière provoquent une dissociation du rétinol de l'opsine. Cette dissociation fait partie des mécanismes permettant de convertir les informations lumineuses en signaux électriques envoyés au cerveau par les nerfs optiques.

### **1.1.2.3 Croissance et différenciation des muqueuses et de la peau**

Un des rôles physiologiques de la vitamine A et ses récepteurs est le maintien de l'intégrité de l'épithélium de la peau et des muqueuses [23]. La peau est formée de deux principales couches : le derme et l'épiderme. Ce dernier constitue la couche superficielle de la peau. Il est composé de plusieurs strates différenciées de cellules appelées kératinocytes. Les kératinocytes se renouvellent constamment à partir de cellules souches situées entre le derme et l'épiderme. L'acide rétinoïque stimule la croissance des kératinocytes à partir des cellules souches et module leur différenciation dans les différentes strates de l'épiderme [24]. Au cours de leur différenciation, les kératinocytes produisent des protéines fibreuses spécifiques à leur stade de différenciation : les kératines. Des analogues de la vitamine A sont utilisés pour le traitement de maladies de la peau comme l'acné kystique et pustuleuse et des formes graves du psoriasis [25, 26]. Contrairement à l'épithélium de la peau, les cellules de l'épithélium des muqueuses (cavité orale, tube digestif, voies respiratoires, tractus uro-génital) ne produisent généralement pas de kératines durant leur différenciation [27]. Plusieurs études réalisées avec des animaux soumis à une diète déficiente en vitamine A ont observé une kératinisation de l'épithélium de plusieurs muqueuses [23]. Cette différenciation aberrante de l'épithélium est en général réversible par des traitements à l'acide rétinoïque.

## **1.2 Les maladies associées aux récepteurs de l'acide rétinoïque et des œstrogènes**

Les récepteurs des œstrogènes et de l'acide rétinoïque sont impliqués dans de nombreuses pathologies. Les œstrogènes participent au développement ou à la progression

de nombreuses maladies dont certains cancers [7]. Les œstrogènes ont des effets protecteurs contre l'ostéoporose, les maladies cardiovasculaires et les maladies neurodégénératives (accident cérébrovasculaire, Parkinson, Alzheimer). Plusieurs études cliniques relient les récepteurs des œstrogènes aux cancers du sein, de l'endomètre, du colon et de la prostate. Les récepteurs de l'acide rétinoïque sont associés à divers cancer et maladies de la peau [28, 29]. De nombreux tissus néoplasiques, dont les poumons et la glande mammaire, perdent sélectivement l'expression de RAR $\beta$ . Une translocation entraînant la fusion des gènes PML et RAR $\alpha$  cause la leucémie aiguë promyélocytaire (APL). L'administration d'acide rétinoïque permet la différenciation des promyélocytes et induit des rémissions complètes dans 75% des cas [30]. Plusieurs études réalisées avec des modèles animaux ont démontré la capacité des rétinoïdes naturels et de synthèse à supprimer la carcinogénèse dans une variété de cancers comprenant le cancer des voies orales, de la prostate et le cancer du sein [29, 31].

Les rôles de ER et RAR dans le cancer du sein, qui constitue le contexte biologique de la plupart de mes recherches, sont détaillés ci-dessous.

### **1.2.1 Rôles de ER et RAR dans le cancer du sein**

Le cancer du sein est la forme de cancer la plus diagnostiquée chez la femme. La majorité des cancers du sein ont une croissance hormonodépendante associée à la surexpression des récepteurs des œstrogènes (ER $\alpha$ ). Selon la technique de détection utilisée, le taux de positivité pour ER varie de 59% à 81% [32]. La concentration locale en œstrogènes est dix fois plus élevée dans les tissus cancéreux que dans les tissus normaux [33, 34]. Les traitements actuels contre le cancer du sein ciblent autant la synthèse locale d'œstrogènes par les tumeurs que les récepteurs des œstrogènes [7]. Des anti-œstrogènes sont utilisés pour inhiber l'activité de ER (ex. tamoxifène et fluvestrant), certains réduisent aussi les niveaux de ER (ex. fluvestrant). Des inhibiteurs de l'aromatase (ex. anastrozole) sont utilisés pour bloquer la synthèse d'œstrogènes.

À l'instar de ce qui est observé dans d'autres cancers, la signalisation par les RAR est altérée dans le cancer du sein. Plusieurs lignées humaines de carcinome mammaire ont une capacité réduite à synthétiser l'acide rétinoïque [13]. Cette baisse de capacité peut entraîner des changements majeurs étant donné que les RAR sont des répresseurs de la transcription en absence de ligand [28]. De plus, deux isoformes de RAR, RAR $\alpha$ 2 et RAR $\beta$ 2, sont inactivés par des modifications épigénétiques de leurs promoteurs [35, 36]. Les promoteurs de ces deux isoformes contiennent des RARE permettant leur autorégulation par RAR [28]. Un isoforme de RAR, RAR $\alpha$ 1 est toutefois induit par les œstrogènes suite à la fixation de ER sur un élément de réponse aux œstrogènes (ERE) dans son promoteur [37]. La croissance des cellules de carcinome mammaire qui expriment RAR $\alpha$  est inhibée par des traitements à l'acide rétinoïque [38].

## **1.3 Dualité de fonction des récepteurs nucléaires : récepteurs et facteurs de transcription**

### **1.3.1 Organisation modulaire des récepteurs**

ER et RAR font partie de la superfamille des récepteurs nucléaires qui compte 48 membres chez l'humain [39]. Tous les récepteurs nucléaires possèdent une structure modulaire constituée de cinq à six domaines (A à F, Figure 3). Les domaines de liaison à l'ADN (DBD, C) et au ligand (LBD, E) sont les deux domaines les plus importants et les plus conservés au cours de l'évolution de la superfamille. Les récepteurs nucléaires ont deux fonctions d'activation de la transcription. La première, AF-1, indépendante du ligand se situe dans le domaine A/B. La longueur et la séquence de ce domaine varient beaucoup d'un récepteur nucléaire à l'autre. La deuxième, AF-2, dépendante du ligand, se situe dans le domaine E. Les fonctions d'activation de la transcription sont impliquées dans le recrutement de diverses protéines co-régulatrices capables de remodeler la chromatine ou modifier les histones de façon post-traductionnelle. Le domaine D, également peu conservé

entre les récepteurs nucléaires, contient le signal de localisation nucléaire et sert de charnière entre le DBD et le LBD. En plus de la fonction AF-2 et du LBD, le domaine E contient l'interface de dimérisation qui permet selon le récepteur de former des homodimères ou des hétérodimères avec les récepteurs rétinoïde X (RXR). Comme les autres facteurs de transcription, les récepteurs nucléaires sont sujets à des modifications post-traductionnelles. Ces modifications comme la phosphorylation, l'acétylation ou l'ubiquitinylation modulent l'activité des fonctions AF-1 et AF2, la stabilité du récepteur et l'interaction avec les co-régulateurs [40].

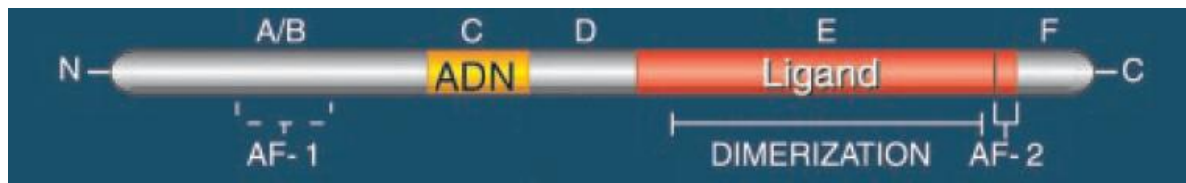


Figure 3 Domaines fonctionnels des récepteurs nucléaires

La figure est adaptée de [41].

Deux récepteurs nucléaires ont été décrits pour les œstrogènes : ER $\alpha$  (ESR1) et ER $\beta$  (ESR2). La séquence d'acides aminés du DBD des deux récepteurs est identique à 97%, leur LBD est identique à 56% et leur portion N-terminale est identique à 24% [42]. La fonction AF-1 de ER $\alpha$  est plus active que celle de ER $\beta$ . Chez l'humain, sept promoteurs ont été caractérisés pour ER $\alpha$  [43] et deux promoteurs pour ER $\beta$  [42]. Les deux récepteurs forment des homodimères ou des hétérodimères.

Deux familles de récepteurs nucléaires comptant chacune trois membres médient les effets physiologiques de l'acide rétinoïque [28, 44]. Les récepteurs de l'acide rétinoïque (RAR) ont comme ligand les isomères tout-trans et 9-cis de l'acide rétinoïque. Les récepteurs rétinoïde X (RXR) ont comme ligand l'isomère 9-cis de l'acide rétinoïque. Chaque RAR a plus d'un isoforme issu de deux promoteurs : RAR $\alpha$  ( $\alpha$ 1 et  $\alpha$ 2), RAR $\beta$  ( $\beta$ 1 à  $\beta$ 4), et RAR $\gamma$  ( $\gamma$ 1 et  $\gamma$ 2). Chaque RXR a deux isoformes : RXR $\alpha$  ( $\alpha$ 1 et  $\alpha$ 2), RXR $\beta$  ( $\beta$ 1 et



$\beta 2$ ), et  $RXR\gamma$  ( $\gamma 1$  et  $\gamma 2$ ). Les isoformes de RAR et RXR diffèrent les uns des autres dans leur région N-terminale. Les isoformes  $\alpha 2$ ,  $\beta 2$ , et  $\gamma 2$  de RAR sont inductibles par les rétinoïdes grâce à des RARE dans leurs promoteurs. RXR forme des homodimères ou des hétérodimères avec RAR.

### 1.3.2 La signalisation des œstrogènes et de l'acide rétinoïque

Les œstrogènes et l'acide rétinoïque influencent la transcription des gènes par plus d'une voie de signalisation (Figure 4). Leurs récepteurs peuvent se fixer sur les séquences régulatrices de leurs gènes cibles (voie classique), ils peuvent être recrutés par d'autres facteurs de transcription (voie alternative) et ils peuvent moduler l'activité de protéines signalisatrices (voie non-génomique).

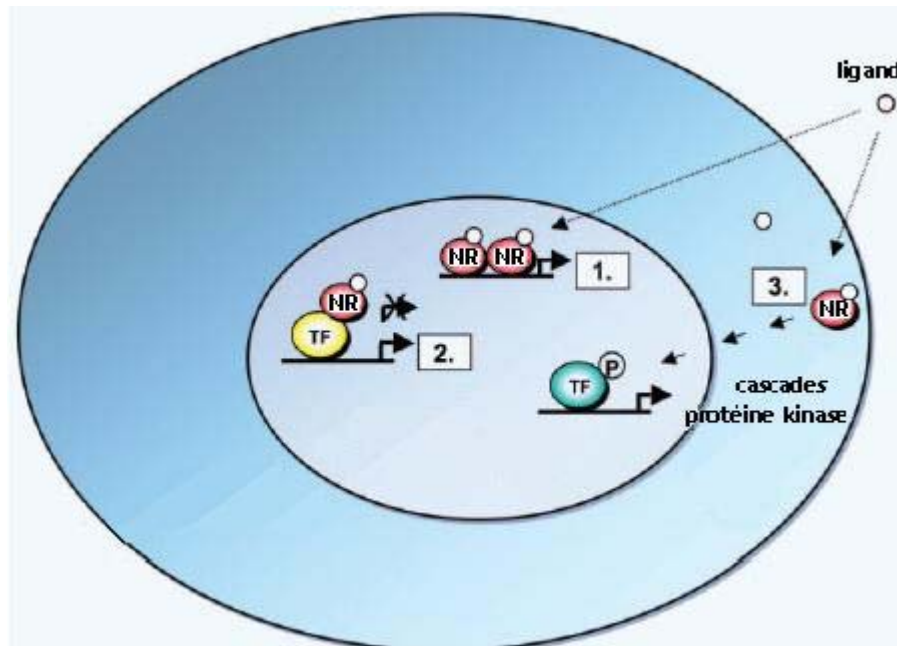


Figure 4 Voies de signalisation des récepteurs nucléaires

1- Signalisation nucléaire classique. 2- Signalisation nucléaire alternative 3- Signalisation non-génomique  
 « NR » représente un récepteur nucléaire. « TF » représente un facteur de transcription. La figure est adaptée de [45].

### 1.3.2.1 Signalisation nucléaire classique

La plupart des récepteurs nucléaires lient l'ADN sous forme de dimères avec d'autres membres de la superfamille des récepteurs nucléaires (Figure 5). Le DBD de chaque monomère reconnaît un demi-site de six nucléotides. Selon les dimères de récepteurs, l'orientation des demi-sites forme des palindromes (« *Inverted Repeat* »), des répétitions directes (« *Direct Repeat* ») ou des palindromes inversés (« *Everted Repeat* ») [41, 46]. La composition des demi-sites et l'espacement entre les deux demi-sites varient aussi entre les dimères de récepteurs. Le récepteur des œstrogènes (ER) et celui des androgènes (AR) reconnaissent des palindromes dont les demi-sites sont espacés de trois nucléotides (IR3), mais les demi-sites pour lesquels ils ont plus d'affinité sont différents. ER reconnaît les demi-sites (A/G)GGTCA et AR reconnaît (A/G)GNACA [47]. Les récepteurs de l'acide rétinoïque (RAR), de la vitamine D (VDR) et de l'hormone thyroïdienne (TR), qui forment des hétérodimères avec RXR, reconnaissent des demi-sites (A/G)G(G/T)TCA [48]. Des analyses de promoteurs de gènes cibles des RAR ont montré que ce récepteur reconnaît des demi-sites directement répétés avec un espacement de un, deux ou cinq nucléotides (DR1, DR2, DR5). VDR et TR reconnaissent les demi-sites directement répétés avec un espacement de trois et quatre nucléotides respectivement (DR3, DR4). Il est à noter que certains arrangements de demi sites sont reconnus par plusieurs récepteurs nucléaires. En plus de RAR, HNF4A, NR2F2, PPAR et des homodimères de RXR peuvent fixer des DR1 [49]. RAR et VDR reconnaissent les mêmes palindromes inversés séparés par huit nucléotides (ER8) *in vitro* et *in vivo* [50].

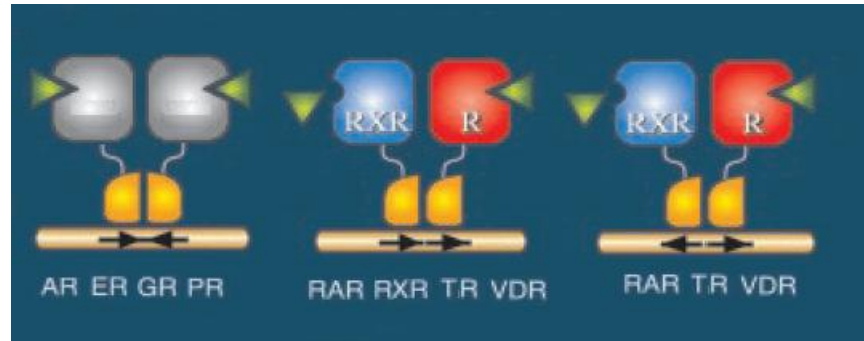


Figure 5 Éléments de réponse des récepteurs nucléaires

Les récepteurs nucléaires reconnaissent des demi-sites de 6 nucléotides arrangés en palindromes ( $\rightarrow\leftarrow$ ), répétitions directes ( $\rightarrow\rightarrow$ ) et palindromes inversés ( $\leftarrow\rightarrow$ ). La figure est adaptée de [41].

En absence d'acide rétinoïque, RAR et RXR se retrouvent principalement dans le noyau des cellules. Des dimères de RXR et RAR peuvent lier des DR1, DR2 ou DR5, aussi appelés éléments de réponse à l'acide rétinoïque (RARE), et réprimer la transcription des gènes par l'intermédiaire de co-répresseurs (SMRT/NCoR) qui recrutent des histones déacétylases [51]. La liaison des ligands entraîne des changements de conformation à RXR et RAR favorisant le relâchement des co-répresseurs et le recrutement de co-activateurs. Les co-activateurs comprennent des complexes avec activité histone acétylase (p160, p300), des complexes de remodelage de la chromatine (SWI/SNF) et des composantes du complexe médiateur (TRAP/DRIP) qui aident à l'assemblage du complexe de pré-initiation de la transcription.

En absence d'œstrogènes, ER est maintenu inactif par des protéines chaperonnes (HSP90) dans le cytoplasme et le noyau des cellules. La liaison d'œstrogènes aux récepteurs entraîne la dissociation des chaperonnes et des changements de conformation favorisant la dimérisation des récepteurs. Les dimères de ER peuvent lier des IR3, aussi appelés éléments de réponses aux œstrogènes, et induire la transcription des gènes en recrutant des co-activateurs [45].

### 1.3.2.2 Signalisation nucléaire alternative

Les récepteurs des oestrogènes et les récepteurs de l'acide rétinoïque peuvent moduler la transcription de gènes indépendamment de leur domaine de liaison à l'ADN en étant recrutés comme co-régulateurs par d'autres facteurs de transcription (Figure 4-2). AP1 (JUN, FOS) et SP1 sont des exemples de facteurs de transcription qui recrutent ER et RAR comme co-régulateurs. ER et RAR peuvent augmenter la transcription de certains gènes cibles de AP1 [52, 53]. En général, ER augmente la transcription de certains gènes cibles de AP1 alors que RAR les réprime [54, 55].

### 1.3.2.3 Signalisation non-génomique

Les récepteurs des oestrogènes et les récepteurs de l'acide rétinoïque peuvent influencer la transcription de gènes de façon non-génomique (Figure 4-3) en modulant des voies de signalisation intracellulaires résultant en la phosphorylation de facteurs de transcription [45, 56, 57]. Plusieurs protéines kinases peuvent être activées par les oestrogènes ou l'acide rétinoïque : PI3K (« *phosphatidylinositol 3-kinase* »), MAP kinases (« *mitogen-activated protein kinases* ») et PKA. L'activation de ces protéines kinases a des répercussions sur l'activité de facteurs de transcription comme AP1, SP1, NFkB et plusieurs récepteurs nucléaires.

## **Chapitre 2 : Rôle de la chromatine dans la régulation de l'expression des gènes**

Les cellules sont influencées par des molécules spécifiques de leur environnement. Plusieurs mécanismes contrôlent la durée et l'amplitude de la réponse à ces signaux. La transcription des gènes est une des étapes importantes dans la régulation de la signalisation des cellules eucaryotes.

Les principes généraux de la transcription ont été établis bien avant que la séquence du génome humain soit connue. Près de 10% des gènes codent des protéines impliquées dans la régulation de la transcription [58]. Ces protéines peuvent être regroupées en trois catégories : les facteurs de transcription capables de lier l'ADN, les protéines associées à l'ARN polymérase II catalysant la synthèse de l'ARN et les complexes de remodelage de la chromatine et modification des histones.

### **2.1 Structure tridimensionnelle et dynamique de la chromatine**

L'ADN contenu dans le noyau des cellules est organisé en une structure nucléoprotéique appelée chromatine. Cette structure est nécessaire pour accommoder les 3 milliards de paires de base du génome humain (environ 2 mètres) dans le noyau qui fait environ 10  $\mu\text{m}$  de diamètre et également contrôler l'accessibilité de l'ADN aux facteurs de transcription [59]. L'empaquetage de l'ADN joue un rôle aussi important que la séquence dans la régulation de l'expression des gènes.

Les nucléosomes représentent le niveau d'organisation de base de la chromatine. Un nucléosome se compose d'un octamère de protéines appelées histone (H2A, H2B, H3, H4) autour duquel 146 paires de base d'ADN sont enroulées. L'histone H1 (histone lien) lie l'ADN et se trouve entre les nucléosomes. Les nucléosomes sont enroulés pour former une

fibre de 30 nm. Ces fibres sont à leur tour enroulées en différentes boucles et super-hélices pour finalement former les chromosomes (Figure 6) [59-61]. La chromatine n'est pas uniforme dans le noyau. Il y a des régions où l'ADN est densément compacté et d'autres où l'ADN est moins compacté comme celles où des gènes sont activement transcrits [61].

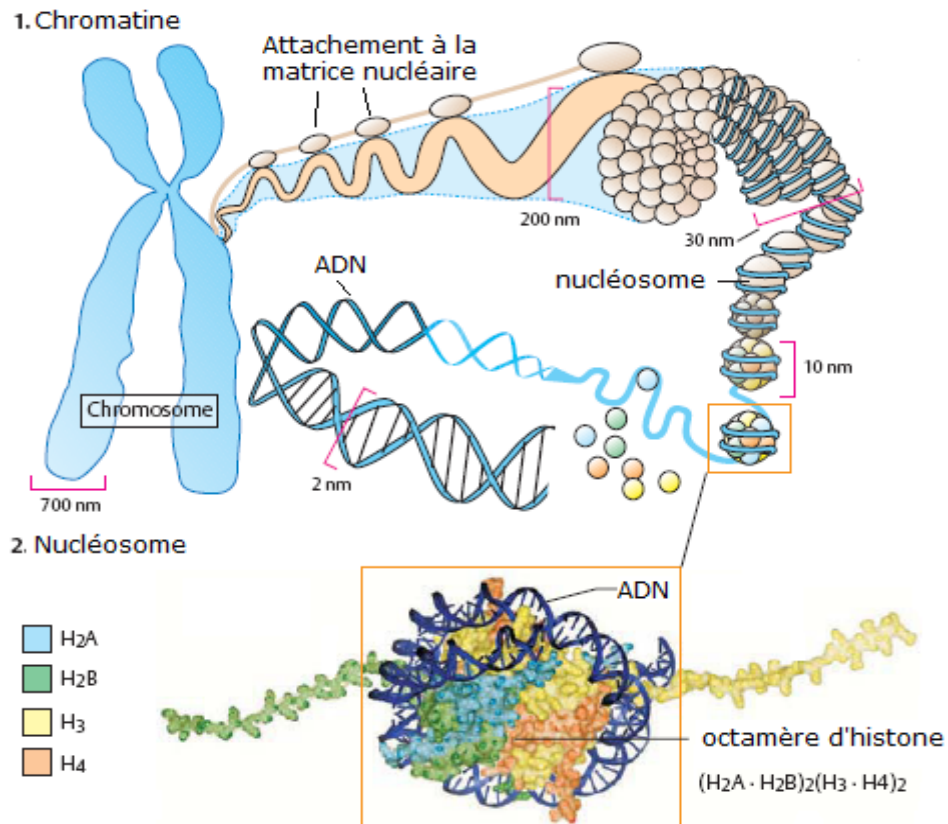


Figure 6 Empaquetage de la chromatine

La figure est adaptée de [62].

Les histones, protéines composant les nucléosomes, peuvent être modifiées de façon post-traductionnelle. Les modifications des histones modulent dynamiquement la structure de la chromatine et sont en général réversibles. Les principales modifications sont l'acétylation, la méthylation, la phosphorylation et l'ubiquitinylation. L'ensemble de ces modifications post-traductionnelles des histones forme un code impliqué non seulement

dans la transcription mais aussi dans la réplication, la recombinaison et la réparation. Les histones H3 (lysine 9) des régions du génomes transcriptionnellement inactives sont en général méthylées tandis que les histones H3 des régions actives sont hyperacétylées [60].

Plusieurs facteurs de transcription dont les récepteurs nucléaires interagissent avec des protéines co-régulatrices capables d'acétyler les histones ou remodeler la chromatine. L'acétylation déstabilise l'interaction entre l'ADN et les histones, ce qui relâche la chromatine. Les complexes de remodelage de la chromatine, comme SWI/SNF, rendent l'ADN plus accessible en facilitant le déplacement des nucléosomes. Ces changements de la structure de la chromatine permettent à la machinerie transcriptionnelle de base d'être recrutée au promoteur minimal [61].

## 2.2 Les composantes des promoteurs

Les gènes sont transcrits par la machinerie transcriptionnelle de base. Cette dernière est recrutée au site d'initiation de la transcription (TSS, +1) suite au remodelage de la chromatine par des co-régulateurs interagissant avec les facteurs de transcription. Plusieurs études ont caractérisé les éléments minimaux nécessaires pour la transcription basale des gènes. Il n'y a pas de promoteur universel chez les eucaryotes mais plutôt un ensemble de séquences pouvant faire partie de différentes classes de promoteurs minimaux [63, 64].

Les éléments qui peuvent être trouvés dans les promoteurs minimaux des gènes sont : les boîtes TATA, l'élément initiateur (Inr), l'élément DPE (« *Downstream promoter element* »), l'élément BRE (« *TFIIB recognition element* ») et les îlots CpG. Ces éléments ne sont pas tous présents dans les promoteurs, la composition de ces éléments varie aussi d'un promoteur à l'autre.

Les îlots CpG comprennent au moins 200 paires de bases riches en dinucléotides CG. Ils ont beaucoup d'influence sur la présence des autres éléments en raison de leur longueur et de leur biais de composition. Les deux principales classes de promoteurs

correspondent à ceux dont le TSS est inclus dans un îlot CpG et ceux dont le TSS est à proximité d'autres éléments (Figure 7). La principale différence entre ces deux classes de promoteurs se situe au niveau du site d'initiation de la transcription. Contrairement aux gènes dont les promoteurs sont composés d'une boîte TATA ou d'un élément DPE, les gènes dont le promoteur comporte un îlot CpG ont tendance à avoir plusieurs TSS répartis sur une centaine de paires de bases. Entre 50% et 75% des promoteurs des gènes humains contiennent un îlot CpG [64, 65].

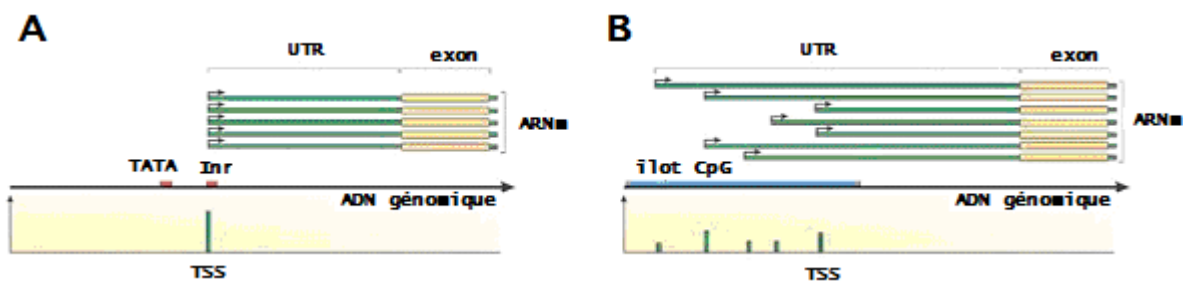


Figure 7 Classes de promoteurs

**A.** promoteur contenant une boîte TATA . **B.** promoteur contenant un îlot CpG. La figure est adaptée de [64].

L'initiation de la transcription nécessite le recrutement de facteurs généraux de transcription (TFIIA, B, D, E, F et H) au TSS. TFIID est composé de TBP (« *TATA-binding protein* ») en plus de 14 facteurs associés [66]. TFIID est capable de reconnaître les boîtes TATA, l'élément initiateur (Inr) et l'élément DPE [63].

La boîte TATA a été nommée d'après sa séquence consensus (TATAA) et elle est l'équivalent eucaryote des boîtes Pribnow des procaryotes (TATAAT). Les boîtes TATA sont en général situées en amont du TSS (-28 pb , -34 pb). Le TSS des promoteurs des gènes ayant une boîte TATA contient souvent un élément initiateur [64].

Bien que l'élément initiateur soit souvent présent avec les boîtes TATA, ces dernières ne sont pas requises. Inr englobe le TSS et sa séquence consensus est YYANWYY (le A correspond à la position +1 du TSS) [64].



L'élément DPE est présent dans les promoteurs n'ayant pas de boîte TATA. Il est également souvent présent avec un site Inr. Sa séquence consensus est RGWYV et il se trouve en aval du TSS (+28 pb, +32 pb) [64].

L'élément BRE est parfois en amont des boîtes TATA. Sa séquence consensus (SSRCGCC) est reconnue spécifiquement par TFIIB. BRE peut à la fois activer et réprimer la transcription basale des gènes [64].

Malgré que le promoteur minimal soit suffisant pour assurer l'initiation de la transcription par la machinerie transcriptionnelle associée à l'ARN polymérase II, l'activation et la répression des gènes requiert bien plus que les facteurs généraux de la transcription.

## 2.3 Régulation de la transcription à longue distance

Le promoteur des gènes se décrit typiquement comme étant une région d'environ 1000 paires de base qui comprend le site d'initiation de la transcription, les éléments du promoteur minimal et un promoteur proximal contenant des séquences reconnues par des facteurs de transcription [63, 67]. La transcription des gènes est également régulée par des sites de fixation de facteurs de transcription spécifiques en aval ou en amont des promoteurs. Les régions où se trouvent ces sites de fixations sont appelées « *enhancer* » lorsque la transcription basale est augmentée ou « *silencer* » lorsque la transcription du gène est réprimée (Figure 8). Les *enhancer* et les *silencer* peuvent fonctionner à de grandes distances du TSS et irrespectivement de leur orientation.

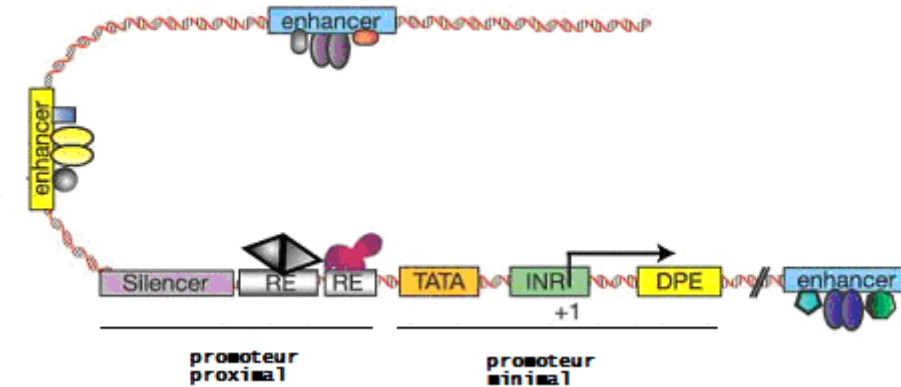


Figure 8 Régions impliquées dans la régulation de la transcription des gènes

Le TSS est marqué par « +1 ». Les sites de fixation de facteur de transcription du promoteur proximal sont marqué par RE (*Response Element*). « *enhancer* » marque les régions activatrices. « *silencer* » marque les régions répressives. La figure est adaptée de [63].

L'impact des sites des facteurs de transcription présents dans les régions activatrices ou répressives sur les gènes dépend des signaux reçus par la cellule. Pour les facteurs de transcription qui dépendent de ligands comme les récepteurs nucléaires, la présence ou l'absence détermine la réponse transcriptionnelle. Par exemple, les oestrogènes induisent des changements de conformation des récepteurs augmentant leur affinité pour les ERE et favorisent le recrutement de coactivateurs et de complexes protéiques capables de remodeler la chromatine et d'acétyler les histones. Les RAR lient l'ADN aussi bien en présence qu'en absence d'acide rétinoïque. En présence de ce dernier, RAR active la transcription de ses gènes cibles alors qu'en absence il réprime la transcription. En absence de ligand, RAR recrute des corépresseurs et des complexes protéiques capables de désacétyler les histones.

L'activité des facteurs de transcription peut également être modulée par des modifications post-traductionnelles. JUN, un membre de la famille AP1, est activé suite à sa phosphorylation par des protéines kinases [68]. En plus des signaux reçus par la cellule,

la transcription des gènes est modulée par des interactions protéine-protéine entre facteurs de transcription. La réponse transcriptionnelle suivant la fixation d'hétérodimères de JUN et FOS sur des sites AP1 peut être influencée par ER et RAR. Les récepteurs des oestrogènes augmentent la transcription des gènes cibles de AP1 alors que les récepteurs de l'acide rétinoïque les répriment [54, 55].

La régulation d'un gène peut impliquer plus d'une région activatrice (« *enhancer* ») ou répressive (« *silencer* ») qui peuvent se situer à plusieurs milliers de paires de bases en amont ou en aval du site d'initiation de la transcription. Plusieurs sites de fixation reconnus par des récepteurs nucléaires ont été caractérisés à de grande distance du TSS. Deux RARE DR5 sont impliqués dans la régulation de CYP26A1, le premier est près du TSS (-200 pb), le second est 2kb en amont [69]. KLK3, un gène cible du récepteur des androgènes, est régulé par trois éléments de réponse aux androgènes. Deux de ces éléments se trouvent dans le promoteur proximal (-630 pb), le troisième est 4kb en amont du TSS [70]. GREB1 est régulé par trois ERE consensus répartis sur une vingtaine de kb en amont du site d'initiation de la transcription [71, 72]. Les éléments distaux participent à la régulation de leurs gènes cibles en formant des boucles de chromatine avec les régions promotrices ou avec des éléments proximaux. Des études de formation de boucles de chromatine par la technique 3C (« *Chromatin Conformation Capture* ») ont déterminé que les éléments distaux des gènes GREB1 et KLK3 s'associent aux éléments proximaux [70, 72].

La structure de la chromatine joue également un rôle dans la régulation des gènes par les régions activatrices et répressives. Les chromosomes sont organisés en territoires chromatiniens appelés euchromatine et hétérochromatine (Figure 9). Les régions d'hétérochromatine sont en général méthylées, très condensées et transcriptionnellement peu actives. Les régions d'euchromatine sont moins condensées, moins méthylées et plus actives transcriptionnellement. Des régions frontières délimitent des territoires transcriptionnels afin de limiter l'effet des régions activatrices ou répressives sur la transcription gènes avoisinant les gènes ciblés [73]. Ces régions frontières peuvent être soit des régions d'attachement à

la matrice nucléaire (« *MAR* ») ou des barrières (« *insulator* ») limitant la propagation des modifications post-traductionnelles des histones. Par exemple, le gène *MYC* se situe dans un territoire transcriptionnel de 160kb délimitée par deux *MAR*. Lorsque transcrit, son promoteur est hyperacétylé, peu méthylé et flanqué de deux barrières démarquant des régions de chromatine plus méthylées [74].

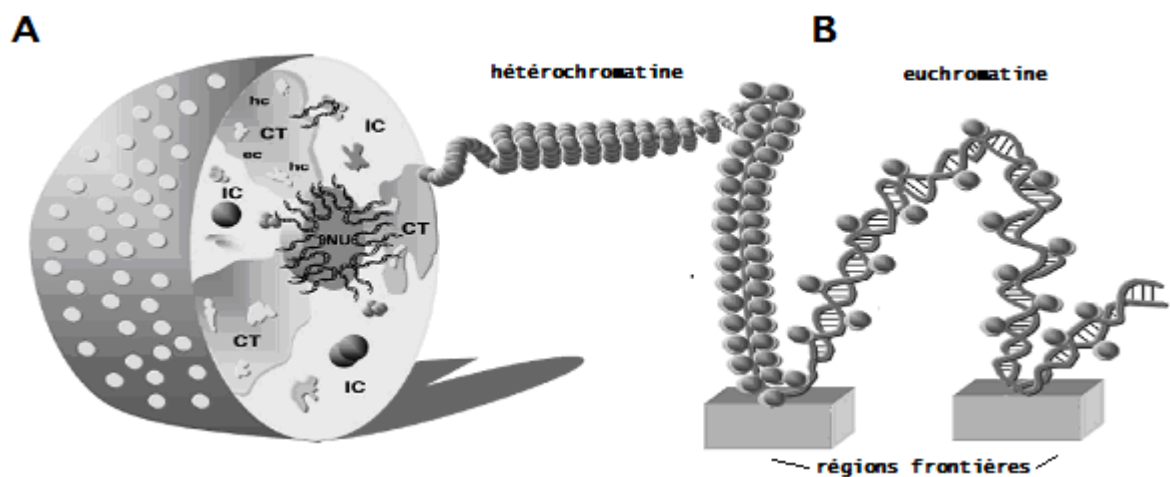


Figure 9 Territoires chromatiniens et régions frontières

**A** Positionnement des chromosomes dans le noyau. « CT » territoire chromatinien. « IC » représente les compartiments interchromosomiques. « hc » hétérochromatine (chromatine plus condensée). « ec » euchromatine (chromatine moins condensée). « NU » nucléole. **B** Contexte linéaire de la chromatine. Les régions frontières sont soit des régions d'attachement à la matrice nucléaire ou des barrières aux modifications post-traductionnelles des histones. La figure est adaptée de [73].

## **Chapitre 3 : Impact des éléments transposables sur le génome humain**

Les éléments transposables (TE) sont des séquences répétées dispersées dans le génome. Ces séquences sont très fréquentes dans le génome des animaux et des plantes. Près de 50% du génome humain correspond à des éléments transposables, chez certaines plantes cette proportion s'élève à 90% [75].

### **3.1 Classes d'éléments transposables**

Il y a quatre grandes classes d'éléments transposables chez les mammifères [58]. Les transposons à ADN forment une de ces classes. Ces éléments se sont insérés dans le génome par un mécanisme de type « couper-coller » à l'aide d'un intermédiaire ADN. Les trois autres classes de TE, aussi appelées rétrotransposons, se sont insérés dans le génome par un mécanisme de duplication. Pendant la duplication, ces éléments sont d'abord transcrits en ARN, puis retranscrits en ADN avant d'être insérés ailleurs dans le génome. Les trois classes de rétrotransposons sont: les LTR (« Long Terminal Repeat »), les LINE (« Long Interspersed Nucleotide Elements ») et les SINE (« Short Interspersed Nucleotide Elements »). Chaque classe peut être subdivisée en famille selon les séquences d'origine et les périodes durant lesquelles les éléments ont été actifs.

#### **3.1.1 LTR**

Les LTR ont une structure similaire aux rétrovirus et leur longueur peut atteindre 11 kpb (Figure 10). Ils contiennent deux gènes, gag et pol, qui codent des enzymes nécessaires à leur réinsertion. Ces deux gènes sont flanqués par deux longues séquences terminales contenant les éléments nécessaires à leur transcription. Les LTR sont subdivisés en familles selon le rétrovirus auquel ils sont similaires [76]. Les LTR ne se transposent

plus chez l'humain depuis environ 40 millions d'années alors que chez la souris plusieurs familles sont toujours actives [58, 76].



Figure 10 Structure des LTR

La figure est provient de [77].

### 3.1.2 LINE

La longueur des éléments LINE varie de 4 à 6 kpb. Ils se composent de deux cadres de lecture, ORF1 et ORF2, précédé d'une partie non traduite (Figure 11). Le premier cadre de lecture code une protéine capable de lier l'ADN, alors que le deuxième code une endonucléase et une transcriptase inverse. La partie non traduite en 5' contient un promoteur nécessaire pour leur transcription par l'ARN polymérase II [75]. Le génome humain et celui des souris contiennent trois familles de LINE : LINE1 (L1), LINE2 (L2) et LINE3 (L3). Seules certaines familles de L1 sont encore actives chez les deux espèces. Il y a environ une centaine d'éléments L1 capables de se transposer chez l'humain [78]. Les L1 sont plus actifs chez la souris que chez l'humain [58, 79]. Bien que les éléments L1 chez les deux espèces semblent provenir d'une même famille de LINE (L1MA6), la région 5' non traduite varie beaucoup entre l'humain et la souris [79]. Chez l'humain, la transcription des L1 peut se faire suite à la fixation des facteurs de transcription RUNX3 ou YY1 dans le promoteur de la région 5' non traduite [80, 81].

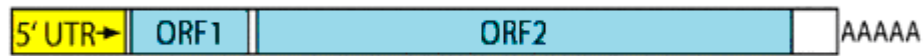


Figure 11 Structure des LINE

La figure est provient de [77].

### 3.1.3 SINE

Les éléments SINE contiennent un promoteur interne pour l'ARN polymérase III et leur longueur varie de 100 à 300 paires de bases selon la famille à laquelle ils appartiennent (Figure 12). Dans le génome humain, on retrouve principalement deux familles : les MIR (« *Mammalian-wide Interspersed Repeat* ») et les Alu. Les MIR sont dérivés d'ARN de transfert (ARNt) alors que les Alu sont dérivés du gène de l'ARN 7SL. Les Alu sont spécifiques aux primates et sont les seuls SINE encore actifs chez l'humain. Les SINE dépendent des protéines codées par les LINE pour la réinsertion de leur intermédiaire ARN dans le génome. Les MIR ont cessé d'être actifs en même temps que les LINE L2 il y a 80 à 100 millions d'années [58]. Les séquences Alu, apparues il y a environ 65 millions d'années, dépendent de LINE L1 pour leur transposition. L'amplification des séquences Alu n'a pas été uniforme au cours de l'évolution. La plupart des insertions remontent à plus de 40 millions d'années avec l'expansion de la famille AluS. Les Alu encore capables d'être transposés appartiennent à la famille AluY. Leur taux d'insertion est estimé à une nouvelle insertion toutes les 200 naissances [82]. Les rongeurs ont une famille de SINE homologues aux Alu, les SINE B1, dérivés également de l'ARN 7SL mais plus courts que les Alu [79].

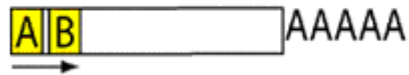


Figure 12 Structure des SINE

La flèche représente le promoteur interne pour l'ARN polymérase III des SINE. Ce promoteur est composé d'une boîte A et une boîte B. La figure est provient de [77].

### 3.2 Distribution dans le génome humain

Près de 50% du génome humain correspond à des éléments transposables SINE, LINE ou LTR (Figure 13). Les LINE comptent pour 20% avec plus de 800 mille éléments. Les SINE ont contribué 13% de la séquence du génome. Les Alu sont les éléments transposables les plus fréquents avec plus d'un million de copies (10% du génome). Les LTR sont les éléments les moins fréquents avec 400 mille copies, correspondant à 8% du génome [58].



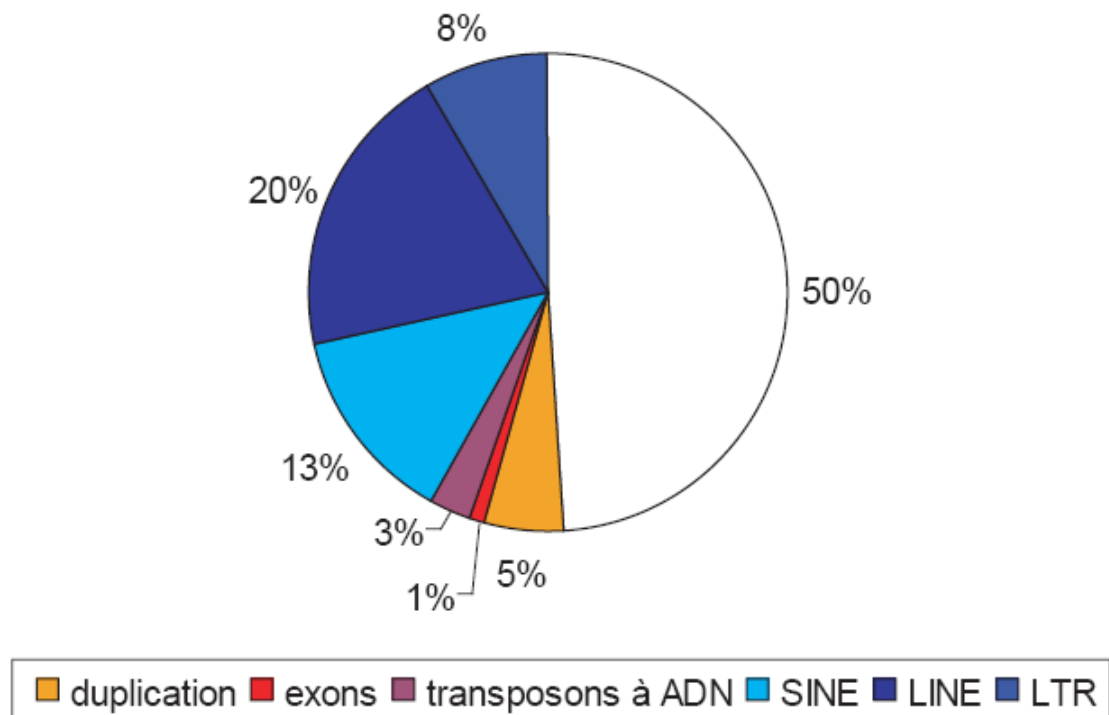


Figure 13 Proportion du génome humain correspondant à des éléments transposés

Les pourcentages sont tirés de [58] et [83].

Les éléments transposables ne sont pas distribués uniformément dans les chromosomes. La fréquence des éléments répétés est influencée par la densité en gènes et en îlots CpG. Les éléments L1 sont plus fréquents dans les régions riches en AT qui sont en général pauvres en gènes et en îlots CpG. Les éléments Alu sont plus fréquents dans les régions riches en GC où les gènes et les îlots CpG sont plus fréquents [58, 84-86]. Les SINE sont les éléments transposables qui se sont insérés le plus fréquemment à proximité des gènes (Figure 14). Globalement, il y a beaucoup moins de TE à proximité du site d'initiation de la transcription, sans doute dû à l'importance de cette région pour la régulation de la transcription des gènes.

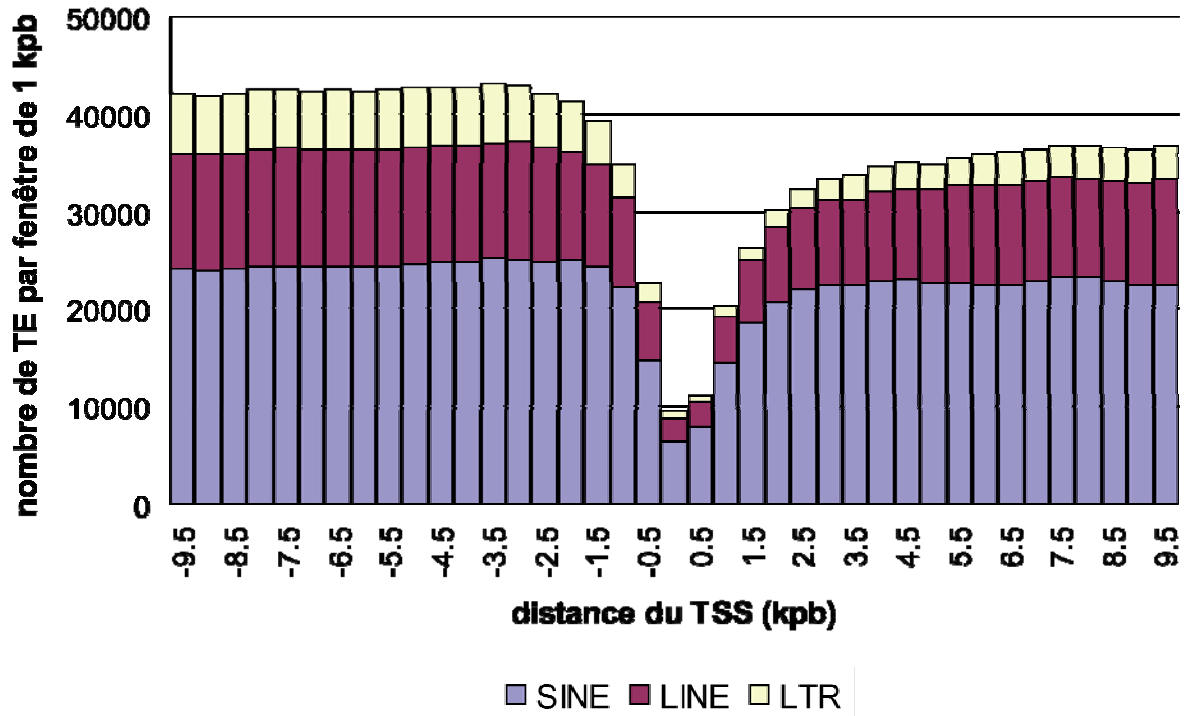


Figure 14 Distribution des TE à proximité des gènes

Les LTR, LINE et SINE ont été comptés avec une fenêtre coulissante de 1000 pb entre + et - 10 kpb du TSS de tous les gènes du génome humain. 293729 SINE, 137285 LINE et 55965 LTR se trouvent à +/- 10 kpb du TSS. La majorité des gènes ont au moins un TE (18454/18500). La figure a été bâtie avec l'annotation des TE du génome humain version hg17 du UCSC [87] et l'annotation des gènes RefSeq de juin 2007.

### 3.3 Impact des éléments transposables sur les gènes

La transposition de TE a généré une portion non négligeable du génome humain. Les TE n'ont pas seulement augmenté la taille du génome au cours de l'évolution, ils ont aussi influencé les séquences environnant leurs sites d'insertion. En raison de la similarité des membres d'une même famille, les TE sont une source de délétion ou duplication de segments par recombinaison homologue. Près de 5% du génome humain correspond à des duplications de blocs d'ADN de plusieurs centaines de kpb. Les éléments Alu sont possiblement à l'origine de 27% des duplications de segment du génome humain [88]. La recombinaison homologue d'éléments Alu a causé la délétion de près de 500 régions du

génomique humain, incluant 3 exons de gènes ayant des orthologues chez les chimpanzés [89].

L'insertion *de novo* d'éléments transposables dans la séquence des gènes peut avoir des effets négatifs comme le décalage du cadre de lecture, la délétion ou la duplication d'exons. À long terme, les TE peuvent aussi avoir des effets positifs sur les gènes. Les TE se trouvant dans les introns des gènes peuvent générer de nouveaux transcrits par épissage alternatif. Les TE dans les promoteurs peuvent être la source de nouveaux sites de facteurs de transcription. Les TE sont des agents de changement à court et long terme sur le génome et les gènes.

### 3.3.1 Mécanismes limitant la transposition des TE

Le premier niveau de contrôle de la transposition des TE est l'ADN. La méthylation de l'ADN est associée à la répression de la transcription. Les cytosines des dinucléotides CpG des TE sont en général méthylées. Les LINE L1 sont inhibés par la méthylation des CpG dans leur promoteurs [90, 91]. Les CpG méthylés ont tendance à être convertis en TpG. La conversion des CpG et l'accumulation de changements de bases ponctuels peuvent également contribuer à l'inactivation des TE en modifiant les séquences promotrices critiques pour leur transcription (Figure 10, 11 et 12).

Le deuxième niveau de contrôle est la transposition des TE. APOBEC3A et APOBEC3B, deux facteurs de résistance aux rétrovirus, inhibent la transposition de transcrits d'éléments LINE L1 et Alu [92]. Les transcrits des LINE L1 sont également susceptibles à la dégradation par interférence par l'ARN [93].

### 3.3.2 Génération de nouveaux transcrits

Les TE sont fréquents à proximité des gènes (Figure 14). Plusieurs études ont noté l'inclusion de fragments d'éléments transposables dans la séquence codante de gènes [94-

97]. Entre 500 et 2000 gènes humains ont des transcrits comportant des exons dérivés de TE. L'incorporation d'exons dérivés de TE dans la séquence d'un gène est appelée « exonisation ». L'exonisation nécessite des mutations ponctuelles dans la séquence TE permettant de changer des sites similaires à des sites d'épissage en sites reconnus par la machinerie d'épissage (GT/AG). Ces sites d'épissage peuvent créer des exons constitutifs présents dans tous les transcrits du gène (Figure 15 A) ou des exons présents seulement dans quelques uns des transcrits (Figure 15 B). Une analyse récente de l'exonisation des TE a identifié 1824 cas chez l'humain et 506 cas chez la souris [98]. Plus de la moitié des cas d'exonisation identifiés chez l'humain sont dus aux éléments Alu spécifiques aux primates. Peu des exons identifiés sont constitutifs (58 chez l'humain et 29 chez la souris). La majorité d'entre eux introduisent un changement dans le cadre de lecture ou un codon stop lorsqu'ils sont inclus dans l'ARN messenger. Les auteurs estiment que seulement 4% des exons identifiés génèrent des transcrits pouvant être traduits en protéines.

Certains TE ont également contribué une partie de leur séquence à des gènes codant pour des ARN. Les microARN sont de courtes molécules d'ARN qui contrôlent la traduction de leurs gènes cibles par interférence par l'ARN. Plus de cinquante microARN validés expérimentalement sont en partie dérivés d'éléments transposables [99, 100]. Ces microARN sont dérivés de toutes les classes de TE, même des séquences spécifiques aux primates. Les séquences Alu peuvent aussi servir de promoteurs à des microARN et permettre à ces derniers d'être transcrits par l'ARN polymérase III [101]. Enfin, les TE peuvent fournir des cibles à certains microARN. Une trentaine de microARN sont complémentaires à des séquences Alu, LINE L2 et MIR présents dans plusieurs ARNm [102].

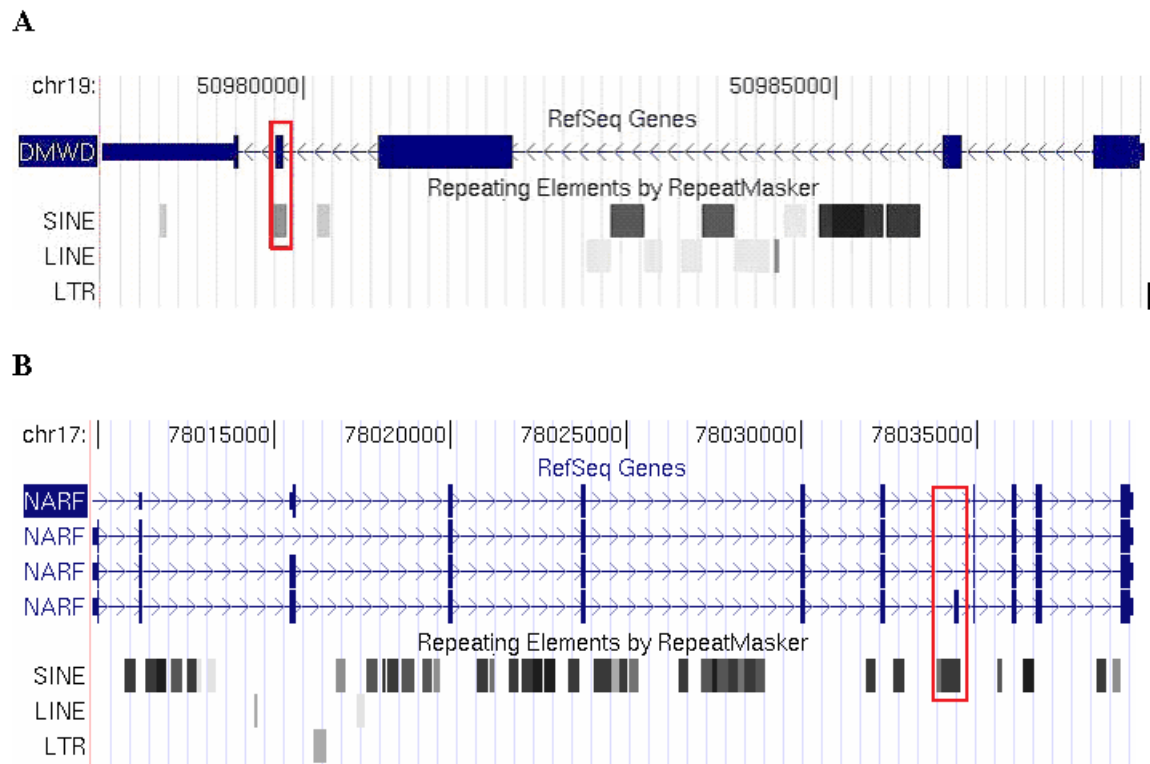


Figure 15 Exemples d'exons dérivés de TE

**A** Exon dérivé d'un élément MIR du gène DMWD (« dystrophia myotonica, WD repeat containing »). L'élément MIR et l'exon sont conservés chez l'humain, la souris et le chien. L'exon a été validé expérimentalement dans [98]. **B** Exon dérivé d'un élément Alu du gène NARF (« nuclear prelamina A recognition factor »). L'exon a été validé expérimentalement dans [103].

### 3.3.2 Génération de nouveaux sites de facteurs de transcription

Plusieurs études *in silico* ont souligné le potentiel des TE à contribuer à la transcription des gènes [104-108]. Ces études ont observé des sites pouvant être reconnus par divers facteurs de transcription dans les TE situés dans les séquences promotrices des gènes. Certaines sous-familles d'éléments Alu contiennent des séquences similaires aux ERE et aux RARE. La capacité de ces éléments à réguler l'expression de gènes a été démontrée par des expériences de retard sur gel et de gènes rapporteurs [109-111]. Il y a

quelques années, une étude a examiné la contribution des TE aux sites de fixation validés expérimentalement catalogués dans la banque de données TRANSFAC. Près de 2.5% des sites de fixation (21 sur 846) sont inclus dans des TE [112]. Toutes les classes de TE (SINE/LINE/LTR) ont contribué des sites de fixation. La littérature rapporte plusieurs exemples de sites de fixation dérivés de TE reconnus par des récepteurs nucléaires (Table 1). La fixation de ces sites par leurs récepteurs respectifs a été validée *in vivo* par immunoprécipitation de la chromatine, sans qu'ils soient nécessairement identifiés comme des sites provenant d'éléments transposables. Parmi ces sites, seul le VDRE du gène CAMP (« *cathelicidin antimicrobial peptide* ») a été identifié comme provenant d'un élément Alu étant donné que le gène est régulé par la vitamine D chez l'humain et non chez la souris.

Table 1 Sites de fixation de récepteurs nucléaires provenant de TE validées par ChIP

facteur	site	séquence	gène	Classe/Famille	référence
AR	ARE	GGAACA TAT TGTAtC	KLK3	LTR/LTR40a	[70, 113]
ER	ERE	AGGTCA TTG TGACaC	IGFBP4	SINE/MIRm	[71]
RAR	DR2	GGaTCA CG AGGTCA AG AGaTCA	BTG2	SINE/AluSc	[114]
RAR	DR2	AGGTCA GG AGTTCA	CASP9	SINE/AluSx	[115]
RAR	DR2	GGaTCA CG AGGTCA GG AGTTCg	KLK10	SINE/AluSg	[116]
RAR, VDR	ER8	TGACCT GACTCAGG GGTTCA	TYROBP	LINE/L2	[50]
VDR	DR3	GGTTCA ATG GGTTCA	CAMP	SINE/AluSx	[117, 118]
VDR	DR3	AGTTCA AGC AGTTct	KLK6	SINE/AluSx	[119]
VDR	DR3	GGTTCA AAC AGTTct	KSR2	SINE/AluSx	[120]

Le nombre de TE dans les promoteurs varie d'un gène à l'autre. Près des gènes KLK, les insertions ont été beaucoup plus fréquentes que près des gènes HOX (Figure 16). Les quinze gènes KLK (« *Kallikrein* ») sont des protéases à sérine secrétées ayant divers rôles physiologiques. Plusieurs KLK sont des marqueurs biologiques de certains cancers dont le cancer de la prostate (KLK3) et le cancer du sein (KLK5, KLK10) [121, 122]. Trois gènes KLK ont des sites de fixation dérivés de TE (Table 1). L'élément de réponse aux

androgènes (ARE) dérivé d'un LTR de KLK3 forme une boucle de chromatine avec des ARE du promoteur proximal suite à la fixation de AR et aux recrutement de co-activateurs [113]. Les KLK sont de bons exemples de gènes pour lesquels certains éléments TE ont acquis des fonctions régulatrices.

