Identification of functional genomic targets of ERα and ERRα

Josée Laganière

A thesis submitted to the faculty of Graduate studies and Research
in partial fulfillment of the requirements for the degree
of
Doctor of Philosophy

Department of Biochemistry
McGill University
Montréal, Québec, Canada

Your file Votre référence
ISBN: 978-0-494-32304-5
Our file Notre référence
ISBN: 978-0-494-32304-5

ABSTRACT

The biological activity of estrogens is mediated by estrogen receptors (ERs), members of the superfamily of nuclear receptors. ERs modulate gene transcription through binding to estrogen response elements (EREs) within regulatory regions of their target genes followed by recruitment of coregulator complexes that modify the chromatin structure and allow general transcription factors to activate transcription. The estrogen-related receptors (ERRs) α, β, and γ are orphan receptors that share sequence similarity with the ERs but are not activated by natural estrogens.

ERα is a major player in breast cancer initiation and progression while ERRα is suspected as an important regulator of energy metabolism. Very few direct target genes of ERα and ERRα have been described to date. The goal of the present work was to use functional genomic tools in order to uncover which cellular pathways the receptors regulate. We utilized chromatin immunoprecipitation (ChIP)-based approaches since they allow the study of direct protein-DNA interactions occurring *in vivo* in a given cellular context. We first developed and used a ChIP-cloning technique to identify ERα *cis*-regulatory sequences. This approach indeed allowed the detailed characterization of one region in particular, which was revealed to be essential for estrogen regulation of the retinoic acid receptor α 1 (*RARα1*) gene. In addition, promoter arrays were used following ChIP (a technique called ChIP-on-chip or location analysis) to identified the pathways directly controlled by ERα. These genome-wide studies revealed novel mechanisms of action for ERα, in particular its dependence on another transcription factor, FOXA1, for its recruitment on a subset of targets. We next used a human genetic screen that led to the discovery of a polymorphic autoregulatory element located within the *ERRα* promoter. In promoter binding profiling, ERRα was shown to control sets of genes involved in mitochondrial respiration and biogenesis. Our studies suggest that ERα and ERRα control distinct genetic pathways despite their high structural and functional similarities.

RÉSUMÉ

L'activité biologique des oestrogènes est médiée par les récepteurs des oestrogènes (ER). Les récepteurs apparentés aux oestrogènes (ERR) α, β et γ sont des récepteurs nucléaires orphelins structurellement et fonctionnellement apparentés aux ERs mais qui ne sont pas activés par les oestrogènes naturels. ERα et ERRα possèdent des domaines de liaison à l'ADN similaires leur permettant de lier des éléments de réponse quasi identiques.

Peu de gènes cible pour ERα ont été identifiés à ce jour, et encore moins ceux de ERRα. Pour mieux caractériser les cibles de ERα et ERRα, nous avons opté pour des techniques de génomique fonctionnelle dérivées de l'immunoprécipitation de chromatine (ChIP), privilégiant cette approche car elle permet l'étude des interactions protéine-ADN *in vivo*.

Nous avons premièrement développé et utilisé la technique de "ChIP-clonage" pour identifier des régions régulatrices contrôlées par ERα. Nous avons identifié une région essentielle à l'expression du gène du récepteur de l'acide rétinoïque (RARα). Ensuite, des études génomiques à haut-débit de ChIP suivies d'hybridation sur puce à ADN (ChIP-chip) ont révélé de nouveaux mécanismes d'action pour ERα, en particulier sa dépendance à un autre facteur de transcription, FOXA1, pour son recrutement à un sous-ensemble de promoteurs. Nous avons par la suite utilisé un criblage génétique humain qui a mené à la découverte d'un élément d'autorégulation dans le promoteur de ERRα. La technique de ChIP standard nous a permis dans ce cas de montrer que ERRα endogène occupe son propre promoteur. Avec des études de ChIP-chip représentant 19,000 promoteurs, plusieurs nouveaux gènes cibles de ERRα ont été identifiés. De cette manière, nous avons identifié que ERRα est un important régulateur de gènes mitochondriaux nécessaires pour la production d'ATP. En dernier lieu, ERα et ERRα semblent contrôler des fonctions cellulaires distinctes suggérant des rôles physiologiques et/ou pathologiques indépendents malgré leurs importantes similarités structurales et fonctionnelles.

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Vincent Giguère, for his great support during my years in the laboratory. With the great balance between guidance and freedom and all the opportunities he provided, I believe that it was a really beneficial training and enjoyable period of my life.

I would like to give special thanks to Geneviève, for her help and sharing of ups and downs, and to Cathy, for her understanding and simply for all the laughs and good times. I would also like to thank Annie, Brian and Majid, as well as all other past and present lab members for discussions. Many thanks to Cathy and Brian for the proofreading of this work. I also want to thank our collaborators François Robert and Alain Bataille for their instructions and advice about the ChIP-chip technology. I would also like to extend my thanks to Drs Ronald Evans and Jun Sonoda for the great collaboration experience.

I would like to thank my family and friends for their love and encouragements. My motivation to succeed was also strengthened by those around me who believed in my capacity as a scientist.

I am thankful to the US Department of Defence Breast Cancer Research Program and MUHC for funding.

PREFACE

The guidelines concerning thesis preparation issued by the Graduate and postdoctoral Studies at McGill University reads as follows:

1. Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis.)

2. The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges preceeding and following each manuscript are mandatory.

3. The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts.

The thesis must include the following:
1 a table of contents;
2 a brief abstract in both English and French;
3 an introduction which clearly states the rational and objectives of the research;
4 a comprehensive review of the literature (in addition to that covered in the introduction to each paper);
5 a final conclusion and summary;
6 a thorough bibliography;
7 Appendix containing an ethics certificate in the case of research involving human or animal subjects, microorganisms, living cells, other biohazards and/or radioactive material.

4. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.

5. In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis

as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contributions of Authors" as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defence. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to clearly specify the responsibilities of all the authors of the co-authored papers.

I have chosen to write my thesis according to these guidelines, with three published manuscripts and one additional chapter describing recent results. The thesis is organized into seven chapters: (I) General introduction and literature review, (II-V) three published and one unpublished manuscripts, each containing their own abstract, introduction, materials and methods, results, discussion and references, (VI) general discussion of all results with references, and (VII) describes my contribution to original research.

PUBLICATIONS

ARISING FROM WORK OF THE THESIS

- **Laganière, J.**, Deblois, G. and V. Giguère. Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor α1 gene in breast cancer cells. *Mol Endocrinol. 2005 (6):1584-92.*

- **Laganière J.**, Deblois, G., Lefebvre, C., Bataille, AR, Robert, F., and V. Giguère. Location Analysis of Estrogen Receptor α reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci. 2005, 102:11651-56.*

- **Laganière J**, Tremblay GB, Dufour CR, Giroux S, Rousseau F, Giguère V. A polymorphic autoregulatory hormone response element in the human estrogen-related receptor α (ERRα) promoter dictates peroxisome proliferator-activated receptor γ coactivator-1α control of ERRα expression. *J Biol Chem. 2004; 279(18):18504-10.*

OTHER PUBLICATIONS:

- **Laganière J**, Deblois G, Giguère V. Nuclear receptor target gene discovery using high-throughput chromatin immunoprecipitation. *Methods Enzymol. 2003;364:339-50.*

- Seely J, Amigh KS, Suzuki T, Mayhew B, Sasano H, Giguère V, **Laganiere J**, Carr BR, Rainey WE. Transcriptional regulation of dehydroepiandrosterone sulfotransferase (SULT2A1) by estrogen-related receptor α. *Endocrinology. 2005;146(8):3605-13.*

- Barry J, **Laganière J.** and Giguère V. A single nucleotide in a estrogen related receptor site can dictate mode of binding and PGC-1 activation of targets promoters. *Mol Endocrinology, 2005;20: 302-10.*

- Blanchette, M., Bataille, AR., Chen, X., Poitras, C., **Laganière, J.,** Lefèbvre, C., Deblois, G., Giguère, V., Ferretti, V., Bergeron, D., Coulombe, B., and Robert F. Genome-wide Computational Prediction of Transcriptional Regulatory Modules Reveal New Insights Into Gene Expression. *Genome Res 2006;16:656-68.*

- **Laganière, J.** Giguère V. Détermination du répertoire des gènes cibles et du mode d'action du récepteur de l'œstrogène par la génomique fonctionnelle. *Bull Cancer. 2006;93:883-7.*

- Sonoda, J., **Laganière, J.**, Mehl, I., Robert F., Giguère V., and R. Evans. Nuclear receptor ERRα is an essential effector of IFN-γ induced host defense. Manuscript will be submitted shortly.

CONTRIBUTION OF AUTHORS

Bataille, Alain R. :	Designed and printed the 19K promoter array, contributed to the optimization of the ChIP-chip technology and to our training.
Deblois, Geneviève :	Participated to the generation and validation of ChIP-cloning libraries and the optimization of the ERα ChIP-on-chip.
Dufour, Catherine :	Performed MEF isolation and transfections and proofreading of the manuscript.
Giguère, V.:	Lab director, general guidance.
Giroux, Sylvie :	Performed the high-throughput genotyping of the *ESRRA23* element in the cohort.
Lefebvre, Céline :	Participated in the chromatin stocks preparation and ChIP validations.
Robert, François :	Designed and printed the 19K promoter array, gave advice about ChIP-on-chips results and validation. Revised the manuscript and wrote section about chip production.
Rousseau, François :	Lab director who helped develop the project.
Tremblay, Gilles :	Cloned the *ESRRA* promoter and performed initial *in vitro* characterization of the polymorphic element.

TABLE OF CONTENTS

LIST OF TABLES

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ACTR	Activator of thyroid and retinoic acid receptor
AD	Activation domain
ADN	Acide Désoxyribonucléique
AF-1	Activation function 1
AF-2	Activation function 2
AIB1	Amplified in breast cancer 1
AKT	v-akt murine thymoma viral oncogene homolog
AP1	Activator Protein-1
AR	Androgen receptor
ATF	Activating transcription factor
BCR	Breakpoint cluster region
bHLH-PAS	Basic helix-loop-helix-Per/ARNT/Sim
BMD	Bone mass density
Bp	Base pair
cAMP	Cyclic adenosine monophosphate
CARM1	Coactivator-associated arginine methyltransferase-1
CAT	Chloramphenicol acetyltransferase
CBP	CREB-binding protein
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CoCoA	Coiled-coil coactivator
CoQ	Coenzyme Q
COUP-TF	Chicken ovalbumin upstream promoter transcription factor
CRE	CREB response element
CREB	cAMP response element binding
CRM	*Cis*-regulatory module

CTE	C-terminal extension
CYP19	Cytochrome P450, family 19
CYP450	Cytochrome P450
DAX-1	Dosage-sensitive sex reversal adrenal hypoplasia congenital critical region on the X chromosome gene 1, NR0B1
DES	Diethylstilbestrol
DNA	Deoxyribonucleic acid
E2	17β-Estradiol
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMSA	Electromobility shift assay
ER	Estrogen receptor
ErbB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
ERE	Estrogen response element
ERK	Extracellular Signal-Regulated Kinase
ERKO	ER knock out
ERR	Estrogen-related receptor
ERRE	ERR response element
ESRRA	ERRα gene
FAO	Fatty acid oxidation
Fos	v-fos FBJ murine osteosarcoma viral oncogene homolog
FSH	Follicule stimulating hormone
FXR	Farnesoid X receptor
GATA	GATA binding protein
GCNF	Germ cell nuclear factor
GO	Gene ontology
GR	Glucocorticoid receptor
GRIP1	Glucocorticoid receptor-interacting protein-1
GTF	General transcription factor

GTP	Guanosine triphosphate
HAT	Histone acetyltransferase
HGP	Human genome project
HMT	Histone methyltransferase
HNF-4	Hepatocyte nuclear factor 4
HPG	Hypothalamic-pituitary-gonadal
HRE	Hormone response element
HRT	Hormone replacement therapy
HSP	Heat shock protein
IFN	Interferon
iNOS	Nitric oxide synthase
Jun	Jun oncogene
KO	Knock out
LH	Luteinizing hormone
LM-PCR	Ligation-mediated PCR
LRH-1	Liver receptor homolog-1
LXR	Liver X receptor
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog
MAPK	Mitogen-activated protein kinase
MCAD	Acetyl-Coenzyme A dehydrogenase, medium chain
MEF2	Myocyte enhancer factor 2
MGC	Mammalian gene collection
MMTV	Mouse mammary tumor virus
MR	Mineralocorticoid receptor
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NCoR	Nuclear receptor corepressor
NDUF	NADH dehydrogenase (ubiquinone)
NFkB	Nuclear factor of kappa light polypeptide gene enhancer in B-cells

NGFI-B	Nerve growth factor-induced clone B
NRF	Nuclear respiratory factor
NRIP	Nuclear receptor interacting protein
OHT	4-hydroxytamoxifen
P/CAF	p300/CBP associated factor
p/CIP	p300/CBP interacting protein
PCR	Polymerase chain reaction
PERC	PGC-1 related Estrogen Receptor Coactivator
PGC	Primordial germ cells
PGC-1	PPAR-gamma coactivator-1
PMW	Position-weight matrix
PNR	Photoreceptor-Specific Nuclear Receptor
PPAR	Peroxisome proliferator-activated receptor
PR	Progesterone receptor
PRC	PGC-1-related coactivator
PRMT1	Protein arginine methyltransferase 1
PXR	Pregnane X receptor
qPCR	Quantitative PCR
Rac3	Receptor associated coactivator 3
RAR	Retinoic acid receptor
RNA	Ribonucleic acid
ROR	Retinoid related orphan receptor
ROS	Reactive oxygen species
RSK	Ribosomal S6 kinase
RXR	Retinoid X receptor
S	Svedberg
SAGE	Serial analysis of gene expression
SERM	Selective ER modulator
SHP	Small heterodimer partner
SLC	Solute carrier
SP1	Specificity protein 1

SRA	Steroid Receptor RNA activator
SRC	Steroid receptor coactivator
TFBS	Transcription factor binding site
TFF	Trefoil factor
TK	Thymidine kinase
TR	Thyroid hormone receptor
TRAM-1	Thyroid hormone receptor interacting protein
TSS	Transcriptional start site
UCSC	University of California Santa Cruz
UDP	Uridine diphosphate
UGT	UDP-glucuronosyltransferase (UGT)
VDR	Vitamin D receptor

CHAPTER I: Literature Review

1.1 The Nuclear Receptor Superfamily

The nuclear receptor superfamily forms the largest eukaryotic group of transcription factors. Inducible by small lipophilic ligands, nuclear receptors control specific transcriptional responses involved in numerous developmental and physiological processes. The diseases engendered by the deregulation of their activity highlights their importance for the maintenance of homeostasis. Nuclear receptors were the first transcription factors for RNA polymerase II to be cloned in the mid 1980's (Green et al., 1986; Hollenberg et al., 1985; Miesfeld et al., 1986). The early identification of nuclear receptors GR (glucocorticoid receptor) and ER (estrogen receptor) allowed the cloning of all other receptors that are part of the superfamily (Mangelsdorf et al., 1995).

1.1.1 Nuclear receptors: a functional classification

The nuclear receptor superfamily is composed of 48 receptors in humans that can be subdivided into 3 categories based on prior knowledge about their ligands: the classic endocrine receptors, the orphan receptors and the adopted orphan receptors. Nuclear receptor ligands are small lipophilic molecules with the innate ability to cross the cellular membranes, making them available for target cells bearing an intracellular receptor. The classic endocrine receptor subfamily comprises the receptors for the steroid hormones; estrogens (ER), androgens (AR), progesterone (PR), glucocorticoids (GR), mineralocorticoids (MR), as well as the steroid-derived vitamin D (VDR). They have served as a basis for establishing the classical model for nuclear receptor activation. Other receptors that are part of this subfamily recognize non-steroid ligands such as vitamin A for the retinoic acid receptor (RAR) and thyroid hormone for thyroid hormone receptor (TR). In contrast, some nuclear receptors were discovered without any associated ligand and were thus called "orphans". Since they were not associated to any hormones, most orphan receptors were identified by

experimental screenings based on their sequence similarity to known classic receptors, such as the ERRs and the RORs, for their similarity to ERs and RARs, respectively (Giguère, 1999). Their structural similarities with classic receptors led to the suggestion that unidentified hormone response systems remained to be discovered. Indeed, some receptors were soon linked to various types of small ligands such fatty acids, oxysterols, bile acids and various xenobiotics. These receptors form the group of adopted orphan receptors (Mangelsdorf et al., 1995).

1.1.2 Nuclear receptors: an official nomenclature

Members of nuclear receptor subfamilies were successively named according to their ligand names, their similarity to another receptor or in relation to their acknowledged role at the time of discovery. Each type of receptor comprises a number of different subtypes, each one being represented by a Greek letter corresponding to a particular gene located on different chromosomes. For instance, three RAR genes form the RAR subfamily, with the different subtypes α, β, and γ. In many cases, a subtype can also form distinct transcripts due to the presence of more than one promoter or due to alternative splicing, leading to specific isoforms of the receptors. The various receptor isoforms are each represented by a Greek letter that follows a number (for example, the *RARα* gene contains two distinct promoters that lead to two distinct isoforms, *RARα1* and *RARα2*). However, this nomenclature encounters various problems: because the receptors were identified in diverse species by different groups and in various cellular environments, each receptor was given many different names according to its action in the context of the discovery. Like for the CYP450 gene family, the nuclear receptor family was reorganized on the basis of their molecular phylogeny and named correspondingly (1999; Laudet, 1997; Nebert et al., 1987). The phylogeny of the nuclear receptor superfamily allows the definition of six subfamilies. In this system, each receptor is designed by the letters "NR" (for nuclear receptors) and a three-digit identifier: Arabic numerals are used to distinguish the receptor's subfamily, the group is designed

by a capital letter, followed by a number that identifies the gene (Gronemeyer et al., 2004).

	Name	Ligand	Official Symbol
Endocrine Receptors	ER α β	Estrogens	NR3A1, NR3A2
	PR	Progestins	NR3C3
	AR	Androgens	NR3C4
	GR	Glucocorticoids	NR3C1
	MR	Mineralocorticoids	NR3C2
	RAR α β γ	Retinoic acids	NR1B1, NR1B2, NR1B3
	TR α β	Thyroid hormone T_3	NR1A1, NR1A2
	VDR	Vitamin D	NR1I1
Adopted Orphan Receptors	RXR α β γ	9-cis retinoic acid	NR2B1, NR2B2, NR2B3
	PPAR α δ γ	Fatty acids	NR1C1, NR1C2, NR1C3
	LXR α β	Oxysterols	NR1H2, NR1H3
	FXR	Bile acids	NR1H4
	PXR	Xenobiotics	NR1I2
	CAR	Xenobiotics	NR1I3
Orphan Receptors	SF1	-	NR5A1
	LRH-1	-	NR5A2
	DAX-1	-	NR0B1
	SHP	-	NR0B2
	TLX	-	NR2E1
	PNR	-	NR2E3
	NGFI-B (α β γ or Nur77, Nurr1, Nor1)	-	NR4A1, NR4A2, NR4A3
	ROR α β γ	-	NR1F1, NR1F2, NR1F3
	ERR α β γ	-	NR3B1, NR3B2, NR3B3
	Rev-Erb α β	-	NR1D1, NR1D2
	GCNF	-	NR6A1
	TR 2, TR4	-	NR2C1, NR2C2
	HNF-4 α γ	-	NR2A1, NR2A2
	COUP-TF α β γ	-	NR2F1, NR2F2, NR2F6

Table 1.1. Members of the Nuclear Receptor Superfamily

Nuclear receptors can be divided into 3 groups: the endocrine receptors, the adopted orphan receptors and the orphan receptors. If they have been identified, the receptor's ligands are listed. The official symbols of the nuclear receptors appear in the last column.

1.2 Estrogens and their receptors

Elwood V. Jensen was the first researcher to effectively use tritiated estradiol and administer it to rats to quantify the amount of radioactivity in the blood and different tissues. With these studies, as well as others using a non-oxidizable estradiol, researchers came to the conclusion that tissues sensitive to estrogen stimulation like the uterus contain a specific component that binds estradiol and retains it in the nucleus of these tissues (and not the other metabolites apparently produced by the liver and kidneys) without changing its chemical properties (Jensen and DeSombre, 1973; Jensen et al., 1967; Toft and Gorski, 1966). Using sucrose-density sedimentation, Jensen and coworkers observed that the nuclear protein complex that binds estradiol sediments at a different sedimentation value (S) than in the presence of the hormone, a 9S (which was later associated to the cytoplasmic unliganded receptor-heat shock protein (HSP) heterocomplex) "transforming" to a 5S form (nuclear liganded receptor dimer complex) (Pratt and Toft, 1997). These studies together with others lead to the hypothesis that the receptor changes conformation upon ligand binding, dimerizes, translocates to the nucleus and recruit other proteins and regulate target genes from which emerged the idea of a "two-step" mechanism for estradiol activation, replacing the model accepted in the 1950s which was an enzymatic one (Jensen, 2004; Jensen et al., 1966; Jensen et al., 1968; MacGregor and Jordan, 1998). This two-step mechanism was soon extended to other nuclear receptors and became the classic model for nuclear receptor activation. In the classic model, the non-liganded or "apo" receptors associate with HSP chaperones within the cytoplasm. As a result of ligand incursion within a cell, the receptor recognizes and binds this specific ligand and changes conformation to reach the "holo" or liganded form, causing the release of HSPs. The receptor can then dimerize, translocate to the nucleus and bind DNA in specific regions of the genome containing hormone response elements (HREs), in turn switching on or off networks of genes in response to a specific

hormonal signal. The regulation of gene expression induced by nuclear receptors is a complex process involving the recruitment of multiple coactivators and chromatin-remodeling complexes, which allow the general transcription factors (GTFs) and RNA polymerase II to join with subsequent modification of gene transcription (Figure 1.1).