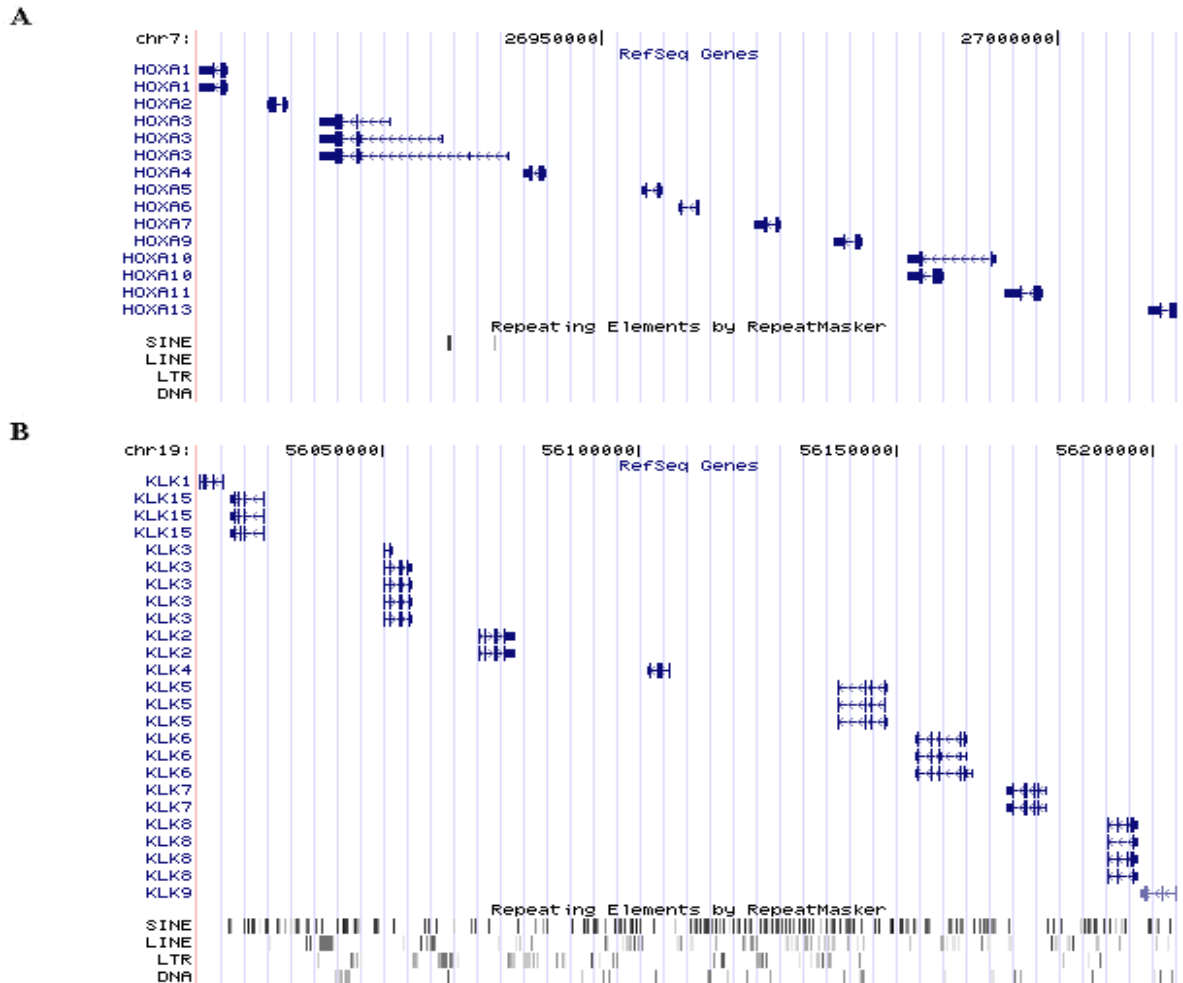


Figure 16 Fréquence des TE près des groupes de gènes HOX et KLK

**A** SINE/LINE/LTR près du groupe de gènes HOXA (Homeobox A). **B** SINE/LINE/LTR près du groupe de gènes KLK (Kallikrein). La figure a été bâtie avec la version hg17 du génome humain du UCSC [87].

## **Chapitre 4 : Méthodes biochimiques et bio-informatiques permettant l'étude de la transcription à large échelle**

Le Projet Génome Humain avait pour buts de séquencer l'ensemble des 3 milliards de paires de bases du génome humain et d'identifier tous les gènes. La partie codante des gènes, les exons, représente moins de 2% du génome. Le reste du génome contient, entre autres, les séquences impliquées dans la régulation de l'expression des gènes. Le séquençage du génome humain et le développement subséquent des micropuces d'ADN ont permis le passage de l'étude de la transcription gène par gène à une étude de la transcription à large échelle.

La technologie des micropuces permet l'étude de l'expression de milliers de gènes en une seule expérience, mais aussi lorsque combinée à une analyse par ChIP, la localisation des sites de facteurs de transcription à l'échelle des promoteurs et du génome. L'analyse bio-informatique des séquences génomiques apporte des méthodes prédictives nécessaires à la caractérisation des sites de facteurs de transcription et des mécanismes sous-jacents aux régulations observées dans ce type d'expériences à large échelle.

Les sections suivantes décrivent les méthodes biochimiques et bio-informatiques qui ont permis d'obtenir les résultats présentés dans les deux articles de ce mémoire. Certaines des figures, réalisées au cours des recherches ayant menés à ces deux articles, sont incluses dans ce chapitre pour les cas où des figures équivalentes pouvaient difficilement être trouvées dans la littérature.



## 4.1 Identification de gènes cibles à l'aide de micropuces

En parallèle au projet du génome humain, plusieurs groupes ont identifié les gènes transcrits dans différents types cellulaires en séquençant des fragments d'ADN complémentaire (ADNc), qui sont des copies des ARN messagers obtenues grâce à l'action d'une enzyme, la transcriptase inverse [123, 124].

Plusieurs techniques exploitant ces banques de transcrits ont été développées afin d'étudier l'expression de milliers de gènes en une seule expérience. Ces techniques se basent sur des approches de séquençage de fragments d'ADNc comme la méthode SAGE (« *Serial Analysis of Gene Expression* ») ou sur l'hybridation des gènes transcrits sur des puces d'ADNc [125] ou d'oligonucléotides [126]. Je vais me limiter aux micropuces d'oligonucléotides pour la suite de cette section étant donné que j'ai seulement utilisé des résultats provenant de puces Affymetrix.

### 4.1.1 Micropuces d'oligonucléotides Affymetrix

Les micropuces Affymetrix contiennent des centaines de milliers de sondes d'oligonucléotides (« *probes* ») synthétisées sur une surface de verre par un procédé alliant la photolithographie et la chimie des phases solides. Chaque transcrit de la puce est représenté par un ensemble de onze à vingt paires de sondes de 25 nucléotides (Figure 17). Chaque paire de sonde comprend 25 nucléotides correspondant exactement à la séquence du transcrit (PM, « *Perfect Match* ») et 25 nucléotides avec un mésappariement de la base centrale (MM, « *MisMatch* ») permettant de mesurer respectivement l'hybridation spécifique et non-spécifique. Un grand nombre de copies de chaque sonde est présente par unité de surface de la puce, permettant une hybridation quantitative à l'ADNc correspondant.

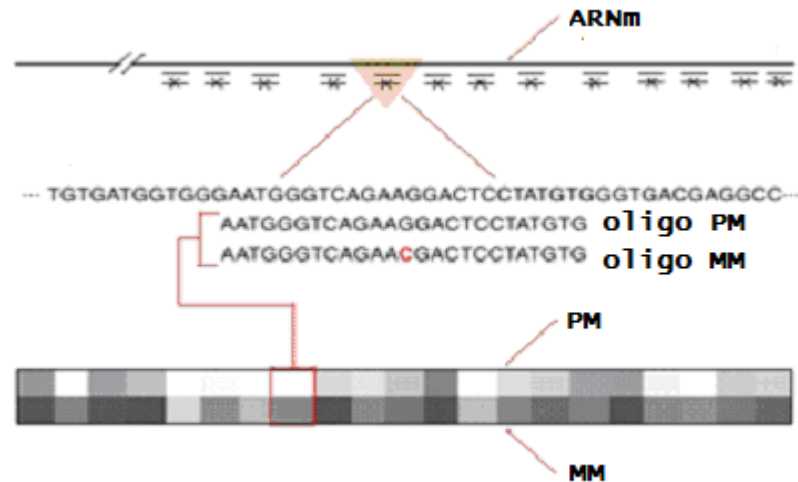


Figure 17 Gènes sur les puces Affymetrix

La figure est adaptée de <http://www-microarrays.u-strasbg.fr/>.

#### 4.1.2 Identification des transcrits régulés

L'identification des gènes régulés nécessite deux principales étapes : la normalisation des données brutes et l'analyse statistique. La liste de gènes régulés est habituellement obtenue en appliquant des seuils sur l'amplitude de la régulation (« *fold change* ») et des tests statistiques. Je vais me limiter à décrire la normalisation et l'analyse des micropuces MCF7 E2 +/- CHX présentés au chapitre 7 [127]. Les méthodes et les seuils qui peuvent influencer les gènes présents dans listes de gènes régulés vont également être mentionnées.

La normalisation des micropuces a été faite avec RMA (« Robust Multi-array Analysis ») [128]. La méthode RMA utilise seulement les valeurs PM. Une correction du bruit de fond est d'abord apportée aux valeurs de PM. Cette correction consiste à soustraire aux valeurs de PM la contribution du bruit optique et de l'appariement non spécifique estimée en supposant que cette contribution a une valeur moyenne commune à toutes les puces. Ensuite les valeurs de PM corrigées sont normalisées par quantile dans le but d'amener chaque puce à la même distribution de valeurs. Finalement un modèle en log

additif est appliqué afin d'obtenir une valeur représentative de l'expression de chaque gène. RMA a été choisie afin d'utiliser la même méthode de normalisation que d'autres études disponibles par l'intermédiaire de la banque de données GEO (« Gene Expression Omnibus »). La méthode de normalisation est une source potentielle de variabilité. Les gènes véritablement régulés devraient, en principe, être peu sensibles à la méthode de normalisation utilisée. En pratique, bien qu'il soit possible d'identifier les gènes régulés en faisant l'intersection des résultats de plusieurs méthodes de normalisation, cette stratégie a tendance restreindre la liste de gènes sans nécessairement l'améliorer [129].

Comme pour la normalisation, plusieurs méthodes d'analyse statistique sont disponibles pour quantifier l'expression différentielle des gènes [130]. Je me limiterai à *limma* (« *Linear Models for Microarray Analysis* ») [131, 132]. Après analyse, cette méthode fournit une amplitude de la régulation  $M$  en log (« *fold* »), une valeur  $A$  représentant la moyenne d'expression sur toutes les puces et la valeur- $p$  d'un test de  $t$  modéré. Les ensembles de gènes régulés identifiés au chapitre 7 [127] ont été établis en fixant des seuils sur l'amplitude de la régulation (« *fold* »  $\geq 1.4$ ), les valeurs de  $A$  ( $A \geq 5$ ) et les valeurs- $p$  ajustés du test de  $t$  modéré ( $p \leq 0.01$ ). Le test de  $t$  modéré a la même interprétation que le test de  $t$  standard mais la déviation standard a été remplacée par une estimation qui prend en compte la variabilité de l'ensemble des gènes [131]. Les valeurs- $p$  du test modéré ont été ajustées pour les tests multiples afin de contrôler le taux de fausse découverte par la méthode de Benjamini et Hochberg [133]. La valeur  $A$  permet d'identifier les valeurs de  $M$  (« *fold* ») dues à de faibles intensités entre les puces traitement et contrôle.

### 4.1.3 Études des voies de signalisation de E2 et RA à l'aide de micropuces

Les œstrogènes stimulent la prolifération des cellules de carcinome mammaire alors que l'acide rétinoïque inhibe leur croissance. Plusieurs études ont utilisé des micropuces Affymetrix afin d'identifier des gènes cibles de ER et RAR dans les cellules MCF7, un modèle de carcinome mammaire surexprimant ER $\alpha$  (Table 2 et 3). Le plan expérimental

diffère entre les études au niveau de la durée des traitements, du nombre de réplicats et du nombre de gènes sur les micropuces. Il y a 12000 transcrits sur les puces HG-U95A et près de 47000 sur les HG-U133 Plus 2.0. Les études utilisant plusieurs temps de traitement ont souvent un seul temps pour le contrôle.

Table 2 Micropuces avec traitement E2

<b>Durée (traitement/contrôle)</b>	<b>réplicat</b>	<b>puce Affymetrix</b>	<b>Gènes régulés</b>	<b>référence</b>
3, 6 et 12h / 0h	3	HG-U133 Plus 2.0	1524	[134]
4, 8, 24 et 48h / 0h	2	HG-U95A	438	[135]
24h / 24h	2	HG-U133A	674	[136]
1 semaine / 1 semaine	3	HG-U133A	213	[137]

Table 3 Micropuces avec traitement RA

<b>Durée (traitement/contrôle)</b>	<b>réplicat</b>	<b>puce Affymetrix</b>	<b>Gènes régulés</b>	<b>référence</b>
4h / 4h	3	HG-U133A/B	825	[115]
6h / 6h	4	HT-HG_U133A	263 *	[138]
24, 48 et 72h / 72h	?	HG-U133 Plus 2.0	316	[139]

\* = micropuces analysées comme pour le chapitre 7 [127].

Plusieurs facteurs de transcription se retrouvent parmi les gènes régulés par E2 ou RA de ces études. FOS, EGR3, MYC et RAR $\alpha$  sont parmi les facteurs induits par E2 [135]. JUN, HOXA1 et PPAR $\gamma$  sont parmi les facteurs induits par RA [139]. La plupart des études n'ont pas regardé quelle voie ou combinaison de voies de signalisation peuvent expliquer la régulation des gènes sur les micropuces. Parmi les études de la table 2 et 3, seule celle de Carroll et al [134] examine les mécanismes de régulation en identifiant des régions liées par ER $\alpha$  à l'aide d'immunoprécipitation de la chromatine (ChIP) et micropuces d'oligonucléotides contenant la portion non-répétitive (excluant les TE) du génome humain.

## 4.2 Identification de régions régulatrices par immunoprécipitation de la chromatine

En complément aux techniques permettant l'étude à large échelle de l'expression des gènes, des techniques de localisation des facteurs de transcription ont été développées. Ces dernières utilisent les fragments d'ADN purifiés par immunoprécipitation de la chromatine (ChIP) à l'aide d'anticorps spécifiques à des facteurs de transcription d'intérêt. Comme les techniques d'étude de l'expression des gènes, certaines sont basées sur le séquençage de l'ADN obtenu par ChIP alors que d'autres techniques hybrident ces séquences sur des puces d'oligonucléotides ou de fragments de promoteurs (« *ChIP-on-chip* ») [140].

### 4.2.1 Techniques de séquençage et micropuces

Une des techniques basées sur les micropuces utilise des oligonucléotides de 25 nucléotides toutes les 30 paires de base (« *tiling array* », Figure 18 A). Habituellement ces micropuces ne couvrent pas les éléments transposables en raison de la similarité de leurs séquences. L'identification des régions fixées par le facteur ciblé nécessite une analyse statistique du signal sur les puces.

Une des techniques basées sur le séquençage, consiste à séquencer en paire le début et la fin (« PET : Paired-End diTags ») des fragments d'ADN obtenus par ChIP (« *ChIP-PET* »). Après séquençage, ces fragments sont localisés sur le génome. Ensuite, les régions fixées par le facteur ciblé sont identifiées grâce aux limites correspondant aux fragments séquencés (Figure 18 B).

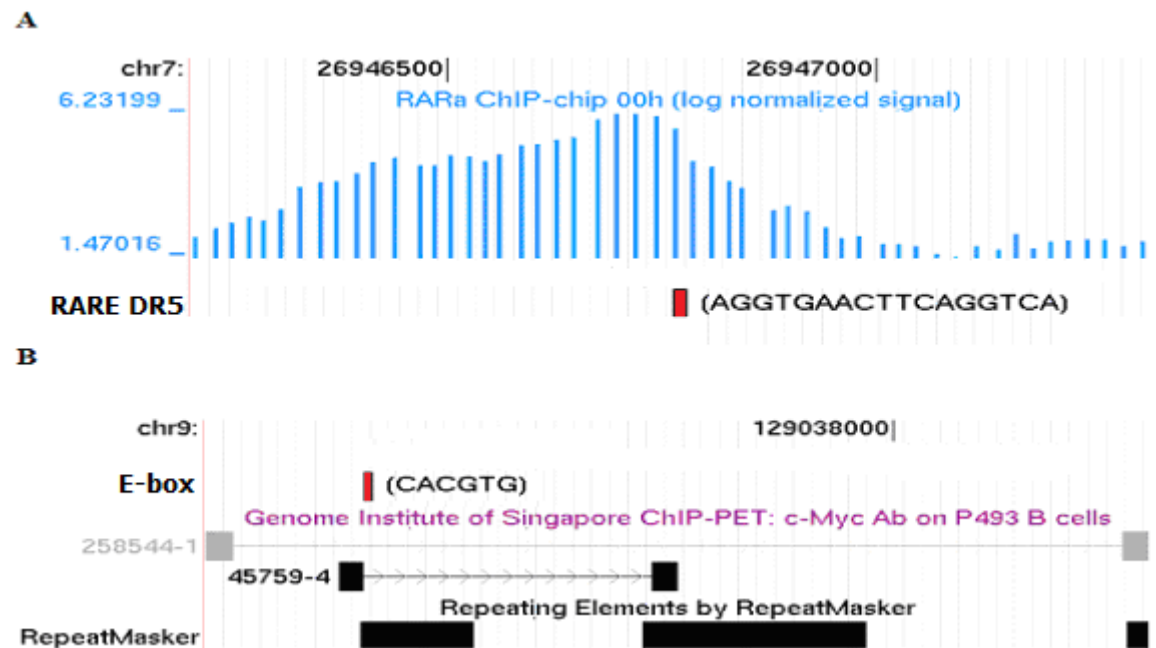


Figure 18 Régions identifiées à l'aide de micropuces et séquençage

**A** Région fixée par RARα avec la technique de *ChIP-on-chip*. **B** Région fixée par MYC avec la technique de *ChIP-PET*. Le site CACGTG se trouve dans un élément transposable et il correspond au consensus des sites reconnus par MYC. La figure a été bâtie avec la version hg17 du génome humain du UCSC [87].

Une limitation commune aux deux techniques vient de la technique de ChIP elle-même, qui génère des fragments de chromatine de 200 pb à 1000 pb. Les sites des facteurs de transcription, qui dépassent rarement 15-20 nucléotides, ne sont donc pas cartographiés précisément. On s'attend à ce que le site expliquant le recrutement du facteur de transcription immunoprécipité se trouve près du centre de la région de *ChIP-on-chip* (Figure 19). Les deux techniques ne distinguent pas *a priori* les cas où le facteur de transcription visé est fixé à l'ADN des cas où il est recruté comme co-régulateur par un autre facteur. Une des différences entre les deux techniques est au niveau des TE : les micropuces de la technique de *ChIP-on-chip* excluent les éléments transposables alors que la technique de *ChIP-PET* permet d'obtenir des régions contenant des éléments transposables (Figure 18B et Figure 20).

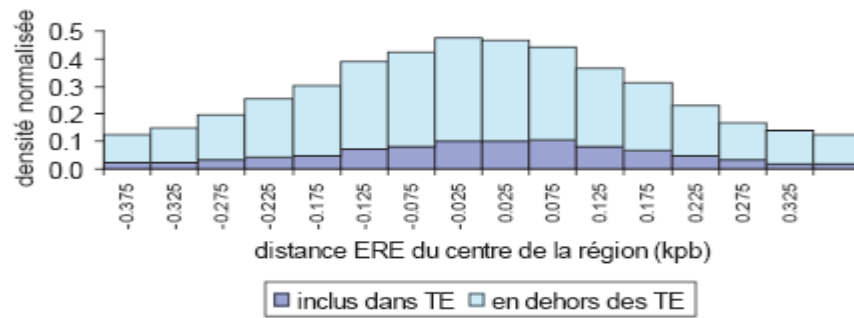


Figure 19 Distribution des ERE dans les régions liées par ER $\alpha$  ChIP-PET

Distribution de 542 ERE sont à +/- 500 pb du centre des régions liées par ER $\alpha$  identifiées par la technique de ChIP-PET. Près de 20% des ERE (105) sont à l'intérieur de séquences répétées. Les ERE ont été comptés à l'aide d'une fenêtre de 200 pb déplacée tous les 50 pb, le nombre de ERE observé dans chacune des fenêtres a été divisé par le nombre total de ERE (542). La figure a été bâtie avec les données de *ChIP-PET* de [141].

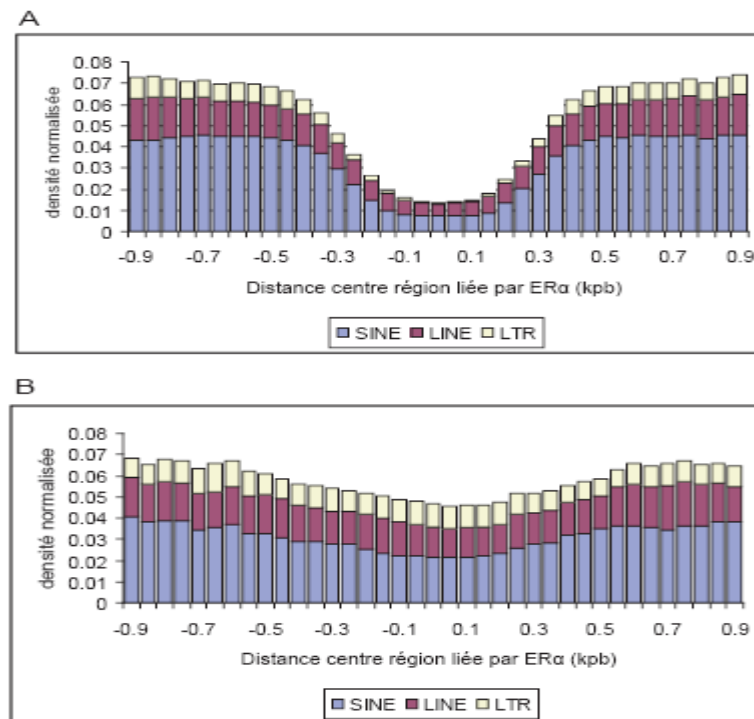


Figure 20 Contribution des TE aux régions liées par ER $\alpha$

Les TE ont été comptés à l'aide d'une fenêtre de 200 pb déplacée tous les 50 pb, le nombre de TE observé dans chacune des fenêtres a été divisé par le nombre total de TE. **A** Régions liées par ER $\alpha$  en ChIP-chip (4245 régions sur 10600 contiennent des TE) **B** Régions liées par ER $\alpha$  en ChIP-PET (900 régions sur 1234 contiennent des TE). La figure a été bâtie avec les données de *ChIP-on-chip* de [134], les données de *ChIP-PET* de [141] et avec l'annotation des TE du génome humain version hg17 du UCSC [87].

#### 4.2.2 Localisation de régions fixée par ER $\alpha$ à l'échelle du génome

Plusieurs études réalisées pendant la durée de mes études ont localisé des régions fixées par ER $\alpha$  à l'aide de micropuces et techniques de séquençage. Les études utilisant des micropuces ont employé une variété de design : banque d'îlots CpG [142], promoteurs [37, 143-145], régions contenant des ERE prédits *in silico* [146] et chromosomes entiers [134, 147, 148]. Les MCF7, un modèle cellulaire de carcinome mammaire surexprimant ER $\alpha$ , ont été utilisées dans la majorité de ces études.



À l'échelle du génome, 3665 régions fixées par ER $\alpha$  ont été identifiées avec un taux de fausse découverte de 1% et 10600 avec un taux de 5% par la technique de *ChIP-on-chip* avec des puces de courts oligonucléotides couvrant les portions non répétitives du génome humain [134]. À l'aide de la technique *ChIP-PET*, 1234 régions fixées par ER $\alpha$  ont été identifiées [141]. Sans doute en raison des différences entre les deux approches, seulement la moitié des régions identifiées par la technique de *ChIP-PET* correspondent à une région identifiée par *ChIP-on-chip*. Malgré cette différence dans les régions identifiées, la localisation de ces dernières par rapport aux gènes est similaire. Près de 95% des régions liées par ER $\alpha$  se situent à plus de 5 kbp du site d'initiation de la transcription d'un gène. Les deux études ont identifié des gènes régulés par les œstrogènes à l'aide de micropuces Affymetrix et ont observé un enrichissement de régions situées jusqu'à 50 kbp des gènes induits par E2. Ces résultats confirment la capacité de ER $\alpha$  de réguler la transcription de ses gènes cibles via des éléments distant du site d'initiation de la transcription.

L'analyse *in silico* de la séquence des régions liées par ER $\alpha$  reproduit les observations d'études précédentes utilisant les chromosomes 21 et 22 et les promoteurs proximaux [37, 147]. Près de 30% des régions se situent près de ERE prédits et les sites pour lesquels ER a plus d'affinité sont plus souvent trouvés près de régions (Table 4). Les sites AP1 sont parmi les sites les plus fréquemment trouvés dans les régions fixées par ER $\alpha$  après les ERE. Près du centre des régions, ces deux facteurs sont souvent mutuellement exclusifs. Les sites reconnus par les facteurs de transcription FOX, OCT et C/EBP ont été fréquemment observés dans les régions ayant un ERE ou un site AP1 suggérant un coopération entre ER $\alpha$  et ces facteurs.

Table 4 Fréquence des ERE à proximité de régions liées par ERα ChIP-on-chip

ERE prédits						régions liées par ERα (10600)	
Score	Génome	ERα FDR 5%	%	attendu	enrichissement	# régions	%
65-100	136533	3684	2.70	1699.83	2.17	3085	29.10
70-100	55247	2135	3.86	687.82	3.10	1935	18.25
75-100	30380	1511	4.97	378.23	3.99	1432	13.51
80-100	10089	837	8.30	125.61	6.66	822	7.75
85-100	3192	412	12.91	39.74	10.37	411	3.88
90-100	1000	239	23.90	12.45	19.20	241	2.27
95-100	689	176	25.54	8.58	20.52	178	1.68

La table a été bâtie avec les données de *ChIP-on-chip* de [134] et les ERE prédit de l'article du chapitre 7 [127]. Les ERE sont à +/- 1 kpb du centre des régions fixées par ERα.

### 4.3 Reconnaissance de sites de facteurs de transcription *in silico*

L'exploitation des données obtenues à l'aide des technologies permettant l'étude de la régulation des gènes à large échelle nécessite l'utilisation d'approches *in silico* pour caractériser les sites de fixation de facteurs de transcription et les mécanismes de régulation. Durant les 25 dernières années, deux grandes catégories complémentaires d'approches ont été développées pour la prédiction de sites de facteurs de transcription dans les séquences génomiques [149-151]. Les approches avec *a priori* utilisant des modèles des sites de facteurs de transcription connus et les approches sans *a priori* avec lesquelles des modèles de facteurs connus ou inconnus sont déduits des séquences par découverte de motifs. Je vais me limiter au consensus et au modèle matriciel, les deux représentations avec *a priori* utilisées durant ma maîtrise.

### 4.3.1 Représentation consensus

Les modèles utilisés pour la reconnaissance de sites de facteurs de transcription tentent de représenter la variabilité des sites reconnus par un facteur donné. La représentation consensus utilise un alphabet étendu (Table 5) permettant d'utiliser un seul symbole pour les positions où on peut observer plus d'un nucléotide. Par exemple, les ERE de haute affinité sont représentés par le consensus RGGTCANNNTGACCY. La représentation consensus a l'avantage d'être simple à manipuler et mettre en œuvre. Les sites correspondant à un consensus peuvent être cherchés dans les séquences génomiques par expression régulière ou un des nombreux algorithmes de recherche de motifs dans un texte [152, 153]. Un des désavantages de la représentation consensus est qu'elle ne permet pas de quantifier les préférences pour les positions où on peut observer plus d'un nucléotide. De plus, si l'on accepte des variations par rapport au consensus (« *mismatch* »), il n'est pas possible de quantifier l'impact de ces variations sur l'affinité du facteur de transcription. Pour les ERE, certaines variations d'un nucléotide par rapport au consensus ont plus d'impact que les autres sur l'affinité des récepteurs des œstrogènes [71].

Table 5 Symboles représentant les acides nucléiques

code	description
A	Adénine
C	Cytosine
G	Guanine
T	Thymine
R	Purine (A ou G)
Y	Pyrimidine (C ou T)
M	C ou A
K	T ou G
W	T ou A
S	C ou G
B	C, T ou G
D	A, T ou G
H	A, T ou C
V	A, C ou G
N	A, C, G ou T

La table est adaptée de [154].

### 4.3.2 Modélisation des sites de fixations par matrice

La représentation par matrice des sites de facteurs de transcription est plus générale que la représentation consensus. Dans les matrices, la probabilité d'observer les nucléotides A, C, G ou T est compilée à chacune des positions d'un site de fixation (Figure 21). Les matrices sont habituellement construites à partir de compilation de sites validés expérimentalement tirés de la littérature où à partir d'expériences de sélection *in vitro*. Il existe plusieurs banques de modèles de sites de facteurs de transcription représentés par des matrices. Entre 100 et 800 matrices sont disponibles dans les banques de données JASPAR [155], MatBase de la compagnie Genomatix [156] et TRANSFAC de la compagnie BIOBASE [157].

	1	2	3	4	5	6
	<b>R</b>	<b>G</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>A</b>
<b>A</b>	39.47%	7.89%	5.92%	5.92%	1.32%	84.21%
<b>G</b>	36.18%	82.24%	86.18%	10.53%	5.92%	7.24%
<b>C</b>	13.82%	5.26%	2.63%	8.55%	88.16%	4.61%
<b>T</b>	10.53%	4.61%	5.26%	75.00%	4.61%	3.95%

Figure 21 Exemple de matrice représentant un demi ERE

Les pourcentages en bleu correspondent aux nucléotides d'un demi ERE de haute affinité (RGGTCA).

La recherche de site de fixation à l'aide de matrice repose sur l'idée suivante : l'affinité entre facteur de transcription et un site potentiel peut être mesurée par un score [149, 158]. Le score des sites potentiels se calcule avec une matrice pondérée correspondant au logarithme du ratio de la fréquence du nucléotide A, C, G ou T dans le modèle par rapport à sa fréquence dans le génome. Un exemple de matrice pondérée est donné à la figure 22. Les positions pour lesquelles A, C, G et T ont la même fréquence ont un poids de 0. Les nucléotides moins fréquents pour une position donnée ont en général un poids inférieur à 0 alors que les plus fréquents un poids supérieur à 0.

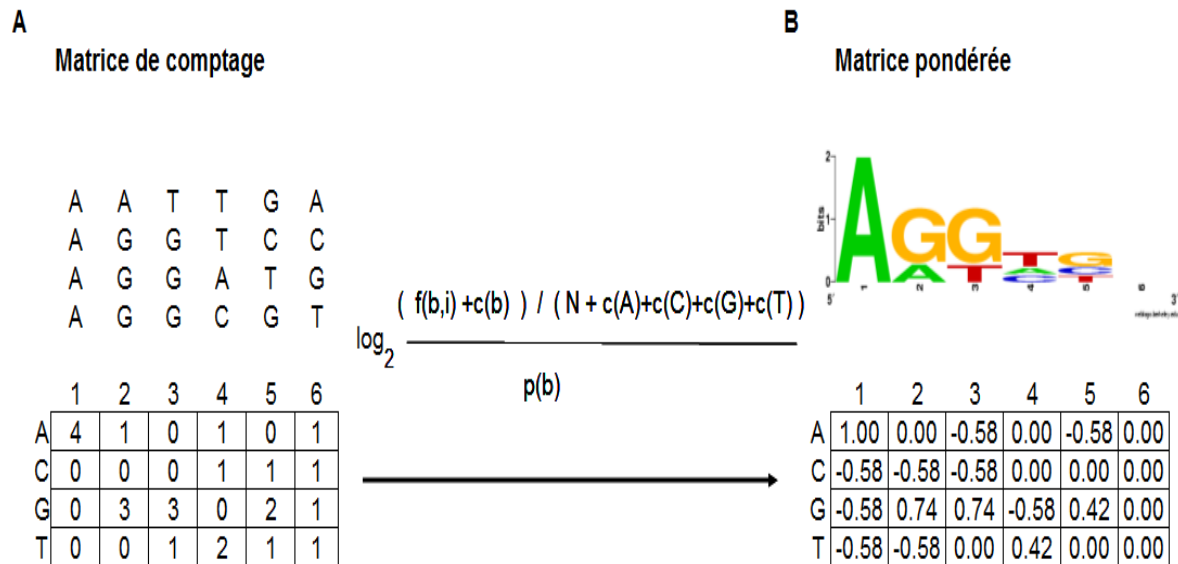


Figure 22 Exemple de matrice pondérée

**A** Matrice de comptage dérivée de quatre séquences. Chaque case correspond au nombre de fois qu'un nucléotide a été observé à cette position. **B** Matrice pondérée dérivée de la matrice de comptage. La formule utilisée pour transformer la matrice de comptage est au dessus de la flèche. Dans cette formule, N représente le nombre total de site dans la matrice de comptage,  $f(b,i)$  correspond à la fréquence du nucléotide b à la position i,  $p(b)$  est la probabilité *a priori* du nucléotide b (0.25 pour tous les nucléotides dans cet exemple) et  $c(b)$  est un pseudo-poid (« *pseudocount* ») servant à corriger les petits échantillons pour un nucléotide b et pour éliminer les valeurs nulles. La matrice pondérée est utilisée pour assigner un score aux sites potentiels. La figure est adaptée de [159].

On obtient le score d'un site potentiel en additionnant le poids de chacun de ses nucléotides dans la matrice pondérée. Ce score est souvent normalisé en une valeur entre 0 et 1. Les deux méthodes les plus populaires de normalisation sont basées sur le ratio du score obtenu par rapport au score maximum de la matrice. Pour le logiciel MatInspector développé par la compagnie Genomatix, le ratio du score obtenu et du score maximum sert à normaliser le score des quatre positions consécutives les moins variables de la matrice [160]. Cette méthode peut aussi être utilisée pour l'ensemble de la matrice si on ignore les scores inférieurs à ceux que donne une matrice non spécifique dont A, C, G et T sont

équiprobable pour toutes les positions (score  $< 0$ ). Les logiciels MATCH [161] et ConSite [162], développés respectivement par la compagnie BIOBASE et par les auteurs de la banque de données JASPAR, normalisent le score en soustrayant le score minimum de la matrice au dénominateur et au numérateur du ratio. Les sites potentiels sont considérés comme des bons candidats si leur score normalisé est supérieur à un seuil imposé par l'utilisateur. Ce seuil permet de se limiter aux sites de haute affinité s'il est élevé (ex. 0.90) et il permet d'identifier des sites plus divergent s'il est plus bas (ex. 0.65).

Comme tous les modèles, les matrices comportent des limitations. Une première limitation provient du calcul du score. L'addition des poids de toutes les positions d'un site potentiel suppose que chacune des positions est indépendante. Une deuxième limitation vient de la rigidité que le modèle matriciel impose aux sites pouvant être reconnus. Les matrices représentent les nucléotides reconnus par un seul domaine de liaison à l'ADN. Il n'est pas possible de représenter par une seule matrice les facteurs comme RAR/RXR qui reconnaissent des demi sites avec un espacement variable. De plus, les matrices modélisant les sites reconnus par des dimères comme ER incluent entre les demi sites des positions ne contribuant pas à l'affinité. Bien qu'il soit possible de minimiser la contribution de ces positions au score en fixant la fréquence de A, C, G et T à 25%, cette correction est rarement appliquée aux banques de matrices.

Plusieurs autres modèles ont été proposés pour répondre aux limitations des matrices. Les matrices de di-nucléotides permettent de tenir compte des dépendances entre les positions dans le calcul des scores [163]. La dépendance entre des positions et un espacement variable entre des demi sites peut s'incorporer à des modèles de Markov caché (hidden Markov model) [164]. L'usage de ces modèles pour la reconnaissance de sites de facteurs de transcription n'est pas aussi répandu que les matrices pour le moment.

### 4.3.3 Matrices de sites reconnus par ER

Plusieurs matrices représentant les sites reconnus par ER ont été proposées dans la littérature. Toutes les matrices proviennent de compilation de ERE validés expérimentalement. La matrice modélisant les ERE présente dans la banque de données JASPAR comporte 9 sites [155], celle de la banque de données TRANSFAC contient 20 sites [157], celle de la revue de O'Lone *et al* [165] est construite à partir de 25 ERE et celle l'article de Jin *et al* [166] est basée sur l'alignement de 48 ERE. La matrice de TRANSFAC est plus souvent utilisée que les autres.

En plus du nombre de sites différents entre ces matrices, leur longueur varie de 14 à 19 nucléotides. La portion des matrices correspondant à chacun des demi ERE diffère dans toutes les matrices. L'écart entre les demi sites est particulièrement marqué dans la matrice de TRANSFAC (Figure 23). Bien que cet écart soit prévisible dans un alignement d'un nombre limité de ERE, il n'est pas nécessairement souhaitable dans un modèle de deux demi sites arrangés en palindromes reconnus par des homodimères ou hétérodimères dont le domaine de liaison à l'ADN est à 97% identique. Une façon simple d'incorporer au modèle le caractère palindromique des ERE consiste rajouter la séquence complémentaire inverse des ERE inclus dans la matrice.



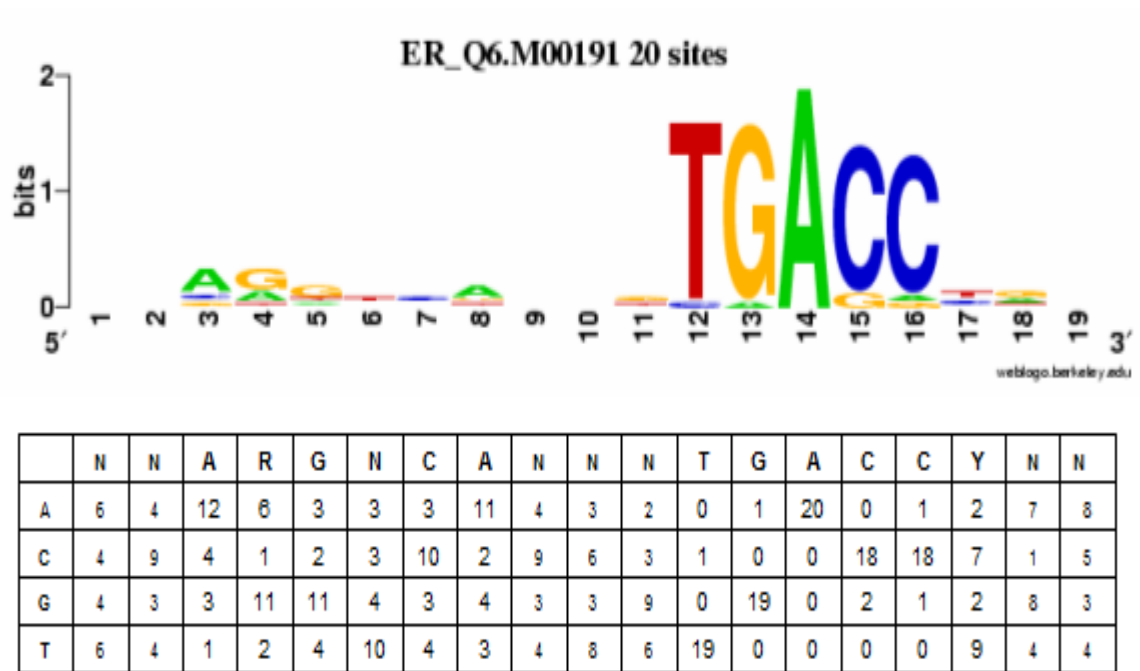


Figure 23 Matrice ERE de TRANSFAC

#### 4.3.4 Évaluation de la surreprésentation de ERE dans les promoteurs de gènes cibles

Une hypothèse de travail couramment utilisée dans l'analyse bio-informatique de gènes co-régulés est qu'ils sont soumis au contrôle du même ensemble de facteurs de transcription. Les motifs d'ADN reconnus par ces facteurs devraient en principe être surreprésentés dans les promoteurs des gènes cibles. Dans la littérature des tests statistiques et des simulations de type Monte Carlo sont employés afin d'identifier les sites de facteurs de transcription significativement surreprésentés.

Lin et collaborateurs [167] ont utilisé des simulations de type Monte Carlo afin d'évaluer l'enrichissement des ERE situé entre -3 kpb et +500 pb du site d'initiation de la transcription de 89 gènes répondants aux œstrogènes et insensibles à des traitements avec

un anti-œstrogènes (ICI) et un inhibiteur de la synthèse protéique (CHX). Chaque simulation évalue si l'écart observé entre la fréquence relative de ERE pour les gènes régulés et celle des gènes non régulés est reproduite par deux ensembles de gènes de même tailles choisis aléatoirement. Les gènes des deux ensembles sont classés du meilleur au pire. L'écart entre les deux ensembles de gènes est comparé à l'aide de l'aire sous la courbe du nombre de ERE par gène allant de un gène au nombre de gène total dans chacun des ensembles. Une valeur p (« *p-value* ») empirique est calculée en divisant le nombre de fois où l'écart a été reproduit par le nombre total de simulation. Après 100 millions de simulations, les auteurs ont obtenu une p-value de 0.0027254. Un enrichissement significatif de ERE par rapport à la moyenne de l'ensemble des gènes du génome, p-value 0.00103 (Figure 24), a aussi été observé pour les 356 gènes induits par les œstrogènes seulement en présence de cycloheximide de l'article présenté au chapitre 7 [127] avec un million de simulations. En raison du grand nombre de simulations nécessaire pour obtenir une p-value, le temps de calcul requis peut devenir prohibitif pour l'évaluation de l'enrichissement de plusieurs centaines de matrices à différents seuils et différentes longueurs de promoteurs.

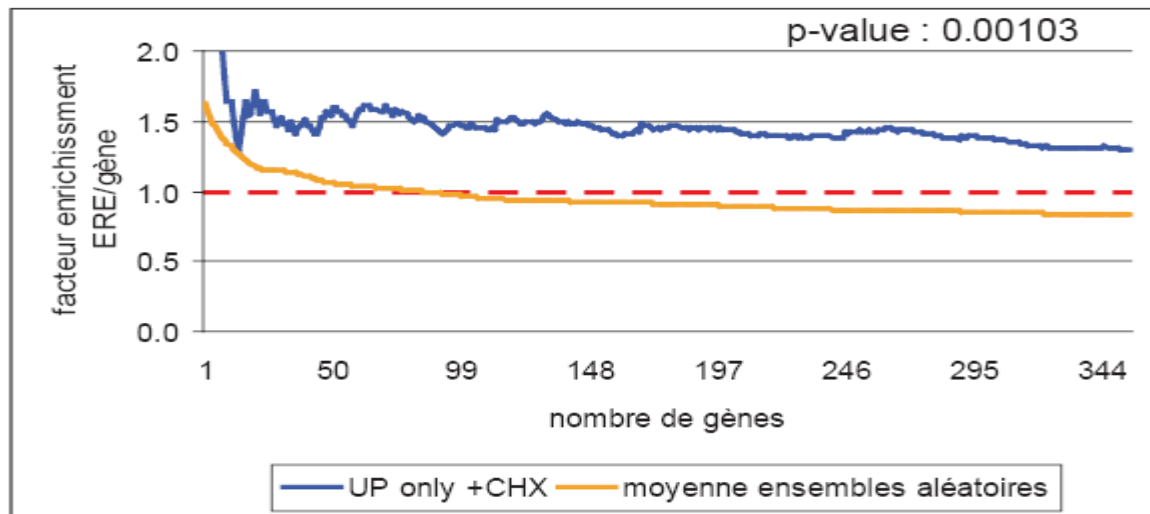


Figure 24 Simulation Monte Carlo des ERE 70% pour les gènes induits par les œstrogènes seulement en présence de cycloheximide

La *p-value* représente le nombre de fois où l'enrichissement des ERE par rapport à la moyenne de l'ensemble des gènes du génome a été reproduit par chance avec une liste de gène choisie au hasard sur 1 million de simulations. La ligne rouge pointillée correspond à un nombre de ERE par gène identique au génome (0.61198 ERE/gène). La ligne bleue correspond ratio du nombre de ERE par gène induit par E2 en présence de cycloheximide et du nombre de ERE par gène du génome. La ligne orange correspond ratio du nombre de ERE par gène moyen des ensembles de gènes choisis aléatoirement et du nombre de ERE par gène du génome. La figure a été bâtie à l'aide des ERE prédits à +/- 10 kpb du TSS et des gènes régulés par E2 de l'article présenté au chapitre 7 [127].

Ho Sui et collaborateurs [168] ont proposé l'utilisation de deux mesures statistiques complémentaires pour évaluer l'enrichissement de sites de facteurs de transcription dans les promoteurs de gènes co-régulés : un Z-score et la p-value d'un test de Fisher exact. Le Z-score sert à identifier les facteurs dont l'occurrence est plus fréquente dans l'ensemble de gènes co-régulés que celle attendu. Le nombre de sites attendus est calculé à l'aide d'un ensemble de gène non-régulés (« *background* ») en supposant une distribution binomiale. Le Z-score correspond à la différence entre le nombre de sites observé et le

nombre attendu divisé par l'écart-type. La p-value du test de Fisher sert à identifier une association non-aléatoire entre les gènes co-régulés et les sites d'un facteur de transcription. Le test de Fisher compare la proportion des promoteurs contenant au moins un site d'un facteur de transcription de l'ensemble de gènes co-régulés à la proportion des gènes du background. Les auteurs suggèrent un seuil de signification minimal de 10 pour le Z-score et un seuil maximal de 0.01 pour la p-value. Ces seuils ont été déterminés de façon empirique à l'aide d'ensembles de gènes cibles et de gènes co-régulés provenant de micropuces.

Tout comme les simulations de Monte Carlo, le Z-score et le test de Fisher permettent de conclure que l'enrichissement des ERE est significatif dans les promoteurs de 356 gènes induits par les œstrogènes seulement en présence de cycloheximide par rapport à l'ensemble des gènes du génome (Figure 25). Cette analyse a pris plusieurs heures de calculs pour le million de simulations pour chacun des seuils sur le score et quelques minutes pour le Z-score et la p-value du test de Fisher. Bien qu'il soit possible de diminuer le temps de calcul des simulations de Monte Carlo en réduisant le nombre d'ensemble de gène à choisir aléatoirement, le Z-score et le test de Fisher ont été privilégié pour l'évaluation de la surreprésentation de sites de facteurs de transcription dans l'article présenté au chapitre 7. Avec les mêmes seuils pour la matrice des ERE et la même taille de promoteurs, les ERE sont significativement enrichis dans les promoteurs de gènes cibles de ER (Figure 26) et ne le sont pas pour les gènes cibles de E2F et SP1 (Figure 27).

**A**

array/cutoff	70	75	80	85
UP only +CHX	0.00103	0.00073	0.00005	0.00620

**B**

UP only + CHX

cutoff	70	75	80	85
# ERE	282	158	78	30
# genes	174	126	73	30
fold	1.2945	1.41522	1.89696	2.24354
Z-score	16.8297	16.9817	22.2568	17.5766
P-value	0.00278	1.69E-05	1.67E-08	2.44E-05
significant	Yes	Yes	Yes	Yes

Figure 25 Évaluation de l'enrichissement des ERE à l'aide de simulations et tests statistiques

**A** Simulation Monte Carlo, chaque case représente le nombre de fois où l'enrichissement des ERE par rapport à la moyenne de l'ensemble des gènes du génome a été reproduit par chance avec une liste de gène choisie au hasard sur 1 million de simulations. **B** *Z-score* et test de Fisher exact. Les seuils de significations utilisés sont *Z-score* >10 et *p-value* < 0.01. La figure a été bâtie à l'aide des ERE prédits et des gènes régulés par E2 de l'article présenté au chapitre 7 [127].

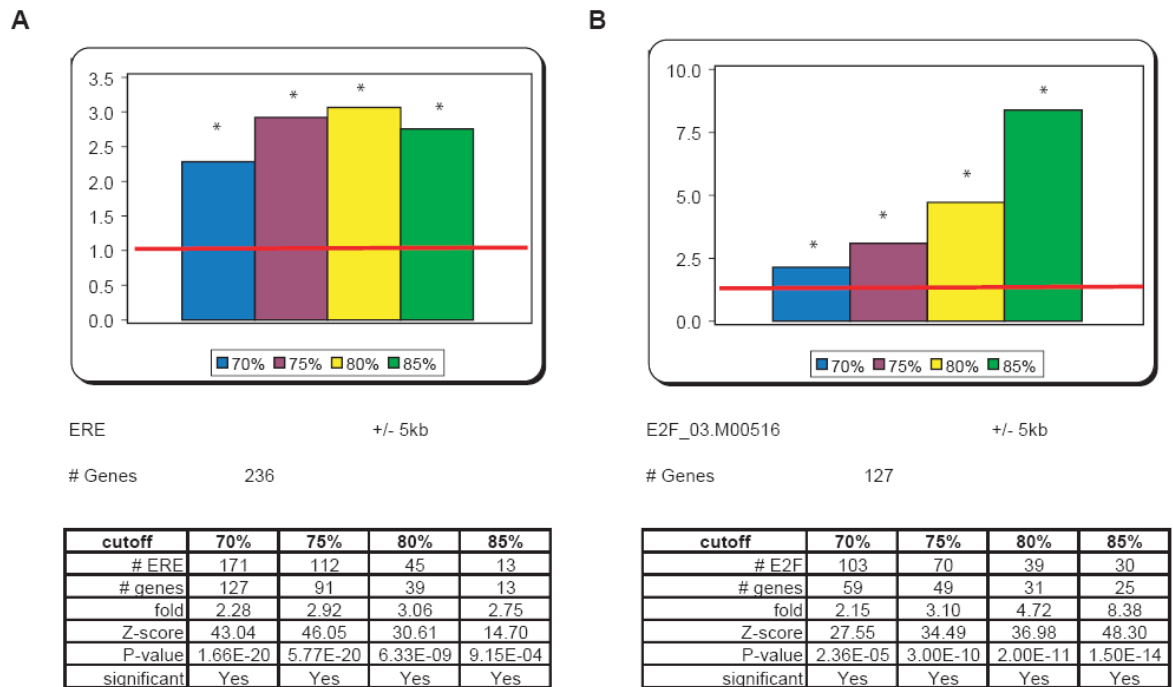


Figure 26 Enrichissement des ERE et des sites E2F dans les promoteurs de gènes cibles

**A** Enrichissement de ERE dans les promoteurs de gènes cibles de ER. **B** Enrichissement de site E2F dans les promoteurs de gènes cibles de E2F. Les gènes cibles de ER viennent de [71] et ceux de E2F de [169].

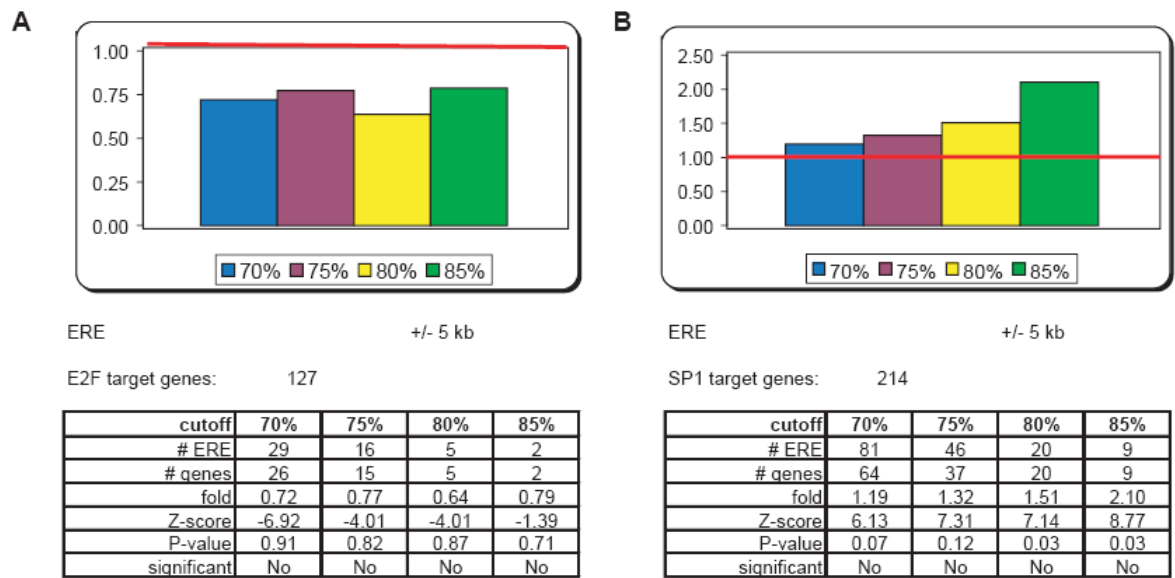


Figure 27 Absence d'enrichissement des ERE dans les promoteurs de gènes cibles E2F et SP1

**A** Absence d'enrichissement de ERE dans les promoteurs de gènes cibles de E2F. **B** Absence d'enrichissement de ERE dans les promoteurs de gènes cibles de SP1. Les gènes cibles de E2F et SP1 viennent de [169].

## Chapitre 5 : Objectifs

Le cancer le plus diagnostiqué chez la femme est le cancer du sein. Environ deux tiers des tumeurs surexpriment ER $\alpha$  et ont une croissance dépendante des œstrogènes (Chapitre 1). Certaines cellules de carcinome mammaire expriment également RAR $\alpha$  et leur croissance est inhibée par des traitements à l'acide rétinoïque.

Durant ma maîtrise, j'ai collaboré à des projets visant à mieux comprendre la régulation de l'expression des gènes par l'acide rétinoïque et les oestrogènes dans le contexte du cancer du sein. L'objectif général de ces projets consiste à identifier des gènes cibles des oestrogènes et de l'acide rétinoïque afin d'améliorer notre compréhension des rôles et des modes d'action de ER et RAR dans le contrôle de la prolifération cellulaire.

Dans un premier temps, dans le but d'avoir une vue à grande échelle des sites de liaisons de haute affinité des RAR, une recherche de sites correspondant au consensus de ces récepteurs a été faite dans les génomes humains et de souris (Chapitre 6). Une recherche similaire pour les ERE avait été effectuée précédemment dans le laboratoire [71].

Dans un second temps, pour identifier des gènes impliqués dans les mécanismes tumoraux propres aux œstrogènes médié par ER, des expériences de micropuces à ADN ont été effectuées dans des conditions permettant de départager les cibles primaires des cibles secondaires (Chapitre 7). La lignée cellulaire utilisée, les MCF7, exprime ER $\alpha$  et RAR $\alpha$ . Afin d'identifier des cibles de RA, les données d'expériences de micropuces utilisant des MCF7 traitées à l'acide rétinoïque d'un autre groupe [138] ont été utilisées (Annexe).

Dans un troisième temps, des modèles matriciels pour les ERE et RARE ont été définis à partir d'éléments caractérisés dans la littérature et d'expériences de *ChIP-on-chip* (Chapitre 7 et Annexe). La fréquence relative des sites de haute affinité et des sites plus divergents du consensus dans les promoteurs de gènes cibles issus des expériences de



micropuces a été comparée à l'ensemble des gènes du génome dans le but de relier la réponse à l'acide rétinoïque et aux œstrogènes à leur récepteurs respectifs.

Enfin, les promoteurs de gènes cibles des œstrogènes ont été criblés avec des modèles matriciels d'autres facteurs de transcription afin d'identifier les facteurs susceptibles de médier la réponse aux oestrogènes par les voies de signalisation alternatives et non-génomique et les facteurs potentiellement responsables de la régulation des cibles secondaires (Chapitre 7).

## **DEUXIÈME PARTIE**

Chapitre 6: Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution.

Chapitre 7: Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells.

Chapitre 8: Discussion et perspectives

## **Chapitre 6: Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution**

Laperriere D, Wang TT, White JH, Mader S.

Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics*. 2007 Jan 19;8:23.

À titre de premier auteur, j'ai réalisé les recherches bio-informatiques dans le génome humain et le génome de souris et l'analyse de la distribution des éléments de réponse dans les séquences répétées. J'ai également participé aux révisions du manuscrit.

## **Abstract**

### **Background**

Nuclear receptors are hormone-regulated transcription factors whose signaling controls numerous aspects of development and physiology. Many receptors recognize DNA hormone response elements formed by direct repeats of RGKTCA motifs separated by 1 to 5 bp (DR1-DR5). Although many known such response elements are conserved in the mouse and human genomes, it is unclear to which extent transcriptional regulation by nuclear receptors has evolved specifically in primates.

### **Results**

We have mapped the positions of all consensus DR-type hormone response elements in the human genome, and found that DR2 motifs, recognized by retinoic acid receptors (RARs), are heavily overrepresented (108,582 elements). 90% of these are present in Alu repeats, which also contain lesser numbers of other consensus DRs, including 50% of consensus DR4 motifs. Few DR2s are in potentially mobile AluY elements and the vast majority are also present in chimp and macaque. 95.5% of Alu-DR2s are distributed throughout subclasses of AluS repeats, and arose largely through deamination of a methylated CpG dinucleotide in a non-consensus motif present in AluS sequences. We find that Alu-DR2 motifs are located adjacent to numerous known retinoic acid target genes, and show by chromatin immunoprecipitation assays in squamous carcinoma cells that several of these elements recruit RARs *in vivo*. These findings are supported by ChIP-on-chip data from retinoic acid-treated HL60 cells revealing RAR binding to several Alu-DR2 motifs.

## Conclusion

These data provide strong support for the notion that Alu-mediated expansion of DR elements contributed to the evolution of gene regulation by RARs and other nuclear receptors in primates and humans.

## Background

Alu repeats are SINEs (Short INterspersed Elements) whose original insertion in genomic sequences appears to have occurred shortly after the dawn of the primate lineage. While most SINEs arose from tRNA genes, Alu elements are derived from the 7SL RNA gene, which encodes a component of the protein signal recognition complex. The structure of modern Alu sequences arose from a duplication of primordial elements, which were approximately 200 bp in length and composed of an RNA polymerase III promoter at one extremity and a polyA tail at the other. Dimeric Alu elements expanded extensively throughout primate genomes as a result of a parasitic relationship with the transposition machinery encoded by L1 retrotransposons, which are LINEs (Long INterspersed Elements) [1,2], and today occupy ~10% of the human genome.

Several lines of evidence indicate that a variety of events associated with Alu transposition, recombination and expansion have contributed to genome evolution and to altering gene expression by several mechanisms. Because of their high CpG content and the tendency of CpG dinucleotides to mutate, and because of their prevalence, Alu repeats contain a substantial percentage of the single nucleotide polymorphisms in the human genome [1]. Their polyA tails act to seed microsatellite formation and expansion [3]. Moreover, Alu transposition has been proposed to underlie the expansion of segmental duplications that comprise approximately 5% of the human genome [4]. Conservative estimates suggest that 0.3–0.5% of human genetic disorders arise from mobile element insertion or from element-driven recombination events [5,6]. There is also evidence that Alu repeats inserted in an antisense orientation to gene transcripts can introduce alternative

splicing sites [7,8], and that insertion of Alu sequences may alter epigenetic regulation of gene expression [9].

In addition to the above, increasing evidence suggests that Alu repeats are a source of elements that regulate transcriptional initiation by RNA polymerase II [10].

We have been interested in genome-wide mapping of hormone response elements in the promoter-proximal regions of human genes [11,12] to study their evolutionary conservation in the vicinity of genes, and their accessibility to cognate members of the nuclear receptor family. Nuclear receptors are the primary targets of a range of lipophilic signaling molecules such as steroid and thyroid hormones, vitamin D, retinoids, specific prostaglandins and cholesterol metabolites that regulate many aspects of physiology and metabolism, as well as embryonic development [13]. They directly regulate gene transcription by ligand-dependent recruitment of so-called coregulatory proteins that carry out the histone modifications, chromatin remodeling and binding of RNA polymerase II and ancillary factors necessary for initiation of transcription [13,14].

Nuclear receptors are composed of a series of domains, with the DNA binding and ligand binding domains being the most highly conserved. Most nuclear receptors function as homo- or heterodimers, and many receptor DNA binding domains recognize variants of consensus RGKTCA motifs arranged as either direct or inverted repeats[15]. Many non-steroid receptors recognize direct repeats in the form of heterodimers with members of the retinoid × receptor (RXR) family. For example, retinoic acid receptors (RARs) heterodimerized with RXRs specifically recognize retinoic acid response elements (RAREs) as RGKTCA direct repeats separated by 1, 2 or 5 bp (DR1, DR2 or DR5 motifs), whereas heterodimerized receptors for vitamin D (VDR) and thyroid hormone (TR) recognize DR3 and DR4 elements, respectively.

Evidence is accumulating that Alu repeats constitute a significant source of hormone response elements [16]. Notably, a motif, AGGTCAnnAGTTTCG, found within

most subclasses of AluS sequences, corresponds to a non-consensus DR2 element recognized by RARs, and has been shown to function as a RARE [17]. Signaling through RARs is of particular interest because retinoids control many aspects of embryonic development in a wide variety of organisms [13,18,19].

Here, we used a genome-wide screen to map DR-type hormone response elements in the human genome, and found that consensus DR2 elements are massively overrepresented relative to DR1 and DR5 RAREs and DRs recognized by other nuclear receptors, due to the presence of consensus DR2 motifs in >100,000 Alu repeats. The vast majority of these correspond to AluS sequences, which drove the Alu element expansion in primate that started 35–40 Myr ago. We also show that these "Alu-DR2" elements are present in retinoic acid regulated genes and bind RARs *in vivo*. Taken together, our data provide strong support for the notion that Alu-mediated expansion of DR elements contributed to the evolution of gene regulation by RARs and other nuclear receptors in primates and humans.

## **Results**

### **Mapping DR-type hormone response elements in the human and mouse genomes**

A number of nuclear receptors recognize direct repeats of RGKTCA motifs spaced between 1 and 5 nucleotides as response elements, and individual receptors distinguish between DR elements based on inter-repeat spacings and subtle differences in motif sequence [15]. Sequencing of entire genomes has opened the way to mapping DNA response elements for nuclear hormone receptors on a genome-wide scale [11,12]. Given that mice have been used extensively as genetic models to study nuclear receptor signaling (e.g. [19]), we compared the frequency and distribution of consensus direct repeat response elements [RGKTCA<sub>*i*</sub>RGKTCA, *i* = 1 to 5] both in the mouse and human genomes (Table 1). Although the frequency of these motifs should be comparable in random DNA

sequences, our screen revealed that DR2 elements are about 10-fold more frequent than other DR sequences in the human genome, and are about 4-fold more common than in the mouse genome (108,582 vs 26,717). In addition, DR1 and DR4 elements occur more frequently than DR3 or DR5 elements (about 2-fold). To investigate the varying frequencies of different DR elements in the human and mouse genomes, we examined the proportions of each of these motifs found in transposable elements (TEs, Table 1). The vast majority of human DR2 motifs map within SINEs, in particular in Alu repeats (102,359/102,564 elements, see Figure 1 and Table 1), while approximately 50% of DR2 elements in the mouse (13,356) are found in retroviral LTRs, which only account for 1064 of the human elements (Table 1). DR4 motifs were also significantly associated within SINEs in the human genome, whereas DR1 motifs in TEs are found mostly in LINES (Table 1). The frequencies of DR elements outside TEs is relatively similar, indicating that incorporation of DR elements within SINEs, LINES or LTRs is the main cause for their variable distributions.