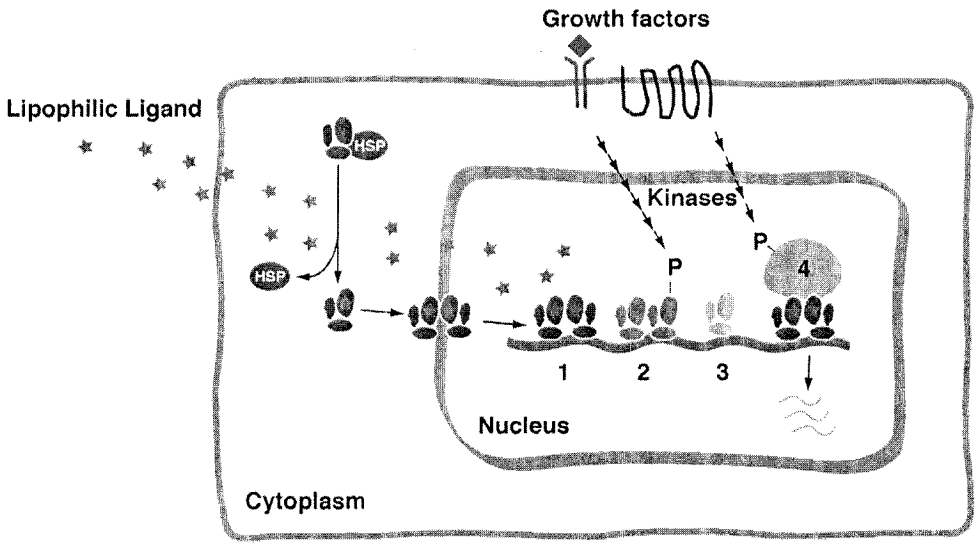


Figure 1.1 Nuclear Receptor activation: the classic model.

When no ligand is available, the nuclear receptor lies in the cytoplasm in an inactive form, in complex with HSP chaperones. When ligands enter the cell and recognize a specific receptor, the receptor changes conformation which allows dimerization, translocation to the nucleus, and recruitment to specific genomic targets, usually containing HREs. Nuclear receptors can homodimerize (1), heterodimerize (2) or act as a monomer (3) to modulate transcription. The receptors can also be modulated via downstream signaling cascades by post-translational modifications. Active receptors can then recruit coregulatory complexes (4), leading to altered RNA polymerase II activity, thus allowing the transcriptional modulation of target genes.

1.2.1 Physiological roles of estrogens

The effects of estradiol are mediated by two estrogen receptors referred to as ERα and ERβ. The ERα was cloned in 1986 in Dr. Chambon's laboratory using the MCF7 breast cancer cell line (Green et al., 1986; Greene et al., 1986). Another form of the receptor, ERβ, was discovered 10 years later and displays 95% and 55% identity to ERα in its DNA binding domain (DBD) and ligand binding domain (LBD), respectively (Enmark et al., 1997; Kuiper et al., 1996; Mosselman et al., 1996). ERα and ERβ bind E2 and a variety of other synthetic compounds with comparable affinities (Kuiper et al., 1997). The hinge region of the two ER subtypes contains nuclear localization signals and displays 30% identity (Enmark et al., 1997). ERα and ERβ can form heterodimers when expressed in the same cell, but it seems that ERα preferentially homodimerizes while ERβ favours heterodimerization with ERα (Cowley et al., 1997; Ogawa et al., 1998; Pettersson et al., 1997; Tremblay et al., 1997). However, RT-PCR and ribonuclease protection assays have indicated that the tissue distribution is quite distinct between ERα and ERβ, minimizing the heterodimerization possibilities in physiological conditions. ERα mRNA is abundant in the uterus, mammary gland, testis, pituitary, liver, kidney, heart and skeletal muscle, whereas ERβ transcripts are present in the ovary and prostate (Byers et al., 1997; Couse et al., 1997; Kuiper et al., 1997). Although some tissues like the ovary, epididymis, thyroid, adrenals, bone and brain express both ERs, detailed studies have often revealed that the two ER subtypes are rarely expressed in the same cells constituting a tissue. For example, in the ovary, ERα is expressed in the theca cells surrounding the granulosa cells of the follicle, where ERβ is merely detectable (Couse et al., 1997; Dotzlaw et al., 1997; Enmark et al., 1997; Hiroi et al., 1999; Sar and Welsch, 1999).

Like other sex steroids, estradiol is derived from the lipid substrate cholesterol. The majority of estrogen production occurs in the theca and

granolusa cells of the ovary, under the control of the luteinizing hormone (LH) and follicule stimulating hormone (FSH) which are released from the pituitary during the menstrual or estrus cycle to regulate ovulation. In men and postmenopausal women, estrogen is not a circulating hormone, but testosterone rather acts like a pro-hormone. Local CYP19 expression, in complex with the NADPH flavoprotein, is responsible for the conversion of testosterone to estradiol in breast, brain, adipose tissue and bone (Kamat et al., 2002; Simpson, 2002; Simpson et al., 2005). Another important precursor of sex hormones derived from cholesterol is the adrenal DHEA, which is produced in greater amount in humans than in other species.

Estrogen receptor gene disruptions and mutations in mice as well as the study of tissue distribution of the levels of receptors expressed have provided colossal information in the understanding of estrogen biology (Dechering et al., 2000; Hewitt et al., 2005). ERα and ERβ KOs (αERKO and βERKO, respectively) exhibit unique phenotypes, which is unsurprising with the examination of their patterns of expression. More profound defects are observed in the αERKO.

ERs have a crucial role in the female reproductive tract and mammary gland (Hewitt et al., 2005; Lubahn et al., 1993). The ablation of ERα in mice leads to infertility in females and more surprisingly, KO males are infertile as well (Couse and Korach, 1999; Hewitt and Korach, 2002). ERβ KO females are subfertile, meaning that they produce smaller litter sizes, but the males are fertile (Hewitt et al., 2005). In brief, the infertile αERKO female mice are anovulatory, but ovulation can be induced by exogenous addition of gonadotropins, indicating that the ovary responds to LH and FSH so these processes do not require ERα, in contrast with the appropriate regulation of the hypothalamic-pituitary-gonadal (HPG) axis, which requires ERα (Couse and Korach, 1999). Proliferation of the epithelial cells of the uterus in preparation to implantation, which usually leads to increased weight of this organ, does not

occur after introduction of exogenous E2 in the αERKO mouse, indicating the need for ERα in this process and explains its immature appearance in the KO.

ERα is needed in somatic cells of the male reproductive tract for proper sperm formation: the αERKO males are characterized by a progressive deterioration of the testicular tissue leading to decreased spermatogenesis. This infertility in males was unsuspected since estrogens were believed to have a role only in female reproduction (Mahato et al., 2000; Mahato et al., 2001).

Optimal fertility also requires the presence of ERβ. This is exemplified by the small litter sizes of the βERKO females, but males preserve complete fertility in contrast with the αERKO. ERβ is highly expressed in the granulosa cells and necessary for effective ovulation, which is inefficient in the KO and hence the cause of their subfertility.

The mammary gland is a highly specialized organ composed of a network of branching epithelial ducts that originate from the nipple with the function of feeding the offspring in mammals. Rudimentary mammary development begins during embryogenesis and is alike in females and males until puberty where the high levels of estrogens and progesterone produced in the ovary cause further growth in females. The αERKO mammary gland fails to develop at puberty. βERKO shows a normal mature and lactating mammary glandular structure, indicating a predominant role for ERα in estradiol-mediated mammary gland function.

1.2.2 Estrogens and diseases

Estrogens have long been associated with the development of several human diseases, namely breast cancer. Nowadays, breast cancer is the most common type of cancer in women, affecting approximately one in nine women in the United States, Canada, and United Kingdom over their life time. Breast cancer

can also occur in males but the incidence is relatively low (less than one in a thousand). Most breast cancers are considered spontaneous, with no clear familial history (Ali and Coombes, 2002).

The mammary gland remains undifferentiated until pregnancy and lactation, where the gland undergoes striking differentiation to produce milk-secreting structures called alveoli. Histologically, the mature mammary gland is comprised of branching ducts lying in a fat pad. The ductal structure is composed of epithelial cells that form the luminal layer and is responsible for milk synthesis. Myoepithelial cells establish an enclosing layer that has the ability to contract the epithelial structure for milk delivery. These ducts are terminated by the alveoli and surrounded by a fibroblast stroma. The αERKO model revealed an important role for E2 in ductal growth given that the mammary gland of the adult female exhibits characteristics similar to that of a wild-type newborn i.e. comprised of the stromal and epithelial portions, connective tissue and a rudimentary ductal tree (Bocchinfuso et al., 1999). Accordingly, it was shown that E2 stimulates cellular proliferation of the mammary duct epithelium and the development of terminal buds through the action of ER, while antiestrogens inhibit this effect (Silberstein et al., 1994). A proportion of epithelial cells of the ducts and stromal cells express ERα in the mammary gland (Ali and Coombes, 2002). However, it seems that the estrogen-induced proliferation occurs in ER-negative epithelial cells and that the ER-positive cells do not divide (Clarke et al., 1997; Russo et al., 1999). The ER-positive cells are thus thought to secrete paracrine factors and promote the proliferation of the ER-negative cells. The most acknowledged idea for spontaneous breast cancer development is that it begins with hyperproliferation of the epithelial cells induced by estrogens, resulting in a mass of proliferating epithelial cells disrupting the highly organized ductal structure of the mammary gland (Ali and Coombes, 2002; Deroo and Korach, 2006). This hyperproliferation is attributed to the binding of estrogen to ERα that augments cell division and DNA synthesis thereby increasing the possibility of

replication errors (Preston-Martin et al., 1990). If not repaired, these errors may affect normal apoptosis, proliferation, as well as DNA repair processes and eventually lead to neoplastic transformation (Hahn and Weinberg, 2002). Another more controversial theory suggests that estradiol metabolites are themselves genotoxic mutagenic carcinogens (Preston-Martin et al., 1990; Yue et al., 2003).

In either way, estrogens have long been recognized to play an important role in the progression of this disease. In the late 1800's, it was observed that in premenopausal women with advanced breast cancer, tumor size increased and decreased during the menstrual cycle. This observation linked estrogens to breast cancer growth and prompted researchers to use oophorectomy (removal of the ovaries) as a means to diminish tumor growth in premenopausal women (Ali and Coombes, 2002). This process was clearly associated with a dramatic decrease in tumor size and improvement of patient prognosis in a proportion of women. However, only about a third of the patients responded to that treatment. Nevertheless, these discoveries paved the way for strategies involving procedures or agents that antagonize the effects of estrogens that soon became a norm in the management of breast cancers (Jensen and Jordan, 2003).

The first clinically useful anti-estrogen for the treatment of breast cancer was tamoxifen (4-hydroxytamoxifen or OHT). Although one of its original proposed applications was to control fertility, it is used in breast cancer management since 1977 and has completely revolutionized the way patients were treated (Lerner and Jordan, 1990). First used only for advanced breast cancers in postmenopausal women as an adjuvant therapy, it was later discovered that tamoxifen was more efficient against the growth of early and ER-positive tumors (Kiang and Kennedy, 1977). Nowadays, it is used as an adjuvant and neoadjuvant therapy and more recently, studied for its efficiency in the prevention of the disease. Maximal benefit is achieved at 5 years of treatment, with a 51% reduction in recurrence and about 28% reduction in

death. These effects are age-independent but of course restricted to ER-positive cancers (1998; Osborne, 1998).

Tamoxifen (brand names are Nolvadex®, Istubal®, and Valodex®) is called a Selective Estrogen Receptor Modulator (SERM) because it displays mixed agonist/antagonist activities (Smith and O'Malley, 2004). For example, tamoxifen is anti-estrogenic in the breast but acts as an estrogen or partial estrogen in bone and uterus (Jordan et al., 1987; Turner et al., 1987; Wolf and Jordan, 1992). This estrogenic effect in the bone is favorable because it contributes to the maintenance of bone density in postmenopausal women. In contrast, post-menopausal patients treated with tamoxifen are considered about 2-3 fold more at risk for endometrial cancer (1998; Fornander et al., 1989).