### **Over 100,000 consensus DR elements are present in Alu repeats**

To assess whether the large over-representation of DR2s in SINEs reflects amplification of a specific sequence, we calculated the frequencies of all motifs corresponding to the DR2 consensus. Of the 102,359 consensus DR2 elements located in Alu repeats, 92,686 corresponded to the sequence AGGTCAnnAGTTCA, and of these, AGGTCAggAGTTCA accounted for 82,319 occurrences, along with 9,877 copies with only one variation in the spacer nucleotides (data not shown). Other consensus DR2 elements containing one variation in the repeated motifs were much less frequent (less than 3,000 copies each for the single variations at the R and K positions, as calculated from the percentages in DR2 elements listed in Table 2). Consensus DR1 or DR3 motifs resulting from removal or addition of one nucleotide from the highly represented DR2 sequence were even more infrequent in Alu repeats (less than 0.2% or about 200 copies, Table 2, line 2).



The 102,359 DR2 elements, which are RAR target sequences, are by far the most frequent consensus DR motif present in Alu repeats (93.4%, Fig. 1), followed by DR4 elements, recognized specifically by TRs (5.7% or 6198 occurrences, Fig. 1). It is noteworthy that the 6198 Alu-DR4 elements represent ~50% of the total consensus DR4 elements in the genome (Fig. 1). Moreover, unlike in the DR2 consensus, RGGTCA motifs are found predominantly in the 3' repeat of the Alu DR4 sequences (Table 2, DR4 column). This position is bound by the TR when RXR-TR heterodimers bind to DR4 motifs. Since the TR has weak affinity for RGGTCA but recognizes RGGTCA motifs [20], these DR4 elements represent optimal TR binding sites.

### **The vast majority of Alu-DR elements are present in AluS sequences**

Further analysis of human DR2-Alu elements revealed that their 5' ends are tightly clustered around positions 66–68 of the Alu repeat (Fig. 2). DR1 and DR3 elements showed largely similar distributions. In contrast, the 5' ends DR4 elements were found predominantly at position 58 of the repeat (Fig. 2). While DR5 elements were somewhat more widely distributed, most motifs started at positions 58, 66 or 68. The results reveal that DR2 and DR4 elements originated from partially overlapping sequences, where the 5' half-site of the consensus DR2 element corresponds to the 3' half-site of the DR4 motif (see below).

The vast majority of consensus DR2 elements (97,779 or 95.5%; Fig. 3A) are present in AluS sequences, which represent just under 56% of total Alu sequences in the human genome. The overrepresentation of DR elements in AluS sequences, which largely drove the Alu repeat expansion 35–40 Myr ago [1,2,4], indicates that the majority of the consensus Alu-DR2 sequences (and related DR elements) present in the human genome multiplied through primate genomes during this AluS expansion. Consensus DR4 elements are similarly distributed primarily in AluS sequences (Fig. 3A). In contrast, consensus DR2 elements are markedly underrepresented in older AluJ and younger AluY sequences (2.3%

and 1.7%, respectively; Fig. 3A). Of the 1.7% of DR2 motifs (1740 in total) present in potentially mobile AluY elements, only 41 motifs were not found to be present in the chimp and rhesus macaque genomes [see Additional file 1], indicating that more recent AluY-mediated transposition was not a major driving force in shaping retinoid-regulated gene expression in the human genome.

### **Consensus Alu-DR2 elements arose predominantly via deamination of a methylated CpG dinucleotide**

The distribution and frequency of DR2 elements in Alu sequences can be explained by the fact that the consensus DR2 RARE differs from the consensus sequence of several AluS subfamilies by a single G to A substitution in the 3'-most nucleotide of the DR2 repeat, whereas generation of an RARE consensus from Alu J or Y sequences would require multiple substitutions (Fig. 3B). Similarly, a single A to G or T substitution in the third position of the upstream GGATCA motif in AluS subfamilies Sx, Sq and Sp would generate a consensus DR4 element. Note however that although two different single mutations can lead to formation of a consensus DR4 versus one single replacement leading to a consensus DR2, the rate of conversion to a consensus DR2 is far superior to that to a DR4 element in all Alu subclasses.

It is noteworthy that generation of the DR2 RARE consensus could occur through deamination of a methylated C residue in the CpG dinucleotide of the AluS consensus, converting it to a T on the complementary strand of the direct repeat. A comparison of the frequencies of all four possible CpN dinucleotides in AGGTCAnnAGTTCN motifs present in AluS sequences indicates that CpA occurs far more frequently than CpT or CpC (Fig. 3C). Similarly, the conversion to T of the C residue of the CpG on the sense strand occurred far more frequently than other substitutions (Fig. 3C). Collectively, these data support the notion that the CpA dinucleotide in the DR2 element arose predominantly via a deamination-driven mechanism rather than by random base substitutions. The data are also

consistent with other observations of high rates of CpG methylation [21] and of CpG dinucleotide decay generally in Alu repeats [22,23]. Further, the rate of conversion to consensus DR2s was similar in different subclasses of AluS elements differing by only one nucleotide (about 15–20% except for the AluSg1, a minor subfamily of AluSg [24], for which the rate of conversion is about 60%, Fig. 3B), suggesting generation of DR2s via the same methylation/deamination mechanism throughout evolution of the family. In addition, it is worth noting that the conversion to a perfect DR2 from the consensus AluS sequence, while introducing a change in the sequence of the box B of the RNA polIII promoter, occurs without altering its consensus (GWTCRANNC), suggesting that Alu elements containing perfect DR2 may still be mobile [25]. In a recent paper Price et al. defined over 200 AluS subfamilies [26], seven of which contain consensus DR2 elements. One of these, AluSg\_14 shares the presence of the consensus RARE DR2 with three subfamilies [see Additional file 4]. These AluS subfamilies are likely examples of DR2 elements amplified by transposition instead of spontaneous mutation of a CpG.

### **Distribution of Alu-DR2 elements relative to the 5'ends of human genes**

Generation of consensus DR2 in Alu elements via a mechanism implicating methylation of CpG dinucleotides, which is associated with transcriptional silencing, raises the question of whether these motifs can be recognized by RARs as functional binding sites. To investigate whether these elements could contribute to gene regulation, we first characterized the distribution of repeats containing DR2 elements relative to the 5' ends of human genes (Fig. 4). The distribution of Alu-DR2 differed from that of Alu repeats ( $\chi^2$  test, p-value = 2.84E-19), with a sharper reduction in the number of elements between -2 and +2 kb of transcriptional start sites. We note that there was a 20–30% further relative reduction in elements containing DR2 motifs within 1 kb of known 5' ends of genes compared to the general distribution of Alu elements, suggestive of a selection pressure against the presence of promoter-proximal Alu-DR2 sequences. However, RAREs are known to function as enhancer elements at distances up to 10–20 kb. Our mapping studies

revealed over 18,000 Alu-DR2 elements lying within -10 kb and +10 kb of the 5'ends of >10,000 genes, [see Additional file 2], representing a substantial proportion of human genes.

### **Association of RARs with Alu-DR2 elements in vivo**

It is highly likely that most of the genes with proximal Alu-DR2 elements are not implicated in RAR-dependent gene regulation in a given cell type, with access to the sites being potentially limited by chromatin structure or binding of other proteins. For example, chromatin immunoprecipitation studies followed by microarray analysis (ChIP-on-chip) of genes regulated by related estrogen receptors in estrogen-sensitive breast carcinoma cells have suggested that only a minority of potential high affinity estrogen response elements of a given chromosome are accessible to receptors [27]. Moreover, as mentioned above, Alu repeat sequences are highly susceptible to methylation [21], which correlates with a closed chromatin structure and transcriptional silencing. However, 363 Alu-DR2 were found near 193 known retinoic acid target genes [see Additional file 2], suggesting that some Alu-DR2 elements can drive retinoid-regulated gene expression. We therefore tested the accessibility of several of these Alu-DR2 sequences to RARs by chromatin immunoprecipitation (ChIP) assay. We chose multiple elements present in the vicinity of 5 genes known to be RA-responsive in cells of squamous epithelial origin, RAI1, GPRC5A (RAI3), SMYD5 (RAI15), RARRES1 and RARRES3, and assessed RAR binding to these elements in SCC25 cells, a relatively well-differentiated human head and neck squamous cell carcinoma (HNSCC) line [28]. SCC25 cells express RARs and are retinoid-responsive [28], and are arrested by RA in the G0/G1 phase of the cell cycle [29]. The genes RAI1 and GPRC5A were originally identified as RA-regulated genes in HNSCC [30]. Similarly, SMYD5 and RARRES 1 and 3 (retinoic acid receptor responder genes 1 and 3) were found to be induced by RA in human keratinocytes [31]. RARRES3 expression is lost in many tumors, and its product inhibits cell proliferation when transiently overexpressed [32].

ChIP analysis with an anti-panRAR antibody revealed RAR binding to all elements tested (Fig. 5), whereas no RAR binding was observed to control sequences lacking DR2 elements (data not shown). In most instances, binding was either not RA-dependent or weakly enhanced by ligand, consistent with capacity of RARs to bind DNA in the absence of ligand [15]. However, immunoprecipitation of RARs associated with elements #3 in GPRC5A and #1 in SMYD5 was consistently ligand-dependent (Fig. 5, and data not shown). These studies are important as they confirm that all of the Alu-DR2 elements tested are in a chromatin environment that renders them accessible to RARs *in vivo*. We also confirmed that an oligonucleotide corresponding to the predominant DR2 element present in Alu repeats drove robust RA-inducible gene expression when cloned upstream of a truncated thymidine kinase promoter (data not shown), supporting its potential to function as an RARE *in vivo*.

In addition, we obtained further evidence for the function of Alu-DR2 elements as RA-dependent enhancer sequences from the publicly-available results of the ENCODE Project Consortium, which mapped transcription factor binding sites including those for RARs in 44 selected regions of the human genome (30 Mb in total) [33]. ChIP-on-chip studies of these regions using retinoic acid-treated human HL60 cells identified binding regions for RAR $\alpha$ , 17 of which encompass Alu elements containing perfect DR2 motifs (Table 3). Of these, we note that expression of caveolin-1, the product of the CAV1 gene, was found to be upregulated in retinoic acid-treated HL60 cells [34]. Taken together, the data of Fig. 5 and Table 3 strongly support the capacity of Alu-DR2 sequences to function as RAREs.

## **Discussion**

Nuclear receptor signaling, particularly that of retinoic acid receptors, has been widely studied in animal models, most notably the mouse. Our findings reveal striking differences in hormone response element frequency and distribution between rodent, and

primate and human genomes. Results of the genome-wide mapping of high affinity hormone response elements for members of the nuclear receptor family of transcription factors presented herein indicate that consensus DR2 retinoic acid response elements are vastly overrepresented relative to other DR response elements in the human genome, and are enriched approximately 4-fold in the human genome relative to the mouse. This enrichment arose as a result of the presence of DR2 motifs in Alu repeats, which are primate-specific. Movement of Alu repeats and other classes of transposable elements has been recognized as a potential mechanism for introduction of RNA polymerase II transcriptional regulatory elements for a variety of transcription factors into promoter regions of genes [10,35,36], thus altering the signal transduction pathways regulating specific genes. For example, a recent study [36] used DNA-binding profiles of representatives of different classes of transcription factors from the TRANSFAC database [37] to identify transcription factor binding sites in transposable elements located in promoter regions. This study did not screen for RAREs, for which matrices are not provided in the TRANSFAC database.

While the presence of imperfect RAREs in Alu elements and their capacity to function as RAREs in reporter assays has been reported previously [17], the main findings from our study are the unexpected large number of consensus DR2 RAREs, and the demonstration that several of these sites are functional RAR binding sites *in vivo*. It is striking that the genome-wide number of imperfect DR2s corresponding to the AGGTCAnnAGTTCG consensus, which matches the consensus sequences of most AluS subclasses, is not much higher (102,657) than that of the consensus DR2s (93,278), underlining the unexpectedly high number of perfect DR2 RAREs in the human genome (Fig. 3C). The generation of consensus DR2 elements by deamination of the CpG dinucleotide within the 3' AGTTCG motif is supported by the comparably high number of AGGTCAnnAGTTTG sequences (80,444), which diverge from a perfect DR2 more than the motif found in consensus AluS sequences but represent the other possible site generated by methylation/deamination at the CpG dinucleotide in the B box of Alu repeats.

Demonstration that Alu DR2 RAREs are functional binding sites in their normal chromatin context is crucial given the high frequency of methylation of Alu sequences in the human genome [21], and the association between DNA methylation and silencing of transcription in general, and of Alu repeats specifically [38]. For instance, expression and retinoic acid responsiveness of one of the genes studied above, RARRES1, is frequently inhibited in malignancies by hypermethylation [39], and the extent of methylation is inversely correlated with the state of differentiation of malignant cells [40].

The capacity of Alu-DR2 motifs to function as RAREs is supported by several lines of evidence. We identified 363 motifs within 10 kb of the 5' ends of 193 known retinoic acid-regulated genes [see Additional file 2]. It is also noteworthy that 8 Alu-DR2 within 10 kb of 9 genes overlap with DNase I-hypersensitive sites from primary human CD4<sup>+</sup> T cells [41] [see Additional file 2]. We confirmed the capacity of several Alu-DR2 elements to bind RARs *in vivo* in retinoic acid-sensitive SCC25 cells by chromatin immunoprecipitation analysis of receptor binding to sequences located proximal to five genes known to be retinoic acid-responsive in squamous carcinoma cells. These studies indicate that access of RARs to the DR2 motifs under study was not impeded by their presence in an Alu repeats in these cells. Moreover, our analyses of ChIP-on-chip studies from the ENCODE consortium showed that several Alu-DR2 elements in human HL60 cells are sites of RA-dependent recruitment of RARs (Table 3), further supporting the function of Alu-DR2 elements as RAREs.

We found that DR2 consensus elements were present in substantial numbers in multiple AluS subfamilies, indicating that they arose throughout the course of AluS expansion. The perfect RARE DR2 motif corresponds to the consensus sequence in seven divergent AluS subfamilies defined by Price et al. [26], in addition to being found sporadically throughout numerous other AluS subfamily sequences (4 and unpublished results). It should be noted that the downstream half-site of the DR2 element lies within box B of the Alu RNA polymerase III promoter, but that the conversion of RGTTCG to

RGTTCA to generate the consensus DR2 motif does not necessarily impair its competence for retrotransposition, consistent with transcriptional activity of Alu RNA polymerase III promoters containing RGTTCAANNC box B motifs [25,42].

Consistent with their presence predominantly in AluS sequences, which represent relatively ancient Alu elements, we found that the vast majority of human Alu-DR2 motifs are present in the chimp and macaque genomes. A few Alu-DR2 elements were apparently unique to the human genome, although in some cases they corresponded to regions that were not sequenced in the simian genomes. While none of these uniquely human Alu-DR2s are adjacent to known retinoid-regulated genes in humans, it will be of interest in the future to characterize the potential of RA to regulate their expression in appropriate cell model systems. It is also noteworthy that one of these genes, CYP2A6, was shown to be regulated by the nuclear receptor hepatic nuclear factor 4 (HNF4) [43], which, like RARs, can recognize DR2 motifs [44].

Our previous work identified a consensus DR3 element recognized by the VDR *in vitro* and *in vivo* in the proximal promoter region of the cathelicidin antimicrobial peptide gene [45], which was found to be strongly regulated by vitamin D in a variety of human cell types. However, neither the regulation nor the response element was conserved in mouse, and studies by others [46] showed the DR3 to be conserved in primates, but not other species, and present in an Alu repeat. This represents a clear example of an Alu-driven divergence in nuclear receptor gene regulation between primates/humans and other species. Results presented here demonstrate that Alu repeat expansion similarly has led to the generation of functional binding sites for retinoic acid receptors. Further studies of DR2-containing Alu repeats will be necessary to determine precisely to which extent these elements have contributed to the primate-specific regulation of target genes by retinoic acid.



## Conclusion

We find that consensus DR2 motifs are heavily overrepresented in the human genome relative to other DR response elements due to their presence in a subset of Alu motifs, in particular in AluS sequences. Although DR4s are found far less frequently in Alu repeats, those present account for 50% of consensus DR4 motifs in the human genome. Consensus Alu-DR2 elements arose predominantly through deamination of a methylated CpG dinucleotide present in AluS elements rather than through random base substitutions. While Alu elements can be transcriptionally repressed by methylation-driven mechanisms, our screen identified Alu-DR2 elements in promoter proximal regions of numerous retinoid regulated genes. Importantly, we found that Alu-DR2s are accessible to RARs in vivo in two RA-responsive human cell lines, indicating that the expansion of DR2 motifs within Alu elements during the course of primate evolution contributed to altering regulation of gene expression by retinoic acid.

## Methods

### Genome-wide identification of DR1-DR5 within Alu sequences

The human (hg17, May 2004) and mouse (mm6, March 2005) reference sequence chromosomes from the UCSC Genome Browser database [47,48] were scanned for RGKTCA motifs separated by 1 to 5 bp (DR1-DR5) using a previously described perl program [11,12]. The RefSeq Genes track [49], downloaded September 21 2005, was used to identify the DR2 located -10 to +10 kb of gene 5'-ends. A custom SQL program used the chromosomal location of Alu sequences from the RepeatMasker track [50] to identify included DR1-DR5. The same program used the chromosomal location of regions bound by RAR $\alpha$  from the ENCODE Affy Sites track to identify DR2 elements within these regions. Another SQL program was used to identify the DR2 in AluY that are not included

within Human/Chimp (hg17 vs panTro1) and Human/Rhesus (hg17 vs rheMac2) pairwise BLASTZ alignment [51]. The  $\chi^2$  test to compare the distribution of Alu-DR2 to Alu repeat was carried out with the `chisq.test` function of the R statistical computing environment [52]. The cladogram of AluSg subfamilies was generated with PHYLIP 3.6 [53] using the consensus of AluS families from Repbase [54] and AluS subfamilies with RARE DR2 in their consensus from Price et al [26]. All programs were run on the bioinformatics cluster of The Quebec Bioinformatics Network (BioneQ) and are available upon request.

## ChIP assays

Approximately  $2 \times 10^7$  SCC25 cells were treated for 1 h with RA 10-6M or vehicle. Then cells were rinsed with PBS and incubated for 10 min in PBS containing 1% formaldehyde at 37°C. Cells were washed with ice-cold PBS and the cell pellets were resuspended in 300  $\mu$ l of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 8.1) and 1 $\times$  protease inhibitor cocktail (Roche Diagnostics, Laval, Que, Canada). Sonication was performed by pulsing three times for 15 s to reduce DNA to fragments of ~300–500 bp in length. Centrifugation was then performed to obtain cleared cell lysates. Lysates were diluted 1:10 in ChIP dilution buffer (1% Triton X-100, 1 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl pH 8.1). A portion of the lysate was taken as INPUT control sample. Pre-clearing was done by incubating the cell lysates with salmon sperm DNA (1  $\mu$ g/ml), BSA (1  $\mu$ g/ml) and protein A-agarose slurry (Santa Cruz Biotechnology, Santa Cruz CA) for 1 h. Cell lysates were then incubated by overnight rotation at 4°C with 4  $\mu$ g of either normal rabbit IgG or primary antibody, followed by rotation-incubation with protein A-agarose for 2 h. The beads were rinsed three times sequentially with TSE I buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 150 mM NaCl; 20 mM Tris-HCl, pH 8.1), TSE II buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 500 mM NaCl; 20 mM Tris-HCl, pH 8.1), LiCl Wash Buffer (0.25 M LiCl; 1% NP-40; 1% deoxycholate (Na salt); 1 mM EDTA; 10 mM Tris-HCl, pH 8.1) and 2 times TE buffer (10 mM Tris-HCl; 1 mM EDTA, pH 8.0). The beads were eluted twice with 250  $\mu$ l elution buffer (1% SDS; 0.1 M NaHCO<sub>3</sub>). To

reverse the cross-linking, the eluates were incubated in a 65°C incubator overnight and then the samples were purified with PCR purification kit (Qiagen, Mississauga, Ont, Canada). PCR primers used for ChIP assays are described in Additional File 3.

## **List of abbreviations**

ChIP, chromatin immunoprecipitation; DR, direct repeat; HNF4, hepatic nuclear factor 4; LINEs, long interspersed elements; RA, retinoic acid; RAR, retinoic acid receptor; RARE, retinoic acid response element; RXR, retinoid × receptor; SINEs, Short Interspersed Elements; TR, thyroid hormone receptor; VDR vitamin D receptor.

## **Authors' contributions**

DL was responsible for the genome-wide analysis of response element distribution in the human and mouse genomes, TTW performed the ChIP assays, JW and SM conceived the study, directed DL and TTW in the analysis of their work and wrote the article. All authors read and approved the final manuscript.

## **Acknowledgements**

This work was supported by an operating grant from the Canadian Institutes of Health Research to JHW and SM (MOP-74571) and by a grant from the Natural Sciences and Engineering Research Council of Canada to SM. JHW and SM are Chercheurs-boursiers of the Fonds de Recherche en Santé du Québec. DL was supported by a scholarship from a CIHR Strategic Training Program Grant in Bioinformatics.

## Figures and Tables

Figure 1 Consensus DR1 to DR5 motifs embedded within Alu elements in the human genome.

The numbers below each DR correspond to the total number of repeats in Alu elements/total number of repeats in the genome.

Figure 2 Positions of DR motifs within Alu sequences in the human genome.

For clarity, only every 10th position is indicated, along with positions 58, 66 and 68. The positions indicated correspond to the 5'nucleotide of the 5'half site. Note the differences in scales of the ordinates.

Figure 3 Consensus DR2 and DR4 elements are found predominantly in AluS sequences.

(A) Distribution of consensus DR2 and DR4 elements within AluJ, S and Y families in the human genome. (B) Consensus sequences of AluJ, AluY and AluS subfamilies within the region containing the DR4 and DR2 element, and frequency of consensus DR2 RARE sequences within each subfamily. The position of box B of the polymerase III promoter is provided above. The sequences of the DR4 and DR2 consensus motifs are provided below. (C) Nucleotide substitution frequencies of the CpG dinucleotide present within AGGTCAnnAGGTCG motifs in AluS sequences

Figure 4 Distribution of Alu sequences within the 10 kb flanking 5' ends of genes.

The 20 kb window for all genes annotated in the human genome were subdivided in 500 bp intervals. The number of Alu sequences and Alu sequences with a DR2 element were counted for each interval. Grey bars represent the ratio of the number Alu repeats within a given interval divided by the total number of Alu repeats within 10 kb of a gene, black bars represent the number DR2-containing Alu repeats within a given interval divided by the total number of Alu repeats within 10 kb of a gene, whereas black squares indicate the number Alu motifs with a DR2 element within a given interval divided by the total number of DR2-containing Alu repeats within 10 kb of a gene.

Figure 5 RARs bind to Alu-DR2 motifs present in retinoic acid responsive genes in vivo.

Upper panel: Schematic representation of the promoter regions of selected known retinoic acid-responsive genes containing multiple consensus DR2 RAREs (AGGTCAggAGTTCA) embedded within an Alu repeat sequence. Lower Panel: Results of ChIP assays for RAR binding to promoter regions containing DR2 motifs. DNA from all immunoprecipitates (RAR-specific and control IgG) was amplified under similar conditions (40–42 cycles of PCR).

Table 1 Consensus RGKTCA direct repeats found in the human and mouse genomes.

Table 2 RGKTCA direct repeats found in Alu elements.

Table 3 Alu-DR2 elements within ENCODE ChIP-on-chip RAR $\alpha$  binding regions.

Figure 1

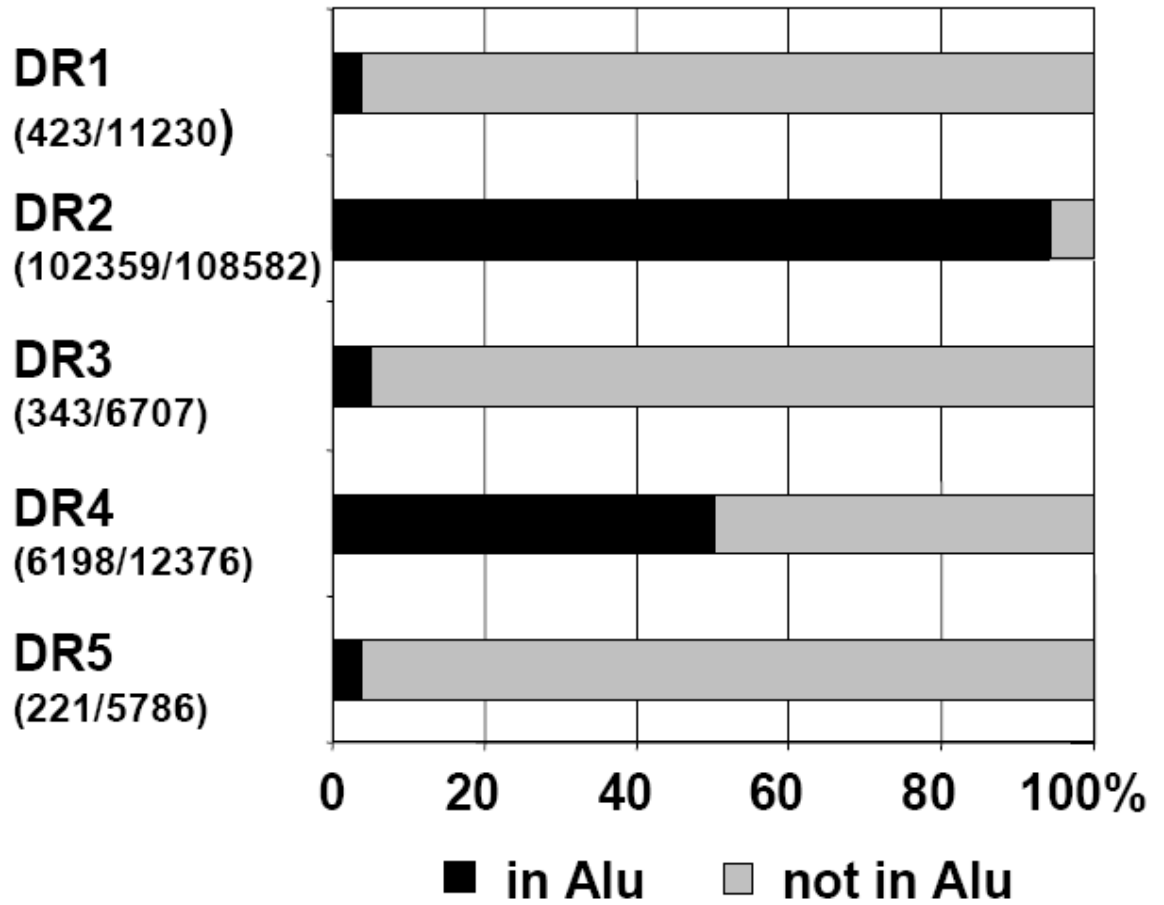


Figure 2

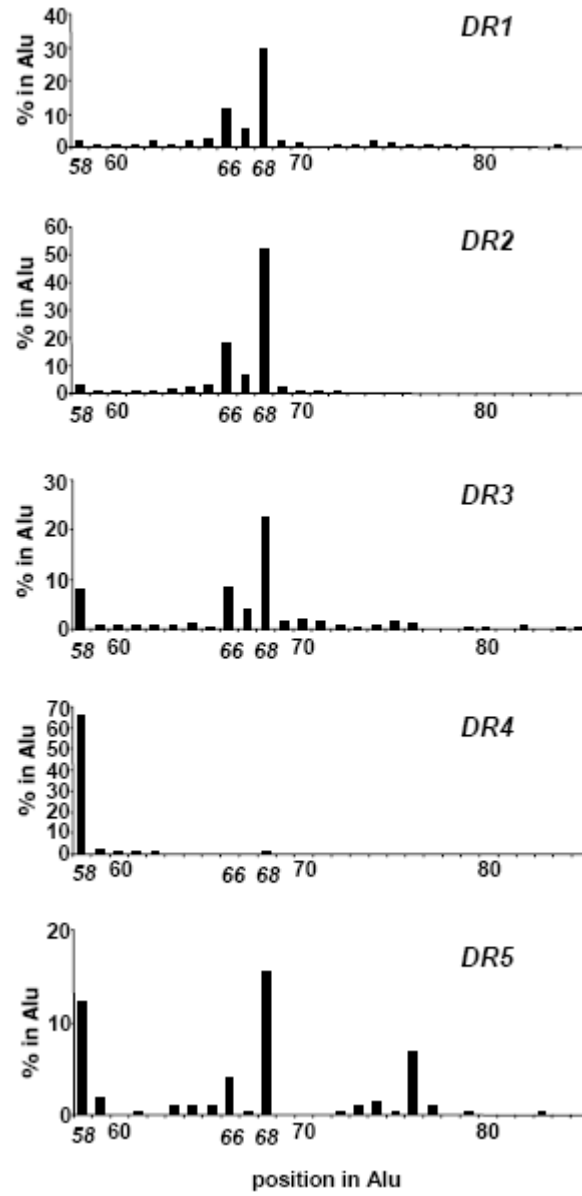


Figure 3

**A**

Alu Family	% Alu in genome	# DR2	% DR2 in Alu	# DR4	% DR4 in Alu
AluJ	24.48	2359	2.30	150	2.42
AluS	55.91	97779	95.53	5997	96.76
AluY	12.67	1740	1.70	23	0.37
other	6.94	481	0.47	28	0.45
	<b>100.00</b>	<b>102359</b>	<b>100.00</b>	<b>6198</b>	<b>100.00</b>

**B**

	B-BOX	GWTCRANNC	% with DR2	% with DR4
7SLRNA	GGTGGGAGGATCGCTTGAGTCCAGGAGTTCTGGGCTGTAG			
AluJb	GGCGGGAGGATCACTTGAGCCCAGGAGTTTCGAGACCAGCC		0.74	0.06
AluSx	GGCGGGCGGATCACCTGAGGTCAGGAGTTTCGAGACCAGCC		16.18	1.29
AluSg	GGCGGGCGGATCAC--GAGGTCAGGAGTTTCGAGACCAGCC		20.67	0.01
AluSg1	GGCGGGCGGATCAC--GAGGTCAGGAGTTTCGAGACCAGCC		59.81	0.09
AluSq	GGCGGGCGGATCACCTGAGGTCAGGAGTTTCGAGACCAGCC		18.97	1.35
AluSp	GGCGGGCGGATCACCTGAGGTCGGGAGTTTCGAGACCAGCC		6.44	0.46
AluSc	GGCGGGCGGATCAC--GAGGTCAGGAGATCGAGACCATCC		1.41	0.01
AluY	GGCGGGCGGATCAC--GAGGTCAGGAGATCGAGACCATCC		1.20	0.01
	** *** ***** * *** * *** ** * *			
DR2		RGKTCANNRGKTCA		
DR4		RGKTCANNNNRGKTCA		

**C**

DR2	genome	in Alu	outside Alu
AGGTCA NN AGTTaA	1182	794	388
AGGTCA NN AGTTac	403	153	250
AGGTCA NN AGTTag	3205	2906	299
AGGTCA NN AGTTat	367	33	334
AGGTCA NN AGTTCA	93278	92686	592
AGGTCA NN AGTTCc	7662	7352	310
AGGTCA NN AGTTcg	102657	102578	79
AGGTCA NN AGTTct	2914	2377	537
AGGTCA NN AGTTgA	1921	1588	333
AGGTCA NN AGTTgc	374	126	248
AGGTCA NN AGTTgg	6984	6627	357
AGGTCA NN AGTTgt	390	80	310
AGGTCA NN AGTTtA	3615	3173	442
AGGTCA NN AGTTtc	1625	1219	406
AGGTCA NN AGTTtg	80444	79882	562
AGGTCA NN AGTTtt	1336	758	578
<b>Total</b>	<b>308357</b>	<b>302332</b>	<b>6025</b>



Figure 4

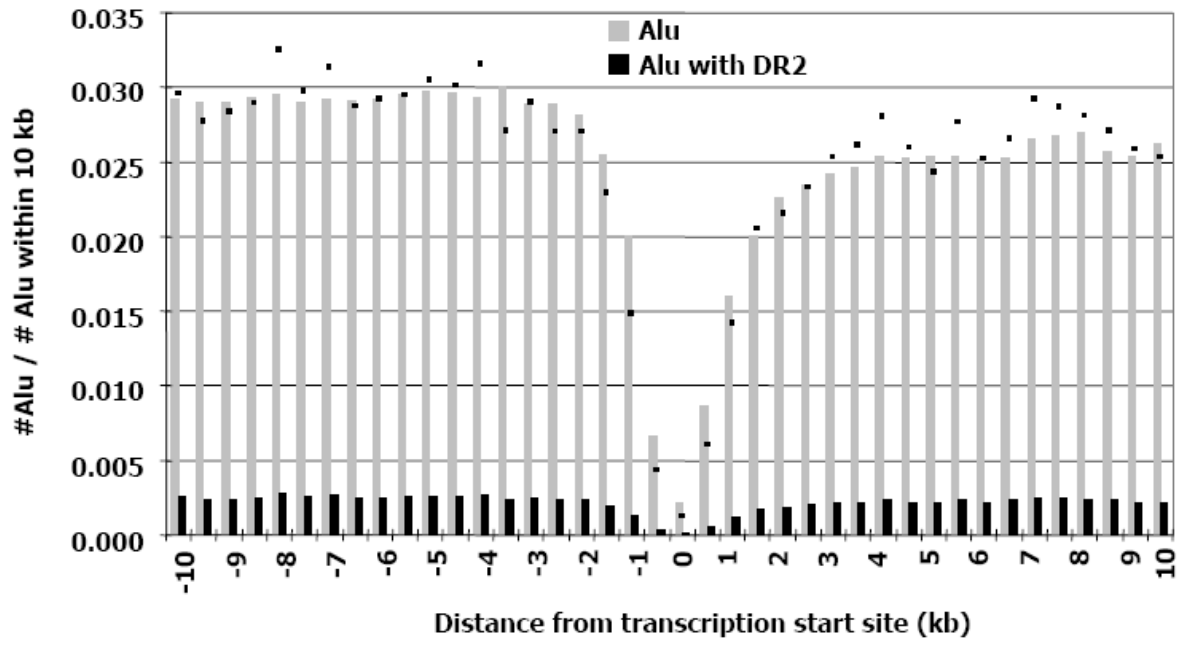


Figure 5

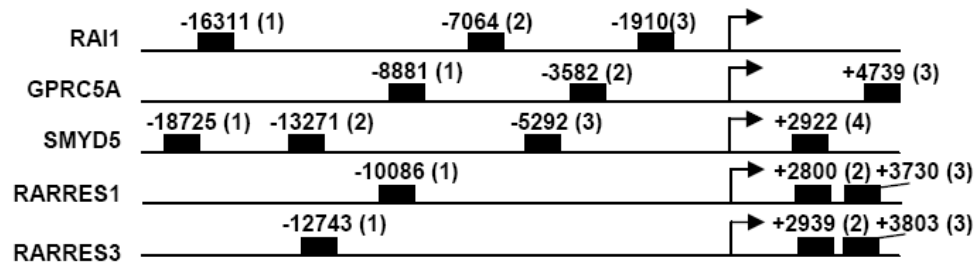


Table 1

	Total		outside transposable elements		in SINEs		in LINEs		in LTRs	
	Human	Mouse	Human	Mouse	Human	Mouse	Human	Mouse	Human	Mouse
<b>DR1</b>	11230	16250	5950	9741	687	3679	3633	788	960	2042
<b>DR2</b>	108582	26717	4186	5623	102564	7157	768	581	1064	13356
<b>DR3</b>	6707	6336	3935	4192	997	571	936	605	839	968
<b>DR4</b>	12376	8893	4397	5549	6602	638	621	457	756	2249
<b>DR5</b>	5786	8846	4337	5338	430	1059	552	1073	467	1376

Table 2

DR sequence	# in Alu	% in DR1	% in DR2	% in DR3	% in DR4	% in DR5
AGGICA AGGICA	2888	0.90	49.90	0.73	47.64	0.83
AGGICA AGTICA	93067	0.19	99.59	0.11	0.06	0.05
AGGICA GGGICA	78	23.08	39.74	6.41	25.64	5.13
AGGICA GGITCA	2212	2.40	96.11	1.04	0.31	0.14
AGITCA AGGICA	524	0.95	46.95	1.72	49.62	0.76
AGITCA AGTICA	2389	1.21	95.10	1.80	0.63	1.26
AGITCA GGGICA	10	20.00	60.00	0.00	20.00	0.00
AGITCA GGITCA	137	12.41	78.10	3.65	2.19	3.65
GGGICA AGGICA	5244	0.08	29.67	0.27	69.68	0.30
GGGICA AGTICA	1285	0.70	91.91	0.47	6.69	0.23
GGGICA GGGICA	94	1.06	27.66	9.57	54.26	7.45
GGGICA GGITCA	44	11.36	68.18	2.27	13.64	4.55
GGITCA AGGICA	1096	0.18	42.43	0.27	55.84	1.28
GGITCA AGTICA	197	5.58	41.11	27.92	13.71	11.68
GGITCA GGGICA	21	9.52	61.91	0.00	19.05	9.52
GGITCA GGITCA	258	24.03	35.66	18.61	7.75	13.95
RGKICA RGKICA	109544	0.39	93.44	0.31	5.66	0.20

Table 3

<b>DR2 position</b>	<b>RAR<math>\alpha</math> binding region</b>	<b>region length</b>	<b>DR2 position in region</b>	<b>nearest gene</b>	<b>distance (kb)</b>
chr11:5234577-5234590	chr11:5234153-5234769	617	437	<i>HBG2</i>	-2
chr11:116438036-116438049	chr11:116437268-116438894	1627	768	<i>KIAA0999</i>	36.1
chr11:5214973-5214986	chr11:5214632-5215086	455	354	<i>HBD</i>	-2.5
chr15:41723046-41723059	chr15:41722559-41723965	1407	500	<i>CATSPER2</i>	5.2
chr15:41832787-41832800	chr15:41832276-41832882	607	524	<i>PDIA3</i>	6.9
chr15:41893221-41893234	chr15:41892245-41893398	1154	989	<i>MFAP1</i>	11
chr19:59507050-59507063	chr19:59506478-59507251	774	572	<i>LILRA5</i>	9.1
chr20:33783677-33783690	chr20:33783609-33783958	350	68	<i>RBM39 (RNPC2)</i>	9.9
chr20:33795208-33795221	chr20:33795121-33795786	666	87	<i>RBM39 (RNPC2)</i>	-1.6
chr21:33890599-33890612	chr21:33890070-33890930	861	529	<i>DONSON</i>	-7.7
chr21:39545848-39545861	chr21:39545687-39546093	407	161	<i>BRWD1</i>	61.5
chr22:30446428-30446441	chr22:30446058-30446529	472	383	<i>C22orf30 (MGC50372)</i>	-13.1
chr6:74021882-74021895	chr6:74021526-74021987	462	369	<i>C6orf148</i>	7.7
chr6:108492975-108492988	chr6:108492625-108493122	498	363	<i>OSTM1</i>	9.6
chr7:115763054-115763067	chr7:115762699-115763680	982	368	<i>CAV1</i>	4.2
chr7:116952987-116953000	chr7:116952795-116953258	464	192	<i>CTTNBP2</i>	154.5
chrX:122568351-122568364	chrX:122567587-122568522	936	777	<i>THOC2</i>	-12.1

## **Additional files**

Additional file 1 A table of DR2 embedded in Alu found in Human but not Chimp and Rhesus genome.

Additional file 2 A table of all DR2 in Alu within 10 kb of human gene 5' ends.

Additional file 3 ChIP Primers for genes with DR2-Alu.

Additional file 4 AluS subfamilies with RARE DR2 in their consensus.

## Additional file 1

Chr.	Str.	DR2	Dist.	mRNA	Gene	EntrezGene (8)	ChimpGene	Alu	Pos.
chr1:116623842-116623855	+	AGGTCACGAGGTCA	-4036	NM_000701	ATP1A1	476	457152	Y	57
chr1:116623842-116623855	+	AGGTCACGAGGTCA	-4036	NM_001001586	ATP1A1	476	457152	Y	57
chr3:9899460-9899473	-	GGGTCATGAGGTCA	-3733	NM_022094	CIDEA	63924	---	Y	57
chr19:46055199-46055212	-	AGGTCAGGAGTCA	-7032	NM_000762	CYP2A6	1548	---	Yb9	67
chr19:51946883-51946896	-	AGGTCAGGAGTCA	5741	NM_024301	FKBP	79147	---	Yb9	67
chr3:9899460-9899473	-	GGGTCATGAGGTCA	-7799	NM_032492	JAGN1	84522	460157	Y	57
chrX:46689602-46689615	+	AGTTCACGAGGTCA	-4427	NM_004683	RGN	9104	465598	Y	56
chrX:46689602-46689615	+	AGTTCACGAGGTCA	-4427	NM_152869	RGN	9104	465598	Y	56
chr19:51946883-51946896	-	AGGTCAGGAGTCA	-5369	NM_013403	STRN4	29888	---	Yb9	67
chr2:108316677-108316690	+	GGGTCATGAGGTCA	-5492	NM_001008743	SULT1C3	442038	470467	Y	53

Chr.	Str.	DR2	Alu	Pos.
chr10:64467365-64467378	+	GGGTCACGAGGTCA	Y	57
chr11:51043146-51043159	+	GGGTCACGAGGTCA	Y	55
chr11:54455200-54455213	-	GGGTCACGAGGTCA	Y	55
chr13:33107905-33107918	+	AGGTCAGGAGTCA	Y	67
chr14:101810268-101810281	-	AGGTCAGGAGGTCA	Y	62
chr14:30757544-30757557	-	GGTTCACGAGGTCA	Y	56
chr17:18602981-18602994	+	AGGTCACGAGGTCA	Y	57
chr17:4231043-4231056	+	GGTTCACGAGGTCA	Y	57
chr19:10794558-10794571	+	AGGTCACGAGGTCA	Y	57
chr19:46055199-46055212	-	AGGTCAGGAGTCA	Yb9	67
chr19:51946883-51946896	-	AGGTCAGGAGTCA	Yb9	67
chr1:116623842-116623855	+	AGGTCACGAGGTCA	Y	57
chr1:1469645-1469658	-	AGTTCAGGAGGTCA	Y	59
chr1:31986497-31986510	-	GGGTCACGAGGTCA	Y	57
chr1_random:224849-224862	-	AGTTCAGGAGGTCA	Y	59
chr1_random:246332-246345	-	GGGTCATGAGGTCA	Y	57
chr20:61882383-61882396	+	AGGTCAGGAGTCA	Y	65
chr22:23610391-23610404	-	AGGTCAGGAGTCA	Y	67
chr2:16198987-16199000	+	GGGTCACGAGGTCA	Y	57
chr2:108316677-108316690	+	GGGTCATGAGGTCA	Y	53
chr3:146448022-146448035	-	AGGTCACAAGGTCA	Y	57
chr3:47544015-47544028	-	AGGTCACGAGGTCA	Y	57
chr3:9899460-9899473	-	GGGTCATGAGGTCA	Y	57
chr4:180467294-180467307	+	GGTTCATGAGGTCA	Y	57
chr4_random:761601-761614	-	GGGTCACGAGGTCA	Y	57
chr7:44728485-44728498	+	GGTTCACAAGGTCA	Y	57
chr8:38067438-38067451	+	GGGTCACGAGGTCA	Y	57
chr9:38780230-38780243	+	AGGTCACGAGGTCA	Y	57
chr9:44047091-44047104	-	AGGTCACGAGGTCA	Y	57
chr9:66132089-66132102	-	AGGTCACGAGGTCA	Y	57
chrX:513704-513717	-	AGGTCAGGAGTCA	Y	90

chrX:2065380-2065393	-	AGTTCACGAGGTCA	Y	57
chrX:46689602-46689615	+	AGTTCACGAGGTCA	Y	56
chrX:2429905-2429918	+	GGGTCACGAGGTCA	Y	57
chrX:90803297-90803310	-	GGGTCATGAGGTCA	Y	41
chrX:73645512-73645525	+	GGTTCACGAGGTCA	Y	57
chrY:13236927-13236940	-	AGGTCACGAGGTCA	Y	38
chrY:513704-513717	-	AGGTCAGGAGTTCA	Y	90
chrY:2065380-2065393	-	AGTTCACGAGGTCA	Y	57
chrY:21139272-21139285	-	AGTTCATGAGGTCA	Y	56
chrY:2429905-2429918	+	GGGTCACGAGGTCA	Y	57

## Additional file 2

DR2	Dist.	mRNA	Gene	Alu	Pos.	note
AGGTCAGGAGTTCA	-8881	NM_003979	<i>GPRC5A</i>	Sp	67	gene has known RARE
AGGTCAGGAGTTCA	-3582	NM_003979	<i>GPRC5A</i>	Sq	67	gene has known RARE
AGGTCAGGAGTTCA	4739	NM_003979	<i>GPRC5A</i>	Y	64	gene has known RARE
AGGTCAGGAGTTCA	-7064	NM_030665	<i>RAI1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-1910	NM_030665	<i>RAI1</i>	Sp	66	gene responsive to RA
AGGTCAGGAGTTCA	3730	NM_002888	<i>RARRES1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3730	NM_206963	<i>RARRES1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2800	NM_002888	<i>RARRES1</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	2800	NM_206963	<i>RARRES1</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	2939	NM_004585	<i>RARRES3</i>	Sp	67	gene has known RARE
AGGTCAGGAGTTCA	3803	NM_004585	<i>RARRES3</i>	Sx	67	gene has known RARE
AGGTCAGGAGTTCA	-5292	NM_006062	<i>SMYD5</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-2561	NM_006062	<i>SMYD5</i>	Sg	63	gene has known RARE
AGGTCAGGAGTTCA	2922	NM_006062	<i>SMYD5</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	5440	NM_000616	<i>CD4</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-503	NM_015308	<i>FNBP4</i>	Sg	50	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	2360	NM_005685	<i>GTF2IRD1</i>	Sg	65	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	2677	NM_016328	<i>GTF2IRD1</i>	Sg	65	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-4554	NM_024048	<i>MGC3020</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-5944	NM_017851	<i>PARP16</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-3183	NM_004195	<i>TNFRSF18</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-3183	NM_148901	<i>TNFRSF18</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-3183	NM_148902	<i>TNFRSF18</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	4240	NM_003327	<i>TNFRSF4</i>	Sx	67	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	-3323	NM_024023	<i>UNKL</i>	Sx	65	DR2 is in DNase hypersensitive site
AGGTCAGGAGTTCA	9993	NM_002291	<i>LAMB1</i>	Sx	67	DR2 is in DNase hypersensitive site ; gene responsive to RA
AGGTCAGGAGTTCA	7745	NM_030568	<i>C6orf148</i>	Sx	70	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	5272	NM_054020	<i>CATSPER2</i>	Sx	66	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	5272	NM_172095	<i>CATSPER2</i>	Sx	66	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	5272	NM_172096	<i>CATSPER2</i>	Sx	66	DR2 within region identified by ChIP-chip with RARA



AGGTCAGGAGTTCA	5272	NM_172097	<i>CATSPER2</i>	Sx	66	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	4277	NM_001753	<i>CAV1</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGAAGTTCA	-7715	NM_017613	<i>DONSON</i>	Sq	61	DR2 within region identified by ChIP-chip with RARA
AGGTCAGAAGTTCA	-7715	NM_145794	<i>DONSON</i>	Sq	61	DR2 within region identified by ChIP-chip with RARA
AGGTCAGAAGTTCA	-7715	NM_145795	<i>DONSON</i>	Sq	61	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-2552	NM_000519	<i>HBD</i>	Sp	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-6927	NM_000559	<i>HBG1</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-2003	NM_000184	<i>HBG2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9130	NM_181879	<i>LILRA5</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9130	NM_181985	<i>LILRA5</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9130	NM_181986	<i>LILRA5</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9171	NM_021250	<i>LILRA5</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9646	NM_014028	<i>OSTM1</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	6918	NM_005313	<i>PDIA3</i>	Sx	53	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9930	NM_004902	<i>RNPC2</i>	Sg	44	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9930	NM_184234	<i>RNPC2</i>	Sg	44	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9930	NM_184237	<i>RNPC2</i>	Sg	44	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9930	NM_184241	<i>RNPC2</i>	Sg	44	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	9930	NM_184244	<i>RNPC2</i>	Sg	44	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-1601	NM_004902	<i>RNPC2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-1601	NM_184234	<i>RNPC2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-1601	NM_184237	<i>RNPC2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-1601	NM_184241	<i>RNPC2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGAGTTCA	-1601	NM_184244	<i>RNPC2</i>	Sx	67	DR2 within region identified by ChIP-chip with RARA
AGGTCAGGGTTCA	5076	NM_000783	<i>CYP26A1</i>	Sx	66	gene has known RARE
AGGTCAGGGTTCA	5511	NM_057157	<i>CYP26A1</i>	Sx	66	gene has known RARE
AGGTCAGAAGTTCA	-8199	NM_000224	<i>KRT18</i>	Sg	65	gene has known RARE

AGTTCAGAAGTTCA	-8011	NM_199187	<i>KRT18</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	7888	NM_000224	<i>KRT18</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	8076	NM_199187	<i>KRT18</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	8820	NM_000224	<i>KRT18</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	9008	NM_199187	<i>KRT18</i>	Sg	65	gene has known RARE
AGGTCAGGAGTTCA	-452	NM_000250	<i>MPO</i>	Sx	67	gene has known RARE
AGGTCAGGAGTTCA	-5339	NM_000250	<i>MPO</i>	Sx	67	gene has known RARE
AGGTCAGGAGTTCA	9401	NM_000930	<i>PLAT</i>	Sx	67	gene has known RARE
AGGTCAGGAGTTCA	9401	NM_000931	<i>PLAT</i>	Sx	67	gene has known RARE
AGGTCAGGAGTTCA	9401	NM_033011	<i>PLAT</i>	Sx	67	gene has known RARE
AGGTCACGAGTTCA	2364	NM_000930	<i>PLAT</i>	Sq	67	gene has known RARE
AGGTCACGAGTTCA	2364	NM_000931	<i>PLAT</i>	Sq	67	gene has known RARE
AGGTCACGAGTTCA	2364	NM_033011	<i>PLAT</i>	Sq	67	gene has known RARE
AGGTCAGAAGTTCA	6684	NM_000392	<i>ABCC2</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	7128	NM_001878	<i>CRABP2</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAAGAGTTCA	3347	NM_001878	<i>CRABP2</i>	Sg	63	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-7915	NM_001878	<i>CRABP2</i>	Sq	59	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-3179	NM_000782	<i>CYP24A1</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-7207	NM_000799	<i>EPO</i>	Sg	64	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-842	NM_000799	<i>EPO</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-6094	NM_005318	<i>HIF0</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	6101	NM_005318	<i>HIF0</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGAAGTTCA	8919	NM_005522	<i>HOXA1</i>	Sq	67	gene has RARE conserved in mouse and rat
AGGTCAGAAGTTCA	8919	NM_153620	<i>HOXA1</i>	Sq	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-9005	NM_002178	<i>IGFBP6</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	4031	NM_002178	<i>IGFBP6</i>	Sg	65	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	6854	NM_002178	<i>IGFBP6</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-8023	NM_001012333	<i>MDK</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-7939	NM_002391	<i>MDK</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-7338	NM_001012334	<i>MDK</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-9174	NM_006617	<i>NES</i>	Sx	67	gene has RARE conserved in mouse and rat

AGGTCAGGAGTTCA	9617	NM_003297	<i>NR2C1</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	9765	NM_001032287	<i>NR2C1</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	4831	NM_000915	<i>OXT</i>	Sx	71	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-176	NM_203289	<i>POU5F1</i>	Sg	64	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	4232	NM_002701	<i>POU5F1</i>	Sg	64	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-1225	NM_203289	<i>POU5F1</i>	Sx	64	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	3183	NM_002701	<i>POU5F1</i>	Sx	64	gene has RARE conserved in mouse and rat
AGTTCAGGAGTTCA	-5157	NM_002737	<i>PRKCA</i>	Sx	66	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-9134	NM_001024809	<i>RARA</i>	Sg	65	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	2164	NM_000964	<i>RARA</i>	Sg	65	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-6108	NM_001024809	<i>RARA</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	5190	NM_000964	<i>RARA</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-6360	NM_000966	<i>RARG</i>	Sq	73	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	1134	NM_000537	<i>REN</i>	Sp	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-1330	NM_000545	<i>TCF1</i>	Sx	67	gene has RARE conserved in mouse and rat
AGGTCAGGAGTTCA	-4029	NM_001088	<i>AANAT</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8571	NM_001088	<i>AANAT</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-8876	NM_007313	<i>ABL1</i>	Sx	45	gene responsive to RA
AGGTCAGGAGTTCA	-7038	NM_007313	<i>ABL1</i>	Jo	84	gene responsive to RA
AGTTCAGGAGTTCA	-5615	NM_007313	<i>ABL1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2448	NM_007313	<i>ABL1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9142	NM_005157	<i>ABL1</i>	Sq	25	gene responsive to RA
AGGTCAGGAGTTCA	5298	NM_001100	<i>ACTA1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3210	NM_001100	<i>ACTA1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3832	NM_001100	<i>ACTA1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9801	NM_001117	<i>ADCYAPI</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	8399	NM_005100	<i>AKAP12</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-4469	NM_144497	<i>AKAP12</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2352	NM_144497	<i>AKAP12</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3660	NM_144497	<i>AKAP12</i>	Sx	55	gene responsive to RA
AGGTCAGGAGTTCA	6245	NM_000478	<i>ALPL</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7523	NM_001647	<i>APOD</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9634	NM_001647	<i>APOD</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-8789	NM_000484	<i>APP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8789	NM_201413	<i>APP</i>	Sx	67	gene responsive to RA

AGGTCAAGAGTTCA	-8789	NM_201414	<i>APP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9381	NM_000044	<i>AR</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4771	NM_004043	<i>ASMT</i>	Sg1	64	gene responsive to RA
AGGTCAGGAGTTCA	-4771	NM_004043	<i>ASMT</i>	Sg1	64	gene responsive to RA
AGTTCATGAGGTCA	-596	NM_004043	<i>ASMT</i>	Sg	57	gene responsive to RA
AGTTCATGAGGTCA	-596	NM_004043	<i>ASMT</i>	Sg	57	gene responsive to RA
AGGTCAGGAGTTCA	8392	NM_004043	<i>ASMT</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8392	NM_004043	<i>ASMT</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9367	NM_004043	<i>ASMT</i>	Sq/x	67	gene responsive to RA
AGGTCAGGAGTTCA	9367	NM_004043	<i>ASMT</i>	Sq/x	67	gene responsive to RA
AGGTCAGGAGTTCA	-7554	NM_000633	<i>BCL2</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-7554	NM_000657	<i>BCL2</i>	Sg	65	gene responsive to RA
GGGTCAGGAGTTCA	-5421	NM_001165	<i>BIRC3</i>	Sq	67	gene responsive to RA
GGGTCAGGAGTTCA	-5421	NM_182962	<i>BIRC3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-1697	NM_001165	<i>BIRC3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1697	NM_182962	<i>BIRC3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4124	NM_001165	<i>BIRC3</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	4124	NM_182962	<i>BIRC3</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-5993	NM_001012270	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5993	NM_001012271	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5993	NM_001168	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4855	NM_001012270	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4855	NM_001012271	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4855	NM_001168	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2197	NM_001012270	<i>BIRC5</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-2197	NM_001012271	<i>BIRC5</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-2197	NM_001168	<i>BIRC5</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	1006	NM_001012270	<i>BIRC5</i>	Sx	33	gene responsive to RA
AGGTCAGGAGTTCA	1006	NM_001012271	<i>BIRC5</i>	Sx	33	gene responsive to RA
AGGTCAGGAGTTCA	1006	NM_001168	<i>BIRC5</i>	Sx	33	gene responsive to RA
AGGTCAGGAGTTCA	2916	NM_001012270	<i>BIRC5</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	2916	NM_001012271	<i>BIRC5</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	2916	NM_001168	<i>BIRC5</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	3212	NM_001012270	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3212	NM_001012271	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3212	NM_001168	<i>BIRC5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8004	NM_004334	<i>BST1</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	4527	NM_000067	<i>CA2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7816	NM_002982	<i>CCL2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5740	NM_057182	<i>CCNE1</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	6400	NM_001238	<i>CCNE1</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-7818	NM_001775	<i>CD38</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8875	NM_001024844	<i>CD82</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8875	NM_002231	<i>CD82</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4441	NM_001793	<i>CDH3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9948	NM_001793	<i>CDH3</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-5348	NM_000389	<i>CDKN1A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5348	NM_078467	<i>CDKN1A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8796	NM_004936	<i>CDKN2B</i>	Sq	56	gene responsive to RA
AGGTCAGGAGTTCA	-8796	NM_078487	<i>CDKN2B</i>	Sq	56	gene responsive to RA

AGGTCAGCAGTTCA	-9130	NM_004936	<i>CDKN2B</i>	Sx	67	gene responsive to RA
AGGTCAGCAGTTCA	-9130	NM_078487	<i>CDKN2B</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8269	NM_000078	<i>CETP</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	5283	NM_000078	<i>CETP</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5591	NM_000078	<i>CETP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6569	NM_000078	<i>CETP</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	-7572	NM_020984	<i>CHAT</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-9598	NM_007096	<i>CLTA</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-9581	NM_001833	<i>CLTA</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-8742	NM_007096	<i>CLTA</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8725	NM_001833	<i>CLTA</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5589	NM_007096	<i>CLTA</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5606	NM_001833	<i>CLTA</i>	Sx	67	gene responsive to RA
AGTTCAGGAGTTCA	-4791	NM_004378	<i>CRABP1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8909	NM_000757	<i>CSF1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8909	NM_172210	<i>CSF1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8909	NM_172211	<i>CSF1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8909	NM_172212	<i>CSF1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-1134	NM_001317	<i>CSH1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1134	NM_022640	<i>CSH1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1134	NM_022641	<i>CSH1</i>	Sx	67	gene responsive to RA
AGGTCAGGGGTTCA	4302	NM_005212	<i>CSN3</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	8279	NM_001908	<i>CTSB</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	8279	NM_147780	<i>CTSB</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	8279	NM_147781	<i>CTSB</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	8279	NM_147782	<i>CTSB</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	8279	NM_147783	<i>CTSB</i>	Sg	65	gene responsive to RA
AGGTCAGGAGGTTCA	1492	NM_001908	<i>CTSB</i>	Sx	27	gene responsive to RA
AGGTCAGGAGGTTCA	1492	NM_147780	<i>CTSB</i>	Sx	27	gene responsive to RA
AGGTCAGGAGGTTCA	1492	NM_147781	<i>CTSB</i>	Sx	27	gene responsive to RA
AGGTCAGGAGGTTCA	1492	NM_147782	<i>CTSB</i>	Sx	27	gene responsive to RA
AGGTCAGGAGGTTCA	1492	NM_147783	<i>CTSB</i>	Sx	27	gene responsive to RA
AGGTCAGGAGTTCA	-7405	NM_001908	<i>CTSB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7405	NM_147780	<i>CTSB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7405	NM_147781	<i>CTSB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7405	NM_147782	<i>CTSB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7405	NM_147783	<i>CTSB</i>	Sq	67	gene responsive to RA
AGGTCAGGGGTTCA	6273	NM_000396	<i>CTSK</i>	Sq	63	gene responsive to RA
AGGTCAGGAGTTCA	-1915	NM_000499	<i>CYP11A1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6768	NM_000499	<i>CYP11A1</i>	Sq	61	gene responsive to RA
AGGTCAGGAGTTCA	344	NM_017460	<i>CYP3A4</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6096	NM_004393	<i>DAG1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2747	NM_004393	<i>DAG1</i>	Sq	67	gene responsive to RA
AGTTCAGGAGTTCA	6855	NM_001922	<i>DCT</i>	Sx	67	gene responsive to RA
GGGTCAGGAGTTCA	2961	NM_001922	<i>DCT</i>	Sq	55	gene responsive to RA
AGGTCAGGAGTTCA	-3450	NM_004939	<i>DDX1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7899	NM_004939	<i>DDX1</i>	Sx	67	gene responsive to RA
AGGTCAGCAGTTCA	2003	NM_030881	<i>DDX17</i>	Sx	63	gene responsive to RA
AGGTCAGGAGTTCA	-1811	NM_006386	<i>DDX17</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	-3588	NM_006386	<i>DDX17</i>	Sx	62	gene responsive to RA

AGGTCAGGAGTTCA	-6558	NM_006386	<i>DDX17</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7487	NM_006386	<i>DDX17</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-8952	NM_006386	<i>DDX17</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-1479	NM_000792	<i>DIO1</i>	Jo	67	gene responsive to RA
AGGTCAGGAGTTCA	-1479	NM_213593	<i>DIO1</i>	Jo	67	gene responsive to RA
AGGTCAGGAGTTCA	-665	NM_000792	<i>DIO1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-665	NM_213593	<i>DIO1</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	3073	NM_000792	<i>DIO1</i>	Sx	65	gene responsive to RA
AGGTCAGAAGTTCA	3073	NM_213593	<i>DIO1</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	6201	NM_000792	<i>DIO1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6201	NM_213593	<i>DIO1</i>	Sx	67	gene responsive to RA
AGTTCAGGAGTTCA	8258	NM_001362	<i>DIO3</i>	FLAM_C	63	gene responsive to RA
GGGTCATGAGGTCA	-6189	NM_001387	<i>DPYSL3</i>	Y	57	gene responsive to RA
AGGTCAGGAGTTCA	5698	NM_001941	<i>DSC3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	5698	NM_024423	<i>DSC3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-2473	NM_004230	<i>EDG5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3305	NM_004230	<i>EDG5</i>	Sg	27	gene responsive to RA
AGGTCAGGAGTTCA	6289	NM_001964	<i>EGR1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8016	NM_001964	<i>EGR1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-1052	NM_002759	<i>EIF2AK2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	838	NM_001423	<i>EMP1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6400	NM_001005862	<i>ERBB2</i>	Sp	61	gene responsive to RA
AGGTCAGGAGTTCA	-5278	NM_001005862	<i>ERBB2</i>	Sg	65	gene responsive to RA
GGGTCACAAGGTCA	-2887	NM_001005862	<i>ERBB2</i>	Y	57	gene responsive to RA
AGGTCAGGAGTTCA	-7153	NM_004448	<i>ERBB2</i>	Sx	70	gene responsive to RA
AGGTCAGGAGTTCA	4708	NM_001005862	<i>ERBB2</i>	Sx	70	gene responsive to RA
AGGTCAGCAGTTCA	4358	NM_004448	<i>ERBB2</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6618	NM_001005915	<i>ERBB3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6618	NM_001982	<i>ERBB3</i>	Sg	65	gene responsive to RA
AGGTCAGGGGTTCA	-5277	NM_001005915	<i>ERBB3</i>	Sq	67	gene responsive to RA
AGGTCAGGGGTTCA	-5277	NM_001982	<i>ERBB3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-2830	NM_001005915	<i>ERBB3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-2830	NM_001982	<i>ERBB3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	4288	NM_001005915	<i>ERBB3</i>	Jb	67	gene responsive to RA
AGGTCAGGAGTTCA	4288	NM_001982	<i>ERBB3</i>	Jb	67	gene responsive to RA
AGGTCAGGAGTTCA	-4862	NM_018948	<i>ERRF1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7711	NM_018948	<i>ERRF1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7660	NM_001993	<i>F3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-969	NM_002002	<i>FCER2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-6131	NM_002002	<i>FCER2</i>	Sx	52	gene responsive to RA
GGGTCACGAGGTCA	-6986	NM_002002	<i>FCER2</i>	Sc	57	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_000141	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022969	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022970	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022971	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022972	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022973	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022974	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022975	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_022976	<i>FGFR2</i>	Sp	67	gene responsive to RA

AGGTCAGGAGTTCA	-3605	NM_023028	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_023029	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_023030	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	-3605	NM_023031	<i>FGFR2</i>	Sp	67	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_000141	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022969	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022970	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022971	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022972	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022973	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022974	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022975	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_022976	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_023028	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_023029	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_023030	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGCAGTTCA	-4179	NM_023031	<i>FGFR2</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-2241	NM_000142	<i>FGFR3</i>	Sx	53	gene responsive to RA
AGGTCAGGAGTTCA	-2241	NM_022965	<i>FGFR3</i>	Sx	53	gene responsive to RA
AGGTCAGGAGTTCA	-4644	NM_022963	<i>FGFR4</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-2014	NM_002011	<i>FGFR4</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-2014	NM_213647	<i>FGFR4</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	4639	NM_005248	<i>FGR</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	-8308	NM_005248	<i>FGR</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6261	NM_016731	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-6226	NM_016729	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-4311	NM_016730	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-4012	NM_000802	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-3793	NM_016724	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-3655	NM_016725	<i>FOLR1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	9933	NM_016731	<i>FOLR1</i>	Sp	66	gene responsive to RA
AGGTCAGGAGTTCA	9968	NM_016729	<i>FOLR1</i>	Sp	66	gene responsive to RA
AGGTCAGGAGTTCA	9614	NM_000803	<i>FOLR2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-8824	NM_005252	<i>FOS</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	-8495	NM_005252	<i>FOS</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3758	NM_021784	<i>FOXA2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-2757	NM_153675	<i>FOXA2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	4515	NM_002033	<i>FUT4</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6402	NM_005971	<i>FXYD3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6402	NM_021910	<i>FXYD3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-1575	NM_153047	<i>FYN</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6850	NM_002037	<i>FYN</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5840	NM_033507	<i>GCK</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5840	NM_033508	<i>GCK</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6375	NM_000515	<i>GHI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6375	NM_022559	<i>GHI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6375	NM_022560	<i>GHI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6375	NM_022561	<i>GHI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6375	NM_022562	<i>GHI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4093	NM_000515	<i>GHI</i>	Sg1	65	gene responsive to RA

AGGTCAGGAGTTCA	4093	NM_022559	<i>GHI</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	4093	NM_022560	<i>GHI</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	4093	NM_022561	<i>GHI</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	4093	NM_022562	<i>GHI</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-1512	NM_000515	<i>GHI</i>	Sx	69	gene responsive to RA
AGGTCAGGAGTTCA	-1512	NM_022559	<i>GHI</i>	Sx	69	gene responsive to RA
AGGTCAGGAGTTCA	-1512	NM_022560	<i>GHI</i>	Sx	69	gene responsive to RA
AGGTCAGGAGTTCA	-1512	NM_022561	<i>GHI</i>	Sx	69	gene responsive to RA
AGGTCAGGAGTTCA	-1512	NM_022562	<i>GHI</i>	Sx	69	gene responsive to RA
AGGTCAGAAGTTCA	-1592	NM_001012512	<i>GRP</i>	Sx	58	gene responsive to RA
AGGTCAGAAGTTCA	-1592	NM_001012513	<i>GRP</i>	Sx	58	gene responsive to RA
AGGTCAGAAGTTCA	-1592	NM_002091	<i>GRP</i>	Sx	58	gene responsive to RA
AGGTCAGGAGTTCA	3535	NM_000852	<i>GSTP1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	6446	NM_000852	<i>GSTP1</i>	Y	65	gene responsive to RA
AGGTCAGAAGTTCA	6224	NM_001945	<i>HBEGF</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6063	NM_002110	<i>HCK</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	8962	NM_002110	<i>HCK</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	7287	NM_001030003	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7287	NM_001030004	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7287	NM_175914	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8901	NM_001030003	<i>HNF4A</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	8901	NM_001030004	<i>HNF4A</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	8901	NM_175914	<i>HNF4A</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-3425	NM_000457	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3425	NM_178849	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3425	NM_178850	<i>HNF4A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5806	NM_000413	<i>HSD17B1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	6622	NM_000413	<i>HSD17B1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8243	NM_000413	<i>HSD17B1</i>	Sx	67	gene responsive to RA
AGGTCAGGGTTCA	6979	NM_002153	<i>HSD17B2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	1435	NM_002162	<i>ICAM3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-6659	NM_002162	<i>ICAM3</i>	Sq	67	gene responsive to RA
AGGTCAGAAGTTCA	6799	NM_000875	<i>IGF1R</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9488	NM_000597	<i>IGFBP2</i>	Sx	67	gene responsive to RA
AGGTCATGAGTTCA	-7359	NM_000597	<i>IGFBP2</i>	Sq	69	gene responsive to RA
AGGTCAGGAGTTCA	-8016	NM_000599	<i>IGFBP5</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-3245	NM_000878	<i>IL2RB</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3056	NM_000600	<i>IL6</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7747	NM_000565	<i>IL6R</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7747	NM_181359	<i>IL6R</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9421	NM_002192	<i>INHBA</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-8283	NM_002209	<i>ITGAL</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7896	NM_002209	<i>ITGAL</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	7133	NM_002209	<i>ITGAL</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	-9325	NM_000632	<i>ITGAM</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2456	NM_002210	<i>ITGAV</i>	Sx	65	gene responsive to RA
AGGTCAGGGTTCA	6558	NM_002210	<i>ITGAV</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-3092	NM_000212	<i>ITGB3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5843	NM_000212	<i>ITGB3</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	9880	NM_000212	<i>ITGB3</i>	Sx	66	gene responsive to RA



AGGTCAGGAGTTCA	-7528	NM_001005619	<i>ITGB4</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-4268	NM_000213	<i>ITGB4</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-4268	NM_001005731	<i>ITGB4</i>	Sq	67	gene responsive to RA
AGGTCAGAAGTTCA	3939	NM_001005619	<i>ITGB4</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	7199	NM_000213	<i>ITGB4</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	7199	NM_001005731	<i>ITGB4</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8177	NM_001005619	<i>ITGB4</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	9092	NM_000526	<i>KRT14</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2218	NM_057088	<i>KRT3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8292	NM_002291	<i>LAMB1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	6296	NM_006762	<i>LAPTM5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5690	NM_003240	<i>LEFTY2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6800	NM_000230	<i>LEP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9735	NM_002307	<i>LGALS7</i>	Sx	54	gene responsive to RA
AGGTCAGGAGTTCA	-1256	NM_002307	<i>LGALS7</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4085	NM_182489	<i>LOC346673</i>	Sx	67	gene responsive to RA
AGGTCAGTAGTTCA	6098	NM_002343	<i>LTF</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-8878	NM_002350	<i>LYN</i>	Sq	65	gene responsive to RA
GGGTCAGGAGTTCA	4524	NM_002350	<i>LYN</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4874	NM_002350	<i>LYN</i>	Sq	64	gene responsive to RA
AGGTCAGGAGTTCA	8430	NM_002745	<i>MAPK1</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	8430	NM_138957	<i>MAPK1</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-856	NM_002745	<i>MAPK1</i>	Sg	63	gene responsive to RA
AGGTCAGGAGTTCA	-856	NM_138957	<i>MAPK1</i>	Sg	63	gene responsive to RA
AGGTCAGGAGTTCA	-5336	NM_002745	<i>MAPK1</i>	Sq	68	gene responsive to RA
AGGTCAGGAGTTCA	-5336	NM_138957	<i>MAPK1</i>	Sq	68	gene responsive to RA
AGGTCAGGAGTTCA	-1886	NM_002386	<i>MC1R</i>	Sx	63	gene responsive to RA
AGGTCAGGAGTTCA	-1405	NM_002386	<i>MC1R</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	-4859	NM_021960	<i>MCL1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4859	NM_182763	<i>MCL1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6051	NM_004527	<i>MEOX1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6051	NM_013999	<i>MEOX1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3410	NM_000900	<i>MGP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2031	NM_005940	<i>MMP11</i>	Y	64	gene responsive to RA
AGGTCAGGAGTTCA	4357	NM_005940	<i>MMP11</i>	Sg1	65	gene responsive to RA
AGGTCAGGGGTTCA	9628	NM_005954	<i>MT3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-5373	NM_001004430	<i>MUC17</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4102	NM_001004430	<i>MUC17</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	9748	NM_001004430	<i>MUC17</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-9500	NM_002457	<i>MUC2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5379	NM_002466	<i>MYBL2</i>	Sq	67	gene responsive to RA
AGGTCAGAAGTTCA	-2381	NM_002466	<i>MYBL2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4924	NM_002466	<i>MYBL2</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	6108	NM_002466	<i>MYBL2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8476	NM_002466	<i>MYBL2</i>	Sx	60	gene responsive to RA
AGGTCAGGAGTTCA	9082	NM_002466	<i>MYBL2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4785	NM_002467	<i>MYC</i>	Sg	46	gene responsive to RA
AGGTCAGGAGTTCA	-3026	NM_005376	<i>MYCL1</i>	Sx	29	gene responsive to RA
AGGTCAGGAGTTCA	4902	NM_000265	<i>NCF1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1044	NM_000265	<i>NCF1</i>	Sp	66	gene responsive to RA

AGGTCAGGAGTTCA	-3830	NM_000265	<i>NCF1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2913	NM_000269	<i>NME1</i>	Sq	62	gene responsive to RA
AGGTCAGGAGTTCA	2913	NM_198175	<i>NME1</i>	Sq	62	gene responsive to RA
AGGTCAGGAGTTCA	6164	NM_017617	<i>NOTCH1</i>	Sx	21	gene responsive to RA
AGGTCAGGAGTTCA	-2824	NM_003298	<i>NR2C2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5950	NM_003298	<i>NR2C2</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	5647	NM_001018074	<i>NR3C1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5647	NM_001018075	<i>NR3C1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5647	NM_001018077	<i>NR3C1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4320	NM_001018074	<i>NR3C1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4320	NM_001018075	<i>NR3C1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4320	NM_001018077	<i>NR3C1</i>	Sx	67	gene responsive to RA
GGGTCACGAGGTCA	-3696	NM_001489	<i>NR6A1</i>	Sg	53	gene responsive to RA
GGGTCACGAGGTCA	-3696	NM_033334	<i>NR6A1</i>	Sg	53	gene responsive to RA
GGGTCACGAGGTCA	-3696	NM_033335	<i>NR6A1</i>	Sg	53	gene responsive to RA
AGGTCAGGAGTTCA	7648	NM_002525	<i>NRD1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5007	NM_002525	<i>NRD1</i>	Sx	67	gene responsive to RA
AGGTCAGCAGTTCA	-6724	NM_002525	<i>NRD1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7090	NM_002543	<i>OLR1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3412	NM_002543	<i>OLR1</i>	Sx	67	gene responsive to RA
AGGTCATGAGGTCA	9754	NM_032971	<i>PCDH11Y</i>	Sc	57	gene responsive to RA
AGGTCAGCAGTTCA	7973	NM_174895	<i>PCP2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2690	NM_174895	<i>PCP2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8973	NM_000442	<i>PECAM1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-3349	NM_000442	<i>PECAM1</i>	Sx	66	gene responsive to RA
AGGTCAGGGTTCA	-4676	NM_000442	<i>PECAM1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5861	NM_000442	<i>PECAM1</i>	Sx	70	gene responsive to RA
AGGTCAGGAGTTCA	-6074	NM_004426	<i>PHCI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5773	NM_004426	<i>PHCI</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9542	NM_004426	<i>PHCI</i>	Sx	67	gene responsive to RA
AGGTCAGAAGTTCA	-4303	NM_002649	<i>PIK3CG</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9655	NM_002649	<i>PIK3CG</i>	Sg	62	gene responsive to RA
AGGTCAGGGTTCA	-3650	NM_002658	<i>PLAU</i>	Sq	23	gene responsive to RA
AGGTCAGGAGTTCA	1504	NM_002658	<i>PLAU</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9459	NM_021132	<i>PPP3CB</i>	Sp	67	gene responsive to RA
AGGTCAGGAGTTCA	6412	NM_032152	<i>PRAMI</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	5311	NM_002738	<i>PRKCB1</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	5311	NM_212535	<i>PRKCB1</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	-5208	NM_000949	<i>PRLR</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-5478	NM_000311	<i>PRNP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5478	NM_183079	<i>PRNP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2462	NM_000311	<i>PRNP</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-2462	NM_183079	<i>PRNP</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	5572	NM_000952	<i>PTAFR</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1813	NM_000952	<i>PTAFR</i>	Sx	67	gene responsive to RA
AGGTCAGCAGTTCA	-4852	NM_000952	<i>PTAFR</i>	Sp	67	gene responsive to RA
AGTTCAGGAGTTCA	-5579	NM_000962	<i>PTGSI</i>	FLAM_C	64	gene responsive to RA
AGTTCAGGAGTTCA	-5579	NM_080591	<i>PTGSI</i>	FLAM_C	64	gene responsive to RA
AGGTCAGGAGTTCA	-8198	NM_000963	<i>PTGS2</i>	Sx	71	gene responsive to RA
AGGTCAGGAGTTCA	-9572	NM_000315	<i>PTH</i>	Sx	65	gene responsive to RA

AGGTCAGGAGTTCA	6363	NM_005607	<i>PTK2</i>	Sp	23	gene responsive to RA
AGGTCAGGAGTTCA	6363	NM_153831	<i>PTK2</i>	Sp	23	gene responsive to RA
AGGTCAGGAGTTCA	-4871	NM_005607	<i>PTK2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-4871	NM_153831	<i>PTK2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3645	NM_006264	<i>PTPN13</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3645	NM_080683	<i>PTPN13</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3645	NM_080684	<i>PTPN13</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3645	NM_080685	<i>PTPN13</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8186	NM_006744	<i>RBP4</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	5288	NM_020216	<i>RNPEP</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	7439	NM_020216	<i>RNPEP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8709	NM_020216	<i>RNPEP</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	-4312	NM_021136	<i>RTN1</i>	Sx	59	gene responsive to RA
AGGTCACAAGTTCA	-6574	NM_006054	<i>RTN3</i>	Sx	61	gene responsive to RA
AGGTCACAAGTTCA	-6574	NM_201428	<i>RTN3</i>	Sx	61	gene responsive to RA
AGGTCACAAGTTCA	-6574	NM_201429	<i>RTN3</i>	Sx	61	gene responsive to RA
AGGTCACAAGTTCA	-6574	NM_201430	<i>RTN3</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	-5349	NM_006054	<i>RTN3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-5349	NM_201428	<i>RTN3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-5349	NM_201429	<i>RTN3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-5349	NM_201430	<i>RTN3</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-2392	NM_006054	<i>RTN3</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	-2392	NM_201428	<i>RTN3</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	-2392	NM_201429	<i>RTN3</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	-2392	NM_201430	<i>RTN3</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	3542	NM_006054	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3542	NM_201428	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3542	NM_201429	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	3542	NM_201430	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9903	NM_006054	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9903	NM_201428	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9903	NM_201429	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9903	NM_201430	<i>RTN3</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7402	NM_172311	<i>SALF</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1688	NM_005063	<i>SCD</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7768	NM_002998	<i>SDC2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	7886	NM_002998	<i>SDC2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9060	NM_000488	<i>SERPINC1</i>	Sq	61	gene responsive to RA
AGGTCAGGAGTTCA	8385	NM_000488	<i>SERPINC1</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	3695	NM_000488	<i>SERPINC1</i>	Sg1	65	gene responsive to RA
AGGTCAGGAGTTCA	-9057	NM_000602	<i>SERPINE1</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6215	NM_000602	<i>SERPINE1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5908	NM_000602	<i>SERPINE1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7819	NM_000542	<i>SFTPB</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	7822	NM_198843	<i>SFTPB</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-7814	NM_000542	<i>SFTPB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7811	NM_198843	<i>SFTPB</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-9248	NM_000193	<i>SHH</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-8778	NM_003055	<i>SLC18A3</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-6381	NM_000340	<i>SLC2A2</i>	Sg	65	gene responsive to RA

GGGTCAGGAGTTCA	7302	NM_001042	<i>SLC2A4</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	1273	NM_000453	<i>SLC5A5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	1568	NM_000453	<i>SLC5A5</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	9526	NM_000453	<i>SLC5A5</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7776	NM_003068	<i>SNAI2</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	8919	NM_000636	<i>SOD2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8919	NM_001024465	<i>SOD2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	8919	NM_001024466	<i>SOD2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	6889	NM_000636	<i>SOD2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	6889	NM_001024465	<i>SOD2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	6889	NM_001024466	<i>SOD2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5529	NM_000636	<i>SOD2</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	5529	NM_001024465	<i>SOD2</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	5529	NM_001024466	<i>SOD2</i>	Sq	65	gene responsive to RA
AGGTCAGGAGTTCA	8186	NM_003113	<i>SPI100</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7131	NM_003123	<i>SPN</i>	Sq	66	gene responsive to RA
AGGTCAGGAGTTCA	-6851	NM_001030288	<i>SPN</i>	Sq	66	gene responsive to RA
AGGTCAGGAGTTCA	-5924	NM_003123	<i>SPN</i>	FRAM	65	gene responsive to RA
AGGTCAGGAGTTCA	-5644	NM_001030288	<i>SPN</i>	FRAM	65	gene responsive to RA
AGGTCAGGAGTTCA	6406	NM_003123	<i>SPN</i>	Sq	68	gene responsive to RA
AGGTCAGGAGTTCA	6686	NM_001030288	<i>SPN</i>	Sq	68	gene responsive to RA
AGGTCAGGAGTTCA	9424	NM_003123	<i>SPN</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	9704	NM_001030288	<i>SPN</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-2008	NM_001005291	<i>SREBF1</i>	Sx	68	gene responsive to RA
AGGTCAGGAGTTCA	-2008	NM_004176	<i>SREBF1</i>	Sx	68	gene responsive to RA
AGGTCAGGAGTTCA	1176	NM_000349	<i>STAR</i>	Sx	57	gene responsive to RA
AGGTCAGGAGTTCA	1176	NM_001007243	<i>STAR</i>	Sx	57	gene responsive to RA
AGGTCAGGAGTTCA	-4792	NM_000349	<i>STAR</i>	Sg	60	gene responsive to RA
AGGTCAGGAGTTCA	-4792	NM_001007243	<i>STAR</i>	Sg	60	gene responsive to RA
AGGTCAGGAGTTCA	-6621	NM_000349	<i>STAR</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6621	NM_001007243	<i>STAR</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7201	NM_000349	<i>STAR</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7201	NM_001007243	<i>STAR</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-5097	NM_177973	<i>SULT2B1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-1356	NM_177973	<i>SULT2B1</i>	Sq	61	gene responsive to RA
AGGTCAGGAGTTCA	3800	NM_177973	<i>SULT2B1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	4564	NM_177973	<i>SULT2B1</i>	Sx	61	gene responsive to RA
AGGTCAGGAGTTCA	5153	NM_177973	<i>SULT2B1</i>	Sq	64	gene responsive to RA
AGGTCAGGAGTTCA	-4307	NM_004605	<i>SULT2B1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	1553	NM_004605	<i>SULT2B1</i>	Sg1	60	gene responsive to RA
AGGTCAGGAGTTCA	7867	NM_004605	<i>SULT2B1</i>	Sx	65	gene responsive to RA
AGGTCAGGAGTTCA	5696	NM_003168	<i>SUPT4H1</i>	Jo	63	gene responsive to RA
AGGTCAGGAGTTCA	1905	NM_003168	<i>SUPT4H1</i>	Sg	64	gene responsive to RA
GGGTCACAAGGTTCA	-7170	NM_003168	<i>SUPT4H1</i>	Y	56	gene responsive to RA
AGGTCAGGAGTTCA	6392	NM_006706	<i>TCERG1</i>	FLAM_C	67	gene responsive to RA
AGGTCAGGAGTTCA	4533	NM_003234	<i>TFRC</i>	Sq	66	gene responsive to RA
AGGTCAGGAGTTCA	-2106	NM_003234	<i>TFRC</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	6419	NM_003236	<i>TGFA</i>	Sg1	66	gene responsive to RA
AGGTCAGGAGTTCA	8629	NM_000660	<i>TGFB1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	8224	NM_000660	<i>TGFB1</i>	Sx	67	gene responsive to RA

AGGTCAGGAGTTCA	6953	NM_000660	<i>TGFB1</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-5019	NM_000660	<i>TGFB1</i>	Sx	69	gene responsive to RA
AGGTCAGGAGTTCA	-6758	NM_004612	<i>TGFBR1</i>	Sx	28	gene responsive to RA
AGTTCAGGAGTTCA	-7149	NM_003251	<i>THRSP</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	2538	NM_003251	<i>THRSP</i>	Sg	65	gene responsive to RA
AGGTCAGGAGTTCA	-4440	NM_001067	<i>TOP2A</i>	Sx	67	gene responsive to RA
AGGTCAGGGGTTCA	-2211	NM_003299	<i>TRA1</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-6119	NM_001001188	<i>TRPM2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-6119	NM_003307	<i>TRPM2</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	7744	NM_001001188	<i>TRPM2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	7744	NM_003307	<i>TRPM2</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	3376	NM_000372	<i>TYR</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	5013	NM_021833	<i>UCP1</i>	Sx	66	gene responsive to RA
AGGTCAGGAGTTCA	-5182	NM_003355	<i>UCP2</i>	Sq	67	gene responsive to RA
AGGTCAGGAGTTCA	-7804	NM_031933	<i>WNT8A</i>	Sx	67	gene responsive to RA
AGGTCAGGAGTTCA	-7804	NM_058244	<i>WNT8A</i>	Sx	67	gene responsive to RA
AGGTCATGAGTTCA	-7034	NM_031933	<i>WNT8A</i>	Sg	64	gene responsive to RA
AGGTCATGAGTTCA	-7034	NM_058244	<i>WNT8A</i>	Sg	64	gene responsive to RA
AGGTCAGGAGTTCA	6816	NM_003422	<i>ZNF42</i>	FLAM_A	66	gene responsive to RA
AGGTCAGGAGTTCA	6816	NM_198055	<i>ZNF42</i>	FLAM_A	66	gene responsive to RA

Le reste de la table est disponible en ligne <http://www.biomedcentral.com/1471-2164/8/23/additional/>.

Les gènes cible de RA (“gene responsive to RA”) viennent de la revue de Balmer et Blomhoff de 2002. Les gènes avec RARE (“gene has known RARE”) viennent principalement de l’article de Balmer et Blomhoff de 2005. Les références utilisées sont indiquées dans la table disponible en ligne.

Balmer, J.E. and Blomhoff, R. *A robust characterization of retinoic acid response elements based on a comparison of sites in three species*. J. Steroid Biochem. Mol. Biol., 2005. **96**, p. 347–354.

Balmer, J.E. and R. Blomhoff, *Gene expression regulation by retinoic acid*. J Lipid Res, 2002. **43**(11): p. 1773-808.

## Additional file 3

**ChIP Primers for genes with DR2-Alu**

Gene			Name of the primer	
1. <i>RAI1</i>	5'	5'-gcttctggcagtacagccca	RAI1D1ChIP5'	
	3'	5'-cttccgcttcagcctcccta	RAI1D1ChIP3'	
	5'	5'-catgatggctcaccactgt	RAI1D2ChIP5'	
	3'	5'-ctaagtgcaactccagcctc	RAI1D2ChIP3'	
	5'	5'-actttgggagtcocgaggcag	RAI1D3ChIP5'	
	3'	5'-cctggagtgcaatggtgca	RAI1D3ChIP3'	
	2. <i>SMYD5</i>	5'	5'-gttgacagcccaagtgacct	SMYD5D1ChIP5'
		3'	5'-cagggcgggttatcccagca	SMYD5D1ChIP3'
		5'	5'-tgcggtggcttacgcctgta	SMYD5D2ChIP5'
3'		5'-ggctggagtgcaatggcact	SMYD5D2ChIP3'	
5'		5'-agtagatgtggccagggtgc	SMYD5D3ChIP5'	
3'		5'-gttgcccaagctggagtgca	SMYD5D3ChIP3'	
5'		5'-tcagccttccgagtagctgg	SMYD5D4ChIP5'	
3'		5'-gcgcagtggtcatgcttgt	SMYD5D4ChIP3'	
3. <i>GPRC5A</i>	5'	5'-agatatcaggccaggcaccg	GPRC5AD1ChIP5'	
	3'	5'-tacaggactgcaccaccagg	GPRC5AD1ChIP3'	
	5'	5'-ttacaggcatgcaccaccac	GPRC5AD2ChIP5'	
	3'	5'-tcctctgaaggaatgtccg	GPRC5AD2ChIP3'	
	5'	5'-ggctcacacctgtaatcccagca	GPRC5AD3ChIP5'	
	3'	5'-ggagtgcaagtggcgtgatct	GPRC5AD3ChIP3'	
4. <i>RARRES1</i>	5'	5'-gagcagctgggattccaggt	RARRES1D1ChIP5'	
	3'	5'-ttgaggccaggcatgatggc	RARRES1D1ChIP3'	
	5'	5'-ggattacaggtgctgcccac	RARRES1D2ChIP5'	
	3'	5'-cgggtggctcacgcctgtaat	RARRES1D2ChIP3'	
	5'	5'-gctcactgcaacctctgcct	RARRES1D3ChIP5'	
	3'	5'-ggctctgcctgtaatccca	RARRES1D3ChIP3'	
5. <i>RARRES3</i>	5'	5'-ggctagagtgcagtgccatg	RARRES3D1ChIP5'	
	3'	5'-gtgcagtggtgacacctgt	RARRES3D1ChIP3'	
	5'	5'-cacgggtggctcatgcctgta	RARRES3D2ChIP5'	
	3'	5'-tgcaatggcacgatctcggc	RARRES3D2ChIP3'	
	5'	5'-aggcacagtggctcactcct	RARRES3D3ChIP5'	
	3'	5'-tgtctcctaggctggagtgc	RARRES3D3ChIP3'	

## Additional file 4

## A

```

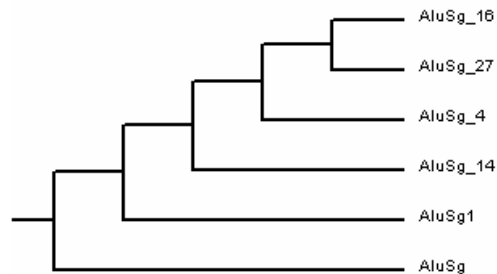
AluSx      61 TCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGT 113
AluSx_4    61 TCACCTGGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGT 113
AluSg      61 TCAC--GAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGT 111
AluSg1     61 TCAC--GAGGTCAGGAGTTCGAGACCAGCCTGGCCAAGATGGTGAAACCCCGT 111
AluSg_4    61 TCAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAAACCCCGT 111
AluSg_14   61 TCAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAAACCCCGT 111
AluSg_16   61 TCAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAAACCCCGT 111
AluSg_24   61 TCAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGT 111
AluSg_27   61 TCAC--GAGGTCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAAACCCCGT 111
AluSq      61 TCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGT 113
AluSq_6    61 TCACCTGGAGGTCAGAAGTTCAAGACCAGCCTGGTCAACATGGTGAAACCCCGT 113
AluSp      61 TCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAACATGGAGAAACCCCGT 113
AluSc      61 TCAC--GAGGTCAGGAGATCGAGACCATCCTGGCCAACATGGTGAAACCCCGT 111
          RGKTCANNRGKTCA

```

## B

Alu Family	# in genome	% with DR2	Rebase Alu family
AluSx_4	485	59.59	AluSx (75%)
AluSg_4	4712	67.32	AluSg1 (76%)
AluSg_14	3587	58.49	AluSg (75%)
AluSg_16	995	68.84	AluSg1 (67%)
AluSg_24	1979	62.81	AluSg (82%)
AluSg_27	300	60.33	AluSg1 (71%)
AluSq_6	1032	59.88	AluSq (85%)

## C



**Supplemental Figure 1** - (A) Alignment of AluS with Alu families from Price et al. [26] that contain consensus RARE DR2. (B) Frequency of RARE DR2 for Alu families from Price et al. (C) Cladogram of AluSg subfamilies

## References

1. Hedges DJ, Batzer MA: From the margins of the genome: mobile elements shape primate evolution. *Bioessays* 2005, 27(8):785-794.
2. Jurka J: Evolutionary impact of human Alu repetitive elements. *Curr Opin Genet Dev* 2004, 14(6):603-608.
3. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA: Alu repeats: a source for the genesis of primate microsatellites. *Genomics* 1995, 29(1):136-144.
4. Bailey JA, Liu G, Eichler EE: An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 2003, 73(4):823-834.
5. Deininger PL, Batzer MA: Alu repeats and human disease. *Mol Genet Metab* 1999, 67(3):183-193.
6. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS: Frequency of recent retrotransposition events in the human factor IX gene. *Hum Mutat* 2001, 17(6):511-519.
7. Sorek R, Ast G, Graur D: Alu-containing exons are alternatively spliced. *Genome Res* 2002, 12(7):1060-1067.
8. Dagan T, Sorek R, Sharon E, Ast G, Graur D: AluGene: a database of Alu elements incorporated within protein-coding genes. *Nucleic Acids Res* 2004, 32(Database issue):D489-92.
9. Greally JM: Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* 2002, 99(1):327-332.
10. Shankar R, Grover D, Brahmachari SK, Mukerji M: Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements. *BMC Evol Biol* 2004, 4(1):37.
11. Bourdeau V, Deschenes J, Metivier R, Nagai Y, Nguyen D, Bretschneider N, Gannon F, White JH, Mader S: Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol Endocrinol* 2004, 18(6):1411-1427.
12. Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemand B, Zhang R, Mader S, White JH: Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin D3 target genes. *Mol Endocrinol* 2005, 19(11):2685-2695.



13. Chawla A, Repa JJ, Evans RM, Mangelsdorf DJ: Nuclear receptors and lipid physiology: opening the X-files. *Science* 2001, 294(5548):1866-1870.
14. McKenna NJ, O'Malley BW: Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* 2002, 108(4):465-474.
15. Claessens F, Gewirth DT: DNA recognition by nuclear receptors. *Essays Biochem* 2004, 40:59-72.
16. Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N: Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene* 1999, 239(2):341-349.
17. Vansant G, Reynolds WF: The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc Natl Acad Sci U S A* 1995, 92(18):8229-8233.
18. Chambon P: A decade of molecular biology of retinoic acid receptors. *Faseb J* 1996, 10(9):940-954.
19. Mark M, Ghyselinck NB, Chambon P: Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annu Rev Pharmacol Toxicol* 2006, 46:451-480.
20. Mader S, Chen JY, Chen Z, White J, Chambon P, Gronemeyer H: The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains. *Embo J* 1993, 12(13):5029-5041.
21. Kondo Y, Issa JP: Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J Biol Chem* 2003, 278(30):27658-27662.
22. Britten RJ, Baron WF, Stout DB, Davidson EH: Sources and evolution of human Alu repeated sequences. *Proc Natl Acad Sci U S A* 1988, 85(13):4770-4774.
23. Labuda D, Striker G: Sequence conservation in Alu evolution. *Nucleic Acids Res* 1989, 17(7):2477-2491.
24. Kapitonov V, Jurka J: The age of Alu subfamilies. *Journal of molecular evolution* 1996, 42(1):59-65.
25. Murphy MH, Baralle FE: Directed semisynthetic point mutational analysis of an RNA polymerase III promoter. *Nucleic Acids Res* 1983, 11(22):7695-7700.
26. Price AL, Eskin E, Pevzner PA: Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res* 2004, 14(11):2245-2252.

27. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M: Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 2005, 122(1):33-43.
28. Hu L, Crowe DL, Rheinwald JG, Chambon P, Gudas LJ: Abnormal expression of retinoic acid receptors and keratin 19 by human oral and epidermal squamous cell carcinoma cell lines. *Cancer Res* 1991, 51(15):3972-3981.
29. Akutsu N, Lin R, Bastien Y, Bestawros A, Enepekides DJ, Black MJ, White JH: Regulation of gene Expression by 1alpha,25-dihydroxyvitamin D3 and Its analog EB1089 under growth-inhibitory conditions in squamous carcinoma Cells. *Mol Endocrinol* 2001, 15(7):1127-1139.
30. Cheng Y, Lotan R: Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor. *J Biol Chem* 1998, 273(52):35008-35015.
31. Nagpal S, Chandraratna RA: Vitamin A and regulation of gene expression. *Curr Opin Clin Nutr Metab Care* 1998, 1(4):341-346.
32. DiSepio D, Ghosn C, Eckert RL, Deucher A, Robinson N, Duvic M, Chandraratna RA, Nagpal S: Identification and characterization of a retinoid-induced class II tumor suppressor/growth regulatory gene. *Proc Natl Acad Sci U S A* 1998, 95(25):14811-14815.
33. Consortium TEP: The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004, 306(5696):636-640.
34. Navakauskiene R, Treigyte G, Gineitis A, Magnusson KE: Identification of apoptotic tyrosine-phosphorylated proteins after etoposide or retinoic acid treatment. *Proteomics* 2004, 4(4):1029-1041.
35. Oei SL, Babich VS, Kazakov VI, Usmanova NM, Kropotov AV, Tomilin NV: Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics* 2004, 83(5):873-882.
36. Thornburg BG, Gotea V, Makalowski W: Transposable elements as a significant source of transcription regulating signals. *Gene* 2006, 365:104-110.
37. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier

- P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006, 34(Database issue):D108-10.
38. Li TH, Kim C, Rubin CM, Schmid CW: K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res* 2000, 28(16):3031-3039.
39. Wu CC, Shyu RY, Chou JM, Jao SW, Chao PC, Kang JC, Wu ST, Huang SL, Jiang SY: RARRES1 expression is significantly related to tumour differentiation and staging in colorectal adenocarcinoma. *Eur J Cancer* 2006, 42(4):557-565.
40. Youssef EM, Chen XQ, Higuchi E, Kondo Y, Garcia-Manero G, Lotan R, Issa JP: Hypermethylation and silencing of the putative tumor suppressor Tazarotene-induced gene 1 in human cancers. *Cancer Res* 2004, 64(7):2411-2417.
41. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, Zhou D, Luo S, Vasicek TJ, Daly MJ, Wolfsberg TG, Collins FS: Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* 2006, 16(1):123-131.
42. Paoletta G, Lucero MA, Murphy MH, Baralle FE: The Alu family repeat promoter has a tRNA-like bipartite structure. *Embo J* 1983, 2(5):691-696.
43. Jover R, Bort R, Gomez-Lechon MJ, Castell JV: Cytochrome P450 regulation by hepatocyte nuclear factor 4 in human hepatocytes: a study using adenovirus-mediated antisense targeting. *Hepatology* 2001, 33(3):668-675.
44. Makita T, Hernandez-Hoyos G, Chen TH, Wu H, Rothenberg EV, Sucov HM: A developmental transition in definitive erythropoiesis: erythropoietin expression is sequentially regulated by retinoic acid receptors and HNF4. *Genes Dev* 2001, 15(7):889-901.
45. Wang TT, Nestel FP, Bourdeau V, Nagai Y, Wang Q, Liao J, Tavera-Mendoza L, Lin R, Hanrahan JW, Mader S, White JH: Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. *J Immunol* 2004, 173(5):2909-2912.
46. Gombart AF, Borregaard N, Koeffler HP: Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3. *Faseb J* 2005, 19(9):1067-1077.

47. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: The UCSC Genome Browser Database. *Nucleic Acids Res* 2003, 31(1):51-54.
48. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ: The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006, 34(Database issue):D590-8.
49. Pruitt KD, Tatusova T, Maglott DR: NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005, 33(Database issue):D501-4.
50. Smit AFA, Hubley R, Green P: Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0. [<http://www.repeatmasker.org>].
51. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: Human-mouse alignments with BLASTZ. *Genome Res* 2003, 13(1):103-107.
52. Team RDC: R: A Language and Environment for Statistical Computing. [<http://www.R-project.org>].
53. Felsenstein J: PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 1989:164-166.
54. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005, 110(1-4):462-467.

## **Chapitre 7: Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells**

Bourdeau V\*, Deschênes J\*, Laperrière D, Aid M, White JH, Mader S.

\* co-premier auteur

Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Res.* 2007 Nov 5;

À titre de deuxième auteur, j'ai réalisé l'analyse des micropuces du laboratoire et des autres études. J'ai compilé la matrice de ERE, j'ai effectué la recherche des sites des 300 facteurs de transcription et évalué leur enrichissement dans les promoteurs des différentes catégories de gènes régulés. J'ai aussi évalué l'enrichissement des sites ERE dans les régions liées ER $\alpha$  en CHIP-on-chip et des sites E2F dans les régions liées E2F1 en CHIP-on-chip dans les promoteurs des différentes catégories de gènes régulés. J'ai également participé aux révisions du manuscrit.

## **ABSTRACT**

Estrogen receptors (ERs), which mediate the proliferative action of estrogens in breast cancer cells, are ligand-dependent transcription factors that regulate expression of their primary target genes through several mechanisms. In addition to direct binding to cognate DNA sequences, ERs can be recruited to DNA through other transcription factors (tethering), or affect gene transcription through modulation of signaling cascades by non-genomic mechanisms of action. To better characterize the mechanisms of gene regulation by estrogens, we have identified more than 700 putative primary and about 1300 putative secondary target genes of estradiol in MCF-7 cells through microarray analysis performed in the presence or absence of the translation inhibitor cycloheximide. Although siRNA-mediated inhibition of ER expression antagonized the effects of estradiol on up- and down-regulated primary target genes, estrogen response elements (EREs) were enriched only in the vicinity of up-regulated genes. Binding sites for several other transcription factors, including proteins known to tether ER, were enriched in up- and/or down-regulated primary targets. Secondary estrogen targets were particularly enriched in sites for E2F family members, several of which were transcriptionally regulated by estradiol, consistent with a major role of these factors in mediating the effects of estrogens on gene expression and cellular growth.

## **INTRODUCTION**

The pleiotropic effects of estrogens in its numerous target tissues, including the reproductive, skeletal, cardiovascular and central nervous systems (1–5) are mediated in large part via ERs (6), which are members of the superfamily of nuclear receptors and function as hormone-dependent transcription factors (7–9). ERs bind DNA directly through their central, conserved DNA-binding domains composed of two zinc fingers of the C4 type (10,11). Cognate DNA-binding motifs, also called estrogen response elements (EREs),

have been characterized in estrogen-responsive promoters (12–15). Consensus EREs derived by compiling natural response elements are 15 bp palindromes of PuGGTCA motifs with a 3 bp spacer and correspond to the highest affinity binding sites for ERs in vitro (16,17). However, natural response elements often deviate from the consensus at one or several positions (14,15). Both estrogen receptors share similar in vitro DNA-binding patterns (18), but their transcriptional activation properties differ (6,19,20), possibly due to differential recruitment of transcriptional coactivator complexes responsible for histone acetylation, chromatin remodeling and enhanced recruitment of the basal transcription machinery (21–27). ER is thought to mediate the proliferative effects of estrogens in breast cancer cells. Indeed, its expression is preserved or increased in two-thirds of breast tumors, correlating with sensitivity of tumors to antiestrogenic treatment (3–5). On the other hand, the expression of ER $\beta$ , which was reported to play a role in terminal differentiation of breast epithelial cells (28), appears to be reduced during tumorigenesis (29,30).

In addition to mediating gene regulation through direct binding to DNA, ERs can regulate gene expression through protein–protein interaction with other transcription factors (tethering). Several transcription factors were shown to mediate positive or negative transcriptional regulation by ERs in the absence of EREs, including AP1, Sp1, NF- $\kappa$ B (15,31–33). In addition, interference between estrogen signaling and other intracellular signaling pathways including the MAPK and PI3K pathways have been widely reported and may result from interactions between ERs and components of these signaling cascades (15,34–38). Finally, it has recently been suggested that estrogens may act also through a membrane receptor member of the GPCR family, GPR30, although the importance of these receptors in breast tumorigenesis remains to be established (39–43). These so-called non-genomic mechanisms of action can lead to rapid kinase-mediated activation of transcription factors and thus modulate gene expression in response to estrogens. Primary gene regulation by estrogen (i.e. genes regulated in the absence of de novo protein synthesis) can therefore result from at least three different mechanisms, including tethering and non-genomic action in addition to classical, ERE-mediated transcriptional regulation.

Better understanding of the mechanisms of action of estrogens in breast tumorigenesis necessitates large-scale identification of estrogen target genes and comprehensive analysis of the mechanisms of target gene regulation to assess the contribution of different regulatory mechanisms and specific targets to the proliferative effects of estrogens. Genome-wide microarray analysis of estradiol (E2) target genes has been performed in ER-positive breast cancer cell lines such as MCF-7, T47D and ZR75 cells, leading to the identification of a large number of target genes (44–50). It is however not always clear to which degree target identification is affected by cell culture conditions, choice of microarray platform and statistical analysis tools. In addition, few studies have used conditions that distinguish between primary and secondary target genes. This may explain why enrichment in EREs in estrogen target genes identified through microarray analysis was not reported in most studies. On the other hand, the promoter regions of 89 E2 target genes regulated in the presence of CHX in T47D cells were found enriched in EREs (46). However, the number of primary E2 target genes identified in that study remains low and the question of how far from the transcriptional start sites of primary target genes enrichment in EREs can be observed remains open given the relatively narrow window used. High-affinity EREs located distally from the start sites of estrogen target genes are functional ER-binding sites in chromatin immunoprecipitation (ChIP) experiments (50,51). Although the role of ERs bound to distal sites in transcriptional regulation remains to be systematically analyzed, chromatin conformation capture assays revealed that ER-bound chromatin regions can act at large distances from regulated genes (48,52). In addition to EREs, binding sites for several other transcription factors were enriched in ER-bound ChIP fragments, suggesting roles for these factors in ER tethering (48).

In order to better study the mechanisms of target gene regulation by estrogens, we aimed to discriminate between primary and secondary E2 target genes in MCF-7 cells through microarray analysis of gene expression patterns in response to E2 either in the presence or the absence of the protein synthesis inhibitor cycloheximide. We verified the



validity of this approach by demonstrating enrichment in EREs in primary, but not secondary up-regulated genes, and identified other transcription factors involved in the regulation of primary and/or secondary estrogen targets. Our results are discussed in the context of previous studies based on large-scale microarray analysis or genome-wide CHIP-on-chip mapping of ER-associated chromatin regions.

## **MATERIALS AND METHODS**

### **Cell culture and treatments**

MCF-7 breast carcinoma cells were maintained in -minimal Eagle's medium (-MEM) (Wisent, St-Bruno, QC, Canada) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich, Oakville, ON, Canada). Three days before experiments, cells were switched to phenol red-free Dulbecco's modified Eagle's medium (DMEM) (Wisent) containing 10% charcoal-treated FBS, 1% sodium pyruvate (Wisent), 1% penicillin/streptomycin (Wisent) and 1% L-glutamine (Wisent). Cells were seeded in 10 cm plates at a density such that near-confluency is obtained at the end of the treatment. The day before hormonal stimulation, the medium was changed to phenol red-free DMEM supplemented with 0.5% charcoal-treated FBS. Pre-treatments with cycloheximide (10 µg/ml, Sigma) were initiated 1 h before hormonal treatment with 17β-estradiol (E2, 25 nM, Sigma), ICI 182,780 (ICI, 100 nM, Sigma) or vehicle (0.1% ethanol) for variable periods of time as indicated in figure legends. Selected siGenome siRNAs (Dharmacon, Chicago, IL, USA) were transfected using siLentFect lipid reagent according to the instruction of the manufacturer (Bio-Rad Laboratories, Hercules, CA, USA). Medium was changed 24 h after transfection with fresh phenol red-free DMEM supplemented with 0.5% charcoal-treated FBS and hormonal treatments were initiated 48 h after transfection. Cells were harvested 24 h after (i.e. 72 h after siRNA transfection).

## **Western analysis**

Whole cell extracts were prepared as described previously (53) using antibodies directed against ER (mouse monoclonal B10 antibody, a kind gift from Prof. P. Chambon) or against  $\beta$ -actin mouse monoclonal antibody (AC-15, Sigma Diagnostics).

## **Growth assays**

MCF-7 breast cancer cells were seeded to low density and treated every 2–3 days with vehicle (0, 0.1% ethanol), 17 $\beta$ -estradiol (E2, 25 nM) or ICI 182-720 (ICI, 100 nM) in 5% charcoal-stripped FBS. Protein concentrations were measured after 9 days as previously described (54). When siRNA were used, cells were re-plated 24 h after transfection at low density and treated twice over a 5-day period with vehicle (0, 0.1% ethanol), 17 $\beta$ -estradiol (E2, 25 nM) or ICI 182-720 (ICI, 100 nM) prior to quantification of total protein concentrations. Experiments shown are representative of two to three independent experiments performed with duplicate samples.

## **Cell cycle analysis**

MCF-7 cells were transfected as previously described. After 24 h, medium was changed for phenol red-free DMEM supplemented with 0.5% charcoal-treated FBS and 24 h later, stimulated with either vehicle (0, 0.1% ethanol) or estradiol (E2, 25 nM) for a last 24 h. Cell cycle values were obtained after ethanol fixation and propidium iodide staining followed by data acquisition on a FACS 'BD LSR II' and analysis using the ModFit LT cell cycle software. Experiments were performed at least twice with duplicate samples.

## **RNA purification and DNA microarrays**

After treatment of MCF-7 cells with estradiol or vehicle in the presence or absence of cycloheximide, medium was completely removed and cells were collected in 1 ml of

TRI-Reagent (Sigma). Total RNA was extracted as recommended by the manufacturer and further purified with RNeasy MinElute Cleanup Kit (QIAGEN, Mississauga, ON, Canada). cRNA synthesis from total RNA, labeling and hybridization to Affymetrix HG-U133 2.0 Plus gene chips were performed at the Genome Quebec and McGill University Innovation Center using standard protocols (<http://www.genomequebec.mcgill.ca/>). A total of 16 chips were hybridized, analyzing quadruplicate samples for each of the four treatment conditions. The microarray analysis was performed with the affyLmGUI Bioconductor package (55) after normalization of probe-level data with RMA (56). Genes deemed significantly regulated were those with 1.4-fold change between vehicle and E2 treatments, average log<sub>2</sub>-expression levels greater than 5 across all samples (A-value) and a P-value for moderated t-statistics (57) smaller than 0.01. Regulated genes were ranked according to  $|\text{fold}| * A * (1 - P\text{-value})$ . The 2144 regulated genes were classified into different subsets as follows. UP-regulated: 1214, DOWN-regulated: 955, UP –CHX: 858, UP only –CHX: 670, UP + and –CHX: 188, UP only +CHX: 356, UP +CHX: 544, DOWN –CHX: 838, DOWN only –CHX: 719, DOWN + and –CHX: 119, DOWN only +CHX: 117, DOWN +CHX: 236. The microarray data is accessible through the GEO accession number GSE8597 [NCBI GEO].

### **Screening for transcription factor binding sites**

The reference human genome sequences (hg17, May 2004) were downloaded from the UCSC Genome Browser database (58). Genomic sequences in windows centered around the transcription start sites (TSS) of all annotated gene in the RefSeq Genes track (59) were extracted from the genome. These sequences were screened with 299 matrices for transcription factor binding sites using a base score cutoff of 65% and 5% increments. The program used for screening implements the formulas described by Wasserman and Sandelin (60); other programs developed for genome-wide screen of nuclear receptor response elements (51,61,62) were used to calculate the distance from the TSS and identify binding sites within transposable elements or regions bound in ChIP-on-chip experiments.

The majority of matrices used was from the public version of TRANSFAC (63) or represented variation of TRANSFAC matrices (i.e. E2F matrix M00516 where only the portion matching TTSGCGC was used). Selected matrices were compiled from the literature including matrices for estrogen response elements (51,64,65) and RXR/RAR (66). Searches were also performed with the same sequences from which transposable elements were removed, or with only sequences present in conserved regions, as described in the 'PhastCons Conserved Elements' track from the UCSC (67), within the same windows. For each transcription factor, four cutoffs were chosen with frequencies in all gene promoters closest to those of EREs (70%: 0.32 ERE/gene, 75%: 0.16 ERE/gene, 80%: 0.06 ERE/gene, 85%: 0.02 ERE/gene). Cutoffs were required to be between 2 sites/gene and 0.01 site/gene to avoid overly abundant sites or rare sites. Z-scores and P-values were determined for each of the selected cutoffs to assess the statistical significance of the observed enrichment in promoters of different sets of regulated genes versus those of all annotated genes. The Z-scores and P-values were calculated with programs adapted from oPOSSUM perl application programming interface (API) using the cutoff recommended by the authors (68). The P-value indicates if the proportion of genes is greater than would be expected by chance in a Fisher exact test. The Z-score evaluates the significance of the rate of occurrence of sites in the test set of regulated genes versus the expected rate estimated from the background set composed of all genes in the genome using a binomial distribution model. Stars indicate the significance of both parameters (Z-scores >10, P-value <0.01). The ER ChIP-on-chip dataset (50) was downloaded from the Brown Lab website (<http://research.dfci.harvard.edu/brownlab/datasets/index.php>) and the E2F1 ChIP-on-chip dataset (69) was downloaded from the NCBI (GSE5175 [NCBI GEO] ).

### **Gene expression quantification**

Total RNAs were extracted as described earlier and aliquots of 2  $\mu$ g were reverse transcribed using the RevertAid H first minus strand cDNA synthesis kit (MBI Fermentas, Burlington, ON, Canada) as recommended by the manufacturer. Reverse transcription

products were diluted 10 times in pure water prior to real-time quantitative PCR. Gene expression levels were determined using primer and probe sets from the Universal Probe Library (<https://www.roche-applied-science.com/>). PCR reactions were performed in 384-well plates using 2  $\mu$ l of cDNA sample, 5  $\mu$ l of TaqMan PCR Master Mix (Applied Biosystems, CA, USA), 2  $\mu$ M of each primer (sequences available upon request) and 1  $\mu$ M of the Universal TaqMan probe in a total volume of 10  $\mu$ l.

The ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) was used to detect amplification levels and was programmed for an initial step at 95°C for 10 min, followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. All reactions were run in duplicates or triplicates and average values were calculated. Quantification was performed with at least two independent experiments. The housekeeping RPLP0 (ribosomal protein, large, P0) gene was used as endogenous control. Relative expression levels of target genes and SD values were determined using the CT method.

### **Chromatin immunoprecipitation**

For chromatin immunoprecipitation (ChIP) assays, chromatin was cross-linked by treating cells with 1.5% formaldehyde for 10 min at room temperature and fragmented by sonication as previously reported (51,52), yielding fragments of average size 400 bp. Immunoprecipitation were conducted as previously described (51,52). Antibodies against a C-terminal epitope of hER (SC-543) and against  $\beta$ -actin (SC-8432) were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA, USA). Detailed conditions for immunoprecipitation, PCR amplification and primer sequences (synthesized by Sigma Genosis) used in ChIP assays are available upon request. Chromatin immunoprecipitation experiments were performed twice with similar results. A representative set of results is shown.

## RESULTS

### Identification of primary target genes of estrogen receptor alpha in MCF-7 cells

MCF-7 cells are one of the most widely used cell models to study estrogen signaling as they are strongly growth-stimulated by 17 $\beta$ -estradiol (E2) in vitro (Figure 1A) and in xenograft experiments (70), and express high levels of estrogen receptor alpha (ER). Transfection of siRNAs directed against ER, but not GAPDH or Luciferase (LUC), strongly suppressed MCF-7 cell growth (Figure 1B), indicating that ER $\alpha$  expression levels are rate limiting for growth of these cells. In the absence of E2, MCF-7 cells cultured in 0.5% charcoal-treated serum were mostly arrested in the G0–G1 phase (>80%), while addition of E2 for 24 h augmented significantly the proportion of cells in S phase (Figure 1C). On the other hand, transfection of siRNAs against ER reduced this proportion both under basal and induced conditions. Thus, it is likely that genes regulated by estrogens under these conditions include a large proportion of cell cycle-regulated genes, some of which may not be primarily regulated by estrogens. Indeed, E2 treatment leads to regulation of transcription factors, cofactors and signaling proteins, which amplifies the genomic effects of estrogens but complicates the analysis of the mechanisms of target gene regulation.

To identify primary estrogen target genes in MCF-7 cells, we performed treatments with E2 for 24 h with or without pre-treatment of cells with the protein synthesis inhibitor cycloheximide (CHX). Gene expression patterns were analyzed using Affymetrix GeneChips HG-U133 2.0 Plus chips representing over 47 000 human transcripts. Differential expression was assessed as the ratio between the average of the four replicates in the presence of estrogen versus its absence, either in the presence of CHX or in its absence. A cutoff of 1.4-fold change in expression was applied, based on the observed

regulation of known ER targets such as TFF1 under these conditions. A total of 2144 significantly regulated genes were identified and ranked according to their combined P-value, fold-regulation and intensity of expression (see the lists of the top 100 genes regulated in the absence or presence of CHX in Table 1). There was a significant overlap in regulated genes with other microarray studies, which was optimal when using similar filtering parameters for the analysis of all the datasets [922/2144 with Carroll et al. (50) and 357/2144 with Rae et al. (71); see Figure 2A]. Statistical analysis by a Fisher exact test indicates that all overlaps between two datasets are significant (P-value lower than  $2.2E-16$  for each pair of microarrays); in addition, much smaller overlaps are found in random simulations performed 1 million times (percentages or number of genes identified by a star in Figure 2A). The large majority (>90%) of genes in the overlapping sets were regulated similarly in pairwise comparisons (see legend of Figure 2A). A total of 203 target genes were identified as estrogen target genes in all three studies. This partial overlap may reflect differences in experimental conditions (e.g. time of treatment, culture medium) or in the type of microarrays (Affymetrix U133A versus U133 Plus 2.0).

Regulated genes included nearly equivalent numbers of up- and down-regulated genes in the absence of CHX (858 up-regulated and 838 down-regulated genes, Figure 2B), but more than twice as many up- versus down-regulated genes in the presence of CHX (544 up-regulated and 236 down-regulated genes, Figure 2B). Only 25 genes had opposite regulation in the absence or presence of CHX. The proportion of genes regulated in the absence (78%) versus the presence (35%) of CHX in this study was similar when considering only genes that were not identified as E2 targets in the two previous large-scale studies in MCF-7 cells (50,71). This indicates that treatment with the protein synthesis inhibitor is unlikely to account for the partial overlap with these studies. Moreover, although it is possible that CHX treatment may mask regulation of some target genes, CHX treatment did not have a pronounced global effect on expression of the group of genes induced only in the absence of CHX (Supplementary Figure 1). Regulation of target gene

expression by E2 in the absence or presence of CHX was confirmed using Q-PCR for 41 putative primary and 22 putative secondary target genes (Supplementary Figure 2).

Surprisingly, the overlap between genes regulated in the presence and absence of CHX was only partial (35% for up-, 50% for down-regulated genes). Regulation of gene expression only in the presence of CHX may be due to several factors, including regulation below the cutoff and/or lack of detection of transcripts in the –CHX category, or suppression of negative feedback mechanisms on ER-mediated gene expression by CHX treatment. Several well-known estrogen target genes were represented in the set regulated only in the presence of CHX, including CCND1, IGFBP4, C3 and MYC. Regulation by E2 of several genes in this category could be detected at earlier times (peak around 8 h after E2 stimulation, Figure 3B), and was also observed in the presence of CHX at 8 h and at earlier time points (data not shown), confirming their status as primary target genes. On the other hand, CHX treatment had an attenuating effect on regulation of other genes such as the immediate early FOS gene (Figure 4A), although regulation of the FOS gene was still clearly observed in the presence of CHX treatment at 1–2 h (data not shown), as previously reported (72).

Representation of genes involved in cell cycle control, including cyclins and other cell cycle regulators, is strikingly different in genes regulated in the absence or presence of CHX (data not shown). This correlates with the fact that CHX treatment led to loss of progression into S phase (Supplementary Figure 3). The group of genes whose induction by E2 was repressed by CHX includes genes associated with the proliferation signature of breast cancer cells, such as members of the E2F family of transcription factors (Figure 4B), of the cell division cycle (CDC) and minichromosome maintenance (MCM) gene families (Supplementary Figure 2). Their regulation reflects the proliferative effects of estrogens and the need for synthesis of cell cycle regulatory proteins that are controlled between cycles by degradation (73).



## **ER participates in both up- and down-regulation of gene expression, but EREs are enriched only in the flanking sequences of up-regulated primary target genes**

To investigate whether ER plays a similar role in the up- and down-regulation of primary target genes, we selected a sample of five primary up- and five down-regulated target genes and assessed the effect of siRNA-mediated ER depletion (Figure 5). Two different siRNAs directed against ER both markedly reduced levels of expression of the receptor and levels of cyclin D1 (CCND1) mRNAs (Figure 5A and B), consistent with their anti-proliferative effects (Figure 1B and C). Modulation of both up- (Figure 5C) or down-regulated (Figure 5D) primary estrogen target genes was markedly attenuated by depletion of ER. Transfection of siRNAs directed against GAPDH or LUC on the other hand had generally little impact on expression of ER target genes in the absence or presence of estradiol, although we observed a reproducible effect of the GAPDH siRNA on lowering TFF1 expression and of the LUC siRNA on increasing expression of ERBB2, possibly due to effects of these siRNAs on unknown factors regulating expression of these genes. Together these results indicate that expression levels of the receptor are a limiting factor for overall target gene regulation in MCF-7 cells. These results also suggest that overexpression of ER during tumorigenesis likely plays an important role in potentiating estrogen signaling.