Tamoxifen has recently been evaluated for its potential to reduce breast cancer incidence in patients at risk for developing that disease. Another SERM, Raloxifene (EVISTA®) was also studied in that objective. Raloxifene has a higher affinity than tamoxifen for the ER. First used as a treatment against osteoporosis, its beneficial effects for the maintenance of bone density are clear (Jones et al., 1984). Raloxifene has the supplemental advantage of not stimulating uterine growth like tamoxifen, thus not increasing the risks for uterine cancer. However, raloxifene has a short biological half-life so dose for dose, tamoxifen is more efficient to reduce breast tumor growth. Unfortunately, neither tamoxifen nor raloxifene have a beneficial effect in the central nervous system. The preliminary results of the clinical trial STAR (Study on Tamoxifen and Raloxifene) indicate that both treatments are as effective in the prevention of breast cancer, reducing the incidence by about 50% for women with high breast cancer risk (Jordan, 2006; Vogel et al., 2006).

Despite the expression of ER in the tumors, a proportion of patients show primary resistance to endocrine therapy. Moreover, most tumors that were initially responsive to anti-estrogens eventually become resistant and

patients require alternative treatments. About two-third of the patients with tamoxifen-resistant disease respond to Faslodex (ICI 182,780 or Fulvestrant®), a pure ER antagonist that works by inhibiting and degrading ER, showing that ER expression is often maintained in these cancers and continues to play a role in tumor growth (Dowsett et al., 2005; Kuukasjarvi et al., 1996; Osborne et al., 2004). Faslodex is used in the treatment of these metastatic ER-positive tumors that have progressed following tamoxifen therapy.

Another way of treating these tamoxifen-resistant tumors in postmenopausal women is using aromatase inhibitors to prevent peripheral androgen conversion to estrogen by the CYP450 (cytochrome p450) aromatase enzyme complex. There are 2 types of aromatase inhibitors: type I irreversible steroidal inhibitors like exemestane (Aromasin®) form a permanent bond with the aromatase enzyme complex, whereas type II non steroidal inhibitors like anastrozole and letrozole (Arimidex® and Femara®) inhibit the enzyme by reversible competition (Miller, 2003).

A subset of cancers has been shown to overexpress the protooncogene human epidermal growth factor receptor 2 (HER-2/neu or ErbB2) (Slamon et al., 1987). ErbB2 is overexpressed in about 30% of breast cancers through gene amplification or augmentation of transcript production. These Her2-overexpressing tumors are typically more aggressive and associated with shorter patient survival (Ross and Fletcher, 1998). ErbB2 is a transmembrane tyrosine receptor kinase member of the erbB family, together with epidermal growth factor receptor (EGFR/HER1), HER3 and HER4. HER family members act through homodimerization or heterodimerization following ligand binding. Dimerization induces tyrosine phosphorylation that results in activation of ras/Raf-mitogen activated protein kinase (MAPK) and phosphoinositide 3-kinase/serine-threonine kinase (PI3K/AKT) regulatory pathways involved in the control of cell proliferation and metastasis. The high levels of ErbB2 in tumors lead to a constitutive activation of the homodimers by phosphorylation

in a ligand-independent manner, resulting in transmission of growth signals in absence of exogenous signals. ErbB2 can also associate with other family members expressed and participate in ligand-dependent signaling. Since ErbB2 overexpression was shown essential for the maintenance of tumor growth in that subset of breast cancers, in addition to the fact that it is found at very low levels in normal adult tissues, ErbB2 was quickly considered an ideal target for therapy (Emens, 2005). Trastuzumab (Herceptin®) has been developed in the view of blocking the receptor's activity for the control of metastasis growth. It is a monoclonal antibody specific for the extracellular domain of ErbB2 that decreases homodimer and heterodimer formation and subsequent activation. Herceptin is a great example of a novel effective targeted therapy thereby illustrating the importance of discovering key players of tumor growth in this heterogenous disease (Le et al., 2005).

Estrogens are important for bone homeostasis in men and women. They increase bone formation and reduce bone resorption. Deficiencies are thus associated with a decreased bone mineral density (BMD) and increased risk for fractures, a disease called osteoporosis, frequent in postmenopausal women. Interestingly, estrogen or hormone replacement therapy (HRT), first introduced as a treatment to reduce symptoms associated with menopause such as hot flushes and vasomotor sweats, as well as raloxifene, are effective in the prevention of osteoporosis (Rossouw et al., 2002).

Lower incidence of strokes is found in premenopausal women compared to men, but the difference is not observed in postmenopausal women compared to men of the same age. Although studies suggest a decreased risk for stroke with HRT in observational studies, a real beneficial effect of estrogens remains controversial due to conflicting results in clinical trials (Brass, 2004). Similarly, since the incidence of cardiovascular disease are low in premenopausal women but higher in men and increased in women after menopause, HRT was thus prescribed for years as a means to protect from cardiovascular diseases.

However, results of recent clinical trials suggest that this procedure may in fact have augmented the number of strokes in patients (Lowe, 2004).

Estrogens are also thought to be neuroprotective, with beneficial effects against Parkinson's and Alzheimer's diseases. Parkinson's disease is more widespread in men than women and some studies indicate a reduced risk in postmenopausal women treated with estrogens. Estrogens would affect the onset and severity of Parkinson's disease by preventing dopaminergic neuron degeneration. However, clinical studies are controversial and some even support an increased severity of the disease with estrogen supplementation (Currie et al., 2004; Saunders-Pullman, 2003; Shulman, 2002). Alzheimer's disease is characterized by an accumulation of hyperphosphorylated tau and β-amyloid proteins that form plaques thought to induce neural cell death. Estrogen might decrease the risk or severity of this disease by augmenting cerebral blood flow and glucose transport and reducing oxidative stress. Estrogens thus seem to reduce risk of developing Alzheimer's disease in women on HRT but have no influence on diminishing the severity of the disease in patients already diagnosed (Pinkerton and Henderson, 2005; Sherwin, 2003).

ERs play a role in the metabolism and location of white adipose tissues and regulate adipogenesis, lipogenesis, lipolysis and adipocyte proliferation (Deroo and Korach, 2006). Estrogen treatments reduce fat accumulations occurring after loss of circulating estrogens in postmenopausal women, thereby implicating estrogens in the control of obesity.

1.2.3 Molecular mechanisms of ERα-mediated gene regulation

1.2.3.1 Transcriptional regulation by ERα

According to the classic model of ER action described above, a multiprotein inhibitory complex in the cytoplasm or nucleus of target cells sequesters

unliganded receptors. The binding of the ligand induces a conformational change within the ER and promotes dimerization and high-affinity binding to the estrogen-response elements (EREs) of target genes followed by recruitment of coregulator proteins. This way, the receptor interacts with coregulatory proteins to recruit the general transcription machinery and/or modify local chromatin structure to positively or negatively modulate gene transcription. This is a model that is still valid in general, but recent studies have highly contributed to detail and update our understanding of ER's mode of action at the molecular level.

One major difference with the classic model is that unliganded ERα is predominantly located in the nucleus and found as monomers and dimers, while the ligand stabilizes the dimer state by an important conformational change (Aumais et al., 1997; Ylikomi et al., 1998). Interestingly, the mRNA of ERα has a half-life of approximately 5 hours, independent of ligand availability (Kenealy et al., 2000). At the protein level, its half-life is of 24 hours in the absence of ligand and about 3-5 hours in the presence of ligand, implicating a ubiquitin-mediated proteasome degradation (Alarid et al., 1999; Nawaz et al., 1999; Preisler-Mashek et al., 2002). Very recently, Frank Gannon's group elegantly demonstrated that ERα mediates ordered, cyclical and combinatorial recruitment of cofactors on the promoter of the *pS2* gene (*TFF1:* trefoil factor 1), a breast cancer biomarker (Metivier et al., 2003). In addition, essential proteasome-mediated degradation of the ER displaces this pre-initiation complex, allowing for promoter clearance and reinitiation of subsequent rounds of transcription.

ER has been shown to interact with numerous coactivators in order to activate gene transcription and they are summarized in Table 1.2. Most cofactors are recruited to the AF2 of ER, like the steroid receptor coactivator (SRC/p160) family of coactivators comprising SRC-1, SRC-2 (TIF2/GRIP1) and SRC-3 (p/CIP: p300/CBP cointegrator-associated protein, RAC3: RAR-

associated coactivator-3, AIB1: Amplified In Breast cancer-1, ACTR and TRAM-1) that share about 40% sequence similarity among each other. SRC-1 was found to interact with almost all nuclear receptors in a ligand-dependent manner. AIB1 was identified in a cancer-relevant context in a study revealing that gene amplification was frequent at the chromosomal region 20q11-12 (Guan et al., 1994). It was later found that AIB1 is a breast and ovarian cancer-amplified and overexpressed ER coactivator (Anzick et al., 1997; Liao et al., 2002). AIB1, like other family members, is composed of several functional domains. The N-terminal basic helix-loop-helix-Per/ARNT/Sim (bHLH-PAS) is involved in dimerization among proteins possessing these motifs while the central region is composed of multiple LXXLL motifs (in which L represents leucine and X represents any amino acid) that mediate interaction with agonist-bound nuclear receptors. An intrinsic activation domain (AD) is located in the C-terminal and was shown to interact with the CBP/p300 cointegrator through LXXLL motifs. This AD domain also displays histone acetyltransferase activities that are however weaker than those of CBP, p300 and p/CAF (Chen et al., 1997; Spencer et al., 1997). High histone acetylation levels are generally associated with increased transcriptional activity of target promoters.

PRMT1 (Protein arginine methyltransferase 1) and CARM1 (coactivator-associated arginine methyltransferase-1) are methyltransferases that can be recruited to promoter regions through the C-terminal region of SRCs and modify the histone methylation state, which in turns activates transcriptional activity of ER in a synergistic manner (Stallcup et al., 2000). CoCoA (coil-coil coactivator) enhances ER target gene expression by associating to p160s through the bHLH-PAS domain (Kim et al., 2003).

Some coactivators like SRA (Steroid Receptor RNA activator) and the p68 RNA helicase are recruited to the AF-1 of ERα and potentiate its activity by interacting with SRC-1 (Deblois and Giguère, 2003; Endoh et al., 1999;

Lanz et al., 1999; Watanabe et al., 2001). Known ERα coactivators are listed in table 1.2.

ER can also recruit corepressor proteins. For instance, NCoR (Nuclear receptor Co-Repressor) is very important for pharmacological actions of anti-estrogens like tamoxifen, leading to chromatin deacetylation through recruitment of histone deacetylases that results in target gene repression. A list of ER corepressors is provided in table 1.3.

1.2.3.2 Alternative mechanisms for regulation by ERα

ERα can also regulate gene expression through alternative regulatory sequences such as AP-1 (Activator Protein-1) and SP1 (Specificity Protein 1). In this situation, ER is tethered to these other transcription factors and acts like a coregulatory protein rather than binding DNA directly.