The classical mechanism of action of ER signaling involves direct binding to cognate DNA sequences, the EREs, whose consensus sequence is a palindrome of PuGGTCA motifs (13,16,17). Surprisingly, there was no difference in the percentage of EREs identified between groups of up-, down- or non-regulated genes in ZR75 cells (47). On the other hand, enrichment in EREs was reported in the promoter sequences of target genes identified in the presence of cycloheximide in T47D cells (46); however, the number

of primary target genes identified was small (89 genes) and enrichment was assessed only in a 3.5 kb window of regulatory sequences. Here, we compared ERE frequencies in genes regulated by estradiol in the presence and/or absence of the protein synthesis inhibitor cycloheximide (CHX) in MCF-7 cells. EREs were identified at variable distances from the transcriptional start sites of genes using a customized position weight matrix, derived by compiling binding motifs from the literature and by symmetrical complementation reflecting the palindromic nature of this site (Figure 6A; note that this matrix yielded higher enrichment in EREs than the TRANSFAC matrix in control datasets corresponding to lists of genes with known EREs or ER-bound ChIP fragments, data not shown). Windows of 20 kb centered on the transcriptional start sites (TSS) of target genes were used with different cutoffs corresponding to frequencies of sites in the genome comprised between 0.1 and 2 sites per gene. This approach detected a significant enrichment in EREs in a test list of genes found to contain EREs in a genome-wide search for high-affinity EREs (51) compared to the whole genome (data not shown). The ratio between the numbers of sites detected in regulated genes versus all genes in the genome remained close to one at all cutoffs when both up- and down-regulated genes were considered together, but increased with higher cutoffs in up-regulated genes while it decreased in down-regulated genes (Figure 6B). However statistical significance (P-value <0.01 and Z-score >10) was reached only at one point in total up-regulated genes in this window. On the other hand, ERE enrichments in E2 target genes up-regulated in the presence of CHX (+CHX group), but not in its absence (-CHX group), were statistically significant. Further subdivision of the +CHX group into only+CHX and -and+CHX indicated statistically significant enrichment in both the only+CHX group and the -and+CHX groups (Figure 6B). No enrichment was seen in the only-CHX group, consistent with the interpretation that most of the genes in this group are secondary target genes. Similarly, while we did not find cutoff-dependent enrichment in EREs in the datasets of up-regulated genes regulated in Carroll et al. (50) or in Rae et al. (71), enrichment in EREs can be observed in any intersection between our set of primary up-regulated genes and these other sets of regulated genes (Supplementary

Figure 4). In addition, our analysis of another published dataset of genes regulated in MCF-7 cells by E2 and CHX at 48 h (74) also indicates enrichment in EREs in up-regulated, but not down-regulated primary targets (Supplementary Figure 5). Thus, the observation that EREs are enriched in primary up-regulated target genes is not dataset dependent. Of note, the observation that enrichment was higher with higher cutoffs indicates that sequences closer to the consensus are more likely to mediate transcriptional regulation, indicating a statistical relationship between *in vitro* affinity and *in vivo* binding to EREs.

Distribution of EREs in sliding 2.5 kb windows with 500 bp increments between –25 and +25 kb of the TSS indicates that enrichment of EREs (identified at a 75% cutoff) in genes up-regulated in the presence of CHX, but not in those up-regulated in its absence, can be observed within –20 to +20 kb of the TSS (Figure 7A). Finally, enrichment in EREs was higher when considering only ER-associated chromatin regions (50) found within a 20 kb window around the TSS of different groups of genes, reflecting the capacity of ChIP to accurately pinpoint the regulatory regions in target gene flanking sequences (Figure 6C). However, no enrichment was observed in the only–CHX group of up-regulated genes even when considering only ER-bound fragments (Figure 6C). Together, these results validate the use of CHX to identify primary E2 target genes.

Surprisingly, no enrichment but rather a statistically significant depletion in EREs was observed in a 20 kb region around the TSS of genes repressed by estrogens (Figures 6B and 7B). We examined whether enrichment in EREs was observed in ER-associated chromatin regions (50) found near genes down-regulated in the absence or presence of CHX (Figure 6C). No enrichment in ERE sequences was observed in the genes down-regulated either in the presence or the absence of CHX. This suggests that binding sites for other transcription factors mediate the down-regulation of genes by E2 in the presence of CHX (see also the next section).

To test whether predicted EREs can bind ER *in vivo*, selected regions containing EREs identified using a 65% cutoff and present at  $\pm 25$  kb of the TSS of target genes up-

regulated in the presence of CHX were amplified by PCR after ChIP. Binding of ER was observed on all predicted elements with the exception of the ERE with two mismatches present at +2.3 kb of the EGR3 TSS (Figure 8). Only two of the nine validated elements (the WISP2 -452 bp ERE and CA12 -6 kb ERE) were detected previously in genome-scale ChIP experiments (50). The intensity of binding to individual elements did not necessarily reflect their degree of similarity to the consensus sequence, as one of the strongest sites bound [at -452 bp of the WISP2 TSS, also observed in Ref. (50)], diverges from the consensus ERE by two mismatches. This may be due to the sequence context of the EREs, which is likely to affect *in vivo* affinity through cooperativity with other transcription factors bound close to the EREs. In addition, the occurrence of several EREs in the flanking regions of primary target genes suggests potential cooperativity between EREs for DNA binding and transcriptional activation through formation of chromatin loops as described for the GREB1 gene (52).

### **Binding sites for other transcription factors are enriched in up- and down-regulated primary estrogen target genes**

The lack of enrichment of EREs in down-regulated target genes suggests transcriptional regulation via other mechanisms than ER $\alpha$  binding to EREs. Several transcription factors have been reported to mediate indirect recruitment of ER to DNA via protein-protein interactions [see refs (15,31-33) for reviews]. In addition, other transcription factors may also mediate estrogenic regulation via non-genomic mechanisms (15,34-38). Binding sites for transcription factors participating in either mechanism may be enriched in the flanking sequences of primary regulated genes, while only enrichment in tethering transcription factors is expected in ER-associated regions identified through ChIP.

To identify transcription factor binding sites enriched in either up- or down-regulated primary targets, we screened TRANSFAC matrices for >1.5-fold enrichment in frequencies within 5 or 10 kb windows around the TSS of different classes of regulated

genes versus all annotated genes. Due to the variable representation of different transcription factor binding sites in low-complexity regions, this analysis was performed considering either whole genomic sequences, sequences without repeated elements or sequences conserved through evolution. Enrichments with Z-scores higher than 10 and P-values  $<0.01$  for at least two different cut-offs of the same matrix were retained as significant. This analysis was performed using either all significantly regulated genes (Supplementary Table 1) or the top 100 regulated genes (Supplementary Table 2) in each category.

Binding sites enriched in primary up-regulated target genes (regulated in the presence of CHX) included, in addition to EREs, sites for AHR/ARNT, AP1, CCAAT box factors, C/EBP homologous protein (CHOP), CHX10 homeobox factor, EGR factors, Maf, GCbox factors and ERE-like sequences with different arrangements of the motifs (Supplementary Tables 1 and 2). On the other hand, binding sites enriched in primary down-regulated genes include sites for the forkhead transcription factor FREAC3 (FOXC1), CHX10 and PAX homeobox factors, EVI1, Ikaros and SRY transcription factors (Supplementary Tables 1 and 2).

### **E2F transcription factors are a major relay for transcriptional regulation by estrogens of secondary target genes in MCF-7 cells**

No enrichment of EREs was observed in genes regulated only in the absence of CHX, consistent with their regulation by other transcription factors. Screening of TRANSFAC matrices indicated enrichment of E2F-binding sites in up-regulated secondary targets, as well as of sites for CCAAT box factors, CREB and Forkhead (HNF-3B) transcription factors (Supplementary Tables 1 and 2). In addition, sites for AP1, CREB, C/EBP, CCAAT-box factors, forkhead factors (FOXA, FOXD3, FOXO4, HFH8), E2F,

GATA factors, GC-box factors, MAZR and PAX factors were enriched in down-regulated secondary target genes (Supplementary Tables 1 and 2).

Because secondary target genes are enriched in cell cycle-associated proteins and E2F factors play well-characterized roles in the regulation of cell proliferation, we examined more closely the possible contribution of these factors in the networks of gene regulation by estrogens. Enrichment in E2F sites within a 10 kb window around transcriptional start sites ( $\pm 5$  kb) was strong in genes up-regulated in the absence of CHX (Figure 9A). It was also observed in genes up-regulated in the presence of CHX, but not in the only+CHX group, suggesting that the –and+CHX group of regulated genes was responsible for this over-representation. A similar enrichment in secondary, but not primary target genes was observed when restricting the analysis for each group to genes found to contain E2F-binding sites in their promoter in ChIP experiments (69), although enrichment in the –and+CHX group was not observed. Enrichment in down-regulated secondary genes was not observed when considering all regulated genes (Figure 9, Supplementary Table 1), but was detected in the top 100 regulated genes (Supplementary Table 2). Several E2F family members in addition to E2F1 were regulated by E2, including E2F2, 6, 7 and 8 (Figure 4B), which may explain the prevalence of the E2F signature in regulated secondary genes.

Genes down-regulated in the absence of CHX were enriched in binding sites for NF-B (Supplementary Table 1), consistent with the repressive role of estrogens on NF-B signaling (15,33) although enrichment of these sites in the absence rather than the presence of CHX was unexpected. Similarly, binding sites for GATA factors were enriched in down-regulated secondary target genes (Supplementary Table 2). Repression of GATA1 transcriptional activity by estrogens has been shown previously (75,76). More surprisingly in view of the reported stimulatory effects of estrogens on Sp1- and AP1-mediated transcription, sites for these transcription factors were also found enriched in down-

regulated secondary target genes (Supplementary Tables 1 and 2). These results suggest that AP1 sites may be involved in mediating gene repression as well as activation (see above for enrichment in up-regulated primary target genes).

Unexpectedly, c-Myc-binding sites were not found enriched in any of the target gene categories, whether considering all sequences, low-complexity sequences or conserved sequences. This is puzzling in view of the observation that induction or repression of MYC expression recapitulates the effects of estrogen and antiestrogens, respectively, on the cell cycle of ER-positive breast cancer cells (77,78). However, comparison of E2 target genes with those of c-Myc target genes (79) and with genes regulated by siRNA-mediated suppression of MYC expression (80) indicates a statistically significant enrichment of c-Myc target genes in genes regulated by E2 (data not shown). C-Myc target genes were distributed in different groups of regulation (up versus down, primary versus secondary), which may explain the lack of significant enrichment in a specific group. Poor performance of the TRANSFAC matrices in identifying c-Myc-binding sites could also play a role in the absence of detected over-representation of these sites in estrogen target genes. Finally, we also monitored whether c-Myc-binding sites may be enriched in primary target genes, consistent with a mechanism of action through non-genomic activation by estrogens. Indeed, estradiol has been shown to stabilize the c-Myc protein for at least 12 h through modulation of its phosphorylation status (81). However, we did not observe significant enrichment in this category of regulated genes either. It remains possible however that identification of E2 target genes at earlier time points may reveal enrichment in c-Myc-binding sites, since MYC is one of the earliest known estrogen target genes (82). Of interest, TIAM1, a repressor of c-Myc that prevents its apoptotic effects (83), is also a primary estrogen target gene and may limit the duration of the induction of some c-Myc target genes.

Together, these data indicate that E2F is a major mediator of propagation of estrogen signaling in MCF-7 breast cancer cells for both up- and down-regulated secondary target genes, and suggest that AP1 may play a role in the down-regulation of secondary target genes.

## **DISCUSSION**

Microarray analyses of estrogen target genes have been performed previously in various ER-positive cell lines. However, only a limited number of genes have been found to be commonly regulated so far (46,71). Although this may be due to intrinsic differences between responses in these cells (71), variability between studies may also take place due to differences in cell culture conditions, microarray platform and selection criteria for significantly regulated genes. In this study, we chose to perform microarray analysis in 0.5% charcoal-treated serum, to minimize unstimulated levels of cellular growth; other studies were performed in 5% (45,50) or 10% serum (70,71). Our control was provided by cells treated with vehicle for the same time as with E2, whereas other studies have compared induction by E2 at different times with the absence of treatment at time 0 (50,70,71). We used the HG-U133 2.0 Plus GeneChips, while most other studies used anterior versions of Affymetrix GeneChips (45,70,71). Finally, we used a combination of three parameters, i.e. fold-change, average expression levels and P-value, to rank regulated genes, with set minima for each parameter (see Materials and Methods section). We identified 2144 significantly regulated genes, of which more than 50% overlapped with at least one of the two previous large-scale microarray studies also performed in MCF-7 cells (50,71). As we did not observe a higher overlap between sets of regulated genes described in these two other studies, we speculate that this partial overlap results from a combination of differences in experimental design and/or microarray platforms, rather than from the use of cycloheximide specific to our study.



While enrichment in different transcription factors has been reported in E2 targets in different studies (47,70), only one study based on microarray analysis exclusively has reported enrichment in EREs in the promoter sequences (–3 to +0.5 kb) of 89 genes identified as primary targets based on lack of effect of cycloheximide on regulation (46). Here, we identified 780 genes significantly regulated by E2 in the presence of CHX (544 up-regulated and 236 down-regulated genes). Overlap with genes regulated in the absence of CHX was partial. This can be explained in part by the fact that regulation of several genes found significantly modulated in the presence of CHX was also observed in its absence at earlier times, but became non-significant at 24 h. Genes in this category may include G1 phase-specific genes, as cycloheximide inhibited the increase in S phase induced by E2. In addition, possible inhibition by cycloheximide of negative feedback regulatory loops may also result in increased gene regulation at 24 h. Because CHX has major effects on gene expression (more than 4000 genes were either up- or down-regulated in our study, data not shown), it remains possible that CHX treatment may mask regulation of some primary target genes. However, we did not find that CHX treatment had a significant global effect on expression of the group of genes induced only in the absence of CHX (Supplementary Figure 1).

Enrichment in EREs was observed only in genes up-regulated in the presence of CHX, including both genes regulated in the absence and presence of CHX and those regulated in the presence of CHX only. These observations validate the characterization of these genes as primary E2 targets, and explain the lack of reported enrichment of EREs in previous microarray studies that did not discriminate between primary and secondary target genes. Enrichment in EREs in genes up-regulated in the presence of CHX was detectable in a  $\pm 20$  kb window centered on the TSS of regulated genes, consistent with a role of distal EREs in transcriptional regulation. This enrichment is not data-dependent, as we observed similar results using the subsets of our primary target genes that are in common with other

microarray analyses, or using a dataset of 130 primary estrogen target genes regulated at 48 h in MCF-7 cells (74) (Supplementary Figures 4 and 5).

Regulatory sequences of up-regulated primary target genes often contained several EREs in a 20 kb window around the TSS. For instance, two functional EREs were validated in ChIP experiments in several selected primary target genes (CDH26, WISP2, CA12). Of interest, only two of the EREs validated in ChIP assays were previously reported in genome-wide ChIP-on-chip studies (50), likely due to the stringency used in this study to minimize false positive rates. While no correlation between identity with the consensus sequence and apparent strength of binding was observed for individual binding sites, general enrichment in EREs were more pronounced for sequences more similar to the consensus sequence, indicating that sequences that are high-affinity binding sites *in vitro* have a higher statistical chance of mediating E2 regulation *in vivo*, although the genomic context of the response element influences affinity at the individual level. The presence of multiple EREs in the vicinity of regulated genes may result in the formation of multiple chromatin loops associating these EREs and the TSS of regulated primary target genes as previously described (52).

Intriguingly, genes down-regulated in the presence of CHX were not enriched in EREs. Rather, a statistically significant depletion in EREs was observed. We also observed a lack of enrichment in EREs in ER-bound chromatin regions located in the vicinity of genes repressed by E2 in Ref. (50). This may indicate that the presence of EREs in the flanking regions of genes confers a likelihood of positive rather than negative regulation. However, down-regulation of gene expression via binding of ER to strategically positioned EREs remains possible, as EREs are still found, although less represented, in the vicinity of repressed primary E2 target genes. Similarly to what was observed in our dataset, we find that EREs were not enriched in down-regulated primary genes identified after 48 h

stimulation with E2 (74). However, it will be of interest in the future to assess the frequency of EREs in the flanking regions of sets of primary genes down-regulated at other, including earlier, time points.

Binding sites for other transcription factors were enriched in primary up- and down-regulated genes. Binding sites enriched in down-regulated primary target genes have not before been associated with E2 action, but include two zinc finger transcription factors participating in corepressor complexes, EVI1 (84) and Ikaros (IKZF1) (85). On the other hand, binding sites enriched in up-regulated target genes include AP1, which have been previously demonstrated to mediate tethering of ERs. In addition, the ARNT gene encodes a protein that forms a complex with the ligand-bound Ah receptor (AhR). Activated AhR/Arnt was shown to act as a coactivator of estrogen receptor-mediated signaling, while ER-AhR/Arnt protein-protein interactions mediate estradiol-dependent transrepression of dioxin-inducible gene transcription (86,87). Enrichment of AhR/Arnt sites in primary up-regulated genes suggests that AhR/Arnt-ER complexes also mediate estradiol-dependent up-regulation of some AhR/Arnt target genes. Finally, EGR genes are, like AP1 components JUN and FOS, immediate early genes activated by a variety of signaling molecules (88,89). It is possible that these factors may contribute to gene regulation via non-genomic activation by estrogen, as suggested before (90).

Sites enriched in primary target genes may mediate regulation either through tethering of ER or through non-genomic effects. Comparison with sites enriched in ER-bound chromatin regions may help to discriminate between the two mechanisms, since only transcription factors mediating ER recruitment via tethering should have sites within CHIP fragments. Enrichment of AP1, Maf, C/EBP, GATA and SOX-binding sites in ER-bound chromatin regions may thus indicate recruitment of ER to these sites. However, choice of background datasets is more complicated in the case of CHIP-on-chip data, and may affect

site enrichment. For instance, enrichment in Egr or Sp1 factors in ER-associated ChIP fragments was dependent on the choice of the reference background (promoter sequences versus flanking sequences of ChIP regions, data not shown). It should also be noted that, since cooperativity between ER–ERE complexes and transcription factors bound to the same promoter may result in indirect association of ER with these transcription factors through chromatin looping, low levels of enrichment of sites for a variety of transcription factors may be expected in ER-associated regions.

Several other transcription factors were found enriched in secondary target genes. Surprisingly, NF-B, AP1, GATA and Sp1 sites were all enriched in secondary down-regulated genes. This observation is consistent with the previous report that AP1 sites were present in ER-associated chromatin regions found close to genes repressed by E2 at late time points, this regulation being dependent on the induction of a co-repressor component such as the NRIP1 ER co-repressor (50). It will be of interest to determine whether NRIP1 induction is also required for regulation of genes repressed by E2 and containing binding sites for NF-B, GATA or Sp1 sites.

One of the most enriched binding sites in up-regulated secondary target genes is that for E2F transcription factors. Enrichment in E2F-binding sites is consistent with the well-known role of this family of transcription factors in the control of cell cycle, with known targets including several minichromosome maintenance (MCM) genes, which are essential for cell cycling due to the function of their products as replication-licensing factors (91). Other common E2F and ER targets involved in cell cycle regulation include the CDC6 and CDC25A cell division cycle genes, the PCNA and POLA2 genes, the replication factor C subunit RFC4, the structural maintenance of chromosome 2 (SMC2) and the protein regulator of cytokinesis one (PRC1) genes. Notably, we observed that not only E2F1, but also E2F2, E2F7 and E2F8 transcription factors were up-regulated by E2. E2F7 and E2F8

can repress E2F-activated transcription (92,93). E2F6 is also an inhibitory E2F family member (94,95) that can repress c-Myc target genes (96), but its induction profile was different from that of other E2F factors. The regulation of this gene may be explained by the fact that it is transcribed in the opposite direction from GREB1, a primary E2 target gene containing three consensus EREs in its flanking sequences (51, 52). Additional studies will be necessary to investigate the mechanisms of the transcriptional regulation of E2F family members by E2, and their exact individual roles in gene expression and cell cycle regulation by E2.

## **ACKNOWLEDGEMENTS**

We are thankful to Pierre Chagnon and Raphaëlle Lambert of the IRIC Genomics platform for help with the TaqMan Q-PCR analysis of estrogen target gene expression. We also thank Patrick Gendron at the IRIC Bioinformatics platform for help with the IRIC computer cluster. This work was supported by grant MOP13147 from the Canadian Institutes for Health Research and by grant 194583-01 from the Natural Sciences and Engineering Council of Canada to S.M. V.B. has been supported by a CIHR post-doctoral fellowship and JD by an FRSQ studentship. S.M. and J.H.W. are holders of FRSQ Chercheur-Boursier National awards, and S.M. holds the CIBC Breast Cancer Research Chair at Université de Montréal. Funding to pay the Open Access publication charges for this article was provided by grant MOP13147 from the Canadian Institutes for Health Research.

Conflict of interest statement. None declared.

## **Footnotes**

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

## FIGURES AND TABLES

Figure 1 Role of ER in mediating the effect of estradiol in MCF-7 breast cancer cells.

(A) MCF-7 breast cancer cells were treated every 2 days with vehicle (0, 0.1% ethanol), 17 $\beta$ -estradiol (E2, 25 nM) or ICI 182 720 (ICI, 100 nM) and protein concentrations were measured after 9 days. (B) Proliferation assays of MCF-7 transfected without (0) or with siRNA against ER, GAPDH or Luciferase (LUC). Treatments were initiated 24 h after transfection and protein concentrations were measured 6 days post-transfection. Inset: western blot analysis of ER and  $\beta$ -actin expression levels in cells transfected with siRNAs directed against ER or in control cells (T: transfection control). (C) Cell cycle analysis performed with cells transfected or not with siRNAs for 72 h and stimulated with vehicle or hormone during the last 24 h.

Figure 2 Cycloheximide-sensitive and resistant E2 target genes in MCF-7 breast cancer cells.

(A) Comparative analysis of regulated genes in different large-scale microarray studies. The number of genes in common between this study and those described in Carroll et al. (50) and Rae et al. (71) or unique to each study is indicated. Regulation of genes common to two different studies is convergent (up- or down-regulation) in 90% of the genes for the overlap between the datasets in Refs (50) and (71), 96% for the overlap between this study and Ref. (71), and 92% for the overlap between this study and Ref. (50). Average overlaps obtained in random simulations performed 1 million times are identified by a star and are expressed as percentages of total genes for pairwise comparisons or as the average number of genes common to all studies. (B) Identification of primary and secondary estrogen target genes in MCF-7 cells. RNA samples were collected 24 h after treatment of MCF-7 cells with vehicle or 17 $\beta$ -estradiol (25 nM). Cells were pre-treated 1 h before E2 stimulation with cycloheximide (CHX, 10  $\mu$ g/ml). Microarray analysis was performed with four biological replicates for each condition (see Material and Methods section). Numbers of genes significantly regulated (1.4-fold change between vehicle and E2 treatments, 5 amplitude of array signals across all conditions and 0.01 P-value based on moderated t-statistics) are indicated for each category.

Figure 3 Expression of several genes induced only in the presence of CHX at 24 h is regulated in its absence at earlier times.

Q-PCR analysis of the effects of E2 (25 nM) with or without CHX (10  $\mu$ g/ml) as a function of time on the expression levels of selected genes identified through microarray analysis. Data is representative of at least two independent experiments and is presented as the ratio between +E2 and -E2 at all time points. (A) Pattern of expression of a gene stably induced by estradiol. (B) Patterns of expression of several genes induced only in the presence of cycloheximide at 24 h in the microarray analysis.

Figure 4 Several transcription factors in the E2F family are target genes of estradiol in MCF-7 breast cancer cells.

Q-PCR analysis of the effects of E2 (25 nM) with or without CHX (10  $\mu$ g/ml) as a function of time on FOS (A) and on E2F transcription factors whose expression was found to be regulated by E2 in the microarray analysis (B). This data is representative of at least two independent experiments and is presented as the ratio between +E2 and -E2 at all time points.

Figure 5 Regulation by estradiol of both up- and down-regulated primary target genes is dependent on ER $\alpha$

Two siRNAs against ER and control siRNAs against GAPDH or Luciferase (LUC) were transiently transfected in MCF-7 cells 48 h before stimulation with E2 (25 nM) or vehicle (0.1% ethanol) for 24 h, and levels of expression of ER (A), CCND1 (B), up-regulated (C) or down-regulated (D) primary target genes were evaluated by using Q-PCR of reverse-transcribed mRNAs.

**Figure 6 Estrogen response elements are enriched in the vicinity of primary up-regulated genes.**

(A) Position weight matrix used for identification of putative EREs in the vicinity of target genes. The frequency of nucleotides at each position of the ERE in the matrix is shown. (B) The ratios of the number of EREs per gene in 20 kb windows ( $\pm 10$  kb) around the transcriptional start sites of regulated genes to the number of EREs per gene in the same windows of all annotated genes in the human genome are shown for various groups of estrogen target genes. (C) The ratios of EREs found in ER-bound chromatin regions (Ref. 50) located within a 20 kb window of the regulated genes identified in our study over EREs found in ER-bound chromatin regions at  $\pm 10$  kb of all genes in the genome are shown for various groups of estrogen target genes. ChIP regions were standardized to 1 kb on either side of the center of the region for this analysis. The numbers of genes in each category analyzed in B and C are indicated in Materials and Methods section.

**Figure 7 Distribution of EREs in a 50 kb window around the TSS of estrogen target genes.**

(A) Number of EREs per gene identified at a 75% cutoff rate using the matrix described in Figure 6A in a 2.5 kb sliding window within 50 kb of genomic regions centered around the transcriptional start sites (TSS) of primary (+CHX) or total (-CHX) up-regulated genes. (B) Number of EREs per gene identified at a 75% cutoff rate in a 2.5 kb sliding window within 50 kb genomic regions centered around the transcriptional start sites (TSS) of primary (+CHX) or total (-CHX) down-regulated genes. Note that all described alternative TSS were considered in this analysis.

**Figure 8 Identification of functional EREs in the flanking regions of primary target genes.**

Chromatin immunoprecipitation was performed using primer pairs specific for selected EREs identified through bioinformatic screens of the flanking regions of primary E2 target genes. The ChIP assay was performed after 45 min stimulation by vehicle (0, 0.1% ethanol), 17 $\beta$ -estradiol (E2, 25 nM) or ICI 182 720 (ICI, 100 nM) using a polyclonal antibody against ER or a negative control antibody against  $\beta$ -actin. NTC: non-template control for PCR amplification. The number of mismatches (m) with the consensus ERE is indicated for each element along with distances (bp) to the most upstream TSS. These results are representative of two independent experiments.



### Figure 9 E2F response elements are enriched in secondary up-regulated genes.

(A) Relative frequencies of E2F-binding sites within conserved regions in a 10 kb window ( $\pm 5$  kb) around the transcriptional start sites of regulated genes as a function of the cutoff rates used for matrix-based identification of sites (see Materials and Methods section). Enrichment is observed in up-regulated secondary target genes (not in the only+CHX category). Enrichment in the +CHX group likely reflects overlap with the -CHX group. (B) Enrichment of E2F sites in a 10 kb window around the transcriptional start sites of genes regulated by E2 in the absence and/or presence of CHX in the current study and reported to contain an E2F1-bound chromatin region in their promoters in Ref. (69). The numbers of genes in each category analyzed in A and B are indicated in Materials and Methods section.

### Table 1 Top hundred genes regulated by E2 in the presence or absence of cycloheximide.

Genes were ranked based on a combination of three parameters, fold-change between E2 and vehicle treatments, amplitude of array signals across all conditions and P-value based on moderated t-statistics. Genes in bold are those found significantly regulated both in the absence and presence of cycloheximide (CHX). Validated genes are those for which regulation was observed in Q-PCR with at least two independent experiments (see also Fig. 3-5 and supplementary Fig.2; DNS, data not shown; -, not tested).

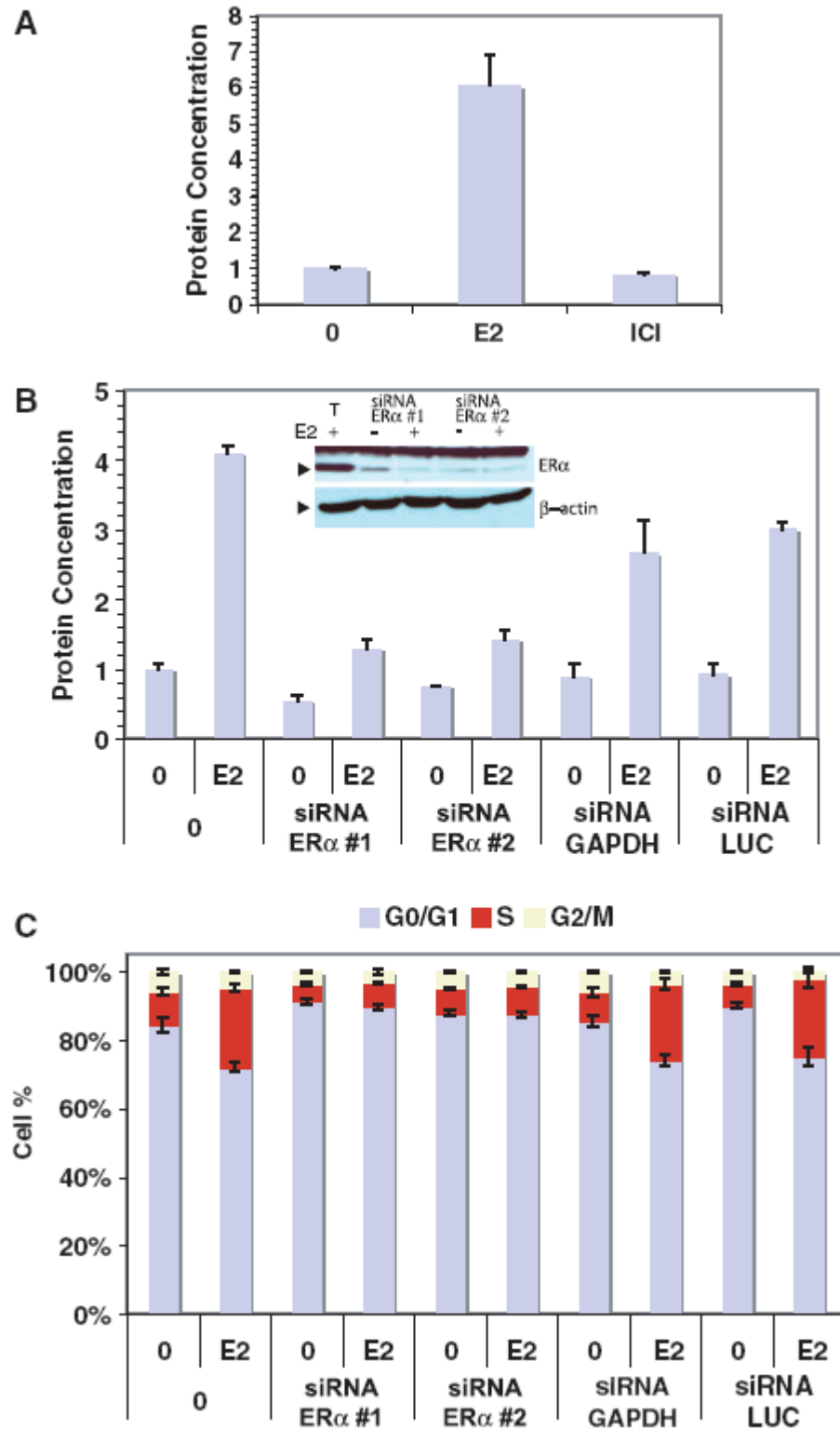


Figure 1.

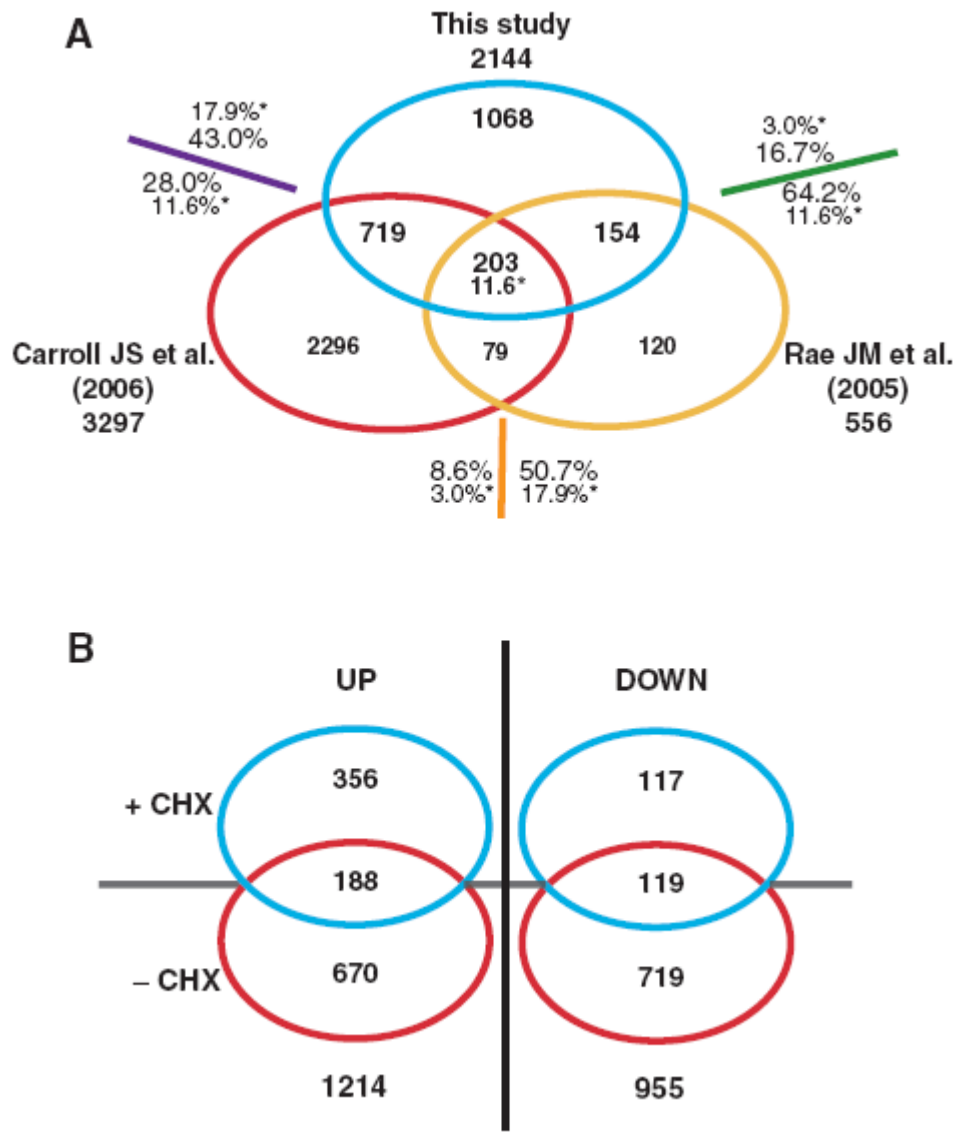


Figure 2.

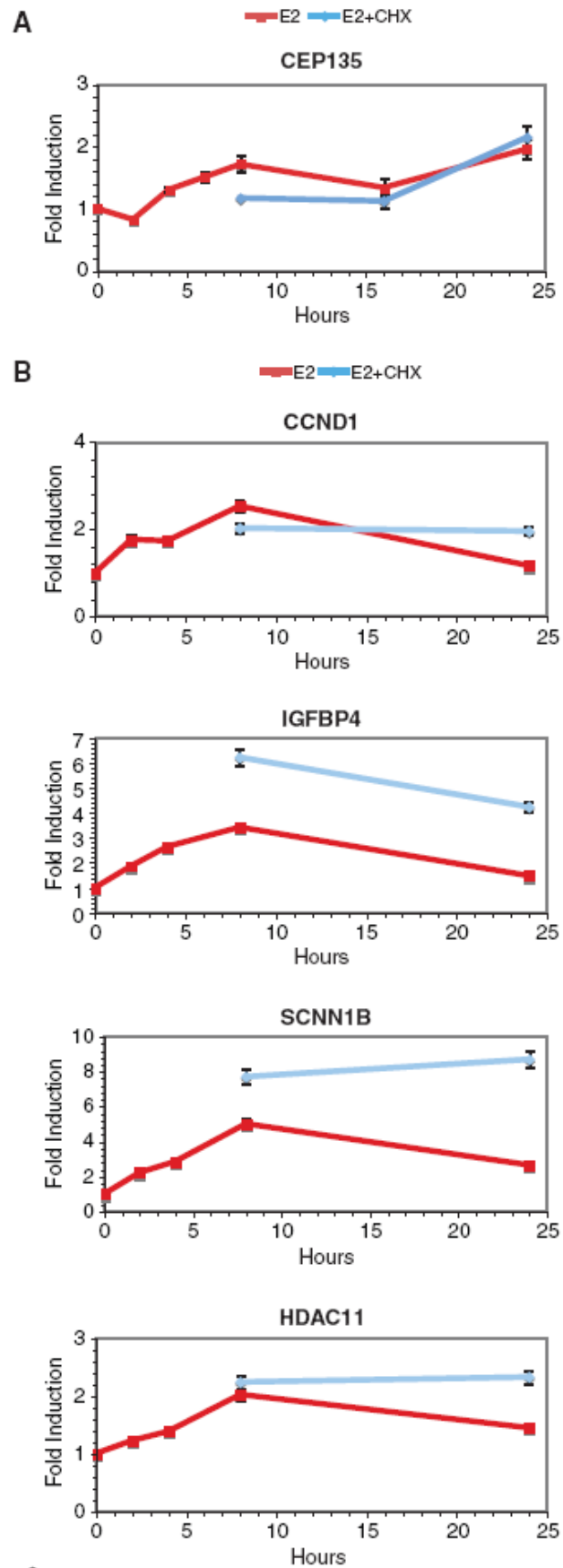


Figure 3.

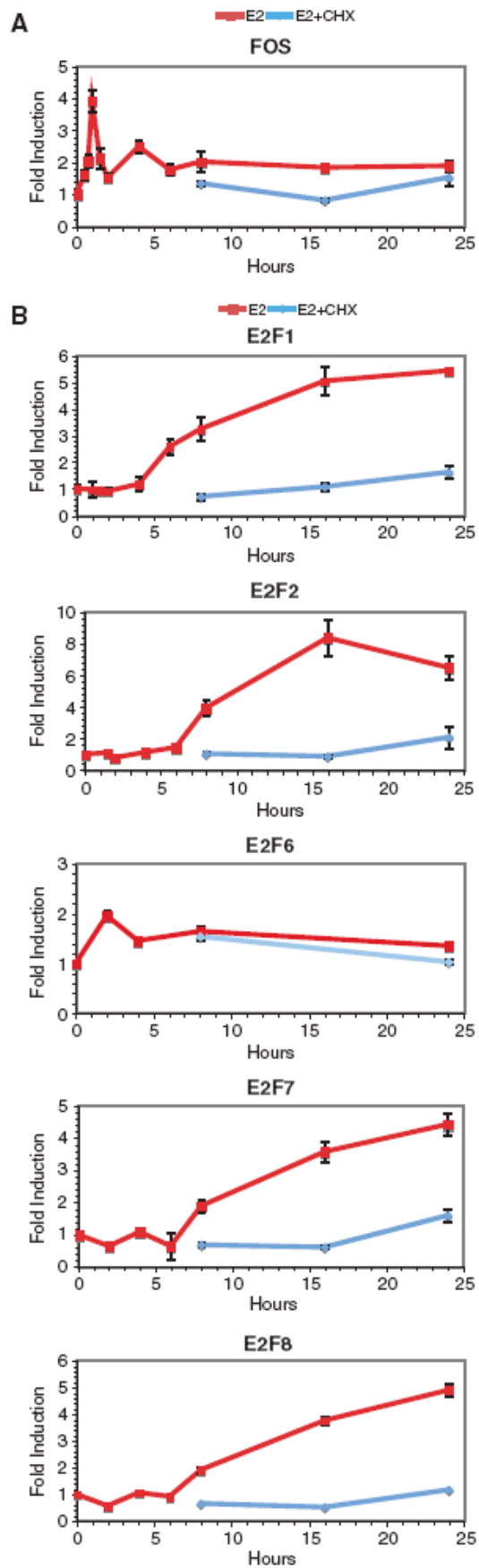


Figure 4.

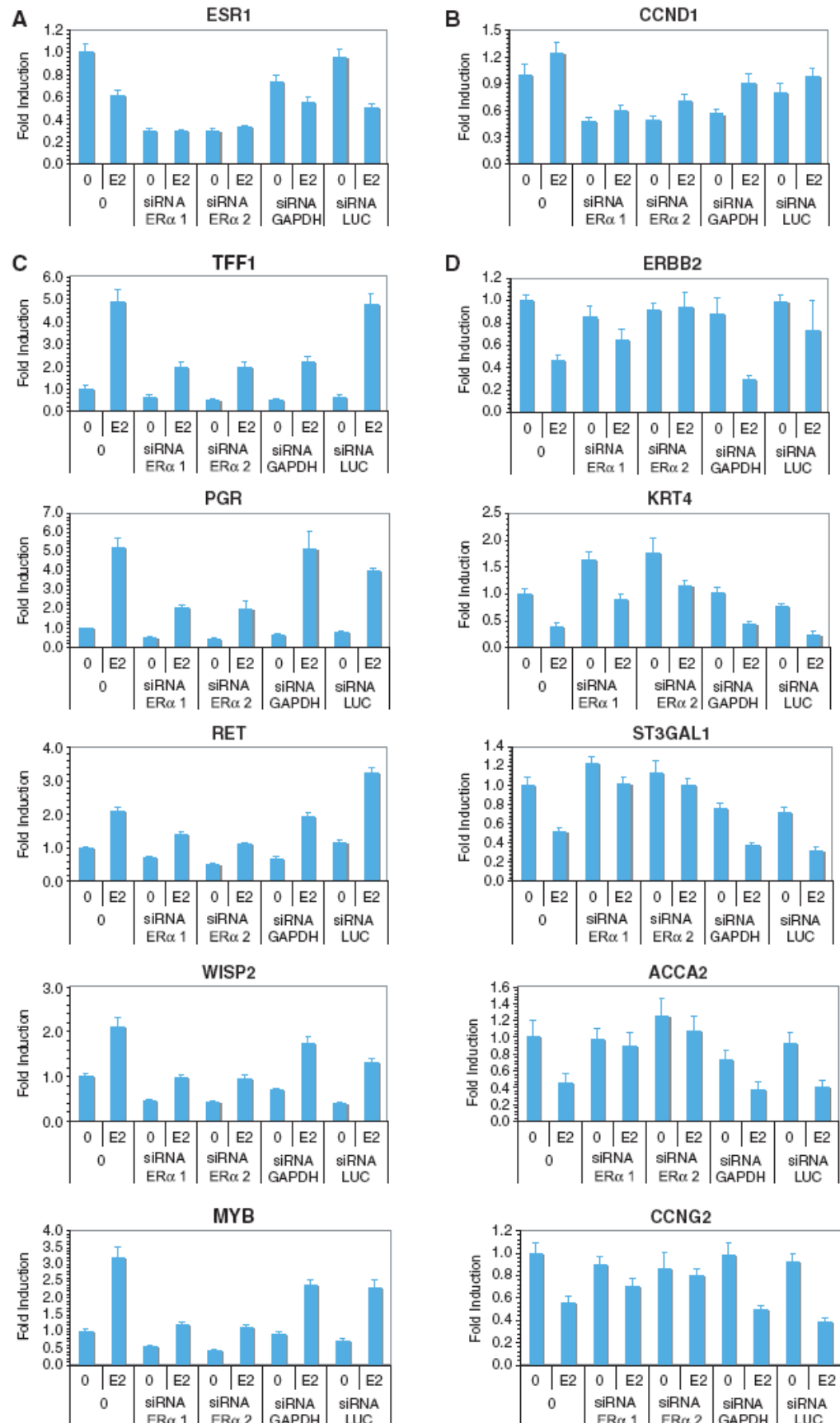


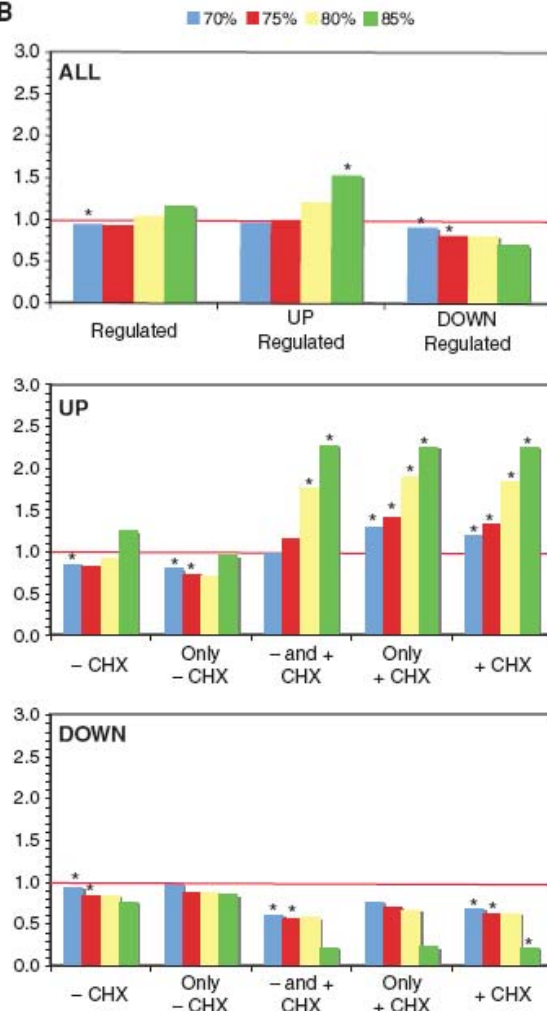
Figure 5.

A



	R	G	G	T	C	A	N	N	N	T	G	A	C	C	Y
A	0.39	0.08	0.06	0.06	0.01	0.84	0.25	0.25	0.25	0.04	0.05	0.75	0.05	0.05	0.11
C	0.14	0.05	0.03	0.09	0.88	0.05	0.25	0.25	0.25	0.07	0.06	0.11	0.86	0.82	0.36
G	0.36	0.82	0.86	0.11	0.06	0.07	0.25	0.25	0.25	0.05	0.88	0.09	0.03	0.05	0.14
T	0.11	0.05	0.05	0.75	0.05	0.04	0.25	0.25	0.25	0.84	0.01	0.06	0.06	0.08	0.39
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

B



C

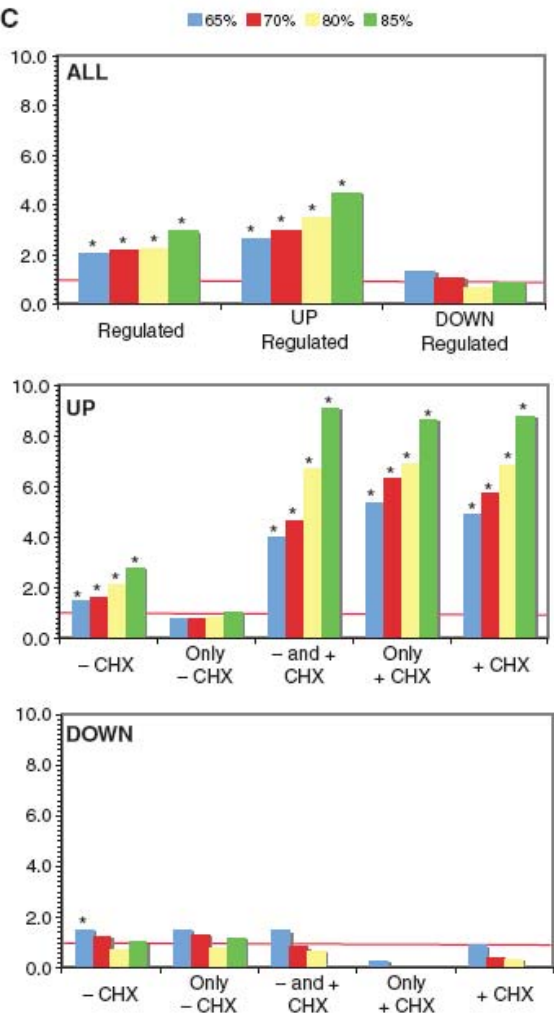


Figure 6.

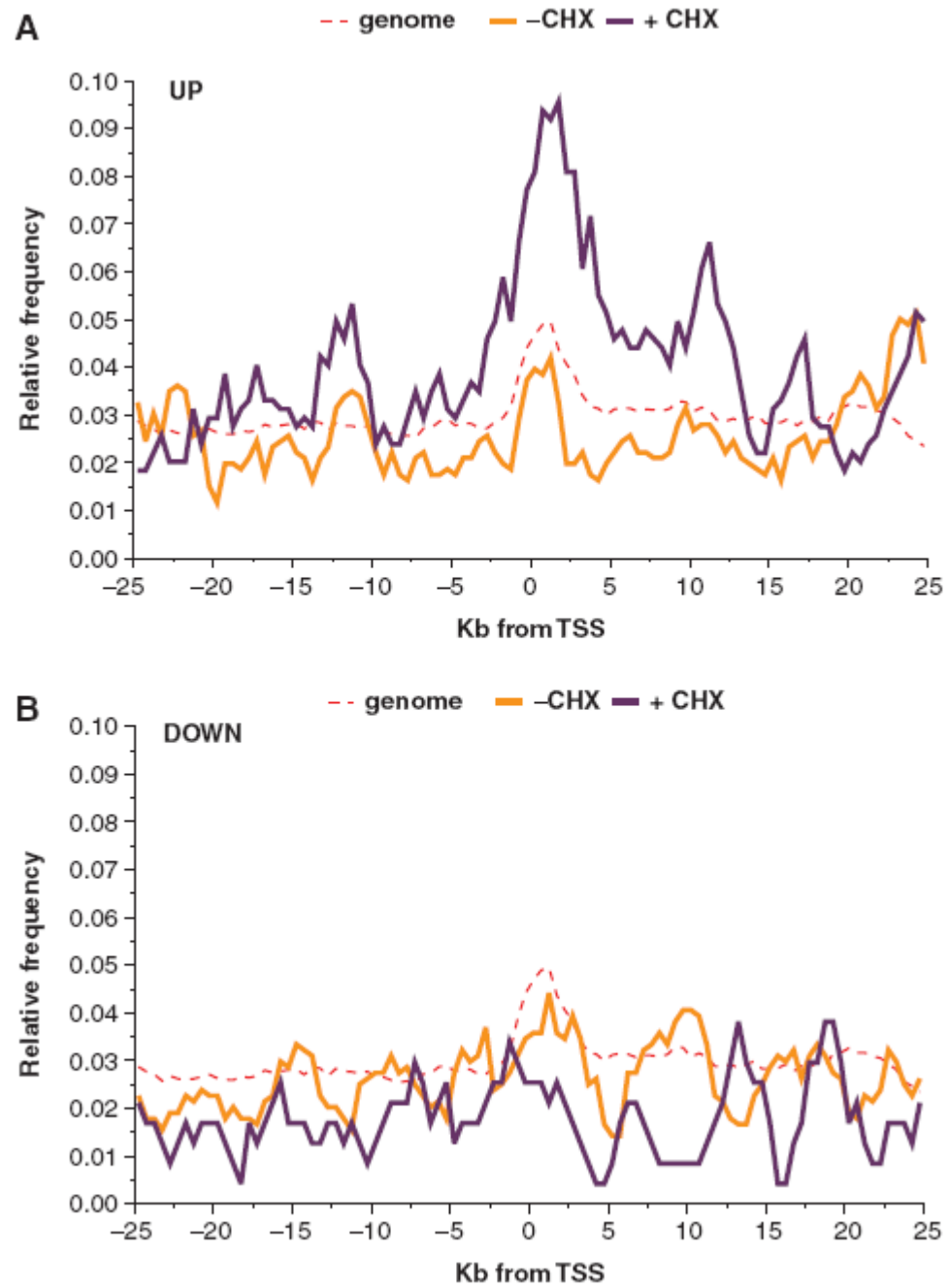


Figure 7.



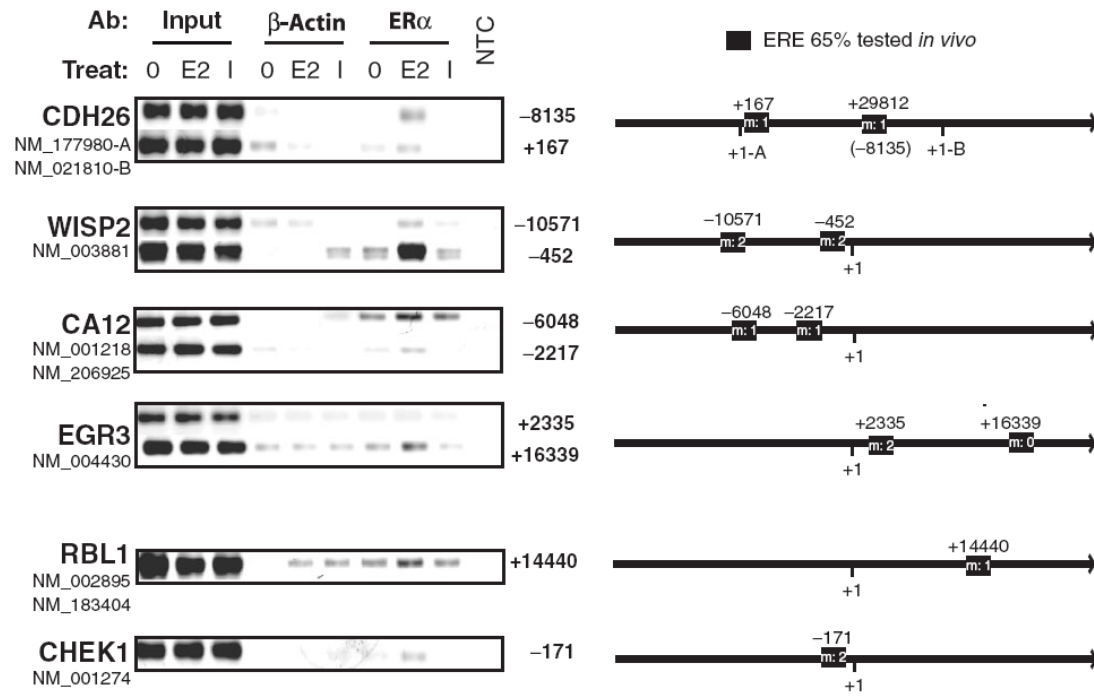


Figure 8.

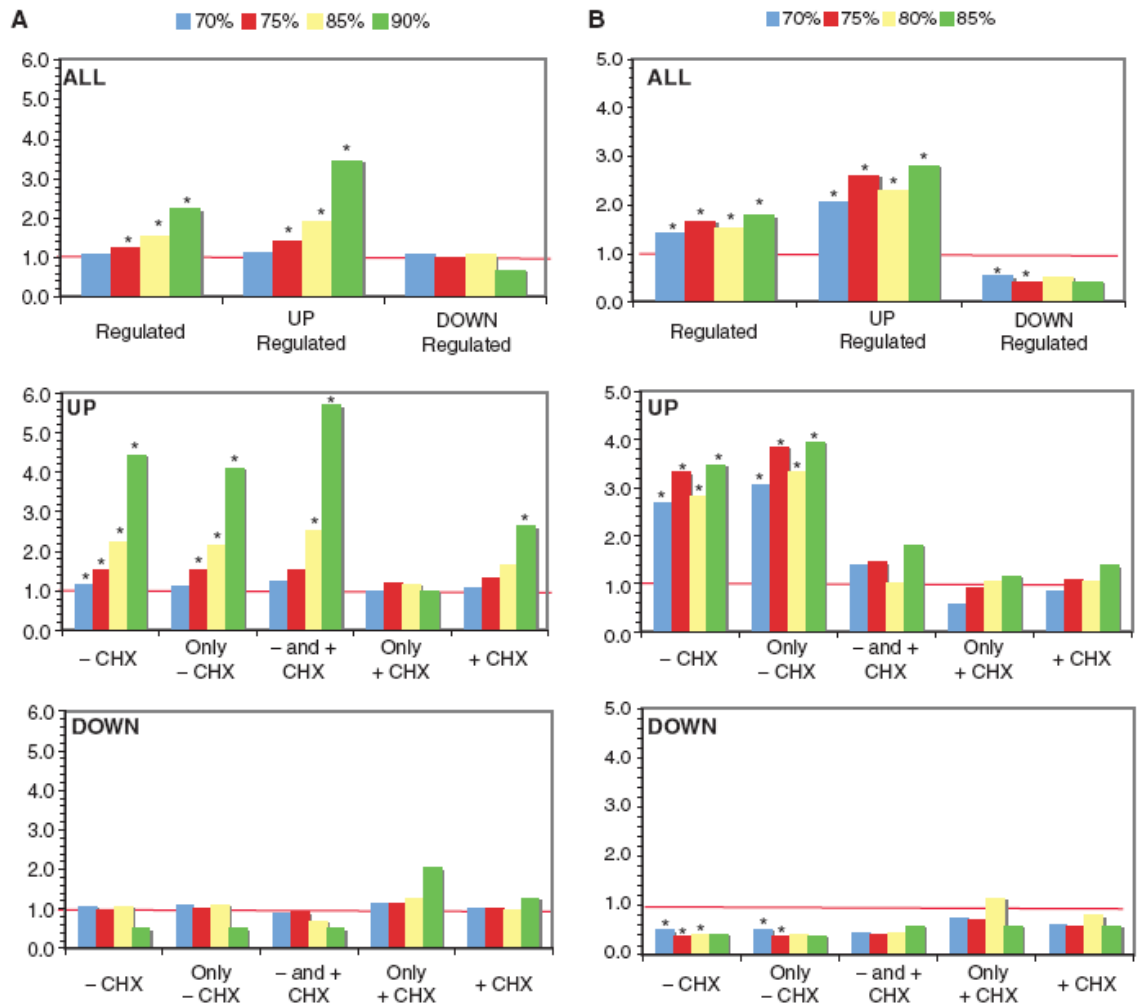


Figure 9.

Table 1. Top hundred genes regulated by E2 in the presence or absence of cycloheximide

Without CHX					With CHX						
RANK	Symbol	Fold	A	P-value	Validated	RANK	Symbol	Fold	A	P-value	Validated
1	MYBL1	13.85	7.23	4.07E-10	yes	1	KRT13	15.90	6.81	3.46E-13	-
2	FLJ30058	14.85	5.66	5.16E-12	-	2	LOC283551	15.04	5.81	3.26E-13	-
3	PKIB	7.09	10.22	5.01E-13	yes	3	FLJ10847	11.11	5.12	1.46E-16	-
4	SYTL4	5.79	7.86	4.39E-10	-	4	CYP26B1	7.46	7.05	5.91E-12	yes (DNS)
5	SGK3	5.99	6.93	2.40E-08	-	5	SGK3	6.89	6.93	1.86E-08	yes (DNS)
6	UGT2B15	6.19	6.45	3.14E-09	-	6	FLJ30058	7.60	5.66	4.93E-11	-
7	PCP4	5.03	7.28	1.14E-11	-	7	SLC26A2	5.40	7.85	1.22E-11	yes
8	TMPRSS3	5.42	6.72	9.36E-12	-	8	MYBL1	5.27	7.23	3.45E-07	yes
9	MCM10	4.96	6.97	8.87E-11	yes	9	RET	4.68	8.12	5.30E-12	yes
10	RRM2	3.44	10.03	9.07E-12	-	10	C8orf46	6.10	5.85	5.78E-11	-
11	MLF1IP	3.89	8.76	1.56E-11	-	11	RBP7	4.53	7.75	5.91E-12	yes
12	ATAD2	3.47	9.45	6.83E-10	-	12	RBM24	4.95	7.02	7.32E-12	-
13	DCC1	4.78	6.64	1.34E-10	-	13	TMEM64	3.26	10.13	3.50E-10	-
14	UHRF1	3.49	9.04	2.63E-11	-	14	CA12	3.43	9.42	5.66E-13	yes
15	TK1	3.58	8.80	4.97E-11	-	15	XBPI	3.13	10.27	4.50E-11	yes
16	IL17RB	4.24	7.38	1.87E-11	-	16	KCNK6	4.33	7.32	5.30E-12	-
17	GREB1	3.14	9.85	5.09E-11	yes	17	DEPDC6	3.75	8.35	8.37E-12	-
18	PBK	3.93	7.85	8.91E-10	-	18	IL17RB	4.08	7.33	2.12E-11	-
19	BRIP1	3.08	9.84	2.39E-07	-	19	OACT1	4.68	6.32	8.37E-12	-
20	TYMS	3.28	9.18	5.91E-11	-	20	CDH26	4.96	5.82	1.26E-07	yes
21	DTL	4.49	6.66	4.39E-10	-	21	CALCR	2.82	9.97	3.18E-11	yes
22	EXO1	4.54	6.58	9.07E-12	-	22	CELSR2	3.22	8.66	1.41E-12	yes (DNS)
23	LOC221981	-2.89	10.32	3.88E-10	-	23	KRT4	-3.91	7.01	1.53E-08	yes
24	ANLN	3.34	8.90	8.05E-09	-	24	PAPSS2	2.60	10.44	1.03E-10	-
25	CCNE2	3.69	7.89	1.95E-09	-	25	DSCAM	4.12	6.57	1.05E-10	yes
26	CDC2	3.10	9.34	1.98E-10	-	26	CARD10	4.04	6.59	1.59E-11	yes (DNS)
27	AREG	3.32	8.69	2.35E-11	yes	27	TPD52L1	2.55	10.14	5.69E-11	-
28	RLN2	3.38	8.36	1.11E-09	-	28	CXCL12	3.62	6.99	8.64E-10	-
29	MGP	3.79	7.37	4.63E-11	-	29	LRIG1	3.82	6.54	5.91E-12	yes
30	CXCL12	3.97	6.99	2.28E-10	-	30	RLN2	2.97	8.36	9.91E-09	-
31	CHEK1	4.18	6.61	7.62E-09	Yes (DNS)	31	ANKH	3.77	6.58	3.18E-11	-
32	TOP2A	2.97	9.24	1.60E-09	-	32	C14orf139	4.71	5.08	5.30E-12	-
33	KIAA0101	2.63	10.41	1.38E-10	-	33	C22orf19	2.90	8.23	2.37E-11	-
34	RERG	3.33	8.15	9.07E-12	-	34	RP13-360B22.2	2.88	8.21	4.39E-11	-
35	CYP26B1	3.85	7.05	6.70E-10	yes	35	PXK	3.59	6.53	4.33E-09	-
36	FSHPRH1	4.41	6.15	1.30E-10	-	36	TUBA3	2.18	10.62	3.25E-09	-
37	Pfs2	3.31	8.19	1.24E-10	-	37	MYB	2.62	8.66	5.91E-12	yes
38	BRCA1	4.15	6.49	2.17E-10	yes	38	C2orf23	3.04	7.43	5.91E-12	-
39	CDKN3	2.90	9.28	1.33E-09	-	39	UGT2B15	3.47	6.45	1.20E-06	-
40	SFXN2	3.15	8.51	1.87E-11	-	40	STC2	2.38	9.31	2.50E-11	-
41	FLJ10719	3.69	7.25	1.60E-09	-	41	P4HA2	2.62	8.29	5.59E-11	-
42	CDC45L	3.82	6.86	1.14E-11	yes	42	FLJ33718	2.73	7.92	1.03E-10	-
43	PCNA	2.61	10.01	2.54E-10	-	43	SMOX	3.20	6.72	2.81E-11	-
44	CDH2	-2.68	9.71	2.48E-11	-	44	GPSM1	2.62	8.13	8.27E-11	-
45	HEY2	3.46	7.51	2.48E-11	-	45	GREB1	2.15	9.85	1.62E-08	yes
46	C10orf3	3.47	7.47	4.06E-08	-	46	ITGA2	3.52	5.99	9.34E-10	-
47	G30	-2.74	9.40	1.21E-11	-	47	FKBP4	1.90	10.90	2.01E-09	-
48	GMNN	2.99	8.60	4.97E-11	-	48	PKIB	2.00	10.22	1.22E-07	yes
49	DEPDC6	3.05	8.35	6.81E-11	-	49	RAB27B	-2.33	8.70	1.78E-08	yes
50	CDC6	4.21	6.05	3.51E-09	-	50	KIAA1324	2.10	9.64	1.79E-10	-
51	KLK11	2.74	9.28	2.48E-11	-	51	PCP4	2.78	7.28	4.14E-09	-
52	FKSG14	4.34	5.82	1.10E-08	-	52	UGCG	2.64	7.56	2.44E-08	-
53	FHL1	2.93	8.63	4.39E-10	-	53	RHOD	2.03	9.83	1.11E-09	-
54	MAD2L1	3.14	8.04	6.55E-09	-	54	ANXA9	2.00	9.92	2.47E-09	-
55	PRIMI	3.22	7.84	4.63E-11	-	55	WFS1	2.43	8.16	7.02E-10	-
56	GLA	2.78	9.06	1.14E-11	-	56	SLC29A1	3.22	6.14	2.05E-10	-
57	CDT1	2.98	8.42	1.49E-11	-	57	SOX9	2.47	8.02	1.70E-09	-
58	MYBL2	3.15	7.89	9.57E-11	yes	58	PLEKHH1	2.88	6.87	8.27E-11	-
59	BCMPH1	3.27	7.57	2.37E-09	-	59	SH2BP1	2.33	8.45	9.91E-09	yes
60	HCAP-G	3.54	6.99	8.72E-10	-	60	WISP2	2.04	9.65	8.39E-10	yes
61	RAD51AP1	3.97	6.20	3.32E-08	-	61	SLC9A3R1	1.76	11.14	7.70E-09	-
62	CDH26	4.20	5.82	2.32E-07	yes	62	RERG	2.41	8.15	2.90E-10	Yes (DNS)
63	ZNF367	3.30	7.33	1.38E-10	-	63	IGFBP4	1.97	9.88	7.97E-09	yes
64	CDCA5	3.06	7.91	4.85E-11	-	64	CD33L3	2.82	6.89	4.17E-10	-
65	ASF1B	2.97	8.10	1.38E-11	-	65	RNF150	3.57	5.41	3.97E-10	-
66	UBE2C	2.38	10.08	4.02E-10	-	66	TGM2	2.97	6.48	1.85E-08	-

(continued)

Table 1. Continued

Without CHX						With CHX					
RANK	Symbol	Fold	A	<i>P</i> -value	Validated	RANK	Symbol	Fold	A	<i>P</i> -value	Validated
67	UBE2T	2.63	<b>9.11</b>	<b>2.48E-11</b>	-	67	SOX2	-2.21	8.69	1.83E-08	-
68	KNTC2	3.32	7.16	2.71E-07	-	68	SCARB1	2.27	<b>8.44</b>	<b>1.86E-10</b>	-
69	MELK	3.07	7.70	1.19E-10	-	69	ABHD2	2.44	7.85	1.10E-09	-
70	SPBC25	4.36	5.42	3.51E-09	-	70	BLVRB	1.97	9.71	8.39E-10	-
71	LOC440687	2.90	8.12	3.59E-09	-	71	IDH1	<b>-2.00</b>	<b>9.54</b>	<b>2.09E-10</b>	yes (DNS)
72	E2F8	3.33	7.09	9.46E-10	yes	72	SYNE2	2.31	<b>8.25</b>	<b>9.16E-09</b>	-
73	KIAA1212	-2.34	10.05	3.07E-11	-	73	SYTL5	1.98	<b>9.55</b>	<b>3.26E-09</b>	yes
74	RAB31	2.77	<b>8.48</b>	<b>3.01E-11</b>	-	74	COL18A1	2.34	8.02	2.34E-10	-
75	FANCD2	3.41	6.87	1.02E-09	-	75	ISG20	2.15	<b>8.65</b>	<b>2.53E-08</b>	-
76	KIF11	3.32	7.04	1.33E-07	-	76	SYTL4	2.36	7.86	1.06E-05	-
77	BRCA2	3.90	<b>5.99</b>	<b>1.48E-09</b>	yes	77	SEPT9	1.72	10.71	1.81E-08	-
78	MCM2	2.64	<b>8.84</b>	<b>7.18E-10</b>	yes	78	CCND1	1.83	9.95	2.55E-07	yes
79	HELLS	3.49	6.67	7.95E-11	-	79	SLC35E2	1.87	9.75	5.74E-09	-
80	MCM6	2.37	<b>9.79</b>	<b>1.10E-09</b>	yes	80	SLC25A24	1.72	10.58	2.16E-07	-
81	TTK	3.19	7.26	8.41E-09	-	81	MYO1B	<b>-1.69</b>	<b>10.73</b>	<b>6.67E-08</b>	yes (DNS)
82	E2F7	3.94	5.88	6.39E-09	yes	82	FER1L3	1.63	11.04	9.86E-08	-
83	PRC1	2.42	9.55	4.49E-10	-	83	ADAMTS19	<b>-2.08</b>	8.64	2.51E-10	-
84	RET	2.84	<b>8.12</b>	<b>4.49E-10</b>	yes	84	ARHGAP26	2.84	6.29	8.64E-10	-
85	PSF1	2.77	8.31	2.48E-11	-	85	KLF4	1.61	10.94	1.69E-07	-
86	XBP1	2.24	<b>10.27</b>	<b>2.59E-09</b>	yes	86	TPBG	1.55	11.39	4.87E-07	-
87	PLAC1	4.21	5.42	2.06E-11	-	87	HIPK2	2.13	8.25	2.16E-09	-
88	LOC441168	-2.11	10.79	2.54E-10	-	88	NPY1R	2.01	<b>8.71</b>	<b>2.56E-08</b>	-
89	CDCA1	3.51	6.46	1.78E-07	-	89	EFEMP1	-1.92	9.09	1.68E-08	-
90	HIST1H4C	2.29	9.88	6.50E-09	-	90	C6orf141	1.81	9.67	1.85E-08	-
91	RAD51	3.78	6.00	2.06E-11	-	91	TST	1.80	9.72	1.25E-08	-
92	KIF23	2.89	7.84	2.78E-07	-	92	MCM6	1.77	<b>9.79</b>	<b>6.11E-07</b>	yes
93	PLK4	3.32	6.79	9.88E-09	yes (DNS)	93	GLA	1.90	9.06	4.33E-09	-
94	FEN1	2.63	<b>8.56</b>	<b>1.04E-10</b>	yes	94	THRAP4	2.51	6.84	6.29E-09	-
95	EGR3	3.95	5.70	1.92E-09	yes (DNS)	95	RBBP8	2.06	<b>8.34</b>	<b>4.62E-07</b>	yes (DNS)
96	DUT	2.36	9.52	4.49E-10	-	96	FLJ20366	2.61	6.57	6.62E-09	-
97	OFCC1	-2.02	11.10	5.46E-10	-	97	LAPT4A	1.47	11.64	7.83E-07	-
98	POLE2	3.32	6.74	3.00E-10	yes	98	SFXN2	2.01	<b>8.51</b>	<b>1.39E-08</b>	-
99	RNASEH2A	2.60	8.61	4.82E-10	-	99	PGR	3.21	5.34	1.48E-07	yes
100	DHFR	2.86	7.78	3.23E-09	-	100	BRP44L	2.10	8.14	5.41E-09	-

Genes were ranked based on a combination of three parameters, fold-change between E2 and vehicle treatments, amplitude of array signals across all conditions and *P*-value based on moderated *t*-statistics. Genes in bold are those found significantly regulated both in the absence and presence of cycloheximide (CHX). Validated genes are those for which regulation was observed in Q-PCR with at least two independent experiments (see also Figures 3–5 and Supplementary Figure 2; DNS, data not shown; -, not tested).

## **SUPPLEMENTARY DATA**

**SUPPLEMENTARY FIGURE 1:** Effect of CHX treatment on expression of genes up-regulated only in the absence of CHX.

Levels of expression of the 924 transcripts in the group of genes up-regulated only in the absence of CHX in the microarray analysis are shown as a boxplot. The boxes represent the 25th to 75th percentiles, the median is marked by a blue dot and the bars represent the 5th and 95th percentiles.

**SUPPLEMENTARY FIGURE 2:** Validation of the estrogen regulation of selected genes at 24 h stimulation.

Validation by Q-PCR analysis of the effects of treatment with E2 (25 nM) and/or CHX (10 µg/ml) at 24 h on selected genes identified through microarray analysis. The data is representative of at least two independent experiments. A, fold regulation in array analysis; P, fold regulation in quantitative PCR analysis.

**SUPPLEMENTARY FIGURE 3:** Estradiol-induced proliferation of MCF-7 breast cancer cells is blocked by cycloheximide.

Cell cycle analysis was performed after a 24 h treatment with vehicle (0) or 17β-estradiol (E2, 25 nM). Cells were pre-treated with vehicle or CHX (10 µg/ml) 1 h before treatment. FACS analysis of propidium iodide stained-MCF-7 cells was performed as in Fig. 1.

**SUPPLEMENTARY FIGURE 4:** Enrichment in EREs in a 10 kb window around the TSS of primary target genes identified through several micrarray analyses.

The relative frequency of EREs in estrogen target genes identified in ref 50 (A) or 71 (B), in the overlap between these groups of genes and target genes identified in our study or in the overlap with primary target genes identified here was analyzed with respect to frequencies in the same windows in all genes in the genome as in Fig. 6. Stars indicate the significance of the enrichment according to both p-value and Z-score.

**SUPPLEMENTARY FIGURE 5:** Enrichment in EREs in a 10 kb window around the TSS of primary target genes identified through an independent microarray analysis in the presence of cycloheximide.

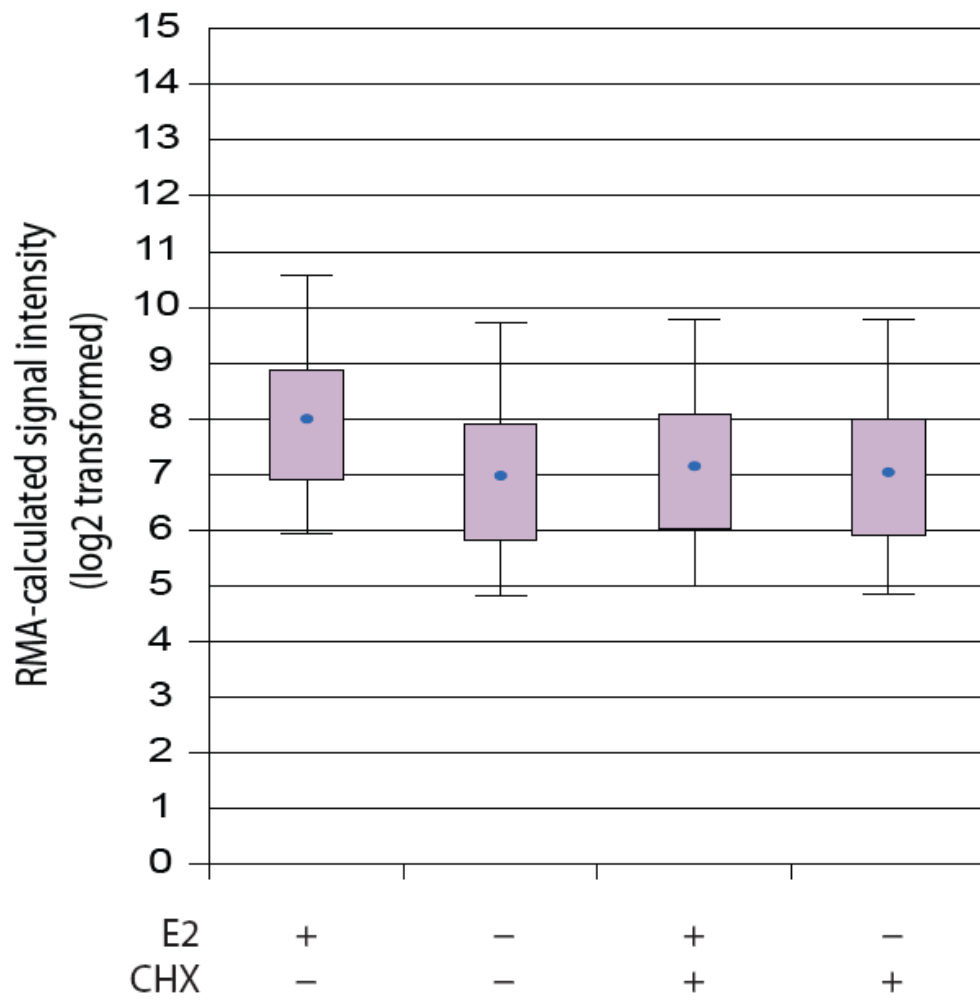
Analysis of enrichment in EREs in genes regulated by E2 at 48 h in the presence of CHX (dataset in ref 74) was performed as in Fig. 6. Stars indicate the significance of the enrichment according to both p-value and Z-score.

**SUPPLEMENTARY TABLE 1:** Transcription factor binding sites enriched in a 5 or 10 kb window around the transcriptional start sites of significantly regulated genes.

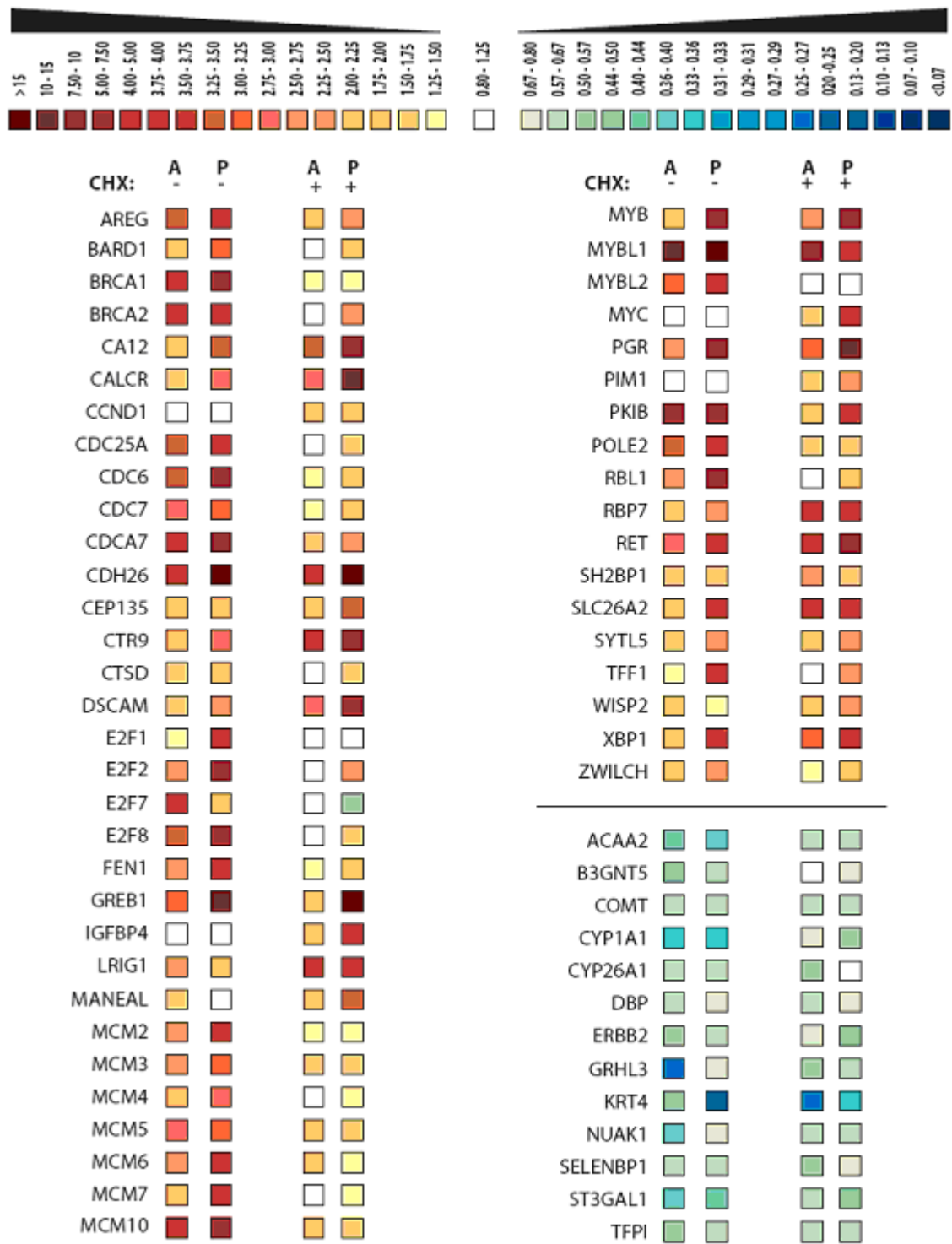
TRANSFAC matrices were used to screen for enrichment in the different categories of regulated genes with respect to their frequencies in all genes in the human genome within 5 kb (within +/- 2.5 kb from the TSS) or 10 kb (within +/- 5 kb of the TSS). Only matrices corresponding to sites enriched more than 1.5-fold with at least two different cut-offs and significant p-values and Z-scores were retained and shown.

**SUPPLEMENTARY TABLE 2:** Transcription factor binding sites enriched in a 5 or 10 kb window around the transcriptional start sites of the top 100 regulated genes in each category.

Screening with TRANSFAC matrices was performed as in supplementary Table 1, but using the top 100 regulated genes instead of all significantly regulated genes in each category.

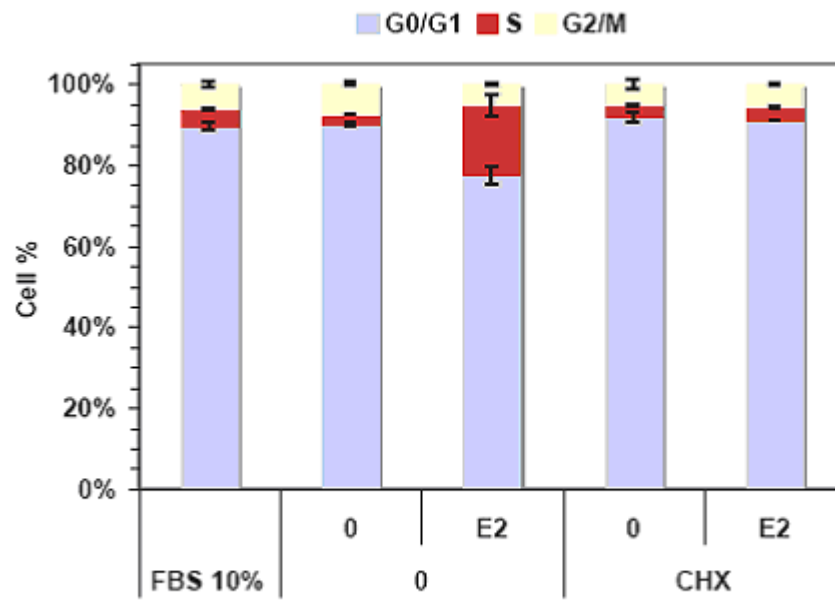


SUPPLEMENTARY FIGURE 1

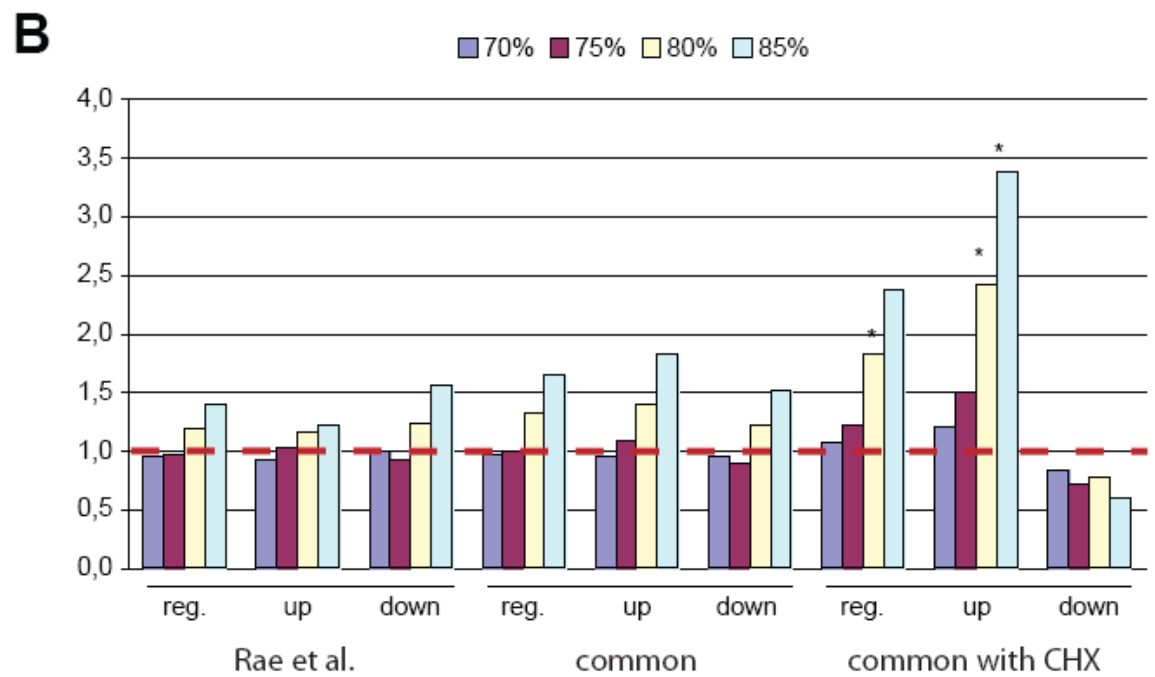
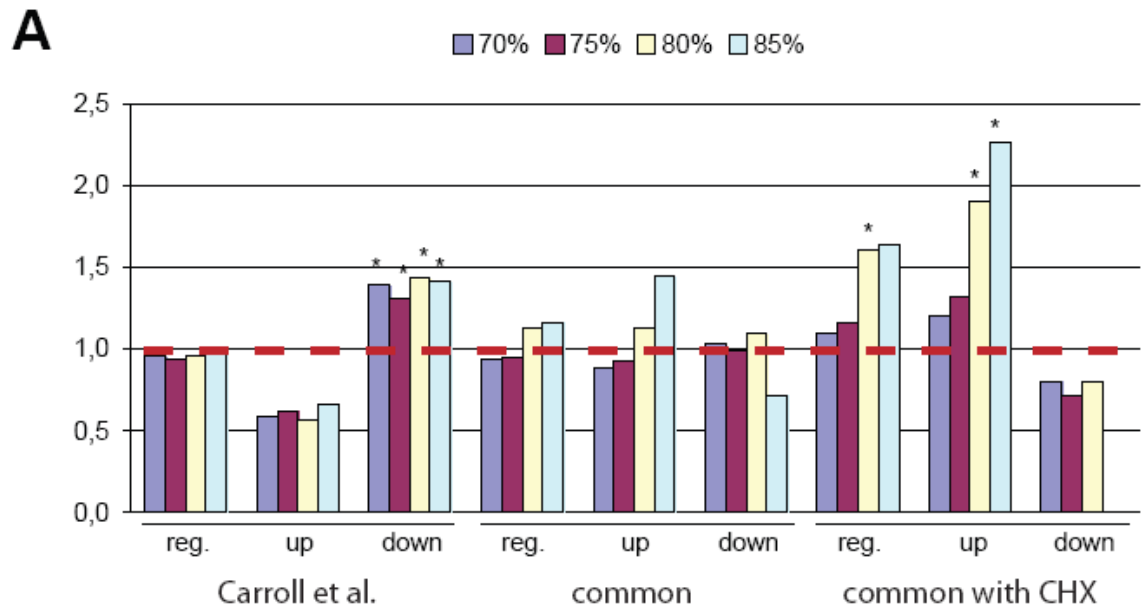


Supplementary Figure 2

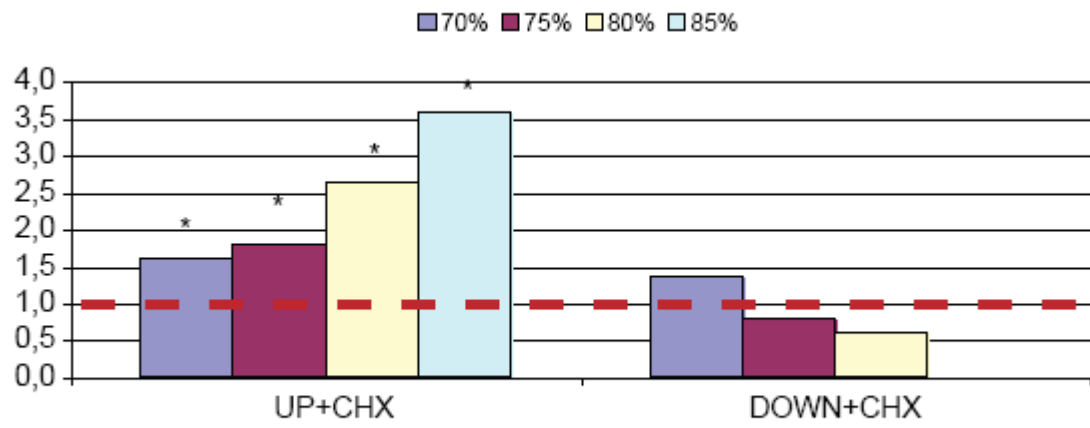




SUPPLEMENTARY FIGURE 3



SUPPLEMENTARY FIGURE 4



SUPPLEMENTARY FIGURE 5

**UP**

Site	Matrix	TRANSFAC	Promoter	+/- 5 kb TSS					+/- 2.5 kb TSS					
				- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	
AHR/ARNT	ARNT_01	M00236	-TE											1,54
CCAAT box	CAAT_01	M00254	-TE		1,70					1,67	2,04			
	NFY_01	M00287	-TE	1,83	2,01					1,93	2,10			
CHOP-C/EBP	NFY_Q6	M00185	conserved								1,58			
	CHOP_01	M00249	-TE											1,65
E2F	E2F_03	M00516		3,10	3,12	3,02				3,64	3,72	3,32		1,64
	E2F_03 core	M00516	conserved	4,44	4,09	5,70				5,48	5,05	7,04		3,11
EGR	EGR1_01	M00243	-TE									2,97		2,06
	EGR2_01	M00246	-TE					1,66				2,87		1,79
	EGR3_01	M00245	-TE					2,07				3,63		2,26
	NGFIC_01	M00244	-TE									3,02		2,26
ERE	ER_Q6	M00191	-TE				3,13	2,50					3,84	2,93
	ERE	-					2,53	2,57					3,58	3,25
ERE-derived	ER0	-	-TE				1,69	1,62						
	ER5	-	-TE				1,84							2,34
GC box	GC_01	M00255	-TE											1,52
homeobox	CHX10_01	M00437	-TE			2,16								

**DOWN**

Site	Matrix	TRANSFAC	Promoter	+/- 5 kb TSS					+/- 2.5 kb TSS					
				- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	
AP1	AP1_Q4	M00188	-TE								1,90			
	AP1_Q6	M00174	conserved							1,62	1,59			
C/EBP	CEBPB_02	M00117	conserved							2,67	3,12			
CCAAT box	NFY_Q6	M00185	conserved								1,53			
	FOXJ2_01	M00130	conserved		1,69									
Forkhead	FOXJ2_01	M00422	conserved								1,53			
	FOXO4_01	M00472	conserved							1,54				
	FREAC3_01	M00291	-TE									2,11		2,07
	HFH8_01	M00294	conserved	1,52	1,50						1,52			
GC box	GC_01	M00255	-TE							1,68				
G-rich	MAZR_01	M00491	-TE								1,85			
	SPZ1_01	M00446	-TE										5,79	
homeobox	CHX10_01	M00437	-TE											2,23
NFkB	NFKAPPAB_01	M00054	-TE		1,71									
	NFKAPPAB65_01	M00052	-TE		1,66									
PAX	PAX4_02	M00377	-TE											1,70
	PAX4_03	M00378	conserved							2,00				
SP1	SP1_Q6	M00196	conserved		1,52					1,54	1,58			

Supplementary Table 1

## UP TOP 100

Site	Matrix	TRANSFAC	Promoter	+/- 5 kb TSS					+/- 2.5 kb TSS				
				- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX
AP1	AP1_01	M00517					2,76						
cAMP-responsive element	CREB_01	M00039	-TE	2,13						2,47	2,47		
	CREB_Q4	M00178									1,89		
CCAAT box	CAAT_01	M00254	-TE							3,42	2,73		
	NF1_Q6	M00193										1,70	
	NFY_01	M00287	-TE	2,69	3,13					3,01	3,27		
	NFY_Q6	M00185	conserved		2,66					2,20			
CHOP-C/EBP	CHOP_01	M00249	-TE				3,71						5,01
E2F	E2F_03	M00516		7,45	5,68					9,36	7,13		
	E2F_03 core	M00516	conserved	12,50	8,33					15,44	10,29		
EGR	EGR1_01	M00243	-TE			1,81						1,95	
	EGR3_01	M00245	-TE					2,30					
	NGFIC_01	M00244	-TE					2,88				1,99	
ERE	ER_Q6	M00191	-TE				8,67					2,39	
	ERE	-				4,00		5,00					4,91 5,89
ERE-derived	ERE_DR5	-					2,73						
	ERE_ER0	-					1,86						
	ERE_ER5	-	-TE				3,24						4,69
	ERE_IR4	-					1,81						
Forkhead homeobox	HNF3B_01	M00131			1,55								
	CHX10_01	M00437	-TE			2,39							
Maf-recognition elements	BACH1_01	M00495					2,16						
PAR	HLF_01	M00260	-TE	3,36									
	VBP_01	M00228	-TE	3,70									

## DOWN TOP 100

Site	Matrix	TRANSFAC	Promoter	+/- 5 kb TSS					+/- 2.5 kb TSS				
				- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX	- CHX	Only - CHX	+ and - CHX	Only + CHX	+ CHX
C/EBP	CEBPB_01	M00109	conserved		1,86								
cAMP-responsive element	ATF_01	M00017	-TE		2,24						2,93		
	CREB_Q4	M00178									3,30		
CCAAT box	NFY_Q6	M00185	conserved		2,50						2,78		
E2F	E2F_03	M00516		2,30	3,55					2,46	3,57		
ERE-derived	ERE_DR6	-	-TE	1,80						2,33	1,91		
EVI	EV11_06	M00011	conserved										2,21
	FOXA	-	conserved							1,75	2,11		
Forkhead	FOXO1_01	M00473	conserved		2,80								
	FOXO4_01	M00472	conserved	2,41	3,86					3,27	4,08		
GATA	GATA1_06	M00347	conserved		1,93								
G-rich	SPZ1_01	M00446										5,51	
homeobox	CHX10_01	M00437	-TE					3,17					3,82
Ikars	IK1_01	M00086	-TE								2,81		
SOX	SRY_02	M00160	conserved					2,53					2,58
SP1	SP1_01	M00008	conserved		1,82					2,02	1,95		

Supplementary Table 2

## REFERENCES

1. Couse, J.F. and Korach, K.S. (1999) Estrogen receptor null mice: what have we learned and where will they lead us? *Endocrine Rev.*, 20, 358–417.
2. McEwen, B.S. and Alves, S.E. (1999) Estrogen actions in the central nervous system. *Endocrine Rev.*, 20, 279–307.
3. Jordan, V.C. (2001) Estrogen, selective estrogen receptor modulation, and coronary heart disease: something or nothing. *J. Natl. Canc. Inst.*, 93, 2–4.
4. Ali, S. and Coombes, R.C. (2002) Endocrine-responsive breast cancer and strategies for combating resistance. *Nat. Rev. Cancer*, 2, 101–112.
5. Simpson, E.R. (2003) Sources of estrogen and their importance. *J. Steroid Biochem. Mol. Biol.*, 86, 225–230.
6. Nilsson, S., Makela, S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G., Enmark, E., Pettersson, K., Warner, M. et al. (2001) Mechanisms of estrogen action. *Physiol. Rev.*, 81, 1535–1565.
7. Robinson-Rechavi, M., Escriva Garcia, H. and Laudet, V. (2003) The nuclear receptor superfamily. *J. Cell Sci.*, 116, 585–586.
8. Kishimoto, M., Fujiki, R., Takezawa, S., Sasaki, Y., Nakamura, T., Yamaoka, K., Kitagawa, H. and Kato, S. (2006) Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications. *Endocr. J.*, 53, 157–172.
9. Metivier, R., Reid, G. and Gannon, F. (2006) Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.*, 7, 161–167.
10. Mader, S., Chambon, P. and White, J.H. (1993) Defining a minimal estrogen receptor DNA binding domain. *Nucleic Acids Res.*, 21, 1125–1132.
11. Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, 75, 567–578.
12. Walker, P., Germond, J.-E., Brown-Luedi, M., Givel, F. and Wahli, W. (1984) Sequence homologies in the region preceding the transcription initiation site of the liver estrogen-responsive vitellogenin and apo-VLDLII gene. *Nucleic Acids Res.*, 12, 8611–8626.

13. Klein-Hitpass,L., Ryffel,G.U., Heitlinger,E. and Cato,A.C. (1988) A 13 bp palindrome is a functional estrogen responsive element and interacts specifically with estrogen receptor. *Nucleic Acids Res.*, 16, 647–663.
14. Klinge,C.M. (2001) Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.*, 29, 2905–2919.
15. Sanchez,R., Nguyen,D., Rocha,W., White,J.H. and Mader,S. (2002) Diversity in the mechanisms of gene regulation by estrogen receptors. *Bioessays*, 24, 244–254.
16. Driscoll,M.D., Sathya,G., Muyan,M., Klinge,C.M., Hilf,R. and Bambara,R.A. (1998) Sequence requirements for estrogen receptor binding to estrogen response elements. *J. Biol. Chem.*, 273, 29321–29330.
17. Kulakosky,P.C., McCarty,M.A., Jernigan,S.C., Risinger,K.E. and Klinge,C.M. (2002) Response element sequence modulates estrogen receptor alpha and beta affinity and activity. *J. Mol. Endocrinol.*, 29, 137–152.
18. Nguyen,D., Bail,M., Pesant,G., Dupont,V.N., Rouault,E., Deschenes,J., Rocha,W., Melancon,G., Steinberg,S.V. et al. (2007) Rational design of an estrogen receptor mutant with altered DNA-binding specificity. *Nucleic Acids Res.*, 35, 3465–3477.
19. Bai,Y. and Giguere,V. (2003) Isoform-selective interactions between estrogen receptors and steroid receptor coactivators promoted by estradiol and ErbB-2 signaling in living cells. *Mol. Endocrinol.*, 17, 589–599.
20. Gustafsson,J.A. (2006) ERbeta scientific visions translate to clinical uses. *Climacteric*, 9, 156–160.
21. Leo,C. and Chen,J.D. (2000) The SRC family of nuclear receptor coactivators. *Gene*, 245, 1–11.
22. Rosenfeld,M.G. and Glass,C.K. (2001) Coregulator codes of transcriptional regulation by nuclear receptors. *J. Biol. Chem.*, 276, 36865–36868.
23. Rachez,C. and Freedman,L.P. (2001) Mediator complexes and transcription. *Curr. Opin. Cell Biol.*, 13, 274–280.
24. Dilworth,F.J. and Chambon,P. (2001) Nuclear receptors coordinate the activities of chromatin remodeling complexes and coactivators to facilitate initiation of transcription. *Oncogene*, 20, 3047–3054.
25. McDonnell,D.P. and Norris,J.D. (2002) Connections and regulation of the human estrogen receptor. *Science*, 296, 1642–1644.
26. McKenna,N.J. and O'Malley,B.W. (2002) Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell*, 108, 465–474.

27. Belandia,B. and Parker,M.G. (2003) Nuclear receptors: a rendezvous for chromatin remodeling factors. *Cell*, 114, 277–280.
28. Forster,C., Makela,S., Warri,A., Kietz,S., Becker,D., Hultenby,K., Warner,M. and Gustafsson,J.A. (2002) Involvement of estrogen receptor beta in terminal differentiation of mammary gland epithelium. *Proc. Natl Acad. Sci. USA*, 99, 15578–15583.
29. Leygue,E., Dotzlaw,H., Watson,P.H. and Murphy,L.C. (1998) Altered estrogen receptor alpha and beta messenger RNA expression during human breast tumorigenesis. *Cancer Res.*, 58, 3197–3201.
30. Roger,P., Sahla,M.E., Makela,S., Gustafsson,J.A., Baldet,P. and Rochefort,H. (2001) Decreased expression of estrogen receptor beta protein in proliferative preinvasive mammary tumors. *Cancer Res.*, 61, 2537–2541.
31. Kushner,P.J., Agard,D.A., Greene,G.L., Scanlan,T.S., Shiau,A.K., Uht,R.M. and Webb,P. (2000) Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.*, 74, 311–317.
32. Safe,S. (2001) Transcriptional activation of genes by 17 betaestradiol through estrogen receptor-Sp1 interactions. *Vitam. Horm.*, 62, 231–252.
33. Kalaitzidis,D. and Gilmore,T.D. (2005) Transcription factor crosstalk: the estrogen receptor and NF-kappaB. *Trends Endocrinol. Metab.*, 16, 46–52.
34. Migliaccio,A., Di Domenico,M., Castoria,G., de Falco,A., Bontempo,P., Nola,E. and Auricchio,F. (1996) Tyrosine kinase/ p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J.*, 15, 1292–1300.
35. Sun,M., Paciga,J.E., Feldman,R.I., Yuan,Z., Coppola,D., Lu,Y.Y., Shelley,S.A., Nicosia,S.V. and Cheng,J.Q. (2001) Phosphatidylinositol-3-OH Kinase (PI3K)/AKT2, activated in breast cancer, regulates and is induced by estrogen receptor alpha (ERalpha) via interaction between ERalpha and PI3K. *Cancer Res.*, 61, 5985–5991.
36. Pedram,A., Razandi,M., Aitkenhead,M., Hughes,C.C. and Levin,E.R. (2002) Integration of the non-genomic and genomic actions of estrogen. Membrane-initiated signaling by steroid to transcription and cell biology. *J. Biol. Chem.*, 277, 50768–50775.
37. Simoncini,T., Fornari,L., Mannella,P., Varone,G., Caruso,A., Liao,J.K. and Genazzani,A.R. (2002) Novel non-transcriptional mechanisms for estrogen receptor signaling in the cardiovascular system. Interaction of estrogen receptor alpha with phosphatidylinositol 3-OH kinase. *Steroids*, 67, 935–939.
38. Pedram,A., Razandi,M. and Levin,E.R. (2006) Nature of functional estrogen receptors at the plasma membrane. *Mol. Endocrinol.*, 20, 1996–2009.



39. Filardo,E.J., Quinn,J.A., Bland,K.I. and Frackelton,A.R.,Jr. (2000) Estrogen-induced activation of Erk-1 and Erk-2 requires the G protein-coupled receptor homolog, GPR30, and occurs via transactivation of the epidermal growth factor receptor through release of HB-EGF. *Mol. Endocrinol.*, 14, 1649–1660.
40. Maggiolini,M., Vivacqua,A., Fasanella,G., Recchia,A.G., Sisci,D., Pezzi,V., Montanaro,D., Musti,A.M., Picard,D. et al. (2004) The G protein-coupled receptor GPR30 mediates c-fos up-regulation by 17beta-estradiol and phytoestrogens in breast cancer cells. *J. Biol. Chem.*, 279, 27008–27016.
41. Revankar,C.M., Cimino,D.F., Sklar,L.A., Arterburn,J.B. and Prossnitz,E.R. (2005) A transmembrane intracellular estrogen receptor mediates rapid cell signaling. *Science*, 307, 1625–1630.
42. Thomas,P., Pang,Y., Filardo,E.J. and Dong,J. (2005) Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology*, 146, 624–632.
43. Funakoshi,T., Yanai,A., Shinoda,K., Kawano,M.M. and Mizukami,Y. (2006) G protein-coupled receptor 30 is an estrogen receptor in the plasma membrane. *Biochem. Biophys. Res. Commun.*, 346, 904–910.
44. Soulez,M. and Parker,M.G. (2001) Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. *J. Mol. Endocrinol.*, 27, 259–274.
45. Frasor,J., Danes,J.M., Komm,B., Chang,K.C., Lyttle,C.R. and Katzenellenbogen,B.S. (2003) Profiling of estrogen up- and downregulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology*, 144, 4562–4574.
46. Lin,C.Y., Strom,A., Vega,V.B., Kong,S.L., Yeo,A.L., Thomsen,J.S., Chan,W.C., Doray,B., Bangarusamy,D.K. et al. (2004) Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol.*, 5, R66.
47. Cicatiello,L., Scafoglio,C., Altucci,L., Cancemi,M., Natoli,G., Facchiano,A., Iazzetti,G., Calogero,R., Biglia,N. et al. (2004) A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive transcriptome. *J. Mol. Endocrinol.*, 32, 719–775.
48. Carroll,J.S., Liu,X.S., Brodsky,A.S., Li,W., Meyer,C.A., Szary,A.J., Eeckhoute,J., Shao,W., Hestermann,E.V. et al. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122, 33–43.
49. Vega,V.B., Lin,C.Y., Lai,K.S., Kong,S.L., Xie,M., Su,X., Teh,H.F., Thomsen,J.S., Yeo,A.L. et al. (2006) Multiplatform genome-wide identification and modeling of functional human estrogen receptor binding sites. *Genome Biol.*, 7, R82.

50. Carroll,J.S., Meyer,C.A., Song,J., Li,W., Geistlinger,T.R., Eeckhoute,J., Brodsky,A.S., Keeton,E.K., Fertuck,K.C. et al. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.*, 38, 1289–1297.
51. Bourdeau,V., Deschenes,J., Metivier,R., Nagai,Y., Nguyen,D., Bretschneider,N., Gannon,F., White,J.H. and Mader,S. (2004) Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol. Endocrinol.*, 18, 1411–1427.
52. Deschenes,J., Bourdeau,V., White,J.H. and Mader,S. (2007) Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 KB of upstream flanking sequences. *J. Biol. Chem.*, 282, 17335–17339.
53. Rocha,W., Sanchez,R., Deschenes,J., Auger,A., Hebert,E., White,J.H. and Mader,S. (2005) Opposite effects of histone deacetylase inhibitors on glucocorticoid and estrogen signaling in human endometrial Ishikawa cells. *Mol. Pharmacol.*, 68, 1852–1862.
54. Dayan,G., Lupien,M., Auger,A., Anghel,S.I., Rocha,W., Croisette,S., Katzenellenbogen,J.A. and Mader,S. (2006) Tamoxifen and raloxifene differ in their functional interactions with aspartate 351 of estrogen receptor alpha. *Mol. Pharmacol.*, 70, 579–588.
55. Wettenhall,J.M., Simpson,K.M., Satterley,K. and Smyth,G.K. (2006) affyLmGUI: a graphical user interface for linear modeling of single channel microarray data. *Bioinformatics*, 22, 897–899.
56. Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, 31, e15.
57. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, 3, Article3.
58. Karolchik,D., Baertsch,R., Diekhans,M., Furey,T.S., Hinrichs,A., Lu,Y.T., Roskin,K.M., Schwartz,M., Sugnet,C.W. et al. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, 31, 51–54.
59. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, 33, D501–D504.
60. Wasserman,W.W. and Sandelin,A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.*, 5, 276–287.
61. Wang,T.T., Tavera-Mendoza,L.E., Laperriere,D., Libby,E., MacLeod,N.B., Nagai,Y., Bourdeau,V., Konstorum,A., Lallemand,B. et al. (2005) Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin D3 target genes. *Mol. Endocrinol.*, 19, 2685–2695.

62. Laperriere,D., Wang,T.T., White,J.H. and Mader,S. (2007) Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genomics*, 8, 23.
63. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M. et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, 34, D108–D110.
64. Jin,V.X., Sun,H., Pohar,T.T., Liyanarachchi,S., Palaniswamy,S.K., Huang,T.H. and Davuluri,R.V. (2005) ERTargetDB: an integral information resource of transcription regulation of estrogen receptor target genes. *J. Mol. Endocrinol.*, 35, 225–230.
65. O’Lone,R., Frith,M.C., Karlsson,E.K. and Hansen,U. (2004) Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol.*, 18, 1859–1875.
66. Balmer,J.E. and Blomhoff,R. (2005) A robust characterization of retinoic acid response elements based on a comparison of sites in three species. *J. Steroid Biochem. Mol. Biol.*, 96, 347–354.
67. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W. et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, 15, 1034–1050.
68. Ho Sui,S.J., Mortimer,J.R., Arenillas,D.J., Brumm,J., Walsh,C.J., Kennedy,B.P. and Wasserman,W.W. (2005) oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res.*, 33, 3154–3164.
69. Jin,V.X., Rabinovich,A., Squazzo,S.L., Green,R. and Farnham,P.J. (2006) A computational genomics approach to identify cisregulatory modules from chromatin immunoprecipitation microarray data – a case study using E2F1. *Genome Res.*, 16, 1585–1595.
70. Creighton,C.J., Cordero,K.E., Larios,J.M., Miller,R.S., Johnson,M.D., Chinnaiyan,A.M., Lippman,M.E. and Rae,J.M. (2006) Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors. *Genome Biol.*, 7, R28.
71. Rae,J.M., Johnson,M.D., Scheys,J.O., Cordero,K.E., Larios,J.M. and Lippman,M.E. (2005) GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res. Treat.*, 92, 141–149.
72. van der Burg,B., van Selm-Miltenburg,A.J., de Laat,S.W. and van Zoelen,E.J. (1989) Direct effects of estrogen on c-fos and c-myc protooncogene expression and cellular proliferation in human breast cancer cells. *Mol. Cell. Endocrinol.*, 64, 223–228.
73. Gutierrez,G.J. and Ronai,Z. (2006) Ubiquitin and SUMO systems in the regulation of mitotic checkpoints. *Trends Biochem. Sci.*, 31, 324–332.

74. Scholtens,D., Miron,A., Merchant,F.M., Miller,A., Miron,P.L., Iglehart,J.D. and Gentleman,R. (2004) Analyzing factorial designed microarray experiments. *J. Multivariate Anal.*, 90, 19–43.
75. Blobel,G.A., Sieff,C.A. and Orkin,S.H. (1995) Ligand-dependent repression of the erythroid transcription factor GATA-1 by the estrogen receptor. *Mol. Cell. Biol.*, 15, 3147–3153.
76. Holth,L.T., Sun,J.M., Coutts,A.S., Murphy,L.C. and Davie,J.R. (1997) Estrogen receptor diminishes DNA-binding activities of chicken GATA-1 and CACCC-binding proteins. *DNA Cell Biol.*, 16, 1477–1482.
77. Prall,O.W., Rogan,E.M., Musgrove,E.A., Watts,C.K. and Sutherland,R.L. (1998) c-Myc or cyclin D1 mimics estrogen effects on cyclin E-Cdk2 activation and cell cycle reentry. *Mol. Cell. Biol.*, 18, 4499–4508.
78. Doisneau-Sixou,S.F., Sergio,C.M., Carroll,J.S., Hui,R., Musgrove,E.A. and Sutherland,R.L. (2003) Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells. *Endocr. Relat. Cancer*, 10, 179–186.
79. Zhao,F., Xuan,Z., Liu,L. and Zhang,M.Q. (2005) TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.*, 33, D103–D107.
80. Cappellen,D., Schlange,T., Bauer,M., Maurer,F. and Hynes,N.E. (2007) Novel c-MYC target genes mediate differential effects on cell proliferation and migration. *EMBO Rep.*, 8, 70–76.
81. Rodrik,V., Gomes,E., Hui,L., Rockwell,P. and Foster,D.A. (2006) Myc stabilization in response to estrogen and phospholipase D in MCF-7 breast cancer cells. *FEBS Lett.*, 580, 5647–5652.
82. Foster,J.S., Henley,D.C., Ahamed,S. and Wimalasena,J. (2001) Estrogens and cell-cycle regulation in breast cancer. *Trends Endocrinol. Metab.*, 12, 320–327.
83. Otsuki,Y., Tanaka,M., Kamo,T., Kitanaka,C., Kuchino,Y. and Sugimura,H. (2003) Guanine nucleotide exchange factor, Tiam1, directly binds to c-Myc and interferes with c-Myc-mediated apoptosis in rat-1 fibroblasts. *J. Biol. Chem.*, 278, 5132–5140.
84. Hirai,H., Izutsu,K., Kurokawa,M. and Mitani,K. (2001) Oncogenic mechanisms of Evi-1 protein. *Cancer Chemother. Pharmacol.*, 48 (Suppl. 1), S35–S40.
85. Koipally,J., Renold,A., Kim,J. and Georgopoulos,K. (1999) Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. *EMBO J.*, 18, 3090–3100.
86. Brunnberg,S., Pettersson,K., Rydin,E., Matthews,J., Hanberg,A. and Pongratz,I. (2003) The basic helix-loop-helix-PAS protein ARNT functions as a potent coactivator of estrogen receptor-dependent transcription. *Proc. Natl Acad. Sci. USA*, 100, 6517–6522.

87. Ohtake,F., Takeyama,K., Matsumoto,T., Kitagawa,H., Yamamoto,Y., Nohara,K., Tohyama,C., Krust,A., Mimura,J. et al. (2003) Modulation of oestrogen receptor signalling by association with the activated dioxin receptor. *Nature*, 423, 545–550.
88. Sukhatme,V.P. (1990) Early transcriptional events in cell growth: the Egr family. *J. Am. Soc. Nephrol.*, 1, 859–866.
89. Khachigian,L.M. and Collins,T. (1998) Early growth response factor 1: a pleiotropic mediator of inducible gene expression. *J. Mol. Med.*, 76, 613–616.
90. Chen,C.C., Lee,W.R. and Safe,S. (2004) Egr-1 is activated by 17beta-estradiol in MCF-7 cells by mitogen-activated protein kinase-dependent phosphorylation of ELK-1. *J. Cell. Biochem.*, 93, 1063–1074.
91. Kato,K., Toki,T., Shimizu,M., Shiozawa,T., Fujii,S., Nikaido,T. and Konishi,I. (2003) Expression of replication-licensing factors MCM2 and MCM3 in normal, hyperplastic, and carcinomatous endometrium: correlation with expression of Ki-67 and estrogen and progesterone receptors. *Int. J. Gynecol. Pathol.*, 22, 334–340.
92. Di Stefano,L., Jensen,M.R. and Helin,K. (2003) E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. *EMBO J.*, 22, 6289–6298.
93. Christensen,J., Cloos,P., Toftegaard,U., Klinkenberg,D., Bracken,A.P., Trinh,E., Heeran,M., Di Stefano,L. and Helin,K. (2005) Characterization of E2F8, a novel E2F-like cell-cycle regulated repressor of E2F-activated transcription. *Nucleic Acids Res.*, 33, 5458–5470.
94. Trimarchi,J.M., Fairchild,B., Verona,R., Moberg,K., Andon,N. and Lees,J.A. (1998) E2F-6, a member of the E2F family that can behave as a transcriptional repressor. *Proc. Natl Acad. Sci. USA*, 95, 2850–2855.
95. Cartwright,P., Muller,H., Wagener,C., Holm,K. and Helin,K. (1998) E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene*, 17, 611–623.
96. Ogawa,H., Ishiguro,K., Gaubatz,S., Livingston,D.M. and Nakatani,Y. (2002) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science*, 296, 1132–1136.

## Chapitre 8 : Discussion et perspectives

### 8.1 L'incorporation de caractéristiques des sites de haute affinité de ER et RAR améliore les modèles matriciels

Le modèle matriciel permet de prédire *in silico* des sites de fixation d'un facteur de transcription à l'aide d'un score approximant l'énergie de liaison ADN-protéine [149]. Étant donné que les sites prédits sont sélectionnés sur la base de ce score, il est impératif que la matrice utilisée représente le plus fidèlement possible l'affinité du facteur pour les sites qu'il reconnaît.

La plupart des recherches utilisant la représentation par matrice des sites de fixation emploient celles des banques de données JASPAR, MatBase ou TRANSFAC [155-157]. Certains facteurs de transcription, comme RAR, sont absents de la plupart des banques de matrices. Seule la banque MatBase a, pour le moment, des matrices modélisant les RARE [170]. Pour d'autres facteurs, comme ER, ces banques de données contiennent des matrices différentes pour le même facteur. En raison de cette variabilité, pour les études ciblant un facteur particulier il vaut la peine d'évaluer si les matrices proposées concordent avec la description des sites modélisés prévalants dans la littérature.

#### 8.1.1 Meilleure détection des ERE à l'aide d'une matrice palindromique

Les sites de haute affinité reconnus par ER sont généralement décrits comme des palindromes composés de deux demi sites espacés de trois nucléotides [171]. Plusieurs matrices bâties à partir de compilation de sites validés expérimentalement ont été proposées dans la littérature [155, 157, 165, 166]. La plupart de ces matrices contiennent moins de trente ERE et présentent des divergences par rapport à la définition usuelle d'un ERE. La matrice de la banque de données TRANSFAC illustre bien ces différences (Chapitre 4,

Figure 23). En effet, les deux demi sites ne forment pas vraiment un palindrome et une position de l'espacement présente une préférence pour un nucléotide.

Dans le cadre du deuxième article, j'ai construit une nouvelle matrice à l'aide de 76 ERE validés dans la littérature (Chapitre 7, Figure 6 A). Afin de renforcer le caractère palindromique des sites reconnus par ER, j'ai additionné la matrice de comptage des 76 ERE à la matrice de comptage de leur séquence complémentaire inverse. La fréquence des nucléotides des positions entre les demi sites a aussi été corrigée pour obtenir 25% pour A, C, G et T afin de minimiser leur contribution au score.

Cette nouvelle matrice donne des scores plus élevés aux ERE de haute affinité que la matrice de TRANSFAC (Annexe, Figure 28). Les palindromes de (A/G)GGTCA obtiennent des scores normalisés entre 90% et 100% avec la nouvelle matrice alors qu'ils obtiennent des scores entre 65% et 85% avec celle de TRANSFAC. L'amélioration des scores de la nouvelle matrice ne se limite pas aux ERE de haute affinité. Les formes avec un nucléotide différent des ERE de haute affinité et la plupart des formes avec deux nucléotides différents ont aussi des scores plus élevés. Bien que les deux matrices permettent d'observer un enrichissement statistiquement significatif de ERE dans les gènes induits en présence d'œstrogènes et de cycloheximide (Chapitre 7, Table supplémentaire 1), la nouvelle matrice trouve en général des ERE dans les promoteurs de plus de gènes. Par exemple, avec un seuil sur le score de 70%, la nouvelle matrice trouve des ERE situés à 5 kbp du TSS de 116 des 356 gènes induits seulement en présence de E2 et cycloheximide (up only E2+CHX) alors que la matrice de TRANSFAC trouve des ERE dans les promoteurs de 84 gènes.

Les deux matrices diffèrent également dans leurs prédictions avec un ensemble de ERE fixés par ER et de séquences similaires à des ERE mais non fixés par ER établi par Vega et collègues[143]. Cet ensemble contient 64 ERE fixés par ER et 68 ERE non fixés par ER. La classification entre fixé et non fixé fait suite à des expériences de retard sur gel, d'immunoprécipitation de la chromatine ou de transfection transitoire à l'aide de gènes

rapporteurs. Il est à noter cependant qu'il est possible que certains ERE non fixés dans les expériences de ChIP soient fixés dans d'autres contextes cellulaires [172]. Sur cet ensemble de ERE, la matrice ERE construite pour le deuxième article a une meilleure sensibilité que la matrice de TRANSFAC car elle détecte en général plus de ERE fixés (Annexe, Figure 29). Avec un seuil de 65% sur le score, le seuil de base utilisé pour la prédiction de ERE dans le deuxième article, la sensibilité de la nouvelle matrice est de 68% contre 50% pour la matrice de TRANSFAC. Par contre, au même seuil, la nouvelle matrice détecte plus de ERE non fixés, sa spécificité est de 84% contre 94% pour la matrice de TRANSFAC. Ces résultats suggèrent que la nouvelle matrice représentant les ERE permet d'obtenir un meilleur compromis entre sensibilité et spécificité que celle proposée par TRANSFAC. L'écart observé entre les prédictions des deux matrices dans les promoteurs de gènes induits par E2 est donc plus attribuable à leur capacité respective de détecter des ERE que de faire de fausses prédictions.

### **8.1.2 Meilleure détection des RARE à l'aide d'une matrice dérivée de ChIP-on-chip**

Moins de matrices ont été proposées pour RAR que pour ER. Parmi les banques de matrices, seule la banque MatBase a des matrices modélisant les RARE [170]. Contrairement aux banques JASPAR et TRANSFAC, les matrices de MatBase ne sont pas accessibles gratuitement. À ma connaissance, la seule alternative venant de la littérature est la matrice représentant un demi RARE proposée par Balmer et Blomhoff [173].

Balmer et Blomhoff ont construit leur matrice à partir de 51 RARE validés expérimentalement dans la littérature. Ils ont seulement considéré les RARE caractérisés chez l'humain, le rat et la souris. Un autre critère utilisé pour l'inclusion des éléments est la conservation des sites entre les espèces, la séquence des RARE étant au moins 75% identique chez les trois espèces. La plupart des 51 RARE inclus sont des éléments dont les demi sites sont directement répétés, incluant 19 DR5, 7 DR2 et 12 DR1.



Les sites de haute affinité reconnus par RAR et RXR sont généralement décrits comme deux demi sites, (A/G)G(G/T)TCA, directement répétés avec un espacement de un, deux ou cinq nucléotides [51]. La matrice proposée par Balmer et Blomhoff représente relativement bien la description typique des demi sites de haute affinité (Annexe, Figure 30). Par contre, le nucléotide A à la première position et le nucléotide G en troisième position sont plus fréquents que ne le suggèrent des expériences de fixation *in vitro*. Hauksdottir et collègues ont évalué l'impact sur la fixation de RXR $\alpha$ /RAR $\alpha$  de mutations ponctuelles sur les deux demi sites AGGTCA d'un DR5 à l'aide d'expériences de retard sur gel [174]. Dans ces expériences, la fixation de RXR $\alpha$ /RAR $\alpha$  sur les DR5 dont le nucléotide A est remplacé par G en première position des deux demi sites ou dont le nucléotide G est remplacé par T à la troisième position correspond à 70% de ce qui est observé pour le site de référence. Dans la matrice de Balmer et Blomhoff, la fréquence du nucléotide G en première position correspond à 40% de la fréquence du nucléotide A et la fréquence du nucléotide T en troisième position correspond à 40% de la fréquence du nucléotide G. La matrice risque de sous-évaluer le score des RARE ayant un G à la première position ou T en position trois. La fréquence de A/G et G/T à ces positions vient d'un biais dans la séquence de l'ensemble de RARE choisis pour construire la matrice qui ne représente pas nécessairement l'affinité de RXR/RAR pour ces positions.