AP-1 is a very complex transcription factor that functions as a dimer formed between the Jun (v-jun avian sarcoma virus 17 oncogene homolog), Fos (v-fos FBJ murine osteosarcoma viral oncogene homolog), MAF (v-maf musculoaponeurotic fibrosarcoma oncogene homolog) and ATF (activating transcription factor) subfamilies to be transcriptionally active. They recognize either 12-*O*-tetradecanoylphorbol-13-acetate (TPA) response elements (5′-TGAG/CTCA-3′) or cAMP (Cyclic adenosine monophosphate) response elements (CRE, 5′-TGACGTCA-3′) (Karin et al., 1997). ER-AP-1 interactions seem to occur through p160 coactivator proteins: for example, the Jun/Fos heterodimer activates gene transcription by recruiting CBP/p300 and associated proteins, that may include the p160s, which also interacts with estrogen-bound ER and makes a bridge between ER and AP-1 complexes. The LBD of ER does not interact with AP-1 directly, but addition of ER/p160 complex to AP-1/CBP/p300 complex seems to increase the AP-1 DNA binding potential and transcriptional stimulation (Kushner et al., 2000). In addition, ERE or half-site-bound ER in proximity to AP1 on a promoter can lead to synergistic activation,

suggesting cooperation between the two complexes.

Recent studies also show that ERα and SP1 physically interact. ERα recognizes the C-terminal DNA-binding domain of SP1 protein. Through direct interaction, ER can contribute to the regulation of GC-rich SP1 target sequences. This mechanism has been proposed for the estrogen regulation of cathepsin D, c-fos, insulin-like growth factor binding protein 4 (IGFBP4), and RARα1 genes (Safe and Kim, 2004) (See table 1.4).

ER can modulate transcription through recruitment to other transcription factors bound on DNA. As illustrated, ER can be recruited to gene regulatory regions indirectly, through coactivators that are shared between the ER and AP-1 transcription factors for instance. In this case, no direct contact exists between ER and AP-1. ER and SP1 factors can associate directly, and ER can also be recruited to SP1 and regulate these targets, without being in direct contact with DNA (Kushner et al., 2000).

1.2.3.3 Regulation by post-translational modifications

ERα is modified by multiple post-translational modifications that contribute to the modulation of its activity in several ways.

As mentioned previously, ERα ubiquitination through the action of ubiquitin-ligases regulates protein stability through the proteasome pathway, as the proteasome inhibitor MG132 inhibits its degradation (Nawaz et al., 1999).

Lysine residues within the hinge and LBD regions of ERα are modified by p300-mediated acetylation, which is believed to reduce hormonal sensitivity and transactivation activity (Kim et al., 2006; Wang et al., 2001).

ERα phosphorylation integrates estrogen signaling into a complex network with other signaling pathways. ERα is phosphorylated at multiple sites, by both serine/threonine and tyrosine kinases. Serine 118 of ERα is phosphorylated downstream of the Ras/MAPK pathway. This modification can lead to ligand-independent transactivation as well as enhanced ligand-induced activity (Joel et al., 1998; Kato et al., 1995). Phosphorylation of serine 167 by the phosphoinositide 3 (PI3)-kinase pathway (RSK and AKT) increases DNA affinity of ligand-bound ERα. It has been proposed that the oncogenic tyrosine kinase c-src regulates the phosphorylation state of tyrosine 537 and affects ER dimerization and DNA binding (Arnold et al., 1995). These modifications are thought to play a role in ER tissue specific activity as well as in cancers and hormone-independent cancers: since specific signaling pathways can be deregulated in cancer, a resulting aberrant ER phosphorylation can for instance lead to an ligand-independent active ERα (Leclercq, 2002).

Coactivator	Full name	Other names	Function	Activity
SRC-1 (p160)	Steroid receptor coactivator-1	NCoA-1	AF-2 Coactivator	HAT
SRC-2 (p160)	Steroid receptor coactivator-2	GRIP-1, TIF-2, NCoA2	AF-2 Coactivator	HAT
SRC-3 (p160)	Steroid receptor coactivator-3	AIB1, ACTR, p/CIP, RAC3, TRAM-1, NCoA-3	AF-2 Coactivator	HAT
CBP/p300	CREB-binding protein		AF-2 Coactivator	HAT
TRAP220, TRAP/DRIP	Thyroid hormone receptor activating protein of 220 kDa	Mediator; PBP	AF-2 Coactivator	
ASC-1	Activating signal cointegrator-1		AF-2 Coactivator	Binds HATs and NR
ASC-2	Activating signal cointegrator-2	RAP250, TRBP, AIB3	AF-2 Coactivator	Binds HATs and NR
SRA	Steroid receptor activator	Splicing	AF-1 Coactivator	Binds ERα
P68	P68 RNA helicase		RNA helicase	Binds ERα
CARM1	Coactivator-associated arginine methyltransferase-1		Secondary coactivator	Binds to p160s Arginine HMT
PRMT1	Protein methyltransferase 1		Secondary coactivator	Binds to p160s HMT
CoCoA	Coiled-coil coactivator		Secondary coactivator	Binds p160s
E6-AP	E6-associated protein		Ubiquitin ligase	Binds ERα
RPF-1	Receptor potentiating factor-1		Ubiquitin ligase	Binds ERα
PGC-1α PGC-1β	PPARγ coactivator 1		Tethering surface for cofactors; splicing	
CAPER-α	Coactivator of (AP1) and ER		Potentiate ER activity	Binds ER indirectly through ASC2
CAPER-β				
CoAA	Coactivator activator			

Table 1.2. List of ER coactivators

Table 1.2. List of ER coactivators.

Adapted from (Hall and McDonnell, 2005). ER coactivators are listed. They potentiate ER activity through direct recruitment to the AF-2 using their LXXLL motifs, to the AF-1, or indirectly. Coactivators can possess HAT (histone acetyltransferase), HMT (histone methylltransferase) or ubiquitin ligase activities, or contribute to the recruitment of factors influencing ER activity.

Corepressor	Full name	Function/Activity	Repression of ER	Interaction with ER
NCoR	Nuclear receptor corepressor	HDAC	Pharmacological	Binds ER's AF-2
SMRT	Silencing mediator for retinoid and retinoid receptors	HDAC	Pharmacological	Binds ER's AF-2
RIP140 (NRIP)	Receptor interacting protein of 140 kDa	Competes with AF2-coactivators; associates with HDACs	Physiological and pharmacological	Binds ER's AF-2
REA	Repressor of ER activity	Interferes with SRC1	Physiological	Indirect
RTA	Repressor of tamoxifen transcriptional activity	Interferes with SRC1; repressed tamoxifen agonist activity	Pharmacological	Binds ER's AF-1
mSiah2	Mammalian homolog of *Drosophila* Seven in absentia (sina)	Mediates cell-specific repression of NRs by targeting NCoR for proteasomal degradation	Pharmacological	Indirect, binds NCoR

Table 1.3. List of ER corepressors

Adapted from (Hall and McDonnell, 2005). Some ER corepressors are recruited in the presence of pharmacological agents like OHT or other SERMs, or in physiological conditions. Repression occurs through HDAC (histone deacetylases) activity or by competing with coactivators for binding.

1.2.4 Estrogen target gene identification: a historical perspective

ERα was one of the first transcription factors to be identified and search for its target genes has since been of profound interest (Green et al., 1986; Greene et al., 1986). Numerous studies were undertaken with the goal of identifying and cloning estrogen-regulated genes. A first wave of experiments involved differential screening of a cDNA induced by estrogen. A library made from MCF-7 breast cancer cells identified a pS2 clone as estrogen-responsive (Masiakowski et al., 1982). This pS2 gene was then sequenced and its estrogen responsiveness further characterized (Brown et al., 1984; Jakowlew et al., 1984). In normal tissue, *TFF1/pS2* is a gene mainly expressed in the stomach epithelium and salivary gland. Its function in breast cancer is not known, but its expression is almost exclusively restricted to ER-positive tumors (96%) and is associated with longer overall and disease-free survival (Ciocca and Elledge, 2000; Foekens et al., 1990; Rio and Chambon, 1990). The cathepsin D lysosomal proteinase was identified as estrogen-responsive at the protein level in uterine cells (Elangovan and Moulton, 1980). Although modulated by estrogens, whether these genes were direct targets of ER remained inconclusive with these experiments.

Early insights concerning ER mode of binding to DNA came from the identification of the minimal ERE core sequence found within the Xenopus vitellogenin A2 gene, a gene previously identified as estrogen-responsive (Klein-Hitpass et al., 1986). They first performed a gene transfer approach and stably transfected the vitellogenin A2 genomic clone containing both 5' and 3' regions into the estrogen-responsive MCF-7 cell line. They then showed by Northern Blot that estrogen treatment could induce expression of this clone (by activation of endogenous ER). In addition, deletions of Vitellogenin A2 flanking sequences cloned upstream of an *Escherichia coli* (*E. coli*) chloramphenicol acetyltransferase (CAT) reporter gene allowed the

determination of the estrogen responsive element in transient transfections. Finally, they defined the sequence 5'GGTCAcagTGACC as an essential feature of the estrogen response.

Following these discoveries, numerous laboratories undertook experiments with the goal of identifying estrogen-regulated genes and to characterize response elements in the 5' flanking region of the regulated gene. The *TFF-1/pS2* promoter was characterized and an imperfect ERE responsible for ER binding and estrogen response was found within its promoter (Berry et al., 1989; Roberts et al., 1988). Perhaps because of its early characterization, presence of a responsive ERE and its nice inducibility in various systems, *TFF1/pS2* became the prototypic estrogen target gene.

In the 1990s, more powerful methods were developed for the identification of regulated genes and were applied to the discovery of estrogen targets. In 1992, the differential display technique was invented to achieve this goal (Liang et al., 1992; Liang and Pardee, 1992). This technique involves a set of oligonucleotide primers, one annealing to the polyadenylate tail of mRNAs, and the other being short and arbitrary in sequence so that it anneals at different positions relative to the first primer for PCR amplification. Another early high-throughput method for the characterization of gene expression patterns was called Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995). This technique consists in the isolation of mRNA and conversion to cDNA, followed by digestion with a restriction enzyme that cuts frequently. Then, linkers are ligated, followed by cloning into a vector and sequenced. Finally, the number of each sequence is computerized, compiled and analyzed to compare gene expression profiles.

Microarrays were developed concomitantly by Patrick Brown and coworkers: in this technique, cDNAs are labeled with fluorescent dyes and hybridized on a glass array containing cDNA probes corresponding to specific genes (Schena et al., 1995). In 1995 the arrays contained only a few dozen

cDNAs. With the improvement in efficiency of DNA synthesis methods in the early 2000's, introduction of arrays made with synthetic oligonucleotides specific for cDNAs made possible the development of high-density oligonucleotide arrays. The arrays now available for expression profiling experiments represent up to 47,000 transcripts (Tavera-Mendoza et al., 2006). Application of differential display, SAGE and microarray high-throughput gene expression profiling allowed the identification of hundreds of genes up- and down-regulated by estrogen and anti-estrogens in breast cancer, uterine cells and in rat and mouse tissues, as well as comparison of the expression patterns of hormone-dependent versus hormone-independent cancers (Abba et al., 2005; Frasor et al., 2003; Frasor et al., 2004; Ghosh et al., 2000; Hoch et al., 1999; Inadera et al., 2000; Nagai et al., 2003; Seth et al., 2002a; Soulez and Parker, 2001; Thompson and Weigel, 1998; Wang et al., 2004). For a complete database see: http://defiant.i2r.a-star.edu.sg/projects/Ergdb-v2/index.htm (Tang et al., 2004). The main advantage of these techniques is that they allow the analysis of expression levels of multiple genes at a time. However, they do not provide direct information on the genomic action of the transcription factor nor the isolation of the regulatory element used to modulate specific gene expression, and thus necessitates the use of complementary techniques to identify the direct target genes among candidates. Like for the *pS2/TFF1* gene, the 5' flanking regions of regulated genes were analyzed to uncover the regulatory region responsible for estrogen response. The promoter regions of regulated genes were studied one by one using CAT and luciferase reporter assays as well as electromobility shift assays (EMSAs) for the characterization of EREs and discovery of other modes of action of ER. A list of human estrogen targets described as direct by the year 2004 along with their apparent regulatory region is provided in table 1.4 (from O'Lone et al., 2004).

The crystal structure of ERα's DBD in the presence of DNA was solved in 1993 by Daniela Rhodes group (Schwabe et al., 1993). This work has revealed how the ERα protein uses key residues to recognize the ERE

consensus. DNA response elements are now considered like allosteric ligands for nuclear receptors, and it is increasingly clear that variation around the consensus ERE can influence ERα structure and function (Hall et al., 2002; Lefstin and Yamamoto, 1998; Wood et al., 1998). Since natural response elements do not always exactly match the ERE consensus, it is conceivable that promoter or enhancer context will dictate specific coregulator recruitment and provide differential response on target genes.

In an attempt to find the estrogen-responsive region within the promoters of estrogen regulated genes lacking an ERE consensus, some groups provided alternative modes of action of ER. With conventional approaches, ER was shown to modulate transcription through the SP1 or AP1 transcription factors, through HRE half sites, or half-sites in proximity of another site like SP1, as described previously. However, some of these mechanisms of action remain controversial since they were most often described in artificial conditions, in the presence of exogenously overexpressed receptor and/or coactivators and in the absence of a chromatin context (Carroll and Brown, 2006). In addition, gene regulation can clearly occur through distal regulatory regions or enhancers. However, the discovery of these potential functional sites located far away from genes necessitates the knowledge of the complete genomic sequence, which was not available in the 1990s.

Consensus ERE: two consensus half sites

Gene	Responsive region	Reference
EBAG9	–60 to –48	(Ikeda et al., 2000b) (Watanabe et al., 1998)
Efp (ZNF147)	3'UTR	(Ikeda et al., 2000a)
COX7RP	+311 to +3274	(Watanabe et al., 1998)

EREs with one consensus half site

TERT	–2677 to –2655	(Kyo et al., 1999)
pS2 (TFF1)	–405 to –393	(Berry et al., 1989)
Lactoferrin	–414 to +69	(Zhang and Teng, 2000)
Keratin 19	+2414 to +2430 (intron 1)	(Choi et al., 2000)
Oxytocin	–164 to –146	(Richard and Zingg, 1990)
Hageman factor XII	–45 to –29	(Citarella et al., 1996)
Complement 3	–235 to –22	(Fan et al., 1996)
Angiotensin	–26 to –10	(Zhao et al., 1999)
Lipocalin 2	–916 to –800	(Seth et al., 2002b)
Cathepsin D	–145 to –1014 (multiple)	(Wang et al., 1997) (Krishnan et al., 1994) (Cavailles et al., 1993)

EREs with no consensus half-site or with unusual spacing

TGF	–252 to –200	(El-Ashry et al., 1996) (Vyhlidal et al., 2000)
VEGF	–1527 to –1511 +395 to +411 3'UTR	(Mueller et al., 2000)

Genes Regulated by ER that Contain Half-ERE or SFRE Sequences

c-H-ras1	+49 to +78	(Pethe and Shekhar, 1999)
ERα (P1 promoter)	-892 to -420	(Treilleux et al., 1997)
Prothymosin	-1051 to -750	(Martini 2001)
NMDA receptor, 2D	3'UTR	(Watanabe et al., 1998) (Watanabe et al., 1999)
Lactoferrin	-414 to +69	(Zhang and Teng, 2000)

Table 1.4. Genes regulated by ER and their response elements

Genes Regulated by ER that Contain Half-ERE Sequences in Proximity to SP1-Binding Sites

TGFα	-625 to -581	(El-Ashry et al., 1996) (Vyhlidal et al., 2000)
Cathepsin D	-199 to -165	(Wang et al., 1997) (Krishnan et al., 1994) (Cavailles et al., 1993)
RARα	-82 to -62	(Rishi et al., 1995)
Progesterone receptor	+565 to +601	(Petz and Nardulli, 2000)

Genes Regulated by ER that Contain SP1 or Non-SP1-Regulatory Sequences through which ER Indirectly Associates

Adenosine deaminase	-79 to -73 (SP1)	(Xie et al., 1999a)
IGFBP-4	-569 to -540 (SP1) -83 to -54 (SP1)	(Qin et al., 1999)
Hsp27	-105 to -84 (SP1)	(Porter et al., 1997)
Thymidylate synthase	-150 to -142 (SP1)	(Xie et al., 2000)
EGFR	-110 to -84 (SP1)	(Salvatori et al., 2000)
LDL receptor	-142 to +35 (SP1)	(Li et al., 2001)
RAGE	-189 to -166 (SP1)	(Tanaka et al., 2000)
c-fos	-1168 to -1161 (SP1)	(Duan et al., 1998)
Progesterone receptor	B-80 to -34 (SP1)	(Schultz et al., 2003)
VEGF	-66 to -47 (SP3, SP1) inh	(Stoner et al., 2000)
Bcl2	-1603 to -1534 (SP1, ATF-1/CREB)	(Dong et al., 1999)
Cyclin D1	-143 to -110 (SP1) -96 to -29 (ATF-2/c-Jun, CREB/ATF-2)	(Castro-Rivera et al., 2001) (Sabbah et al., 1999)
IGF-1	-324 to -124 (AP-1)	(Umayahara et al., 1994)

Table 1.4. (Continued) Genes regulated by ER and their response elements

This table summarizes the direct human estrogen regulated genes studied by 2004 and their proposed mechanisms of action. It is adapted from table 1 to 4 of (O'Lone et al., 2004).

1.3 The Estrogen-Related Receptors

The first orphan nuclear receptors were identified based on their sequence similarity with the ERs, and were referred to as the estrogen-related receptors (ERRs) (Giguere et al., 1988). There are three isoforms encoded by different genes in this subgroup, called ERRα, β and γ (NR3B1, NR3B2, NR3B3) which display an almost identical DNA-binding domain among each other and exhibit 68% amino acid identity with that of ERα. ERRα and β were the first ERRs to be cloned using a low-stringency screening of a human kidney library with an ERα DBD probe (Giguere et al., 1988). A decade after, the ERRγ gene was identified as part of the critical genomic region linked to Usher syndrome but was later shown to play no role in this disease; the ERRγ protein was also isolated in a yeast two-hybrid screen using the SRC family coactivator GRIP1 (SRC2) as a bait (Eudy et al., 1998; Hong et al., 1999).

1.3.1 Tissue expression of the ERRs and *in vivo* function

The ERRs' physiological roles are starting to be defined with the use of various experiments such as tissue mRNA quantification and gene disruption in mice. ERRα is the most ubiquitously expressed among the ERRs. During development, its expression begins at the chorioallantois fusion in the placenta and is later detected in the heart, intestine, brain, spinal cord, brown fat and bone (Giguere, 2002; Giguere et al., 1988; Vanacker et al., 1998). In the adult, ERRα expression is very broad, with higher levels in skeletal and cardiac muscles, brain, liver, colon and kidney (Giguere et al., 1988; Sladek et al., 1997b). ERRα has been shown to be involved in bone development and important for intestine function (Bonnelye et al., 2001; Bonnelye et al., 1997; Carrier et al., 2004). ERRα knockout mice are viable with no flagrant defect with the exception of a reduced body weight and fat deposits and resistance to high-fat diet induced obesity (Luo et al., 2003).

ERRβ expression is restricted to very low levels in adult liver, muscle and heart. During embryogenesis, ERRβ expression is restricted to extra-embryonic tissues and peaks in the cell population fated to produce the chorion, from day 5.5 until day 8.5 post-coitum (Luo et al., 1997; Pettersson et al., 1996). Indeed, Luo et al. have shown that ERRβ knockout embryos die at 10.5 days post-coitum due to a major defect in early placental formation (Luo et al., 1997). ERRβ is also expressed in primordial germ cells (PGC). After placental complementation of the null mutants with wild-type embryos, it was shown that ERRβ was involved in the proliferation of gonadal germ cells, as both males and females had a reduced PGC cell number. At adulthood, these rescued βERRKO were shown to exhibit behavior abnormalities (Mitsunaga et al., 2004).

ERRγ is most highly expressed in the nervous system both during development and in the adult, where it has specific patterns of expression (Hermans-Borgmeyer et al., 2000; Lorke et al., 2000). At adulthood, the gene is ubiquitously expressed in human and mouse tissues.

1.3.2 Molecular properties of the ERRs and similarity to the ERs

ERRs are orphan nuclear receptors and do not recognize E2 despite their important overall sequence similarity with the ERs. The lower degree of conservation of their LBDs (about 35% identity) appears consistent with the incapacity of ERRs to bind E2. In spite of an intensive hunt for a physiological ligand undertaken in several laboratories, none has been identified yet and ERRs are thus generally considered as constitutive orphan transcriptional activators (Vanacker et al., 1999a).

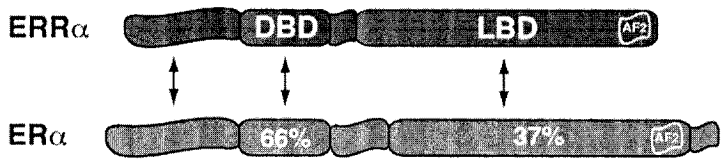


Figure 1.2. ERα and ERRα sequence comparison

ERα and ERRα display 66% and 37% identify in their DBD and LBD, respectively.

Interestingly, high doses of the synthetic OHT has been shown to inhibit ERRβ and ERRγ activities, but has no influence on ERRα (Coward et al., 2001; Greschik et al., 2004; Tremblay et al., 2001a). Similarly, the estrogenic compound diethylstilbestrol (DES), a full ERα agonist, inactivates all three ERR subtypes by dissociating their interaction with coactivators (Coward et al., 2001; Tremblay et al., 2001b). These experiments supported the belief that ERR constitutive activity could be abrogated by ligands. Even though the ERRs are unable to bind E2, it is remarkable that other synthetic estrogenic analogues like SERMs and DES can interact with the ERRs and modulate their activity, showing that ERs and ERRs are somewhat related in their LBDs.

Instead of being activated by a ligand, ERRs rather display a constitutive transcriptional activity that seems dependent on the amount and nature of coregulatory proteins present in the cell. Like ERs and many other nuclear receptors, the ERRs can bind the general coactivators of the SRC family, which in turns provide augmentation of their transcriptional activity (Hong et al., 1999; Lu et al., 2001; Tremblay et al., 2001a; Xie et al., 1999b). They do so in a constitutive manner, in contrast with the ERs that need the presence of their ligand to recruit coactivators. This can be explained by the transcriptionally active conformation that the ERRs adopt in the absence of ligand, a structure that is typically adopted by agonist-bound receptors (Greschik et al., 2002).