J'ai construit une nouvelle matrice pour les RARE à partir de DR5 trouvés près du centre de régions liées par RAR $\alpha$  dans des expériences de ChIP-on-chip. Ces données de ChIP-on-chip ont été générées par Affymetrix dans le cadre projet ENCODE sur 1% du génome humain [83]. La nouvelle matrice a été bâtie avec les demi sites de 126 DR5 sélectionnés à l'aide des scores assignés par une matrice représentant exactement les demi sites de haute affinité (ex. en position 1, A = 50% et G = 50%). Ces DR5 incluent tous ceux dont les demi sites comportent au plus un nucléotide différent de (A/G)G(G/T)TCA et la moitié des sites avec deux nucléotides différents. La matrice représentant les sites de haute affinité a permis d'atténuer les écarts entre les fréquences de A/G et G/T pour les positions

un et trois respectivement. La fréquence du nucléotide G correspond à 60% de la fréquence du nucléotide A à la première position et le nucléotide G et T ont approximativement la même fréquence en position trois (Annexe, Figure 31).

Cette nouvelle matrice donne des scores plus élevés aux RARE de haute affinité que la matrice de Balmer et Blomhoff (Annexe, Figure 32). Les DR5 de (A/G)G(G/T)TCA obtiennent des scores normalisés entre 90% et 100% avec la nouvelle matrice alors qu'ils obtiennent des scores entre 70% et 100% avec celle de Balmer et Blomhoff. L'amélioration des scores de la nouvelle matrice ne se limite pas aux RARE correspondant exactement à la représentation des sites de haute affinité. La plupart des formes avec un ou deux nucléotides différents ont aussi des scores plus élevés. Les deux matrices permettent d'observer un enrichissement statistiquement significatif de RARE DR5 dans les promoteurs de gènes cible, mais la nouvelle matrice prédit plus de sites pour plus de gènes (Annexe, table 6). Comme pour les ERE, l'écart entre les prédictions des deux matrices est attribuable à leur capacité respective de détecter des RARE. Avec un seuil de 65% sur le score, la sensibilité de la nouvelle matrice est de 63% contre 49% pour la matrice de Balmer pour un ensemble de 500 RARE générés aléatoirement (Annexe, Figure 33).

### 8.1.3 Limitations des nouvelles matrices

L'incorporation aux matrices de caractéristiques *in vitro* des sites de haute affinité de ER et RAR a permis d'augmenter les scores assignés aux séquences présentant ces caractéristiques et permet une meilleure détection des sites reconnus par ces deux facteurs dans les promoteurs de leurs gènes cibles. Comme les matrices proposées dans la littérature, les nouvelles laissent également place à l'amélioration. Certaines des positions des nouvelles matrices ne représentent pas parfaitement l'impact d'un changement de base sur la liaison des récepteurs.

Pour ER, la présence d'une cytosine aux positions deux ou trois nuit plus à la fixation du récepteur que les autres nucléotides dans des expériences de retard sur gel. Le

remplacement des bases de ces positions par une cytosine réduit la liaison du récepteur à moins de 40% de ce qui est observé pour un ERE de référence [71]. Dans la matrice, la présence d'une adénosine ou d'une thymidine à ces positions a le même impact sur le score qu'une cytosine (Chapitre 7, Figure 6A).

Pour RAR, la présence d'une thymidine à la position trois réduit la liaison du récepteur à moins de 75% de ce qui est observé pour un RARE ayant une guanosine à cette position [174]. Dans la matrice, la présence d'une thymidine à cette position donne un score un peu plus élevé qu'avec une guanosine (Annexe, Figure 31).

Dans le futur, il serait possible de construire d'autres matrices en se basant sur les résultats d'affinité entre les récepteurs et les variants de ERE et RARE afin de voir si une meilleure modélisation de l'affinité *in vitro* permet d'améliorer la détection des sites reconnus par ERE et RAR. Une autre voie à explorer est d'utiliser d'autres modèles comme les matrices de di-nucléotides et les HMM pour prendre en compte des dépendances entre certains nucléotides. Ces modèles ont le potentiel d'améliorer les prédictions étant donné que les structures cristallines de ER et RAR présentent des interactions entre certains acides aminés de leur domaine de liaison à l'ADN et plus d'une base de leurs sites de fixation respectifs [175, 176].

## **8.2 Les éléments transposables ont contribué des éléments de réponse de ER et RAR au cours de l'évolution**

Occupant la moitié du génome humain, les éléments transposables (TE) ont le potentiel d'altérer l'expression des gènes en contribuant des sites de fixation pour des facteurs de transcription précédemment absents. Une analyse de plus de 800 sites validés expérimentalement catalogués dans la banque de données TRANSFAC a observé que près de 2.5% d'entre eux font partie de TE (Chapitre 3). La séquence de la plupart des TE est méthylée et *a priori* inaccessible aux facteurs de transcription étant donné que la

méthylation de l'ADN conduit à la condensation de la chromatine et interfère avec la fixation des facteurs de transcription.

Quelques études à large échelle ont identifié des TE non méthylés. Une étude des régions sensibles à la DNase dans des lymphocytes, cellules T CD4+, a montré que près de 23% des 14000 régions identifiées contiennent des TE [177]. Concordant avec un rôle de ces régions dans la régulation de gènes, l'expression relative sur des micropuces Affymetrix des gènes présentant une région sensible à la DNase à moins de 1 kpb de leur site d'initiation de la transcription est plus élevée que les gènes n'ayant pas de région. Une autre étude récente portant sur la méthylation des éléments Alu, le TE le plus fréquent dans le génome humain, a observé que les cellules épithéliales saines du colon contiennent en moyenne 25000 Alu non méthylés alors que les cellules tumorales contiennent en moyenne 41000 Alu non méthylés [178]. Cette augmentation du nombre d'éléments Alu non méthylés n'est pas spécifique au cancer du colon puisque l'altération globale de la méthylation de l'ADN fait partie des changements épigénétiques associés à la carcinogénèse.

Le développement et la progression de plusieurs cancers, dont le cancer du sein, est caractérisé par l'hyperméthylation d'îlots CpG et l'hypométhylation des éléments transposables [179, 180]. La méthylation d'îlots CpG entraîne la répression de la transcription de certains gènes alors que la déméthylation des TE cause une instabilité génomique due à des recombinaisons et à leur transcription. Plusieurs cas de réarrangements causés par l'insertion *de novo* d'éléments Alu ont été documentés pour le gène BRCA1 (*breast cancer 1, early onset*) chez des familles ayant une haute incidence de cancers du sein [181-185].

Les TE accessibles ou devenus accessibles suite à leur déméthylation peuvent également fournir des sites de facteurs de transcription. Un exemple vient de la réanalyse par Wang et collègues [186] des régions fixées par le facteur de transcription et suppresseur de tumeur p53 par la technique *ChIP-PET* dans la lignée cellulaire de cancer du colon

HCT116. Cette étude a montré que 35% (89/250) des régions couvertes par 3 fragments (PET3+) ont un site p53 prédit faisant partie d'éléments LTR. La fixation de p53 a été confirmée sur 4 des sites dérivés de LTR par immunoprécipitation de la chromatine.

ER et RAR sont deux facteurs qui ont été associés aux éléments transposables. Bien que la capacité des ERE et RARE dérivés d'éléments Alu à agir comme sites de fixation ait été démontrée dans la littérature à l'aide de gènes rapporteurs [109, 110], le recrutement de ER et RAR *in vivo* sur des sites provenant de TE n'avait pas été examiné. La localisation de sites validés expérimentalement par CHIP indique cependant que certains de ces sites coïncident avec une séquence Alu (Chapitre 3, Table 1). À l'échelle du génome, les TE contribuent une proportion non négligeable des sites prédits pouvant être reconnus par ces deux récepteurs (Annexe, Figure 35). Les travaux présentés aux chapitre 6 et 7 appuient la contribution d'au moins certains de ces éléments à la régulation de gènes cibles de ces récepteurs.

### 8.2.1 Amplification des RARE par les éléments transposables

L'association entre RAR et les TE a été initialement suggérée d'après les observations que des sites ressemblant à deux RARE DR2 partageant un demi site, GGaTCA CG AGGTCA GG AGaTCg, sont présents dans un élément Alu du promoteur de KRT18 et que des sites similaires sont présents dans la représentation consensus de plusieurs sous-familles des Alu [110]. Un DR2 correspondant à un RARE de haute affinité, AGGTCA GG AGTTCA, situé dans un élément Alu du promoteur proximal du gène MPO (*myeloperoxidase*) a également été caractérisé à l'aide d'expériences de retard sur gel et de gènes rapporteurs [111]. La fixation *in vivo* de RXR/RAR a aussi été observée par CHIP pour quelques exemples de DR2 dans des séquences Alu (Chapitre 3, Table 1), cependant ces sites n'avaient pas été identifiés comme étant dérivés de TE.

Deux analyses à petite échelle de séquences génomiques ont observé que les RARE DR2 se retrouvent fréquemment dans les éléments Alu. Babich et collègues [104] ont fait

cette observation suite à la recherche de sites DR2, définis comme AGGT(C/G)(A/G) NN AG(A/T)TC(A/G), sur un ensemble de 16 séquences de tailles variant entre 50 kpb et 200 kpb. Shankar et collègues [106] ont fait cette observation suite à recherche de sites DR2, définis comme (A/G)G(G/T)CA NN (A/G)G(G/T)CA, sur un ensemble de 500 séquences Alu pour chacune des sous-familles.

À l'échelle du génome, mes recherches de sites de haute affinité montrent que les DR2 sont les plus fréquents des RARE. Dans le génome humain, les DR2 dérivés de Alu représentent plus de 90% des 108582 sites trouvés (Chapitre 6, Table 1). Il y a également quatre fois plus de ces éléments dans le génome humain que celui de souris. La majorité des DR2 trouvés dans le génome de souris, dans lequel on ne retrouve pas d'éléments Alu, sont aussi dérivés de TE. Les LTR contribuent 50% des 26717 sites et les SINE contribuent 27% des sites (Chapitre 6, Table 1).

L'amplification des DR2 trouvés dans le génome humain est attribuable aux sites correspondant à AGGTCA NN AGTTCA. Cette forme de DR2 comporte un seul nucléotide différent de la représentation consensus de la plupart des sous-familles AluS qui contribuent 95% des DR2 dérivés de Alu (Chapitre 6, Figure 3). La conversion du site présent dans les séquences consensus des sous-familles AluS en celui trouvé fréquemment dans les éléments Alu est possible par méthylation/déamination du dinucléotide CpG en 3 prime. En accord avec ce mécanisme, les DR2 se trouvent en général aux positions 66 à 68 dans les Alu et l'autre site pouvant être généré par déamination, AGGTCA NN AGTTTG, se retrouve avec une fréquence comparable dans les Alu (Chapitre 6, Figure 2 et Figure 3). Le dinucléotide CpG dont la conversion génère les DR2 de haute affinité est situé dans la boîte B du promoteur interne pol III des Alu. Cette conversion respecte la séquence consensus de ce promoteur, suggérant que certains Alu-DR2 ont pu être propagés par transposition. En concordance avec ce mécanisme, plusieurs sous-familles de AluSg ont le dinucléotide CpA au lieu de CpG dans leur promoteur pol III (Chapitre 6, Fichier supplémentaire 4).

Étant donné que la structure de chromatine dans laquelle se trouvent la plupart de ces éléments limite l'accès des facteurs de transcription, il est difficile d'évaluer la proportion des Alu contenant des DR2 pouvant participer à la régulation des gènes en considérant uniquement la séquence génomique. Toutefois, l'accessibilité de certains Alu-DR2 est appuyée par la présence de 8 éléments dans des régions sensibles à la DNase identifiées dans les cellules T CD4<sup>+</sup> et par la présence de 11 éléments dans des régions fixées dans les cellules HL60 par RAR $\alpha$  identifiées en ChIP-on-chip à l'aide d'une puce couvrant 1% du génome (Chapitre 6, Fichier supplémentaire 4). La présence de 363 Alu-DR2 dans les 20 kpb flanquant le site d'initiation de la transcription de 193 gènes cibles de l'acide rétinoïque supporte la notion d'une contribution d'éléments de réponse par les Alu au cours de l'évolution (Chapitre 6, Fichier supplémentaire 4). Pour 5 de ces gènes, la fixation de RAR $\alpha$  *in vivo* sur 15 Alu-DR2 a été confirmée par *ChIP* (Chapitre 6, Figure 5). Des expériences de *ChIP-PET* et dans une moindre mesure *ChIP-on-chip* à l'échelle génomique permettront d'évaluer le pourcentage d'éléments fixés.

### 8.2.2 Les éléments transposables contribuent des ERE

L'association entre les ER et les TE vient de la description d'un site ressemblant à un ERE, tGGTCA GGC TGgtCT, dans un élément Alu du promoteur proximal de BRCA1[109]. Ce gène est habituellement décrit dans la littérature comme une cible secondaire des œstrogènes, son induction étant fortement diminuée en présence de l'inhibiteur de la synthèse protéique cycloheximide (Chapitre 7, Figure supplémentaire 2). Le promoteur de BRCA1 contient également un site pour le facteur AP1 et deux sites XRE (« *xenobiotic-responsive elements* »), situés à moins de 500 paires de bases du ERE-Alu, fixés *in vivo* en ChIP par JUN/FOS et AHR respectivement [187, 188]. Suite à leur fixation sur le promoteur de BRCA1, AHR et JUN/FOS recrutent ER $\alpha$  comme co-facteur. Bien que l'induction de FOS et JUN par les œstrogènes [189, 190] puisse expliquer l'effet de la

cyclohexime sur l'induction de BRCA1, la proximité du site AP1 de l'élément Alu n'exclut pas totalement le recrutement du récepteur sur le ERE contient.

D'autres TE que les Alu peuvent aussi contenir des ERE. Un des ERE prédits de gène IGFBP4 (« *insulin-like growth factor binding protein 4* ») provient d'un élément MIR (« *Mammalian-wide Interspersed Repeat* ») situé à -4 kpb du site d'initiation de la transcription. IGFBP4 est induit par les œstrogènes en présence de cycloheximide tant sur les micropuces que par PCR quantitatif (Chapitre 7, Figure supplémentaire 2). L'élément de réponse, AGGTCA TTG TGACaC, est plus semblable à un ERE de haute affinité que celui de BRCA1 et il est fixé par ER $\alpha$  en ChIP [71].

À l'échelle du génome, près de 40% des ERE prédits avec un seuil de 65% proviennent des TE (Annexe, Figure 35). Dans les 20 kpb flanquant le site d'initiation de la transcription des gènes, les éléments prédits dérivés de TE comptent pour entre 10% et 36% de sites selon le seuil sur le score utilisé. Cette proportion d'éléments dérivés de TE n'empêche pas l'observation d'un enrichissement statistiquement significatif de ERE dans les promoteurs de cibles primaires des œstrogènes (Chapitre 7, Figure 6). Un des ERE prédits provenant d'un élément MIR, situé à -8 kpb du transcrit NM\_021810 du gène cible primaire CDH26, a été validé en ChIP (Chapitre 7, Figure 8).

Les 1234 régions liées par ER $\alpha$  identifiées par Lin et collègues [141] à l'aide de la technique *ChIP-PET* permettent d'aborder la question de la contribution des TE aux ERE en se limitant à des éléments situés dans régions de chromatine accessibles aux récepteurs. Près de 20% des ERE (105/542) prédits avec un seuil sur le score de 65% à plus ou moins 500 pb du centre de ces régions proviennent de TE (Chapitre 4, Figure 19). Les MIR sont les éléments SINE contribuant le plus de ERE (25), les L2 sont les éléments LINE contribuant le plus de ERE (16) et les ERV1 sont les éléments LTR contribuant le plus de ERE (29). Cet ensemble d'éléments inclut le site dérivé d'un MIR du gène CDH26 mentionné plus tôt. Bien que l'importance de ces sites dans des régions fixées par ER $\alpha$



doive être confirmée expérimentalement, elle supporte la notion d'une contribution d'éléments de réponse par les TE au cours de l'évolution.

### 8.3 ER et RAR régulent leur gènes cibles sur de longues distances

La régulation transcriptionnelle des gènes peut impliquer des régions activatrices (enhancer) ou répressives (silencer) situées à plusieurs milliers de paires de bases du site d'initiation de la transcription. Les récepteurs nucléaires font partie des facteurs de transcription capables de réguler des gènes via des éléments distants du TSS [69-72]. La longueur des promoteurs utilisés pour la recherche de sites de facteurs de transcription *in silico* varie d'une étude à l'autre. Beaucoup d'études se sont concentrées sur la région proximale du TSS. À titre d'exemple, les études de Bajic et Kamalakaran [191, 192] ont prédit des éléments de réponse aux œstrogènes entre -3 kpb et +1 kpb du TSS et -2 kpb et +250 pb respectivement. La prédiction de ERE et RARE à l'échelle du génome en combinaison avec les données de micropuces permettent d'évaluer la distribution des éléments par rapport au TSS de leur gènes cibles.

Les études localisant ER $\alpha$  à l'échelle du génome par les techniques de *ChIP-on-chip* et *ChIP-PET* ont observé que 95% des régions liées sont situées à plus de 5 kpb du site d'initiation de la transcription confirmant que l'action des récepteurs ne se limite pas aux promoteurs proximaux de ses gènes cibles [134, 141]. Cette observation a également été faite dans l'analyse des ERE prédits pour les gènes cibles primaires des œstrogènes dans notre étude (Chapitre 7). En effet, avec un seuil sur le score de 65%, seulement 10% des éléments prédits se situent à 5 kpb et moins du TSS des gènes. Des travaux effectués précédemment dans le laboratoire sur les ERE de haute affinité trouvés dans le génome humain montrent qu'ils sont plus fréquents dans les 2 kpb autour du TSS de l'ensemble des gènes [71]. Cependant, l'identification des gènes cibles primaires des œstrogènes dans le deuxième article a permis d'observer que les ERE sont fortement enrichis dans les 20kb flanquant le site d'initiation de la transcription de ces gènes (Chapitre 7, Figure 7).

Plusieurs sites prédits dans cet intervalle ont été validés par immunoprécipitation de la chromatine : un élément à -8.1 kpb pour le gène CDH26, un élément à -10.5 kpb pour le gène WISP2, deux éléments à -2.2 kpb et -6 kpb pour le gène CA12 et un élément à +16.3 kpb pour le facteur de transcription EGR3 (Chapitre 7, Figure 8).

Pour RAR, les éléments DR5 sont aussi fortement enrichis dans les 20kb flanquant le site d'initiation de la transcription de gènes cibles (Annexe, Figure 34). Dans une moindre mesure, les gènes induits par RA dans les micropuces de Lamb et collègues [138] sont aussi plus fréquents que pour l'ensemble des gènes du génome dans cet intervalle. Plusieurs éléments DR2 distants du TSS prédits dans le cadre du premier article ont été validés par immunoprécipitation de la chromatine : GPRC5A, RAI1, RARRES1, RARRES3 et SMYD5 ont tous au moins un élément fixé par RAR $\alpha$  entre -16 kpb et +4.6 kpb du TSS (Chapitre 6, Figure 5).

L'élucidation des éléments situés à plusieurs dizaines de milliers de paires de base nécessite la combinaison d'approches bio-informatique et biochimiques. L'augmentation du nombre de sites prédits à considérer plus l'on s'éloigne du TSS et la proximité des gènes rend plus difficile l'association d'un site prédit à un seul gène. Pour les ERE, avec un seuil de 65%, près de 31% (42281) des sites prédits dans le génome humain se retrouvent à 25 kpb du TSS et 90% des gènes ont au moins une prédiction. Les techniques de *ChIP-on-chip* ou *ChIP-PET* permettent d'identifier les sites situés dans régions de chromatine accessibles aux facteurs de transcription et les micropuces permettent d'identifier les gènes susceptibles d'être régulés par ces prédictions. La cartographie de régions frontières (« *insulator* ») capables de bloquer l'effet d'éléments distaux comme CTCF [193] pourrait également faciliter l'assignation des sites prédits à leur gènes cible potentiels.

## 8.4 Réseaux de gènes régulés par ER et RAR

La signalisation résultant de la liaison des œstrogènes et de l'acide rétinoïque avec leurs récepteurs varie suivant les tissus cibles et peut être altérée lors de processus pathologiques. Les œstrogènes sont impliqués dans la régulation de la prolifération des cellules épithéliales mammaires normales et cancéreuses. À l'inverse, des traitements à l'acide rétinoïque inhibent la croissance des cellules de carcinome mammaire exprimant RAR $\alpha$ . La caractérisation des gènes et des mécanismes impliqués dans le contrôle de la prolifération cellulaire par les œstrogènes et l'acide rétinoïque est primordiale pour le développement de nouvelles avenues thérapeutiques contre le cancer du sein.

Durant ma maîtrise, j'ai collaboré à des projets visant une meilleure compréhension des mécanismes de régulation de l'expression des gènes par les œstrogènes et l'acide rétinoïque. Dans le cadre des travaux présentés au chapitre 7, une étude d'expression des gènes par les œstrogènes dans les MCF7, une lignée cellulaire dérivée de carcinome mammaire, a été réalisée dans des conditions expérimentales propices à la prolifération cellulaire et permettant de départager les cibles primaires et secondaires (Chapitre 7, Figure 1 et Figure supplémentaire 3). Les 2144 gènes cibles obtenus ont été classés selon l'effet des œstrogènes sur leur transcription (« up » ou « down ») et la sensibilité de cette régulation à la cycloheximide (Chapitre 7, Figure 2). Pour 63 de ces gènes, la régulation par les œstrogènes en présence ou en absence de cycloheximide a été confirmée par PCR quantitatif (Chapitre 7, Figure supplémentaire 2). Parmi ces gènes, 17 sont impliqués dans le cycle cellulaire et 12 sont impliqués dans la prolifération d'après l'annotation des processus biologiques obtenue avec le logiciel DAVID (Database for Annotation, Visualization, and Integrated Discovery) [194]. Certains de ces gènes se retrouvent dans les deux processus, comme les facteurs de transcription MYC et la famille E2F. Plusieurs autres facteurs de transcription font partie des gènes régulés par les œstrogènes sur les micropuces (Annexe, Table 7).

Afin d'identifier les facteurs de transcription susceptibles de co-réguler des sous-ensembles de gènes et de générer des hypothèses quant aux mécanismes de régulation, la séquence des promoteurs de ces classes de gènes a été criblée à l'aide d'une banque de 300 matrices. L'identification des sites de facteurs de transcription surreprésentés a été faite dans une fenêtre de 10 kpb flanquant le TSS, en considérant ces régions au complet, ou en excluant les éléments transposables, ou encore en se limitant aux régions partiellement conservées dans des alignements du génome humain avec ceux de 16 autres espèces de vertébrés. Les éléments transposables sont fréquemment dans une structure de chromatine non permissive à la fixation des facteurs de transcription et peuvent amplifier le nombre de sites trouvés pour certains facteurs au point de masquer leur surreprésentation dans un ensemble de promoteurs. Les régions conservées peuvent, si la conservation entre les espèces reflète une importance fonctionnelle, correspondre à des régions régulatrices comme des « enhancer » et « silencer ». La majorité des facteurs surreprésentés ont été identifiés en excluant les éléments transposables et en se limitant aux régions conservées (Chapitre 7, Table supplémentaire 1 et 2). Parmi les facteurs sur-représentés, on retrouve des facteurs induits par les voies non-génomiques des oestrogènes comme les EGR (*early growth response*) [195, 196], des facteurs induits par les voies génomiques comme E2F1 [197] et des facteurs capables de recruter ER $\alpha$  par interactions protéiques comme AHR, AP1 et SP1.

Dans les sections suivantes, l'enrichissement des ERE dans les promoteurs des cibles primaires et l'enrichissement de sites E2F dans les promoteurs de cibles secondaires vont être discutés plus en détail. Finalement, l'impact de l'acide rétinoïque sur les gènes régulés par les œstrogènes va être discuté dans le contexte de résultats de micropuces utilisant les cellules MCF7 publiés par autre groupe [138].

### 8.4.1 Convergence de résultats de ChIP-on-chip et des prédictions de ERE pour les cibles primaires des œstrogènes

Parmi les 2144 gènes régulés par les œstrogènes sur les micropuces, 86 font partie d'une compilation de 236 gènes cibles tirés de la littérature [71] pour lesquels on observe un enrichissement statistiquement significatif d'ERE dans leurs promoteurs. Bien que ce nombre de gènes en commun semble restreint à première vue, il est trois fois plus grand que ce que l'on peut s'attendre par chance en choisissant au hasard deux ensembles de gènes de même taille parmi les gènes du génome (28).

Les ERE sont enrichis dans les promoteurs des gènes cibles primaires, induits par les œstrogènes en présence de cycloheximide (Chapitre 7, Figure 6B). Selon les seuils sur le score, entre 45 et 255 des 544 gènes induits en présence de cycloheximide (« *Up E2+CHX* ») ont au moins un ERE dans les 20 kpb flanquant leur site d'initiation de la transcription. Un enrichissement similaire d'ERE a également été observé dans une autre étude ayant utilisé la cycloheximide pour identifier des gènes cibles primaires (Chapitre 7, Figure supplémentaire 5).

Afin d'apporter un support expérimental à la présence d'un ERE potentiel et la fixation du récepteur *in vivo* dans les promoteurs de gènes cibles sur les micropuces, j'ai utilisé les régions liées par ER $\alpha$  identifiées en *ChIP-on-chip* par Carroll et collègues [134]. Bien que les différences de conditions expérimentales (temps de traitement, concentration de sérum, concentration d'hormone) entre ces deux études compliquent la comparaison, ceci permet d'obtenir une idée de la proportion d'ERE présents dans des structures de chromatine accessibles dans les cellules MCF7. Les données de *ChIP-on-chip* concordent avec les prédictions dans les promoteurs et montrent un enrichissement d'ERE pour les gènes cibles primaires (Chapitre 7, Figure 6C). Selon les seuils sur le score, entre 27 et 106 des 544 gènes induits en présence de cycloheximide (« *Up E2+CHX* ») ont au moins un ERE dans les 20 kpb flanquant leur site d'initiation de la transcription. À un seuil de 70%

sur le score, 24% des 395 ERE prédits sont à proximité d'une région liée par ER $\alpha$  en ChIP-on-chip, alors qu'à un seuil de 85% sur le score, 61% des 46 ERE prédits se trouvent près de ces régions. Parmi les 106 gènes ayant des ERE dans des régions fixés par ER $\alpha$  en ChIP-on-chip, 8 sont impliqués dans la prolifération et 4 sont impliqués dans la régulation de la croissance (Annexe, Table 8). Ces gènes ne correspondent pas à un enrichissement statistiquement significatif de gènes avec ces fonctions par rapport à l'ensemble des gènes du génome (test de Fisher exact,  $p > 0.01$ ). Pour plusieurs d'entre eux, la régulation par les œstrogènes a été confirmée par PCR quantitatif (Chapitre 7, Figure supplémentaire 2).

L'absence de régions de *ChIP-on-chip* à proximité d'une prédiction n'exclut pas la fixation du récepteur. Par exemple, un seul des deux ERE prédits et validés en ChIP dans le promoteur du gène WISP2 se retrouve dans ces régions (Chapitre 7, Figure 8). De plus, un site qui n'est pas fixé par ER $\alpha$  dans un type cellulaire peut l'être dans un autre modèle cellulaire [172]. Malgré ces différences, la convergence des prédiction *in silico*, des données de *ChIP-on-chip* et des données de régulation de l'expression des gènes fait des gènes cibles primaires des candidats pour un régulation dépendant d'éléments de réponse aux œstrogènes.

#### **8.4.2 Amplification du signal des œstrogènes par la famille E2F**

Plusieurs membres de la famille de facteurs de transcription E2F, connus pour leur implication dans le contrôle du cycle cellulaire, sont régulés par les œstrogènes (Chapitre 7, Figure 4). L'induction de ces facteurs de transcription entre 5 et 10h suggère une amplification du signal œstrogénique par leurs gènes cibles. De plus, l'abolition de l'induction des E2F et le blocage de prolifération par la cycloheximide soulignent l'importance de ces facteurs pour la croissance des MCF7 (Chapitre 7, Figure supplémentaire 3). Dans les MCF7, ER $\alpha$  est recruté dans une région contenant des sites SP1 située à proximité des sites E2F du promoteur proximal de E2F1 [197]. Puisque l'induction de E2F1 par E2 est abolie par la cycloheximide, un facteur de transcription

autre que SP1 est sans doute responsable de l'induction de E2F1 et les autres membres de la famille E2F par les œstrogènes. L'oncogène et facteur de transcription MYC, dont l'induction par E2 est insensible à la cycloheximide, est impliqué dans la régulation de la transcription de plusieurs membres de la famille E2F dont E2F1 [198-202]. Quelques études examinant l'expression des gènes dans des tumeurs mammaires ont identifié E2F1 comme un biomarqueur de la prolifération dont le niveau d'expression varie selon le pronostic [203-205].

Parmi les 2144 gènes régulés par les œstrogènes sur les micropuces, 71 font partie d'une compilation de 127 gènes cibles des E2F tirés de la littérature [169] pour lesquels on observe un enrichissement statistiquement significatif de sites E2F dans leurs promoteurs. Ce nombre de gènes en commun est cinq fois plus grand que ce que l'on peut s'attendre par chance en choisissant au hasard deux ensembles de gènes de même taille parmi les gènes du génome (15). Pour 7 des gènes cibles en commun avec les micropuces, le recrutement de E2F sur leurs promoteurs a été montré en ChIP dans d'autres types cellulaires [206-208].

Les sites E2F situés dans des régions conservées sont enrichis dans les promoteurs des gènes cibles secondaires, induits par les œstrogènes en absence de cycloheximide (Chapitre 7, Figure 9A). Selon les seuils sur le score, entre 52 et 374 des 858 gènes induits en absence de cycloheximide (« *Up E2-CHX* ») ont au moins un site E2F. Afin d'apporter un support expérimental à l'accessibilité et la fixation de certains de ces sites par E2F, j'ai utilisé un ensemble de régions liées par E2F1 identifiées par *ChIP-on-chip* dans les MCF7 par Jin et collègues [209]. Deux différences avec notre étude limitent cependant la comparaison : leurs MCF7 n'ont pas été traitées avec des œstrogènes et les micropuces utilisées couvrent 5 kpb des promoteurs des gènes situés sur les chromosomes 11 à X. Malgré ces différences, les sites E2F proches de régions fixés par E2F1 sont enrichis dans les mêmes catégories de gènes que les sites E2F trouvés dans les promoteurs (Chapitre 7, Figure 9B). En considérant l'ensemble des sites E2F situés à proximité de régions fixés par E2F1, entre 34 et 151 des 858 gènes *Up -CHX* ont au moins un site E2F à proximité d'une

région fixée par E2F1 en *ChIP-on-chip*. Parmi ces gènes, 28 sont impliqués dans le cycle cellulaire et 11 sont impliqués dans la prolifération (Annexe, Table 9). Ces gènes correspondent à un enrichissement statistiquement significatif de gènes avec ces fonctions par rapport à l'ensemble des gènes du génome (test de Fisher exact,  $p < 0.01$ ). Pour les gènes CHEK2, MCM5 et RBBP8, le recrutement de E2F1 sur leurs promoteurs a été montré en ChIP dans d'autres types cellulaires [206-208]. Collectivement, les données de régulation de l'expression des gènes et la convergence des prédictions *in silico* et des données de *ChIP-on-chip* supporte l'amplification du signal des œstrogènes par la famille E2F.

### 8.4.3 Interférence entre les gènes régulés par E2 et RA

Les œstrogènes et l'acide rétinoïque ont des effets contraires sur la prolifération des cellules de carcinome mammaires (Chapitre 1). Parmi les gènes réprimés par les œstrogènes sur les micropuces, on retrouve des gènes cibles induits par les RAR dont plusieurs facteurs de transcription. Parmi ces facteurs, on retrouve EGR1, FOXA1, STAT1 et les récepteurs nucléaires RAR $\gamma$  et NR2C1. Pour tous ces facteurs des RARE ou des séquences contenant au moins un demi site ont été identifiés dans leurs promoteurs [173, 210-212]. Deux gènes cibles des RAR qui interagissent avec l'acide rétinoïque, CRABP2 et CYP26A1, font aussi partie des gènes réprimés par les œstrogènes. CRABP2 est impliqué dans le transport de l'acide rétinoïque entre le cytosol et le noyau [213]. CYP26A1 est une enzyme de la famille P450 impliqué dans le processus de dégradation de l'acide rétinoïque [214]. Il est à noter que l'amplitude de la régulation n'est pas du même ordre pour les deux molécules, E2 réprime CYP26A1 par -1.81 alors que RA l'induit par 83.83 (Annexe, Table 10). Bien que la répression de CYP26A1 par E2 pourrait favoriser la réponse à l'acide rétinoïque, ce bénéfice est possiblement atténué par l'induction par E2 de CYP26B1 une autre enzyme également impliqué dans la dégradation de RA (Chapitre 7, Table 1).



Afin d'avoir un portrait plus exhaustif des gènes régulés par les œstrogènes et l'acide rétinoïque dans les MCF7, j'ai réanalysé les micropuces de Lamb et collègues [138] de la même façon que pour les micropuces de l'article du chapitre précédent. Leur étude comporte deux principales différences avec notre étude des gènes cibles des œstrogènes. Leurs cellules MCF7 ont été traitées moins longtemps avec l'acide rétinoïque (6 heures) et les micropuces Affymetrix utilisées (HT-HG-U133A) contiennent la moitié des transcrits de nos micropuces (HG-U133 Plus 2.0). Sur leurs micropuces, 200 gènes sont induits par l'acide rétinoïque et 63 gènes sont réprimés. Les RARE DR5 prédits à plus ou moins 10 kpb du site d'initiation de la transcription des gènes induits sont enrichis de façon statistiquement significative avec des seuils sur le score de 75% et 80% (Annexe, Figure 36). Bien qu'il soit nécessaire d'enlever les DR5 présents dans les éléments transposables pour observer cet enrichissement, la participation de certains d'entre eux à la régulation de ces gènes n'est pas exclue. Dans les promoteurs (+/- 10 kb du TSS) de gènes induits suite à des traitements à l'acide rétinoïque et la cycloheximide dans les cellules SKBR3, des cellules de carcinome mammaire ER-négatives qui expriment aussi RAR $\alpha$ , on peut observer un enrichissement des DR5 avec et sans les éléments transposables (Annexe, Figure 37). Quelques uns des gènes régulés sur les micropuces de Lamb et collègues ont des RARE DR5 et DR1 caractérisés dans leurs promoteurs : l'enzyme CYP26A1 et les facteurs de transcription FOXA1, HOXA1, HOXA4, HOXA5 et NR2C1 [69, 173, 212, 215]. Il y a aussi quelques gènes avec des DR2-Alu validés expérimentalement : BTG2, CASP9 et GPRC5A (Chapitre 3, Table 1 et Chapitre 6, Figure 5). BTG2, HOXA5 et NR2C1 ont été décrits dans la littérature comme des médiateurs de l'inhibition de la croissance cellulaire par l'acide rétinoïque [114, 215, 216].

Parmi les transcrits communs aux micropuces de Lamb et celles du chapitre précédent, 106 gènes sont régulés par les œstrogènes et l'acide rétinoïque dans les MCF7 (Annexe, Figure 38). 70 de ces gènes présentent une régulation opposée par les deux molécules (Annexe, Table 10). La plupart de ces gènes sont réprimés par les œstrogènes et

induits par l'acide rétinoïque. Parmi ces gènes, 36 sont aussi régulés par l'acide rétinoïque dans les SKBR3 en présence ou en absence de cycloheximide (Annexe, Table 11). Plus de la moitié de ces gènes sont des cibles primaires des l'acide rétinoïque et/ou des œstrogènes. Ces gènes et les facteurs de transcription régulés par l'acide rétinoïque suggèrent une interférence sur ER $\alpha$  et ses gènes cibles.

Dans les MCF7, la transcription et la traduction de ER $\alpha$  sont réprimées par l'acide rétinoïque [217]. Certains des facteurs régulés sur les micropuces de Lamb et collègues (Annexe, Table 12) ont été associés à régulation de la transcription de ER $\alpha$  dans la littérature. TFAP2C (AP2 $\gamma$ ) est induit par les œstrogènes via un ERE dans son promoteurs et il est recruté sur un site AP2 du promoteur de ER $\alpha$  [218, 219]. L'abolition de TFAP2C par d'interférence par l'ARN réduit à 16% les niveaux d'ARNm et de protéine de ER $\alpha$  dans les MCF7 [219]. TFAP2C réprime l'expression de ce gène dans les cellules MCF7. Chez la souris, TFAP2C est induit par l'acide rétinoïque dans les cellules de carcinome embryonnaire P19 [220]. Nous avons caractérisé un DR5 (AtGTCA CCTGG AGTTaA) situé à moins de 600 paires de bases du ERE de TFAP2C. Ce site, bien qu'il ne soit pas présent chez la souris, pourrait expliquer la répression observée sur les micropuces de Lamb et collègues. Certains RARE ont le double rôle d'activateur et répresseur de la transcription selon les co-régulateurs recrutés [212]. Le co-répresseur NRIP1, qui est induit par l'acide rétinoïque dans les MCF7, pourrait rendre cet élément répresseur [221, 222]. Cependant, peu de RARE répresseurs et activateurs sont décrits dans la littérature, il est possible que l'effet de l'acide rétinoïque sur la transcription de TFAP2C passe par d'autres mécanismes. GATA3 est un autre facteur impliqué dans la régulation de la transcription de ER $\alpha$  via deux sites GATA [223]. ER $\alpha$  est également impliqué dans l'induction de l'expression de GATA3 [223] tandis que GATA3 est réprimé par l'acide rétinoïque dans les cellules MCF7. La région où ER $\alpha$  est recruté près du gène GATA3 contient deux sites DR1 pouvant expliquer la répression observée sur les micropuces de Lamb et collègues. Il est à noter cependant que l'acide rétinoïque induit GATA3 dans les cellules SKBR3.

Finalement, NR2C1, qui est régulé par les RAR, interagit avec ER $\alpha$  et interfère avec son homodimérisation [212, 216].

L'acide rétinoïque peut également avoir un effet sur les cibles secondaires des œstrogènes par l'intermédiaire des facteurs de transcription EGR3 et HES1. La répression de EGR3 par l'acide rétinoïque pourrait avoir un impact sur les gènes induits par les œstrogènes pour lesquels nous avons observé un enrichissement de sites EGR (Chapitre 7, Table supplémentaire 1). Les deux DR5 avec une position différente d'un RARE de haute affinité se trouvant dans les 5 kpb en amont du site d'initiation de la transcription de EGR3 constituent peut-être des éléments répresseurs annulant l'effet des ERE situés en aval (Chapitre 7, Figure 8). L'induction du facteur de transcription HES1 par l'acide rétinoïque et sa répression par les œstrogènes a déjà été observée dans MCF7 [224]. L'introduction d'un dominant négatif de HES1 abolit l'effet antiprolifératif de l'acide rétinoïque sur ces cellules. De plus, HES1 est un répresseur de E2F1 [225]. Par l'intermédiaire de HES1, l'acide rétinoïque peut influencer la régulation des cibles secondaires des œstrogènes qui dépendent de E2F1.

## 8.5 Conclusion

Les œstrogènes et l'acide rétinoïque sont des régulateurs de la prolifération des cellules de carcinome mammaire. Les travaux présentés ont permis d'identifier des gènes cibles potentiels de ces récepteurs et des éléments de réponse pouvant expliquer leur régulation. Plusieurs sites prédits ont été validés expérimentalement dans les promoteurs de nouveaux gènes cibles et de cibles connues. Dans le futur, il serait intéressant d'aller plus loin dans l'analyse des gènes cibles des facteurs de transcription régulés les œstrogènes et l'acide rétinoïque afin de déterminer les réseaux de régulation impliquant ces facteurs. À cette fin, il serait possible d'utiliser des modules transcriptionnels, régions contenant plusieurs sites de facteurs de transcription, prédits avec les matrices de TRANSFAC sur l'ensemble du génome humain [226]. En complément à l'analyse des modules prédits, il

serait également possible d'identifier les régions fixées par ER $\alpha$  en ChIP-on-chip [134] contenant les sites de plusieurs de ces facteurs. L'étude des réseaux de régulation pourrait être étendue aux 48 récepteurs nucléaires. Par exemple, des rétinoïdes de synthèse spécifiques aux RXR suppriment la formation de tumeurs mammaires dans des modèles animaux [227]. PPAR et VDR, qui hétérodimérisent avec les RXR comme les RAR, inhibent également la croissance des cellules de carcinome mammaire [228, 229]. Finalement, il serait possible d'utiliser les gènes faisant partie de ces réseaux pour identifier dans les données d'expression venant de patientes des marqueurs de pronostic ou des cibles pour le traitement du cancer du sein.

## Bibliographie

1. Poutanen, M., et al., *Regulation of oestrogen action: role of 17 beta-hydroxysteroid dehydrogenases*. *Ann Med*, 1995. **27**(6): p. 675-82.
2. Simpson, E., et al., *The role of local estrogen biosynthesis in males and females*. *Trends Endocrinol Metab*, 2000. **11**(5): p. 184-8.
3. Manolagas, S.C., S. Kousteni, and R.L. Jilka, *Sex steroids and bone*. *Recent Prog Horm Res*, 2002. **57**: p. 385-409.
4. Vanderschueren, D., et al., *Androgens and bone*. *Endocr Rev*, 2004. **25**(3): p. 389-425.
5. Farhat, M.Y., M.C. Lavigne, and P.W. Ramwell, *The vascular protective effects of estrogen*. *Faseb J*, 1996. **10**(5): p. 615-24.
6. Rosano, G.M., C. Vitale, and S. Lello, *Postmenopausal hormone therapy: lessons from observational and randomized studies*. *Endocrine*, 2004. **24**(3): p. 251-4.
7. Deroo, B.J. and K.S. Korach, *Estrogen receptors and human disease*. *J Clin Invest*, 2006. **116**(3): p. 561-70.
8. Couse, J.F. and K.S. Korach, *Estrogen receptor null mice: what have we learned and where will they lead us?* *Endocr Rev*, 1999. **20**(3): p. 358-417.
9. Nussey, S. and S. Whitehead, *Endocrinology: An Integrated Approach*. 2001: BIOS Scientific Publishers Ltd.
10. Thibault, C. and M.-C. Levasseur, *La reproduction chez les mammifères et l'homme*. 2001: Editions Quae.
11. Nilsson, S., et al., *Mechanisms of estrogen action*. *Physiol Rev*, 2001. **81**(4): p. 1535-65.
12. Blomhoff, R. and H.K. Blomhoff, *Overview of retinoid metabolism and function*. *J Neurobiol*, 2006. **66**(7): p. 606-30.
13. Parisotto, M., et al., *[Retinoid metabolism and cancer]*. *Med Sci (Paris)*, 2006. **22**(12): p. 1101-6.
14. Ross, S.A., et al., *Retinoids in embryonal development*. *Physiol Rev*, 2000. **80**(3): p. 1021-54.
15. Niederreither, K., et al., *Embryonic retinoic acid synthesis is essential for early mouse post-implantation development*. *Nat Genet*, 1999. **21**(4): p. 444-8.
16. Simeone, A., et al., *Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells*. *Mech Dev*, 1991. **33**(3): p. 215-27.
17. Langston, A.W. and L.J. Gudas, *Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene Hox-1.6*. *Mech Dev*, 1992. **38**(3): p. 217-27.
18. Popperl, H. and M.S. Featherstone, *Identification of a retinoic acid response element upstream of the murine Hox-4.2 gene*. *Mol Cell Biol*, 1993. **13**(1): p. 257-65.

19. Ogura, T. and R.M. Evans, *A retinoic acid-triggered cascade of HOXB1 gene activation*. Proc Natl Acad Sci U S A, 1995. **92**(2): p. 387-91.
20. Packer, A.I., et al., *Expression of the murine Hoxa4 gene requires both autoregulation and a conserved retinoic acid response element*. Development, 1998. **125**(11): p. 1991-8.
21. Kim, M.H., et al., *Retinoic acid response element in HOXA-7 regulatory region affects the rate, not the formation of anterior boundary expression*. Int J Dev Biol, 2002. **46**(3): p. 325-8.
22. Luo, T., et al., *Retinoids, eye development, and maturation of visual function*. J Neurobiol, 2006. **66**(7): p. 677-86.
23. De Luca, L.M., *Retinoids and their receptors in differentiation, embryogenesis, and neoplasia*. Faseb J, 1991. **5**(14): p. 2924-33.
24. Fisher, G.J. and J.J. Voorhees, *Molecular mechanisms of retinoid actions in skin*. FASEB J., 1996. **10**(9): p. 1002-1013.
25. Heller, E.H. and N.J. Shiffman, *Synthetic retinoids in dermatology*. Can Med Assoc J, 1985. **132**(10): p. 1129-36.
26. Thacher, S.M., J. Vasudevan, and R.A. Chandraratna, *Therapeutic applications for ligands of retinoid receptors*. Curr Pharm Des, 2000. **6**(1): p. 25-58.
27. Lotan, R., *Suppression of squamous cell carcinoma growth and differentiation by retinoids*. Cancer Res, 1994. **54**(7 Suppl): p. 1987s-1990s.
28. Germain, P., et al., *International Union of Pharmacology. LX. Retinoic acid receptors*. Pharmacol Rev, 2006. **58**(4): p. 712-25.
29. Fields, A.L., D.R. Soprano, and K.J. Soprano, *Retinoids in biological control and cancer*. J Cell Biochem, 2007. **102**(4): p. 886-898.
30. Lo Coco, F., E. Ammatuna, and M.A. Sanz, *Current treatment of acute promyelocytic leukemia*. Haematologica, 2007. **92**(3): p. 289-91.
31. Sun, S.Y. and R. Lotan, *Retinoids and their receptors in cancer development and chemoprevention*. Crit Rev Oncol Hematol, 2002. **41**(1): p. 41-55.
32. Domont, J., et al., *[Neoadjuvant endocrine therapy for breast cancer: an overview]*. Bull Cancer, 2004. **91**(1): p. 55-62.
33. van Landeghem, A.A., et al., *Endogenous concentration and subcellular distribution of estrogens in normal and malignant human breast tissue*. Cancer Res, 1985. **45**(6): p. 2900-6.
34. Brodie, A., Q. Lu, and J. Nakamura, *Aromatase in the normal breast and breast cancer*. J Steroid Biochem Mol Biol, 1997. **61**(3-6): p. 281-6.
35. Farias, E.F., et al., *Retinoic acid receptor alpha2 is a growth suppressor epigenetically silenced in MCF-7 human breast cancer cells*. Cell Growth Differ, 2002. **13**(8): p. 335-41.
36. Sirchia, S.M., et al., *Endogenous reactivation of the RARbeta2 tumor suppressor gene epigenetically silenced in breast cancer*. Cancer Res, 2002. **62**(9): p. 2455-61.

37. Laganiere, J., G. Deblois, and V. Giguere, *Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor alpha1 gene in breast cancer cells*. Mol Endocrinol, 2005. **19**(6): p. 1584-92.
38. Fitzgerald, P., et al., *Retinoic acid receptor alpha expression correlates with retinoid-induced growth inhibition of human breast cancer cells regardless of estrogen receptor status*. Cancer Res, 1997. **57**(13): p. 2642-50.
39. Germain, P., et al., *Overview of nomenclature of nuclear receptors*. Pharmacol Rev, 2006. **58**(4): p. 685-704.
40. Rochette-Egly, C., *Nuclear receptors: integration of multiple signalling pathways through phosphorylation*. Cell Signal, 2003. **15**(4): p. 355-66.
41. Olefsky, J.M., *Nuclear Receptor Minireview Series*. J. Biol. Chem., 2001. **276**(40): p. 36863-36864.
42. Dahlman-Wright, K., et al., *International Union of Pharmacology. LXIV. Estrogen receptors*. Pharmacol Rev, 2006. **58**(4): p. 773-81.
43. Kos, M., et al., *Minireview: genomic organization of the human ERalpha gene promoter region*. Mol Endocrinol, 2001. **15**(12): p. 2057-63.
44. Germain, P., et al., *International Union of Pharmacology. LXIII. Retinoid X receptors*. Pharmacol Rev, 2006. **58**(4): p. 760-72.
45. Bjornstrom, L. and M. Sjoberg, *Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes*. Mol. Endocrinol., 2005. **19**(4): p. 833-842.
46. Glass, C.K., *Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers*. Endocr Rev, 1994. **15**(3): p. 391-407.
47. Nguyen, D., et al., *A G577R mutation in the human AR P box results in selective decreases in DNA binding and in partial androgen insensitivity syndrome*. Mol Endocrinol, 2001. **15**(10): p. 1790-802.
48. Claessens, F. and D.T. Gewirth, *DNA recognition by nuclear receptors*. Essays Biochem, 2004. **40**: p. 59-72.
49. Nakshatri, H. and P. Bhat-Nakshatri, *Multiple parameters determine the specificity of transcriptional response by nuclear receptors HNF-4, ARP-1, PPAR, RAR and RXR through common response elements*. Nucleic Acids Res, 1998. **26**(10): p. 2491-9.
50. Tavera-Mendoza, L., et al., *Convergence of vitamin D and retinoic acid signalling at a common hormone response element*. EMBO Rep, 2006. **7**(2): p. 180-5.
51. Bastien, J. and C. Rochette-Egly, *Nuclear retinoid receptors and the transcription of retinoid-target genes*. Gene, 2004. **328**: p. 1-16.
52. Suzuki, Y., et al., *Physical interaction between retinoic acid receptor and Sp1: mechanism for induction of urokinase by retinoic acid*. Blood, 1999. **93**(12): p. 4264-76.
53. Kim, K., et al., *Analysis of estrogen receptor alpha-Sp1 interactions in breast cancer cells by fluorescence resonance energy transfer*. Mol Endocrinol, 2005. **19**(4): p. 843-54.

54. Webb, P., et al., *The Estrogen Receptor Enhances AP-1 Activity by Two Distinct Mechanisms with Different Requirements for Receptor Transactivation Functions*. Mol Endocrinol, 1999. **13**(10): p. 1672-1685.
55. Schule, R., et al., *Retinoic Acid is a Negative Regulator of AP-1-Responsive Genes*. Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6092-6096.
56. Crowe, D.L., R. Kim, and R.A. Chandraratna, *Retinoic acid differentially regulates cancer cell proliferation via dose-dependent modulation of the mitogen-activated protein kinase pathway*. Mol Cancer Res, 2003. **1**(7): p. 532-40.
57. Bastien, J., et al., *The phosphoinositide 3-kinase/Akt pathway is essential for the retinoic acid-induced differentiation of F9 cells*. Oncogene, 2006. **25**(14): p. 2040-7.
58. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
59. Luger, K., *Dynamic nucleosomes*. Chromosome Res, 2006. **14**(1): p. 5-16.
60. Ray-Gallet, D., et al., *[Variations on the topic of the "histone code"]*. Med Sci (Paris), 2005. **21**(4): p. 384-9.
61. Razin, S.V., et al., *Chromatin domains and regulation of transcription*. J Mol Biol, 2007. **369**(3): p. 597-607.
62. Koolman, J. and K.-H. Rohm, *Color Atlas Of Biochemistry*. 1996: Thieme Medical Publishers.
63. Levine, M. and R. Tjian, *Transcription regulation and animal diversity*. Nature, 2003. **424**(6945): p. 147-51.
64. Sandelin, A., et al., *Mammalian RNA polymerase II core promoters: insights from genome-wide studies*. Nat Rev Genet, 2007. **8**(6): p. 424-436.
65. Butler, J.E. and J.T. Kadonaga, *The RNA polymerase II core promoter: a key component in the regulation of gene expression*. Genes Dev, 2002. **16**(20): p. 2583-92.
66. Muller, F. and L. Tora, *The multicoloured world of promoter recognition complexes*. Embo J, 2004. **23**(1): p. 2-8.
67. Bucher, P. and E.N. Trifonov, *Compilation and analysis of eukaryotic POL II promoter sequences*. Nucleic Acids Res, 1986. **14**(24): p. 10009-26.
68. Deng, T. and M. Karin, *c-Fos transcriptional activity stimulated by H-Ras-activated protein kinase distinct from JNK and ERK*. Nature, 1994. **371**(6493): p. 171-5.
69. Loudig, O., et al., *Transcriptional co-operativity between distant retinoic acid response elements in regulation of Cyp26A1 inducibility*. Biochem J, 2005. **392**(Pt 1): p. 241-8.
70. Shang, Y., M. Myers, and M. Brown, *Formation of the androgen receptor transcription complex*. Mol Cell, 2002. **9**(3): p. 601-10.
71. Bourdeau, V., et al., *Genome-wide identification of high-affinity estrogen response elements in human and mouse*. Mol Endocrinol, 2004. **18**(6): p. 1411-27.



72. Deschenes, J., et al., *Regulation of GREB1 transcription by estrogen receptor alpha through a multipartite enhancer spread over 20 kb of upstream flanking sequences.* J Biol Chem, 2007. **282**(24): p. 17335-9.
73. Alvarez, M., S.J. Rhodes, and J.P. Bidwell, *Context-dependent transcription: all politics is local.* Gene, 2003. **313**: p. 43-57.
74. Gombert, W.M., et al., *The c-myc Insulator Element and Matrix Attachment Regions Define the c-myc Chromosomal Domain.* 2003. **23**(24): p. 9338-9348.
75. Kazazian, H.H., Jr., *Mobile elements: drivers of genome evolution.* Science, 2004. **303**(5664): p. 1626-32.
76. Nelson, P.N., et al., *Human endogenous retroviruses: transposable elements with potential?* Clin Exp Immunol, 2004. **138**(1): p. 1-9.
77. Han, J.S. and J.D. Boeke, *LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?* Bioessays, 2005. **27**(8): p. 775-84.
78. Brouha, B., et al., *Hot L1s account for the bulk of retrotransposition in the human population.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5280-5.
79. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.
80. Yang, N., et al., *An important role for RUNX3 in human L1 transcription and retrotransposition.* Nucleic Acids Res, 2003. **31**(16): p. 4929-4940.
81. Athanikar, J.N., R.M. Badge, and J.V. Moran, *A YY1-binding site is required for accurate human LINE-1 transcription initiation.* Nucleic Acids Res, 2004. **32**(13): p. 3846-3855.
82. Deininger, P.L. and M.A. Batzer, *Alu repeats and human disease.* Mol Genet Metab, 1999. **67**(3): p. 183-93.
83. Consortium, I.H.G.S., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
84. Federico, C., et al., *Gene density in the Giemsa bands of human chromosomes.* Chromosome Res, 2000. **8**(8): p. 737-46.
85. Jabbari, K. and G. Bernardi, *CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families.* Gene, 1998. **224**(1-2): p. 123-7.
86. Pavlicek, A., et al., *Similar integration but different stability of Alus and LINEs in the human genome.* Gene, 2001. **276**(1-2): p. 39-45.
87. Kuhn, R.M., et al., *The UCSC genome browser database: update 2007.* Nucleic Acids Res, 2007. **35**(Database issue): p. D668-73.
88. Bailey, J.A., G. Liu, and E.E. Eichler, *An Alu transposition model for the origin and expansion of human segmental duplications.* Am J Hum Genet, 2003. **73**(4): p. 823-34.
89. Sen, S.K., et al., *Human genomic deletions mediated by recombination between Alu elements.* Am J Hum Genet, 2006. **79**(1): p. 41-53.
90. Hata, K. and Y. Sakaki, *Identification of critical CpG sites for repression of L1 transcription by DNA methylation.* Gene, 1997. **189**(2): p. 227-34.

91. Yu, F., et al., *Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription*. Nucleic Acids Res, 2001. **29**(21): p. 4493-4501.
92. Bogerd, H.P., et al., *Cellular inhibitors of long interspersed element 1 and Alu retrotransposition*. Proceedings of the National Academy of Sciences 2006. **103**(23): p. 8780-8785.
93. Soifer, H.S., et al., *A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon*. Nucleic Acids Res, 2005. **33**(3): p. 846-856.
94. Makalowski, W., G.A. Mitchell, and D. Labuda, *Alu sequences in the coding regions of mRNA: a source of protein variability*. Trends Genet, 1994. **10**(6): p. 188-93.
95. Nekrutenko, A. and W.H. Li, *Transposable elements are found in a large number of human protein-coding genes*. Trends Genet, 2001. **17**(11): p. 619-21.
96. Sorek, R., G. Ast, and D. Graur, *Alu-containing exons are alternatively spliced*. Genome Res, 2002. **12**(7): p. 1060-7.
97. Britten, R., *Transposable elements have contributed to thousands of human proteins*. Proceedings of the National Academy of Sciences, 2006. **103**(6): p. 1798-1803.
98. Sela, N., et al., *Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome*. Genome Biol, 2007. **8**(6): p. R127.
99. Smalheiser, N.R. and V.I. Torvik, *Mammalian microRNAs derived from genomic repeats*. Trends Genet, 2005. **21**(6): p. 322-6.
100. Piriyaongsa, J., L. Marino-Ramirez, and I.K. Jordan, *Origin and evolution of human microRNAs from transposable elements*. Genetics, 2007. **176**(2): p. 1323-37.
101. Borchert, G.M., W. Lanier, and B.L. Davidson, *RNA polymerase III transcribes human microRNAs*. Nat Struct Mol Biol, 2006. **13**(12): p. 1097-101.
102. Smalheiser, N.R. and V.I. Torvik, *Alu elements within human mRNAs are probable microRNA targets*. Trends Genet, 2006. **22**(10): p. 532-6.
103. Lev-Maor, G., et al., *RNA-editing-mediated exon evolution*. Genome Biol, 2007. **8**(2): p. R29.
104. Babich, V., et al., *Association of some potential hormone response elements in human genes with the Alu family repeats*. Gene, 1999. **239**(2): p. 341-9.
105. Oei, S.L., et al., *Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters*. Genomics, 2004. **83**(5): p. 873-82.
106. Shankar, R., et al., *Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements*. BMC Evol Biol, 2004. **4**(1): p. 37.
107. Polak, P. and E. Domany, *Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes*. BMC Genomics, 2006. **7**: p. 133.

108. Thornburg, B.G., V. Gotea, and W. Makalowski, *Transposable elements as a significant source of transcription regulating signals*. *Gene*, 2006. **365**: p. 104-10.
109. Norris, J., et al., *Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers*. *J Biol Chem*, 1995. **270**(39): p. 22777-82.
110. Vansant, G. and W.F. Reynolds, *The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element*. *Proc Natl Acad Sci U S A*, 1995. **92**(18): p. 8229-33.
111. Piedrafita, F.J., et al., *An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element*. *J Biol Chem*, 1996. **271**(24): p. 14412-20.
112. Jordan, I.K., et al., *Origin of a substantial fraction of human regulatory sequences from transposable elements*. *Trends Genet*, 2003. **19**(2): p. 68-72.
113. Wang, Q., J.S. Carroll, and M. Brown, *Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking*. *Mol Cell*, 2005. **19**(5): p. 631-42.
114. Donato, L.J., J.H. Suh, and N. Noy, *Suppression of mammary carcinoma cell growth by retinoic acid: the cell cycle control gene Btg2 is a direct target for retinoic acid receptor signaling*. *Cancer Res*, 2007. **67**(2): p. 609-15.
115. Donato, L.J. and N. Noy, *Suppression of mammary carcinoma growth by retinoic acid: proapoptotic genes are targets for retinoic acid receptor and cellular retinoic acid-binding protein II signaling*. *Cancer Res*, 2005. **65**(18): p. 8193-9.
116. Zeng, M., et al., *The human kallikrein 10 promoter contains a functional retinoid response element*. *Biol Chem*, 2006. **387**(6): p. 741-7.
117. Wang, T.T., et al., *Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression*. *J Immunol*, 2004. **173**(5): p. 2909-12.
118. Gombart, A.F., N. Borregaard, and H.P. Koeffler, *Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3*. *Faseb J*, 2005. **19**(9): p. 1067-77.
119. Wang, T.T., et al., *Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin D3 target genes*. *Mol Endocrinol*, 2005. **19**(11): p. 2685-95.
120. Wang, X., et al., *Expression of human kinase suppressor of Ras 2 (hKSR-2) gene in HL60 leukemia cells is directly upregulated by 1,25-dihydroxyvitamin D(3) and is required for optimal cell differentiation*. *Exp Cell Res*, 2007. **313**(14): p. 3034-45.
121. Borgono, C.A. and E.P. Diamandis, *The emerging roles of human tissue kallikreins in cancer*. *Nat Rev Cancer*, 2004. **4**(11): p. 876-890.
122. Paliouras, M., C. Borgono, and E.P. Diamandis, *Human tissue kallikreins: The cancer biomarker family*. *Cancer Letters*, 2007. **249**(1): p. 61-79.
123. Adams, M.D., et al., *Complementary DNA sequencing: expressed sequence tags and human genome project*. *Science*, 1991. **252**(5013): p. 1651-6.