The PGC-1 (PPARγ-coactivator 1) family of transcriptional coactivators function as inducible regulators of cellular energy metabolic pathways. It is composed of 3 coregulators, PGC-1α, PGC-1β (also called PERC) and PGC-1-related coactivator (PRC). PGC-1α was first identified from a brown fat cDNA library based on its interaction with PPARγ in a yeast two-hybrid system (Puigserver et al., 1998), but has been shown to bind other nuclear receptors such as ER, RAR and TR in an ligand-dependent manner as well as other

transcription factors like nuclear respiratory factor 1 (NRF-1) and members of the MEF2 (myocyte enhancing factor 2) family (Tcherepanova et al., 2000; Vega et al., 2000; Wu et al., 1999). It was later observed that PGC-1α is a strong modulator of ERRα activity (Schreiber et al., 2003). PGC-1α is remarkable for its tissue selective expression as well as its induction by different physiological states of an animal. It is found at high levels in heart, brown fat, kidney and muscle, tissues with a high energy expenditure. It is induced in a tissue specific manner by signals that relay metabolic needs: exposure to cold induces PGC-1α in brown fat and skeletal muscle whereas starvation augments it in heart and liver (Herzig et al., 2001; Lehman et al., 2000; Yoon et al., 2001).

Interestingly, the expression patterns of PGC-1α correspond to that of ERRα. Moreover, a group has shown that PGC-1α expression is sufficient to induce ERRα expression in cell-based assays (Schreiber et al., 2003). Since PGC-1α controls numerous metabolic functions like mitochondrial biogenesis, oxidative phosphorylation and CREB cycle, and ERRα activity is strongly enhanced by this coactivator, it is apparent that ERRα could participate in the control of metabolic activities.

Because ERR DBD's are similar in sequence to that of ERs, several groups were interested in determining whether they could control the same DNA sequences. Indeed, it was shown that ERRs can also recognize ERE sequences as dimers (Johnston et al., 1997; Vanacker et al., 1999a; Vanacker et al., 1998; Vanacker et al., 1999b). Most often, ERRα was found as a weak transcriptional activator of these ERE-based promoters (like the *pS2/TFF1*, lactoferrin and osteopontin promoters) or to interfere with ERα activity when co-expressed, pointing to a potential role for ERRα in estrogenic pathways (Kraus et al., 2002; Lu et al., 2001; Vanacker et al., 1998; Yang et al., 1996).

An unbiased screen for the determination of the ERRα consensus binding site has revealed that the ERRα monomer recognized a nine base pair DNA sequence (TNAAGGTCA), referred to as ERR response element (ERRE) (Sladek et al., 1997a). The same study showed that the medium-chain acyl coenzyme A dehydrogenase gene (MCAD) promoter contains a perfect ERRE that is actually regulated by ERRα. MCAD is an enzyme that catalyses the first step of the mitochondrial fatty-acid β-oxidation pathway, another early hint potentially implicating ERRα in the control of energy metabolism.

1.3.3 ERRα and disease

Several independent studies have suggested that ERRα could play a role in the breast cancer disease. Because they share important characteristics with ERα, an acknowledged key player in the development and progression of breast cancer, a first area of investigation was dedicated to deciphering the potential role of ERRs in this context. An early finding described ERRα as the most expressed ERR subtype in breast cancer cells and tumors (Giguere, 2002; Lu et al., 2001). Importantly, it has been demonstrated by two different groups that ERRα, in contrast to ERα, is a negative prognostic factor for breast cancer disease-free survival (Ariazi et al., 2002; Suzuki et al., 2004). From a molecular point of view, this is coherent with the observation that increased ERRα and ErbB2/HER2 expression levels are positively associated with advanced, tamoxifen-resistant, and/or ERα-negative tumors. A compilation of breast tumor microarray data divulged by Oncomine (www.oncomine.org) illustrates these correlations. ERRα was also shown to be phosphorylated downstream the EGF (epithelial growth factor) pathway (Barry and Giguere, 2005).

Together, these observations suggest a potential role for ERRα in breast cancer progression and evoke novel avenues towards potential alternative treatments that could provide ERRα specific ligands. Since ERRα is expressed in late breast cancers in contrast with ERα, these ERR-specific modulators

could become a treatment option for more aggressive and anti-estrogen unresponsive cancers, which is a big challenge in the management of breast cancer.

Studies performed in mice indicated that ERRα is expressed during bone development and that it controls bone formation (Bonnelye et al., 2001; Bonnelye et al., 1997). These observations could indicate a role for ERRα in bone homeostasis in the human population. In addition, since a subset of ERRα functions could be related to that of PGC1α, it would be interesting to study whether some metabolic diseases could be controlled by these players.

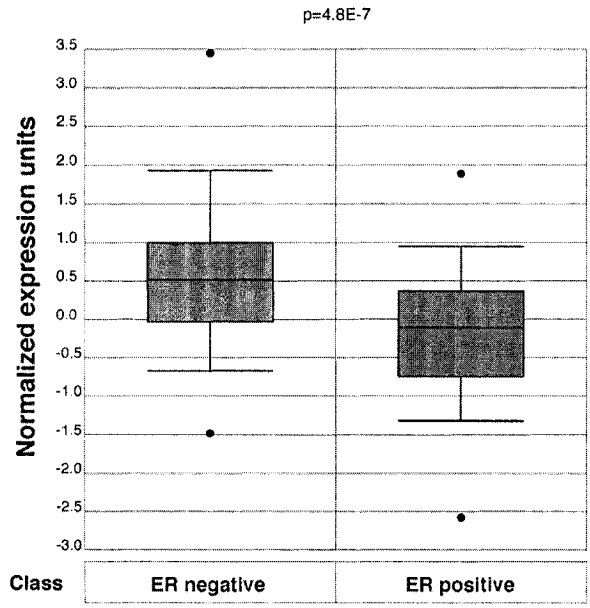


Figure 1.3. ERRα expression inversely correlates with that of ERα in breast tumors

ERRα is expressed to higher levels in ER negative breast cancers compared to ER-positive breast cancers. These results were taken from the Oncomine database (www.oncomine.org). Results are expressed in ERR normalized expression units.

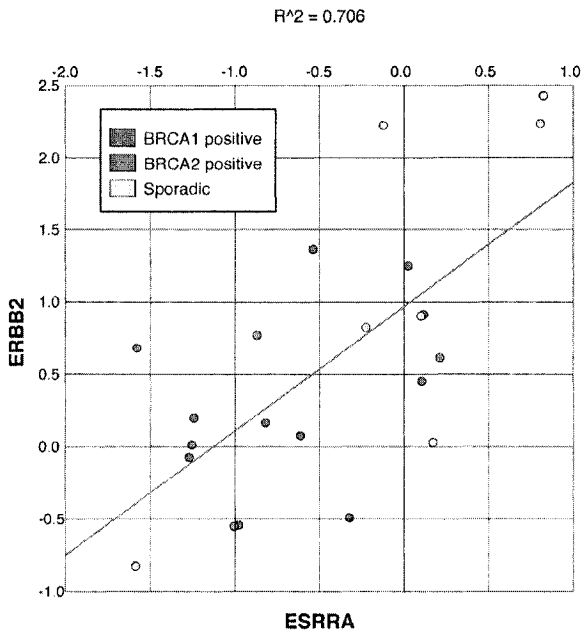


Figure 1.4. ERRα expression correlates with that of ErbB2 in breast tumors

Relative expression levels of ERRα and ErbB2 are compared in sporadic, as well as in BRCA1 and BRCA2 positive tumors.

1.4 Unraveling genomes: impact on target gene discovery

The interest for genome sequences emerged with the discovery of the structure of DNA by James Watson and Francis Crick in 1953 (Watson and Crick, 1953). The first methods for sequencing DNA were published in 1977 and prompted the discussions about the potential benefits of sequencing the entire human genome (Maxam and Gilbert, 1977; Sanger et al., 1977). The Human Genome Project (HGP) worldwide efforts officially started in October 1990, with the goal of sequencing the whole human genome as well as four other model organisms, including *Saccharomyces cerevisiae*, completed in October 1996 (Goffeau et al., 1996). In February 2001, the first draft of the human genome sequence was concomitantly published by the public HGP and the Celera private project, offering boundless potential to scientists in various fields, particularly for the identification and annotation of functional gene regulatory regions (Lander et al., 2001; Venter et al., 2001). The mouse genome was completed in 2002 and is expected to provide a key experimental research tool (Waterston et al., 2002).

1.4.1 Binding site prediction

One of the biggest breakthroughs provided by the availability of the genome sequence is the possibility to study intergenic sequences and their regulatory potential. One way is by searching transcription factor binding sites (TFBS) for a factor of interest. With the advantage of identifying potential response elements at the genome-wide scale, a great deal of interest is dedicated to the mapping of potential regulatory elements for various transcription factors. However, they display several disadvantages. Since most transcription factors including nuclear receptors have great tolerance for variation in their binding sites, the *bona fide* sites are very difficult to decipher. In addition, the example that ER can function indirectly through other factors like SP1 adds to the complexity in predicting functional binding sites. Moreover, these predictions

are made without considering the repressive effect of chromatin structure on transcription factor binding to DNA. Therefore, the predictions made remain hypothetical and may not be functional *in vivo*. Nevertheless, intensive work has been dedicated to develop powerful bioinformatics methods in an attempt to identify binding sites with as much accuracy as possible. Based on the fact that many functional regulatory sequences are expected to be conserved in evolution, bioinformatics methods often include the use of comparative genomics as a first genome screening step for identification of *cis*-regulatory modules (King et al., 2005; Kolbe et al., 2004). One main advantage of this approach is the prediction of many potential binding sites at a genome-wide level, without being limited to a specific cell context (Tavera-Mendoza et al., 2006).

In silico approaches have recently been used for the mapping of ER binding sites in the whole human genome (Bourdeau et al., 2004) or limited to all known promoter sequences (Kamalakaran et al., 2005). The main issue in these procedures is how to define an ERE, since ER is capable of binding to EREs quite variable in sequence, in addition to having the capacity to be recruited to other transcription factors without binding DNA directly.

1.4.2 ChIP-based approaches

Chromatin immunoprecipitation or "ChIP" is a recently developed technique allowing the study of protein-DNA interactions occurring *in vivo*, on a native chromatin structure and a given cellular context.

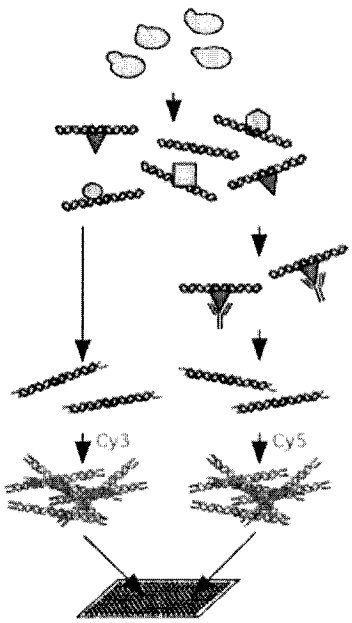
First developed in the 1990s for the study of histone modifications occurring at diverse genomic sites in *Saccharomyces cerevisiae* (Kuo and Allis, 1999), ChIP has been recently adapted to the study of more complex genomes. Briefly, the technique usually begins with a gentle formaldehyde (HCHO) crosslinking of living cells to avoid the loss of factors during the procedure. The fixation of protein-protein and protein-DNA interactions occurs between

the exocyclic amino groups and the endocyclic imino groups of DNA bases in addition to the side-chain nitrogens of lysine, arginine, and histidine (as well as the α-amino groups of all amino acids) (McGhee and von Hippel, 1975a; McGhee and von Hippel, 1975b). Then, cells are lysed and the chromatin is fragmented using sonication or enzymatic digestion. Next, soluble chromatin is immunoprecipitated using a specific antibody against a chromatin-binding factor of interest or, alternatively, against a specific histone modification to enrich in specific protein-DNA complexes. After several washings of the precipitate, crosslinking reversal is carried out, releasing protein-DNA interactions and allowing the analysis of both proteins and DNA. Most often in standard ChIP, only the DNA is purified and analyzed in the aim of determining whether a factor or modification was bound to a specific genomic region, using a non-enriched genomic DNA sample for comparison. This is possible through the use of slot blot, Southern blot or more frequently, quantitative PCR. Thus, standard ChIP offers a great way for *in vivo* validation of putative targets of a factor isolated through other methods like sequence prediction or the gene expression profiling. ChIP was successfully used to characterize ER binding to the *pS2/TFF1* promoter following MCF7 cell treatment with estrogen (Shang et al., 2000).