124. Okubo, K., et al., *Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression*. Nat Genet, 1992. **2**(3): p. 173-179.
125. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. **270**(5235): p. 467-70.
126. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays*. Nat Biotech, 1996. **14**(13): p. 1675-1680.
127. Bourdeau, V., et al., *Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells*. Nucleic Acids Res, 2007.
128. Irizarry, R.A., et al., *Summaries of Affymetrix GeneChip probe level data*. Nucleic Acids Res, 2003. **31**(4): p. e15-.
129. Millenaar, F.F., et al., *How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results*. BMC Bioinformatics, 2006. **7**: p. 137.
130. Steinhoff, C. and M. Vingron, *Normalization and quantification of differential expression in gene expression microarrays*. Brief Bioinform, 2006. **7**(2): p. 166-77.
131. Smyth, G.K., *Linear models and empirical Bayes methods for assessing differential expression in microarray experiments*. Statistical Applications in Genetics and Molecular Biology, **2004**. **3**(No. 1, Article 3).
132. Smyth, G.K., *Limma: linear models for microarray data.*, in *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, V.C. R. Gentleman, S. Dudoit, R. Irizarry, W. Huber, Editor. 2005, Springer. p. 397-420.
133. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
134. Carroll, J.S., et al., *Genome-wide analysis of estrogen receptor binding sites*. Nat Genet, 2006. **38**(11): p. 1289-1297.
135. Frasor, J., et al., *Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype*. Endocrinology, 2003. **144**(10): p. 4562-74.
136. Rae, J.M., et al., *GREB 1 is a critical regulator of hormone dependent breast cancer growth*. Breast Cancer Res Treat, 2005. **92**(2): p. 141-9.
137. Itoh, T., et al., *Letrozole-, anastrozole-, and tamoxifen-responsive genes in MCF-7aro cells: a microarray approach*. Mol Cancer Res, 2005. **3**(4): p. 203-18.
138. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease*. Science, 2006. **313**(5795): p. 1929-35.
139. Chen, Y., et al., *Agonist and antagonist of retinoic acid receptors cause similar changes in gene expression and induce senescence-like growth arrest in MCF-7 breast carcinoma cells*. Cancer Res, 2006. **66**(17): p. 8749-61.
140. Hudson, M.E. and M. Snyder, *High-throughput methods of regulatory element discovery*. Biotechniques, 2006. **41**(6): p. 673, 675, 677 passim.

141. Lin, C.-Y., et al., *Whole-Genome Cartography of Estrogen Receptor Binding Sites*. PLoS Genetics, 2007. **3**(6): p. e87.
142. Cheng, A.S., et al., *Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor-alpha responsive promoters*. Mol Cell, 2006. **21**(3): p. 393-404.
143. Vega, V.B., et al., *Multipplatform genome-wide identification and modeling of functional human estrogen receptor binding sites*. Genome Biol, 2006. **7**(9): p. R82.
144. Kininis, M., et al., *Genomic Analyses of Transcription Factor Binding, Histone Acetylation, and Gene Expression Reveal Mechanistically Distinct Classes of Estrogen-Regulated Promoters*. Mol Cell Biol., 2007. **27**(14): p. 5090-5104.
145. Kwon, Y.-S., et al., *Sensitive ChIP-DSL technology reveals an extensive estrogen receptor {alpha}-binding program on human gene promoters*. Proc Natl Acad Sci U S A, 2007. **104**(12): p. 4852-4857.
146. Blanchette, M., et al., *Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression*. Genome Res, 2006. **16**(5): p. 656-68.
147. Carroll, J.S., et al., *Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1*. Cell, 2005. **122**(1): p. 33-43.
148. Gao, H., et al., *Genome-wide identification of estrogen receptor {alpha} binding sites in mouse liver*. Mol. Endocrinol., 2007: p. me.2007-0121.
149. Stormo, G.D., *DNA binding sites: representation and discovery*. Bioinformatics, 2000. **16**(1): p. 16-23.
150. Bichet, A., D. Polverari, and A. Malpertuy, [*Gene regulation and bioinformatics*]. Bull Cancer, 2005. **92**(1): p. 97-107.
151. GuhaThakurta, D., *Computational identification of transcriptional regulatory elements in DNA sequence*. Nucleic Acids Res, 2006. **34**(12): p. 3585-98.
152. Gonzalo, N., *A guided tour to approximate string matching*. ACM Computing Surveys (CSUR), 2001. **33**(1): p. 31-88.
153. Gonzalo, N., *NR-grep: a fast and flexible pattern-matching tool*. Software: Practice and Experience, 2001. **31**(13): p. 1265-1312.
154. Cornish-Bowden, A., *Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984*. Nucleic Acids Res, 1985. **13**(9): p. 3021-30.
155. Sandelin, A., et al., *JASPAR: an open-access database for eukaryotic transcription factor binding profiles*. Nucleic Acids Res, 2004. **32**(Database issue): p. D91-4.
156. Cartharius, K., et al., *MatInspector and beyond: promoter analysis based on transcription factor binding sites*. Bioinformatics, 2005. **21**(13): p. 2933-42.
157. Matys, V., et al., *TRANSFAC and its module TRANSCmpel: transcriptional gene regulation in eukaryotes*. Nucleic Acids Res, 2006. **34**(Database issue): p. D108-10.
158. Wasserman, W.W. and A. Sandelin, *Applied bioinformatics for the identification of regulatory elements*. Nat Rev Genet, 2004. **5**(4): p. 276-87.

159. Hertz, G.Z. and G.D. Stormo, *Identifying DNA and protein patterns with statistically significant alignments of multiple sequences*. Bioinformatics, 1999. **15**(7-8): p. 563-77.
160. Quandt, K., et al., *MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data*. Nucleic Acids Res, 1995. **23**(23): p. 4878-84.
161. Kel, A.E., et al., *MATCH: A tool for searching transcription factor binding sites in DNA sequences*. Nucleic Acids Res, 2003. **31**(13): p. 3576-9.
162. Lenhard, B., et al., *Identification of conserved regulatory elements by comparative genome analysis*. J Biol, 2003. **2**(2): p. 13.
163. Benos, P.V., A.S. Lapedes, and G.D. Stormo, *Is there a code for protein-DNA recognition? Probab(ilstical)ly*. Bioessays, 2002. **24**(5): p. 466-75.
164. Sandelin, A. and W.W. Wasserman, *Prediction of Nuclear Hormone Receptor Response Elements*. Mol. Endocrinol., 2005. **19**(3): p. 595-606.
165. O'Lone, R., et al., *Genomic targets of nuclear estrogen receptors*. Mol Endocrinol, 2004. **18**(8): p. 1859-75.
166. Jin, V.X., et al., *Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray*. Nucleic Acids Res, 2004. **32**(22): p. 6627-35.
167. Lin, C.Y., et al., *Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells*. Genome Biol, 2004. **5**(9): p. R66.
168. Ho Sui, S.J., et al., *oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes*. Nucleic Acids Res, 2005. **33**(10): p. 3154-64.
169. Zhao, F., et al., *TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies*. Nucleic Acids Res, 2005. **33**(Database issue): p. D103-7.
170. Genomatix. *Matrix Families (Matrix Library 7.0, October 2007)*. 2007 [cited; Available from: [http://www.genomatix.de/online\\_help/help\\_matinspector/matrices.html](http://www.genomatix.de/online_help/help_matinspector/matrices.html)].
171. Klinge, C.M., *Estrogen receptor interaction with estrogen response elements*. Nucleic Acids Res, 2001. **29**(14): p. 2905-19.
172. Denger, S., et al., *Transcriptome profiling of estrogen-regulated genes in human primary osteoblasts reveals an osteoblast-specific regulation of the insulin-like growth factor binding protein 4 gene*. Mol Endocrinol, 2008. **22**(2): p. 361-79.
173. Balmer, J.E. and R. Blomhoff, *A robust characterization of retinoic acid response elements based on a comparison of sites in three species*. J Steroid Biochem Mol Biol, 2005. **96**(5): p. 347-54.
174. Hauksdottir, H. and M.L. Privalsky, *DNA recognition by the aberrant retinoic acid receptors implicated in human acute promyelocytic leukemia*. Cell Growth Differ, 2001. **12**(2): p. 85-98.

175. Schwabe, J.W., et al., *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements*. Cell, 1993. **75**(3): p. 567-78.
176. Rastinejad, F., et al., *Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DRI*. Embo J, 2000. **19**(5): p. 1045-54.
177. Marino-Ramirez, L. and I.K. Jordan, *Transposable element derived DNaseI-hypersensitive sites in the human genome*. Biol Direct, 2006. **1**: p. 20.
178. Rodriguez, J., et al., *Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells*. Nucleic Acids Res, 2008. **36**(3): p. 770-84.
179. Carnell, A.N. and J.I. Goodman, *The long (LINEs) and the short (SINEs) of it: altered methylation as a precursor to toxicity*. Toxicol Sci, 2003. **75**(2): p. 229-35.
180. Wilson, A.S., B.E. Power, and P.L. Molloy, *DNA hypomethylation and human diseases*. Biochim Biophys Acta, 2007. **1775**(1): p. 138-62.
181. Puget, N., et al., *A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17*. Cancer Res, 1997. **57**(5): p. 828-31.
182. Montagna, M., et al., *Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families*. Oncogene, 1999. **18**(28): p. 4160-5.
183. Rohlf, E.M., et al., *An Alu-mediated 7.1 kb deletion of BRCA1 exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10*. Genes Chromosomes Cancer, 2000. **28**(3): p. 300-7.
184. Yap, K.P., et al., *Detection of a novel Alu-mediated BRCA1 exon 13 duplication in Chinese breast cancer patients and implications for genetic testing*. Clin Genet, 2006. **70**(1): p. 80-2.
185. Teugels, E., et al., *De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes*. Hum Mutat, 2005. **26**(3): p. 284.
186. Wang, T., et al., *Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53*. Proceedings of the National Academy of Sciences, 2007. **104**(47): p. 18613-18618.
187. Jeffy, B.D., et al., *An estrogen receptor-alpha/p300 complex activates the BRCA-1 promoter at an AP-1 site that binds Jun/Fos transcription factors: repressive effects of p53 on BRCA-1 transcription*. Neoplasia, 2005. **7**(9): p. 873-82.
188. Hockings, J.K., et al., *The ligand status of the aromatic hydrocarbon receptor modulates transcriptional activation of BRCA-1 promoter by estrogen*. Cancer Res, 2006. **66**(4): p. 2224-32.
189. Hyder, S.M., et al., *The protooncogene c-jun contains an unusual estrogen-inducible enhancer within the coding sequence*. J Biol Chem, 1995. **270**(15): p. 8506-13.
190. Duan, R., W. Porter, and S. Safe, *Estrogen-induced c-fos protooncogene expression in MCF-7 human breast cancer cells: role of estrogen receptor Sp1 complex formation*. Endocrinology, 1998. **139**(4): p. 1981-90.

191. Bajic, V.B., et al., *Dragon ERE Finder version 2: A tool for accurate detection and analysis of estrogen response elements in vertebrate genomes*. Nucleic Acids Res, 2003. **31**(13): p. 3605-7.
192. Kamalakaran, S., S.K. Radhakrishnan, and W.T. Beck, *Identification of estrogen-responsive genes using a genome-wide analysis of promoter elements for transcription factor binding sites*. J Biol Chem, 2005. **280**(22): p. 21491-7.
193. Kim, T.H., et al., *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome*. Cell, 2007. **128**(6): p. 1231-45.
194. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery*. Genome Biol, 2003. **4**(5): p. P3.
195. Pedram, A., et al., *Integration of the non-genomic and genomic actions of estrogen. Membrane-initiated signaling by steroid to transcription and cell biology*. J Biol Chem, 2002. **277**(52): p. 50768-75.
196. Chen, C.C., W.R. Lee, and S. Safe, *Egr-1 is activated by 17beta-estradiol in MCF-7 cells by mitogen-activated protein kinase-dependent phosphorylation of ELK-1*. J Cell Biochem, 2004. **93**(5): p. 1063-74.
197. Tu, Z., et al., *IKK alpha regulates estrogen-induced cell cycle progression by modulating E2F1 expression*. J Biol Chem, 2006. **281**(10): p. 6699-706.
198. Dubik, D. and R.P. Shiu, *Transcriptional regulation of c-myc oncogene expression by estrogen in hormone-responsive human breast cancer cells*. J Biol Chem, 1988. **263**(25): p. 12705-8.
199. Sears, R., K. Ohtani, and J.R. Nevins, *Identification of positively and negatively acting elements regulating expression of the E2F2 gene in response to cell growth signals*. Mol Cell Biol, 1997. **17**(9): p. 5227-35.
200. Adams, M.R., et al., *Complex transcriptional regulatory mechanisms control expression of the E2F3 locus*. Mol Cell Biol, 2000. **20**(10): p. 3633-9.
201. Fernandez, P.C., et al., *Genomic targets of the human c-Myc protein*. Genes Dev, 2003. **17**(9): p. 1115-29.
202. Leung, J.Y., et al., *A role for Myc in facilitating transcription activation by E2F1*. Oncogene, 2008. **27**(30): p. 4172-9.
203. Zhang, S.Y., et al., *E2F-1: a proliferative marker of breast neoplasia*. Cancer Epidemiol Biomarkers Prev, 2000. **9**(4): p. 395-401.
204. Baldini, E., et al., *Cyclin A and E2F1 overexpression correlate with reduced disease-free survival in node-negative breast cancer patients*. Anticancer Res, 2006. **26**(6B): p. 4415-21.
205. Vuaroqueaux, V., et al., *Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome*. Breast Cancer Res, 2007. **9**(3): p. R33.
206. Kel, A.E., et al., *Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors*. J Mol Biol, 2001. **309**(1): p. 99-120.
207. Blais, A., et al., *Regulation of the human cyclin-dependent kinase inhibitor p18INK4c by the transcription factors E2F1 and Sp1*. J Biol Chem, 2002. **277**(35): p. 31679-93.



208. Oberley, M.J., D.R. Inman, and P.J. Farnham, *E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9*. J Biol Chem, 2003. **278**(43): p. 42466-76.
209. Jin, V.X., et al., *A computational genomics approach to identify cis-regulatory modules from chromatin immunoprecipitation microarray data--a case study using E2F1*. Genome Res, 2006. **16**(12): p. 1585-95.
210. Suva, L.J., et al., *Characterization of retinoic acid- and cell-dependent sequences which regulate zif268 gene expression in osteoblastic cells*. Mol Endocrinol, 1994. **8**(11): p. 1507-20.
211. Kolla, V., X. Weihua, and D.V. Kalvakolanu, *Modulation of interferon action by retinoids. Induction of murine STAT1 gene expression by retinoic acid*. J Biol Chem, 1997. **272**(15): p. 9742-8.
212. Gupta, P., et al., *Orphan nuclear receptor TR2, a mediator of preadipocyte proliferation, is differentially regulated by RA through exchange of coactivator PCAF with corepressor RIP140 on a platform molecule GRIP1*. Nucleic Acids Res, 2007. **35**(7): p. 2269-82.
213. Sessler, R.J. and N. Noy, *A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II*. Mol Cell, 2005. **18**(3): p. 343-53.
214. White, J.A., et al., *cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RAI) identifies a novel family of cytochromes P450*. J Biol Chem, 1997. **272**(30): p. 18538-41.
215. Chen, H., et al., *HOXA5 acts directly downstream of retinoic acid receptor beta and contributes to retinoic acid-induced apoptosis and growth inhibition*. Cancer Res, 2007. **67**(17): p. 8007-13.
216. Hu, Y.C., et al., *Suppression of estrogen receptor-mediated transcription and cell growth by interaction with TR2 orphan receptor*. J Biol Chem, 2002. **277**(37): p. 33571-9.
217. Rubin, M., et al., *9-Cis retinoic acid inhibits growth of breast cancer cells and down-regulates estrogen receptor RNA and protein*. Cancer Res, 1994. **54**(24): p. 6549-56.
218. Jin, V.X., et al., *ERTargetDB: an integral information resource of transcription regulation of estrogen receptor target genes*. J Mol Endocrinol, 2005. **35**(2): p. 225-30.
219. Woodfield, G.W., et al., *TFAP2C Controls Hormone Response in Breast Cancer Cells through Multiple Pathways of Estrogen Signaling*. Cancer Res, 2007. **67**(18): p. 8439-8443.
220. Oulad-Abdelghani, M., et al., *AP-2.2: a novel AP-2-related transcription factor induced by retinoic acid during differentiation of P19 embryonal carcinoma cells*. Exp Cell Res, 1996. **225**(2): p. 338-47.
221. Kerley, J.S., et al., *Transcriptional activation of the nuclear receptor corepressor RIP140 by retinoic acid: a potential negative-feedback regulatory mechanism*. Biochem Biophys Res Commun, 2001. **285**(4): p. 969-75.

222. White, K.A., et al., *Limiting effects of RIP140 in estrogen signaling: potential mediation of anti-estrogenic effects of retinoic acid*. J Biol Chem, 2005. **280**(9): p. 7829-35.
223. Eeckhoutte, J., et al., *Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer*. Cancer Res, 2007. **67**(13): p. 6477-83.
224. Muller, P., et al., *The anti-estrogenic effect of all-trans-retinoic acid on the breast cancer cell line MCF-7 is dependent on HES-1 expression*. J Biol Chem, 2002. **277**(32): p. 28376-9.
225. Hartman, J., et al., *HES-1 inhibits 17beta-estradiol and heregulin-beta1-mediated upregulation of E2F-1*. Oncogene, 2004. **23**(54): p. 8826-33.
226. Ferretti, V., et al., *PreMod: a database of genome-wide mammalian cis-regulatory module predictions*. Nucleic Acids Res, 2007. **35**(Database issue): p. D122-6.
227. Howe, L.R., *Rexinoids and breast cancer prevention*. Clin Cancer Res, 2007. **13**(20): p. 5983-7.
228. Welsh, J., et al., *Impact of the Vitamin D3 receptor on growth-regulatory pathways in mammary gland and breast cancer*. J Steroid Biochem Mol Biol, 2002. **83**(1-5): p. 85-92.
229. Girroir, E.E., et al., *Peroxisome proliferator-activated receptor-beta/delta (PPARbeta/delta) ligands inhibit growth of UACC903 and MCF7 human cancer cell lines*. Toxicology, 2008. **243**(1-2): p. 236-43.
230. Balmer, J.E. and R. Blomhoff, *Gene expression regulation by retinoic acid*. J Lipid Res, 2002. **43**(11): p. 1773-808.

## ANNEXE : Figures et tableaux supplémentaires

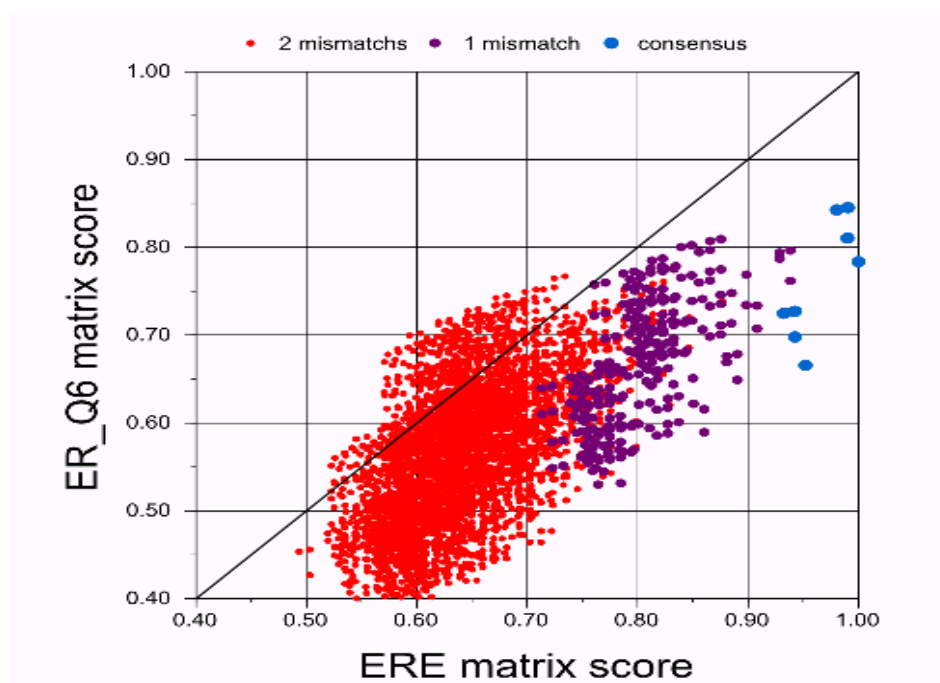


Figure 28 Comparaison des scores des matrices ERE

La matrice ERE correspond à celle utilisée dans l'article du chapitre 7 [127]. La matrice ER\_Q6 provient de TRANSFAC. Chaque point correspond aux scores assignés par les deux matrices à un ERE comptant de 0 à 2 nucléotides différent d'un ERE de haute affinité (RGGTCAnnnTGACCY).

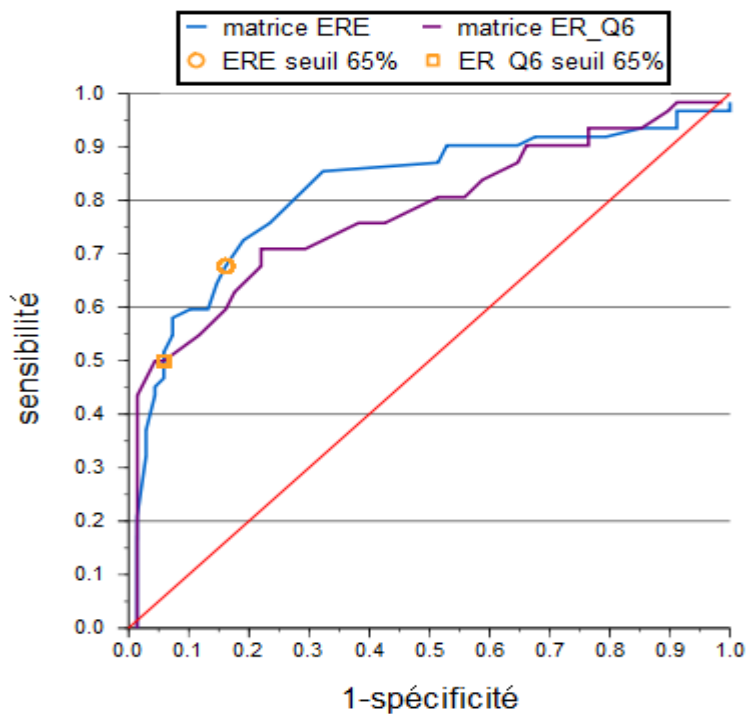


Figure 29 courbe ROC des matrices ERE

La matrice ERE correspond à celle utilisée dans l'article du chapitre 7 [127]. La matrice ER\_Q6 provient de TRANSFAC. La sensibilité et la spécificité des deux matrices ont été calculé à l'aide 64 ERE validés expérimentalement et 68 ERE non fixé par ER [143]. La sensibilité a été calculée à l'aide de la formule  $TP/(TP+FN)$  et la spécificité à l'aide de la formule  $TN/(TN+FP)$ . TN = vrai négatif, TP = vrai positif, FN = faux négatif et FP = faux positif.

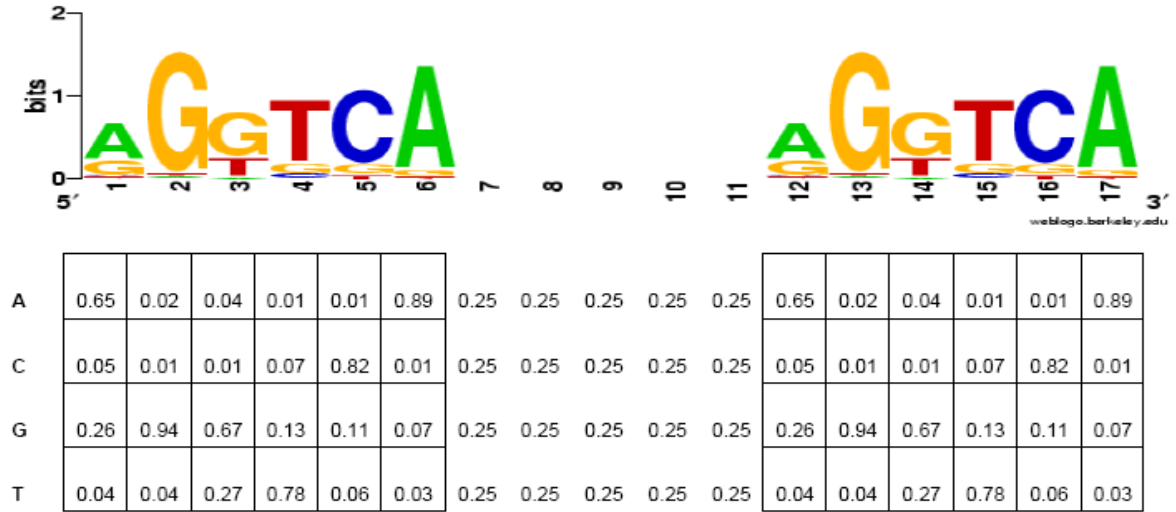


Figure 30 Matrice RARE DR5 provenant d’une compilation de la littérature

La matrice est dérivée de 51 RARE validés expérimentalement et conservés chez l’humain, le rat et la souris [173].

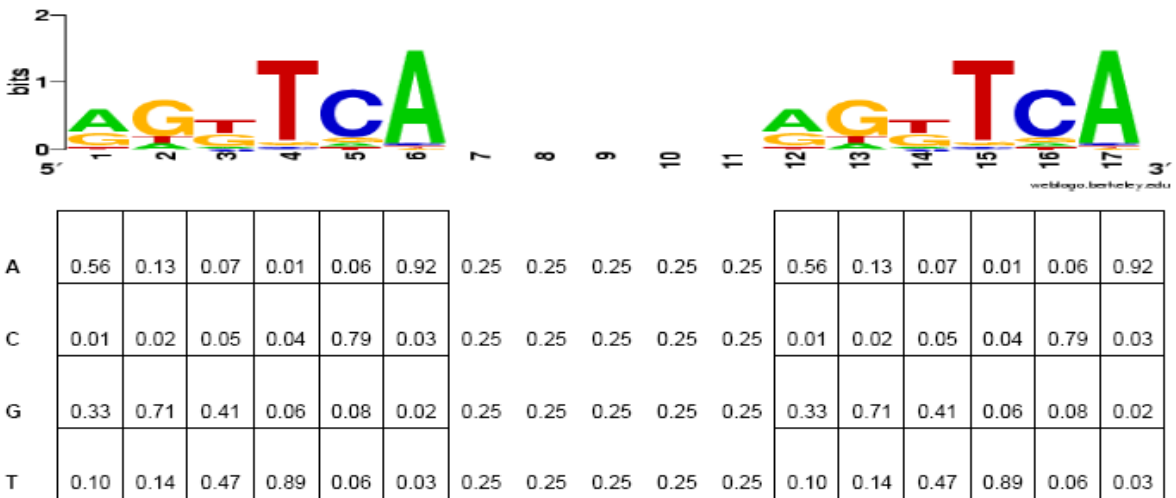


Figure 31 Matrice RARE DR5 dérivée de données de *ChIP-on-chip*

La matrice a été dérivée à partir de DR5 trouvé dans des régions liées par RAR $\alpha$  en ChIP-on-chip. La matrice contient 126 DR5. Les DR5 ayant au plus un nucléotides différents des RARE DR5 de haute affinité RGKTCAnnnnnRGKTCa ont tous été inclus dans la matrice. La moitié des DR5 avec deux nucléotides différents du consensus de haute affinité ont été inclus. La sélection des DR5 avec deux différences a été faite à l’aide des score assignés par une matrice représentant exactement les demi sites de haute affinité (ex. en position 1, A = 50% et G = 50%). Les données de ChIP-on-chip proviennent du projet ENCODE [83].

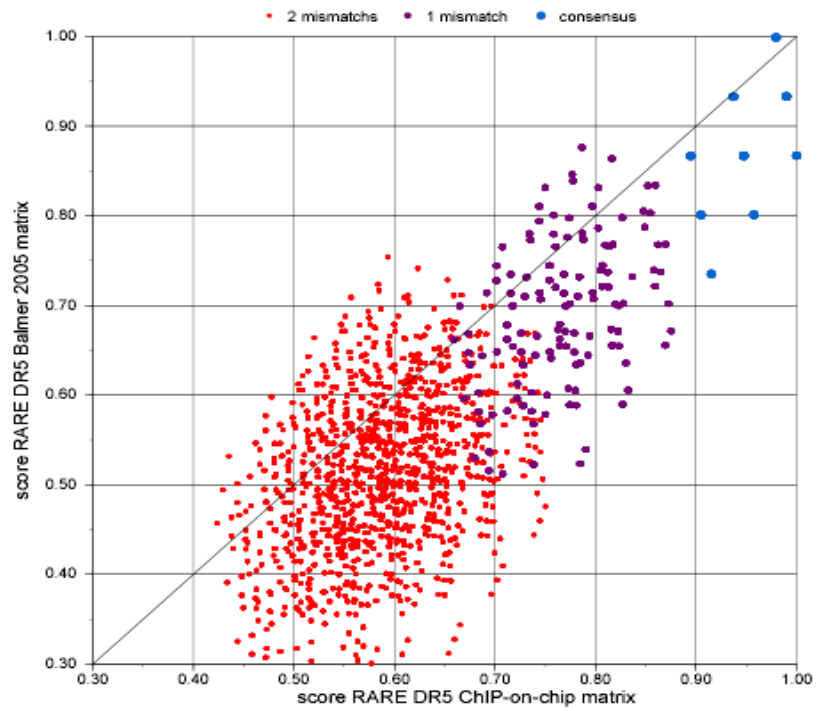


Figure 32 Comparaison des scores assignés par les matrices RARE DR5

Chaque point correspond aux scores assignés par les deux matrices (Figure 30 et 31) à un RARE comptant de 0 à 2 nucléotides différent d'un RARE DR5 de haute affinité (RGKTCAnnnnnRGKTCA).

Table 6 Enrichissement de RARE DR5 dans les promoteurs de gènes cibles observé avec les deux matrices

**Matrice RARE DR5 dérivée de donnée de ChIP-on-chip**

<b>cutoff</b>	<b>0.70</b>	<b>0.75</b>	<b>0.80</b>	<b>0.85</b>
# DR5	121	76	35	17
# genes	81	61	31	15
fold	1.66	1.83	2.60	4.79
Z-score	23.37	22.11	24.12	29.39
P-value	4.11E-07	2.21E-07	1.54E-06	1.96E-06
significant	Yes	Yes	Yes	Yes

**Matrice RARE DR5 compilée de la littérature**

<b>cutoff</b>	<b>0.70</b>	<b>0.75</b>	<b>0.80</b>	<b>0.85</b>
# DR5	97	44	25	10
# genes	65	36	22	10
fold	2.12	2.21	3.23	3.65
Z-score	31.19	22.24	25.51	18.03
P-value	3.01E-08	1.24E-05	3.61E-06	0.00033
significant	Yes	Yes	Yes	Yes

Les RARE DR5 ont été identifiés par les deux matrices (Figure 30 et 31) à +/- 10kb du TSS de 125 gènes cibles des RAR [230].

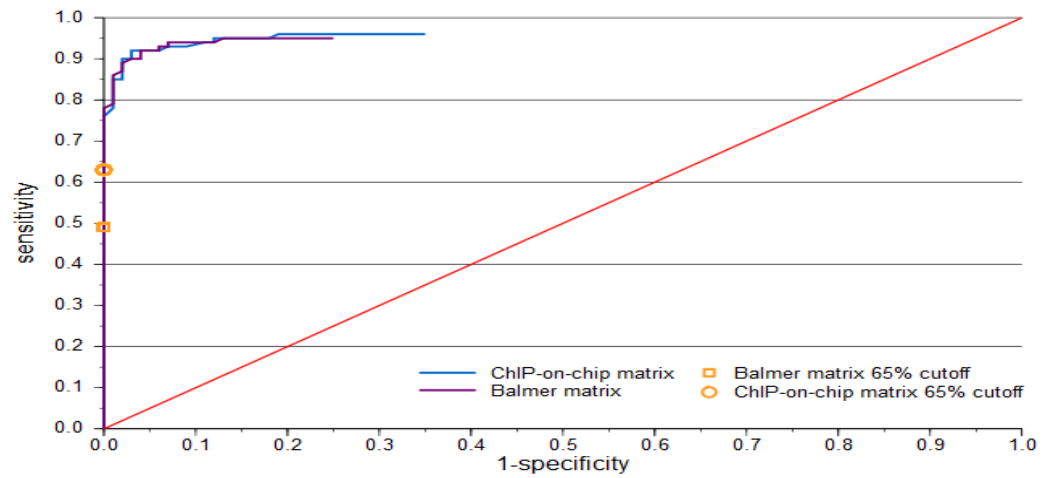


Figure 33 Courbes ROC des matrices RARE DR5

La sensibilité et la spécificité des deux matrices RARE (Figure 30 et 31) ont été calculées à l'aide 500 RARE générés aléatoirement à partir d'une matrice représentant les RARE DR5 de haute affinité RGKTCAnnnnnRGKTCA et 500 séquences aléatoire. La sensibilité a été calculée à l'aide de la formule  $TP/(TP+FN)$  et la spécificité à l'aide de la formule  $TN/(TN+FP)$ . TN = vrai négatif, TP = vrai positif, FN = faux négatif et FP = faux positif.



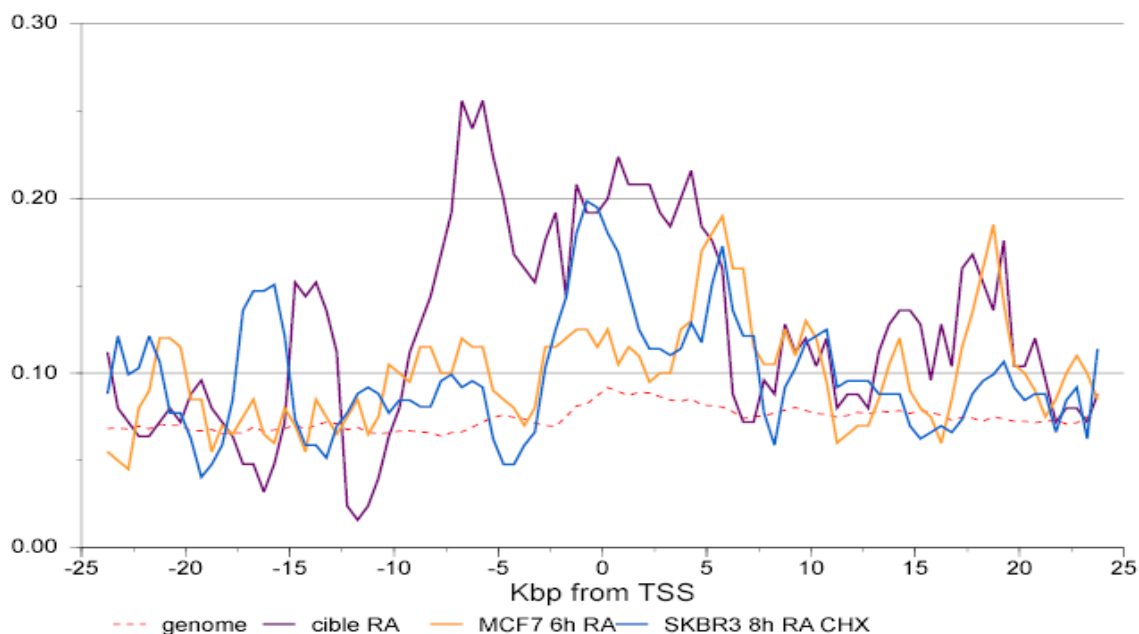


Figure 34 Distribution des RARE DR5 à +/- 25 kpb du TSS de gènes régulés par RA

Les RARE DR5 identifiés avec un seuil de 75% sur le score ont été comptés avec une fenêtre coulissante de 2.5 kpb tous les 500 pb. Les cibles de RA viennent de la revue de Balmer et Blomhoff [230], les gènes « MCF7 » viennent des micropuces de Lamb et collègues [138] et les gènes « SKBR3 » viennent de recherches en cours du laboratoire.

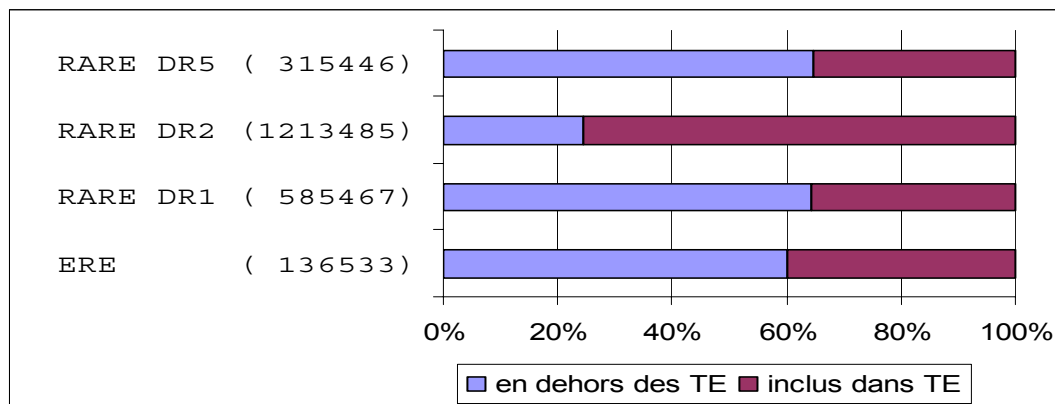


Figure 35 Contribution des TE aux ERE et RARE prédits dans le génome humain

Le nombre de sites prédits dans le génome humain est indiqué entre parenthèse. Un seuil de 65% a été utilisé pour chacune des matrices. La figure a été bâtie avec l'annotation des TE du génome humain version hg17 du UCSC [87]

Table 7 Facteurs de transcription régulés par E2 ayant des matrices disponibles dans TRANSFAC

Gène	régulation E2	site	Matrices TRANSFAC
AHR	Down -CHX	AHR/ARNT	AHR_01 AHRARNT_01 AHRARNT_02 ARNT_01 ARNT_02
FOS	Up -CHX	AP1	AP1_01 AP1_Q2 AP1_Q4 AP1_Q6 AP1FJ_Q2
JUNB	Down - CHX		
TFAP2C	Up + CHX	AP2	AP2_Q6
NFIC	Up +/- CHX	CCAAT box	CAAT_01 NF1_Q6 NFY_01 NFY_Q6
E2F1	Up - CHX	E2F	E2F_01 E2F_02 E2F_03
E2F2	Up - CHX		
E2F5	Down - CHX		
E2F7	Up - CHX		
MYC	Up + CHX	E-box	MYCMAX_01 MYCMAX_02 MYCMAX_03
EGR1	Down - CHX	EGR	EGR1_01
EGR3	Up +/- CHX		EGR3_01
FOXC1	Up +/- CHX	Forkhead	FREAC3_01
FOXO3	Down - CHX		FOXO3_01
FOXA1	Down - CHX		HNF3B_01
ZIC2	Up + CHX	GT-rich	ZIC2_01
MYB	Up +/- CHX	Myb-recognition elements	MYB_Q6
PPARG	Down - CHX	PPRE	PPARG_01 PPARG_02 PPARG_03
NFATC2	Up +/- CHX	NFAT	NFAT_Q6
SOX9	Down - CHX	SOX	SOX9_B1
STAT1	Down - CHX	STAT	STAT_01 STAT1_01
STAT3	Down - CHX		STAT3_01
XBP1	Up +/- CHX	X box	XBP1_01

Table 8 Gènes Up E2+CHX avec des ERE prédits impliqués dans la prolifération et la régulation de la croissance

Processus biologique	Gènes	Nombre	Pourcentage	p-value
prolifération cellulaire	ISG20, TCIRG1, CDK2, PIM1, DLG5, RERG, IGFBP4, AREG	8	7.62%	0.023
régulation de la croissance	CISH, WISP2, RERG, IGFBP4	4	3.81%	0.050

L'annotation des gènes ayant des ERE prédits près de régions lies par ERa en CHIP-on-chip a été faite avec l'outil DAVID (« *Database for Annotation, Visualization, and Integrated Discovery* ») [194]. Les p-value correspondent à un test de Fisher exact comparant le nombre de gènes ayant une annotation fonctionnelle à l'ensemble des gènes du génome.

Table 9 Gènes Up E2-CHX avec des sites E2F prédits impliqués dans la prolifération et le cycle cellulaire

Processus biologique	Gènes	Nombre	Pourcentage	p-value
cycle cellulaire	HSPA2, INCENP, TOP3A, TPX2, KNTC1, BRCA2, DDX11, CDKN3, RBBP8, KIF23, BUB1B, CCDC5, PRC1, CDC45L, ESPL1, KIF22, CNAP1, MIS12, DLG7, BRIP1, CIT, MCM5, CHEK1, PCNA, CHAF1B, NME1, E2F7, CHEK2	28	20.44%	1.243E-13
prolifération cellulaire	BUB1B, TPX2, CSE1L, CHEK1, PCNA, EMP2, NME1, CDKN3, NAB2, DLG7, CKLF	11	8.03%	0.003

L'annotation des gènes ayant des site E2F prédits près de régions lies par E2F1 en CHIP-on-chip a été faite avec l'outil DAVID (« *Database for Annotation, Visualization, and Integrated Discovery* ») [194]. Les p-value correspondent à un test de Fisher exact comparant le nombre de gènes ayant une annotation fonctionnelle à l'ensemble des gènes du génome.

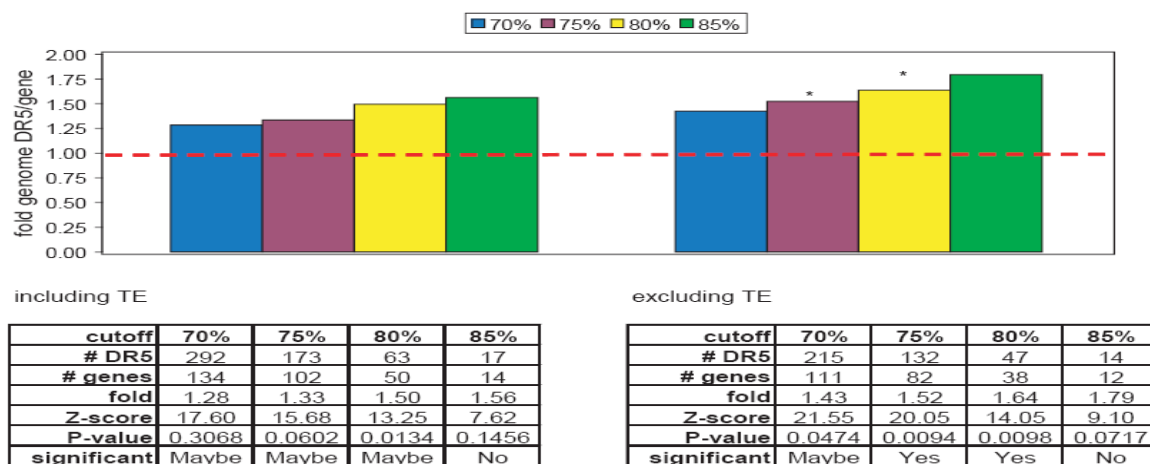


Figure 36 Enrichissement DR5 gènes induits par RA dans les MCF7

L'enrichissement des DR5 a été calculé d'après les sites trouvés à +/- 10 kpb du TSS de 200 induits par RA en présence de cycloheximide dans les cellules MCF7 sur des micropuces Affymetrix. Les micropuces viennent de Lamb *et al* [138]. L'analyse des micropuces a été faite avec les mêmes critères que l'article présenté au chapitre 7 [127].

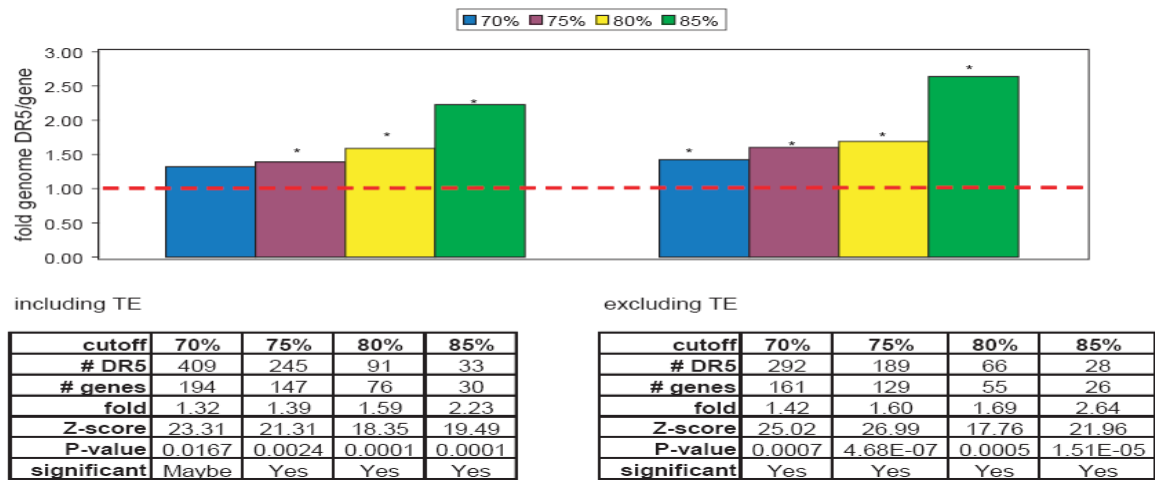


Figure 37 Enrichissement DR5 gènes induits par RA en présence de cycloheximide dans les SKBR3

L'enrichissement des DR5 a été calculé d'après les sites trouvés à +/- 10 kpb du TSS de 272 induits par RA en présence de cycloheximide dans les cellules SKBR3 sur des micropuces Affymetrix. Les micropuces font partie de recherches en cours dans le laboratoire. L'analyse des micropuces a été faite avec les mêmes critères que l'article présenté au chapitre 7 [127].

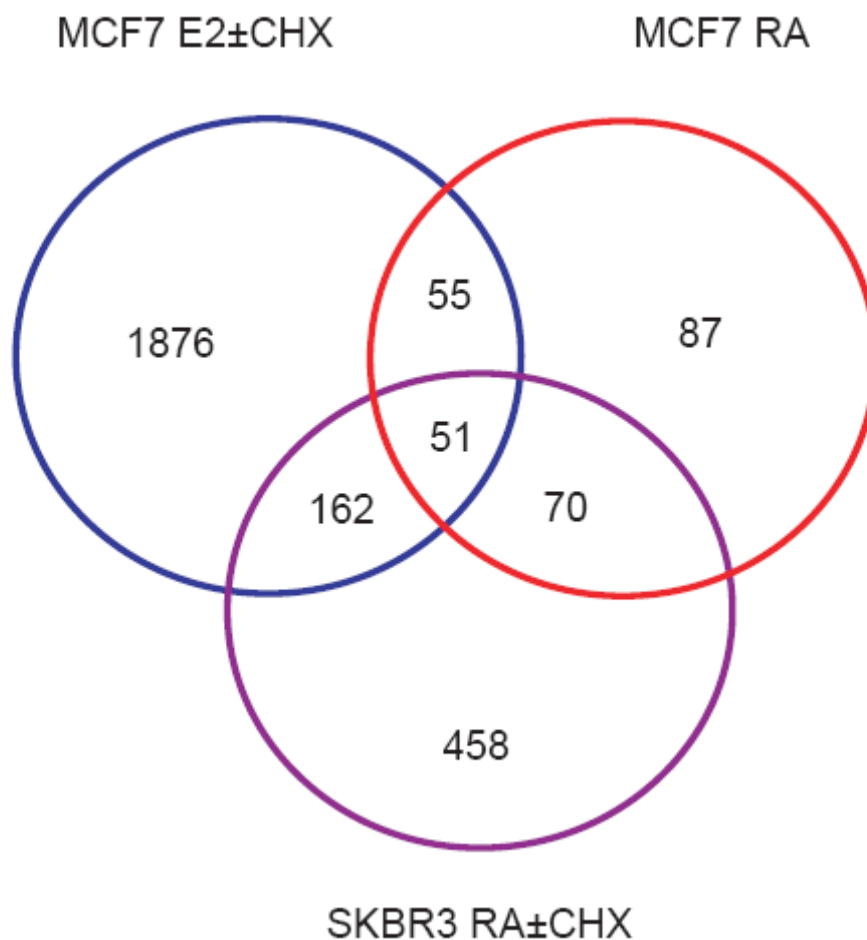


Figure 38 Gènes régulés par E2 et RA

Les gènes pour les traitements E2 viennent de l'article présenté au chapitre 7 [127], ceux pour les traitements avec RA dans les MCF7 viennent de Lamb et al [138] et ceux avec traitements RA dans les SKBR3 viennent de travaux en cours dans le laboratoire. Tous les gènes ont un fold > 1.4 , une valeur de A > 5 et p-value < 0.01 test de t modéré.

Table 10 Gènes ayant des réponses différentes suite à des traitements avec E2 et RA

probeset	Gène	E2-CHX 24h	E2+CHX 24h	RA 6h
218795_at	ACP6	-1.61	-1.14	1.53
213245_at	ADCY1	1.41	1.99	-1.61
202820_at	AHR	-1.42	-1.00	1.67
203180_at	ALDH1A3	-2.75	-1.40	3.10
39248_at	AQP3	-1.75	-1.07	1.61
205431_s_at	BMP5	1.89	1.19	-1.56
201236_s_at	BTG2	-2.43	1.02	7.54
201235_s_at	BTG2	-1.64	-1.01	1.90
217873_at	CAB39	-1.46	-1.46	1.45
218309_at	CAMK2N1	-1.98	-1.13	1.74
203984_s_at	CASP9	-1.47	-1.16	1.89
203757_s_at	CEACAM6	-1.81	-1.01	1.61
221042_s_at	CLMN	-1.51	-1.28	2.60
221541_at	CRISPLD2	-1.59	-1.17	2.70
204971_at	CSTA	-2.68	-1.09	4.94
202295_s_at	CTSH	-1.44	-1.13	1.52
206424_at	CYP26A1	-1.54	-1.81	83.83
202481_at	DHRS3	-2.37	-1.22	10.43
218726_at	DKFZp762E1312	1.98	-1.01	-1.61
204014_at	DUSP4	-1.70	1.03	2.07
204015_s_at	DUSP4	-1.51	-1.03	1.56
202668_at	EFNB2	-1.96	-1.25	1.51
206115_at	EGR3	3.95	2.92	-3.28
210827_s_at	ELF3	-1.86	-1.01	4.70
201510_at	ELF3	-1.75	1.28	4.62
200878_at	EPAS1	-2.68	-1.12	6.16
219694_at	FAM105A	-1.44	-1.51	1.61
211719_x_at	FN1	-1.42	-1.30	1.42
221245_s_at	FZD5	-1.72	-1.33	2.16
221577_x_at	GDF15	-1.72	-1.02	2.92
203394_s_at	HES1	-2.20	1.01	1.42
215933_s_at	HHEX	-1.52	-1.53	2.28
204689_at	HHEX	-1.34	-1.44	1.97
218507_at	HIG2	-1.47	-1.64	1.62
219832_s_at	HOXC13	-2.01	-1.14	2.59
211538_s_at	HSPA2	2.51	1.24	-1.50
201193_at	IDH1	-1.45	-2.00	1.51
212143_s_at	IGFBP3	-2.43	-1.46	1.62
210095_s_at	IGFBP3	-2.42	-1.54	3.07
213839_at	KIAA0500	-1.81	-1.20	4.86
217906_at	KLHDC2	-1.46	-1.38	1.65
212658_at	LHFPL2	-1.51	1.00	1.44
217892_s_at	LIMA1	-2.16	-1.47	2.41
51158_at	LOC400451	-1.68	-1.05	2.84
221880_s_at	LOC400451	-1.51	-1.05	2.86
218729_at	LXN	-1.94	-1.34	1.56
212096_s_at	MTUS1	-1.44	-1.14	1.69
213906_at	MYBL1	13.85	5.27	-1.83
205440_s_at	NPY1R	2.38	2.01	-1.53
204791_at	NR2C1	-1.42	-1.07	1.45
213131_at	OLFM1	1.59	1.40	-1.60
212858_at	PAQR4	1.60	-1.13	-1.80
202464_s_at	PFKFB3	-1.60	1.33	2.25
207222_at	PLA2G10	-1.73	-1.09	2.59
202545_at	PRKCD	-1.57	1.14	2.36
208447_s_at	PRPS1	-1.30	-1.79	1.43
212124_at	RAI17	-1.50	-1.01	1.59
203344_s_at	RBBP8	1.93	2.06	-1.46
221215_s_at	RIPK4	-1.57	-1.10	1.92
210592_s_at	SAT1	-1.73	-1.12	1.86
203453_at	SCNN1A	-2.51	-1.46	2.92
219344_at	SLC29A3	-1.45	-1.11	1.45
218284_at	SMAD3	-1.90	-1.07	2.19
205397_x_at	SMAD3	-1.60	-1.04	1.90
205398_s_at	SMAD3	-1.56	-1.09	2.28
202307_s_at	TAP1	-1.40	-1.33	1.48
205287_s_at	TFAP2C	1.14	1.48	-1.51
201506_at	TGFBI	-2.24	-1.27	4.40
202039_at	TIAF1	-1.47	1.01	3.19
211689_s_at	TMPRSS2	-2.70	-1.35	3.37
217853_at	TNS3	-3.26	-1.32	1.79
35254_at	TRAFD1	-1.76	-1.42	1.61
202837_at	TRAFD1	-1.56	-1.37	1.50
202241_at	TRIB1	-1.58	-1.19	1.51
215111_s_at	TSC22D1	-1.44	-1.08	1.41
208763_s_at	TSC22D3	-2.08	-1.28	2.77
207001_x_at	TSC22D3	-1.76	-1.32	2.18
211965_at	ZFP36L1	-1.63	1.24	1.50
211962_s_at	ZFP36L1	-1.47	1.20	1.52
214670_at	ZKSCAN1	-1.52	1.14	2.37
39891_at	ZNF710	-1.61	-1.03	2.60
213542_at	ZNF710	-1.42	-1.05	2.50

Les facteurs d'enrichissement (fold) pour les traitements E2 viennent de l'article présenté au chapitre 7

[127]et ceux pour les traitements avec RA viennent de Lamb *et al* [138]. Les fold en noir sont statistiquement significatifs (p-value < 0.01 test de t modéré).

Table 11 Gènes ayant des réponses différentes suite à des traitements avec E2 dans les MCF7 et RA dans les SKBR3

probeset	Gène	MCF7			SKBR3		cible
		E2-CHX	E2+CHX	RA-CHX	RA-CHX	RA+CHX	
206424_at	CYP26A1	-1.54	-1.81	83.83	27.10	32.58	primaire E2 & RA
203453_at	SCNN1A	-2.51	-1.46	2.92	2.01	1.41	primaire E2 & RA
201193_at	IDH1	-1.45	-2.00	1.51	1.86	-	primaire E2
1555037_a_at	IDH1	-1.44	-2.04	-	1.98	-	primaire E2
222457_s_at	LIMA1	-2.70	-1.44	-	1.50	-	primaire E2
217892_s_at	LIMA1	-2.16	-1.47	2.41	1.59	-	primaire E2
201236_s_at	BTG2	-2.43	-	7.54	2.57	1.42	primaire RA
218309_at	CAMK2N1	-1.98	-	1.74	1.64	1.49	primaire RA
228302_x_at	CAMK2N1	-1.72	-	-	-	1.51	primaire RA
203984_s_at	CASP9	-1.47	-	1.89	1.51	1.82	primaire RA
202481_at	DHRS3	-2.37	-	10.43	3.84	2.15	primaire RA
238032_at	DHRS3	-1.50	-	-	4.33	2.74	primaire RA
204014_at	DUSP4	-1.70	-	2.07	1.55	2.42	primaire RA
204015_s_at	DUSP4	-1.51	-	1.56	1.69	2.05	primaire RA
202668_at	EFNB2	-1.96	-	1.51	2.72	1.50	primaire RA
202669_s_at	EFNB2	-1.89	-	-	1.98	1.51	primaire RA
210827_s_at	ELF3	-1.86	-	4.70	1.98	1.71	primaire RA
201510_at	ELF3	-1.75	-	4.62	1.67	2.00	primaire RA
221577_x_at	GDF15	-1.72	-	2.92	2.34	1.84	primaire RA
219832_s_at	HOXC13	-2.01	-	2.59	1.64	1.64	primaire RA
218729_at	LXN	-1.94	-	1.56	2.16	1.42	primaire RA
204791_at	NR2C1	-1.42	-	1.45	1.92	1.59	primaire RA
207222_at	PLA2G10	-1.73	-	2.59	1.88	2.67	primaire RA
212124_at	RAI17	-1.50	-	1.59	1.92	1.75	primaire RA
218284_at	SMAD3	-1.90	-	2.19	2.58	1.73	primaire RA
239448_at	SMAD3	-1.89	-	-	-	1.89	primaire RA
205397_x_at	SMAD3	-1.60	-	1.90	2.89	1.74	primaire RA
205398_s_at	SMAD3	-1.56	-	2.28	2.56	1.81	primaire RA
201506_at	TGFBI	-2.24	-	4.40	2.20	1.68	primaire RA
202039_at	TIAF1	-1.47	-	3.19	2.46	1.51	primaire RA



211689_s_at	TMPRSS2	-2.70	-	3.37	5.23	2.42	primaire RA
226553_at	TMPRSS2	-1.92	-	-	4.53	2.35	primaire RA
39891_at	ZNF710	-1.61	-	2.60	1.83	1.58	primaire RA
213542_at	ZNF710	-1.42	-	2.50	1.90	1.47	primaire RA
203757_s_at	CEACAM6	-1.81	-	1.61	1.79	-	secondaire E2 & RA
211657_at	CEACAM6	-1.69	-	-	1.84	-	secondaire E2 & RA
225757_s_at	CLMN	-1.51	-	-	1.63	-	secondaire E2 & RA
221042_s_at	CLMN	-1.51	-	2.60	1.85	-	secondaire E2 & RA
200878_at	EPAS1	-2.68	-	6.16	1.53	-	secondaire E2 & RA
213839_at	KIAA0500	-1.81	-	4.86	1.77	-	secondaire E2 & RA
217906_at	KLHDC2	-1.46	-	1.65	2.23	-	secondaire E2 & RA
212096_s_at	MTUS1	-1.44	-	1.69	2.08	-	secondaire E2 & RA
212858_at	PAQR4	1.60	-	-1.80	-1.42	-	secondaire E2 & RA
202464_s_at	PFKFB3	-1.60	-	2.25	2.09	-	secondaire E2 & RA
221215_s_at	RIPK4	-1.57	-	1.92	1.46	-	secondaire E2 & RA
219344_at	SLC29A3	-1.45	-	1.45	1.67	-	secondaire E2 & RA
202307_s_at	TAP1	-1.40	-	1.48	1.65	-	secondaire E2 & RA
217853_at	TNS3	-3.26	-	1.79	2.01	-	secondaire E2 & RA
202241_at	TRIB1	-1.58	-	1.51	1.46	-	secondaire E2 & RA
211965_at	ZFP36L1	-1.63	-	1.50	1.73	-	secondaire E2 & RA
211962_s_at	ZFP36L1	-1.47	-	1.52	1.62	-	secondaire E2 & RA

Les facteurs d'enrichissement (fold) pour les traitements E2 viennent de l'article présenté au chapitre 7 [127], ceux pour les traitements avec RA dans les MCF7 viennent de Lamb *et al* [138] et ceux avec traitements RA dans les SKBR3 viennent de travaux en cours dans le laboratoire. Les fold sont statistiquement significatifs (p-value < 0.01 test de t modéré).

Table 12 Facteurs de transcription régulés par RA dans les MCF7

<b>probeset</b>	<b>Gene</b>	<b>fold</b>	<b>A</b>	<b>adj. P-value</b>
202820_at	AHR	1.67	8.79	0.0051
206115_at	EGR3	-3.28	8.52	7.11E-009
210827_s_at	ELF3	4.70	8.86	9.96E-012
200878_at	EPAS1	6.16	8.47	2.37E-013
204667_at	FOXA1	1.53	9.05	6.32E-006
204131_s_at	FOXO3A	1.99	10.32	7.36E-007
209603_at	GATA3	-1.48	11.01	0.0003
203394_s_at	HES1	1.42	9.76	0.0013
214639_s_at	HOXA1	9.88	6.27	2.37E-013
206289_at	HOXA4	2.27	6.03	4.08E-008
213844_at	HOXA5	16.05	8.13	8.06E-018
205453_at	HOXB2	1.84	5.66	4.46E-006
206745_at	HOXC11	1.83	7.41	4.48E-005
219832_s_at	HOXC13	2.59	9.40	1.19E-008
207233_s_at	MITF	1.81	6.81	0.0017
205555_s_at	MSX2	-1.58	8.60	0.0033
204791_at	NR2C1	1.45	8.55	0.0036
209262_s_at	NR2F6	1.47	10.71	0.0088
212148_at	PBX1	1.43	11.19	0.0035
218284_at	SMAD3	2.19	8.43	6.50E-008
205286_at	TFAP2C	-1.50	11.02	0.0003
218502_s_at	TRPS1	-1.43	9.77	0.0013
215111_s_at	TSC22D1	1.41	11.00	8.67E-005