Since the DNA obtained after the ChIP is enriched in genomic targets of a specific transcription factor, it can be used as a way to identify chromatin-bound transcription factors targets. A genome-wide approach was undertaken and published in December 2000 for the location analysis of yeast Gal4 and Ste12 transcription factors. This work combined ChIP and microarray, where the DNA obtained after Gal4 or Ste12 ChIP was hybridized on a DNA microarray representing the yeast genome for the mapping of transcription factor binding *in vivo* and was thus termed ChIP-on-chip (Ren et al., 2000). The same group published an integrated *in vivo* binding map of most of the *S. Cerevisiae* transcription factors, showing the power of this technology for the understanding of transcriptional regulatory networks (Lee et al., 2002). Such

studies in the human genome would also fill an important gap in the understanding of transcriptional mechanisms but comprises a non negligible additional hurdle brought by the nature of the human genome: with a sequence of $3x10^9$ bp long, the human genome is 250 times bigger than the complete yeast genome and composed at 97% of non-coding sequences containing numerous potential as regulatory sequences. The representation of the whole-genome on arrays is thus an important technical obstacle that needs to be overcome for whole-genome studies of transcription factor binding sites in a single experiment.

The advantage of this approach is the possibility to identify and characterize transcription factor regulatory regions in their native chromatin context and truly occurring in their natural cellular environment. One limitation includes the dependence on very good and specific antibodies. Although this problem can be circumvented by the use of tagged proteins, this latter choice disrupts the principal advantage of ChIP of studying the action of a transcription factor in its natural conditions. Another limitation is the necessity to choose which genomic regions will be represented on the array, until the representation of whole complex genomes is made technically possible.



- Crosslink protein to DNA *in vivo* with formaldehyde

- Break open cells and shear DNA

- Immunoprecipitate

- Reverse-crosslinks, blunt DNA and ligate to unidirectional linkers
- LM-PCR
- Label unenriched DNA with Cy3 and enriched DNA with Cy5

- Hybridize to array

Figure 1.5. ChIP-on-chip or location analysis technique

ChIP begins with the direct crosslinking of cells with formaldehyde, followed by cell lysis and DNA shearing. Then, the chromatin preparation is immunoprecipitated with a specific antibody, the crosslink reversed and the DNA purified. This DNA enriched in genomic targets of a transcription factors is then labelled using by *Klenow* incorporation of Cy5-nucleotides. The sample is hybridized to a microarray together with an unenriched genomic DNA sample that is labelled with the Cy3 dye. Arrays are scanned to evaluate the Cy5/Cy3 ratios and data are analyzed. The figure was adapted from http://www-dsv.cea.fr/art/images/Pascal/bio28_uk_04.gif.

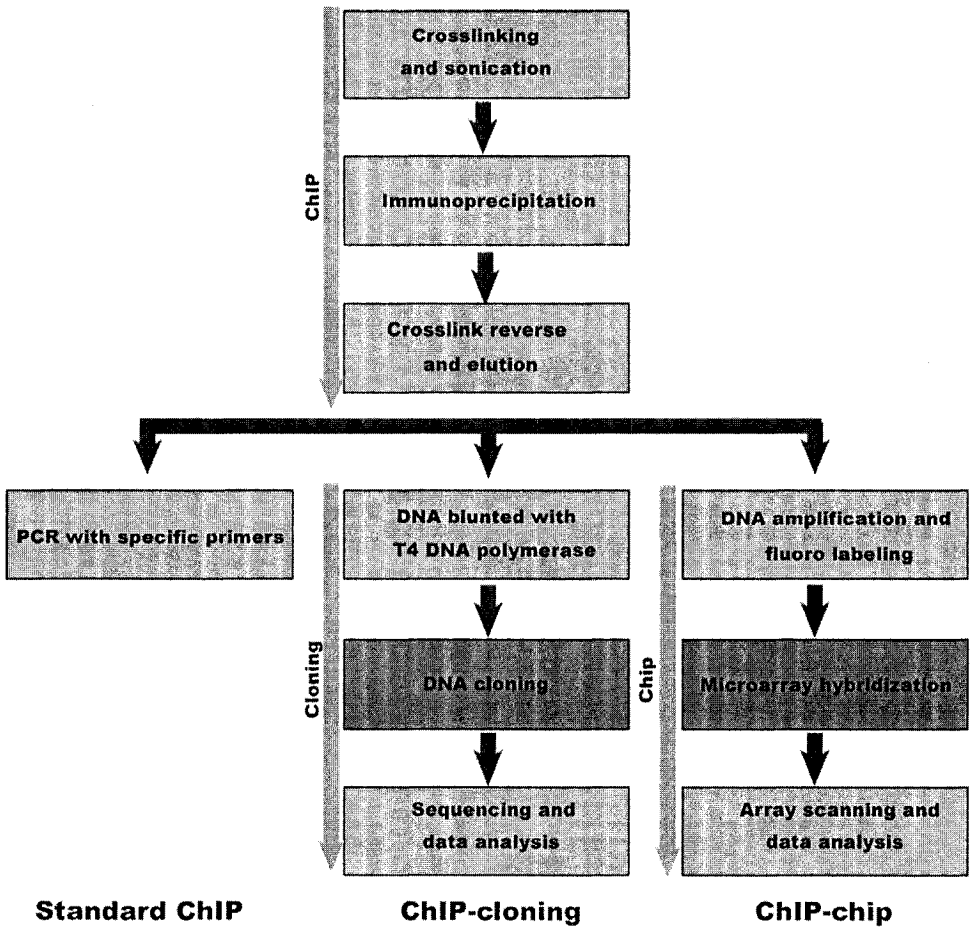


Figure 1.6. ChIP based methods for the identification of direct target genes
Comparison of the ChIP-derived approaches used for the characterization of transcription factor binding to genomic targets *in vivo*. Standard ChIP is used for the verification of binding of a transcription factor to a hypothesized target whereas ChIP-cloning and ChIP-chip are used to discover novel target regions a transcription factor.

1.5 Goals of this thesis

Although nuclear receptor molecular mechanisms of action are quite well understood, especially in the case of ERα, very few direct target genes were identified to date. Since ERRα has been discovered 10 years later, even fewer target genes are known for this receptor. Our comprehension of ERα and ERRα functions is only partial without the knowledge of their binding sites and target genes.

An important challenge is to discover how transcription factors interact with the chromatin within the genome to regulate specific target genes. Since most of the work performed so far in that objective was executed in artificial contexts, our goal was to study ERα and ERRα direct protein DNA interactions truly occurring *in vivo*, which we believe is essential to an accurate cataloguing of their bona fide target genes.

The general goal of this thesis work was to characterize ERα and ERRα actions at the genome level with the special interest of identifying their functional *cis*-regulatory targets. In the aim of better understanding how the receptors interact with chromatin and how they modulate target gene transcription, we developed ChIP and ChIP-derived approaches. Another goal was to compare ERα and ERRα target genes since they possess multiple characteristics in common including the sharing of response elements *in vitro* and reporter assays. In addition, since transcription factors rarely act on their own on their regulatory elements, we paid particular attention to the identification of potential transcriptional partners of ERα and/or ERRα.

In the second chapter, we developed and used a ChIP-cloning approach to uncover novel ERα target regulatory regions (the technique is extensively described in our Methods in Enzymology manuscript figuring in appendix 1).

In the third chapter, we deciphered the direct ERα target promoters in a high-throughput fashion with the development of ChIP-on-chip experiments using a 19K promoter array and discovered a partner of ERα for the regulation of a subset of its target genes. Appendix 2 corresponds to a collaborative work on a bioinformatics prediction of genomic ERα target regulatory modules and our contribution in the ChIP-on-chip verification of these predictions. In the fourth chapter, we used a combination of a human genetic screen together with ChIP and conventional approaches to identify and characterize a novel ERRα functional genomic target. We also participated in another collaborative study where we investigated the role of ERRα in mouse macrophages, again using ChIP followed by hybridization on a mouse 19000 promoter array: chapter five is the summary of our experiments that contributed to characterize the genomic action of ERRα.

BIBLIOGRAPHY

(1998). Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet *351*, 1451-1467.

(1999). A unified nomenclature system for the nuclear receptor superfamily. Cell *97*, 161-163.

Abba, M. C., Hu, Y., Sun, H., Drake, J. A., Gaddis, S., Baggerly, K., Sahin, A., and Aldaz, C. M. (2005). Gene expression signature of estrogen receptor alpha status in breast cancer. BMC Genomics *6*, 37.

Alarid, E. T., Bakopoulos, N., and Solodin, N. (1999). Proteasome-mediated proteolysis of estrogen receptor: a novel component in autologous down-regulation. Mol Endocrinol *13*, 1522-1534.

Ali, S., and Coombes, R. C. (2002). Endocrine-responsive breast cancer and strategies for combating resistance. Nat Rev Cancer *2*, 101-112.

Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X.-Y., Sauter, G., Kallioniemi, O.-P., Trent, J. M., and Meltzer, P. S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science *277*, 965-968.

Ariazi, E. A., Clark, G. M., and Mertz, J. E. (2002). Estrogen-related receptor α and estrogen-related receptor γ associate with unfavorable and favorable biomarkers, respectively, in human breast cancer. Cancer Res *62*, 6510-6518.

Arnold, S. F., Obourn, J. D., Jaffe, H., and Notides, A. C. (1995). Phosphorylation of the human estrogen receptor on tyrosine-537 *in vivo* by *Src* family tyrosine kinases *in vitro*. Mol Endocrinol *9*, 24-33.

Aumais, J. P., Lee, H. S., Lin, R., and White, J. H. (1997). Selective interaction of hsp90 with an estrogen receptor ligand-binding domain containing a point mutation. J Biol Chem *272*, 12229-12235.

Barry, J. B., and Giguere, V. (2005). Epidermal growth factor-induced signaling in breast cancer cells results in selective target gene activation by orphan nuclear receptor estrogen-related receptor alpha. Cancer Res *65*, 6120-6129.

Berry, M., Nunez, A.-M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci U S A *86*, 1218-1222.

Bocchinfuso, W. P., Hively, W. P., Couse, J. F., Varmus, H. E., and Korach, K. S. (1999). A mouse mammary tumor virus-Wnt-1 transgene induces mammary gland hyperplasia and tumorigenesis in mice lacking estrogen receptor-alpha. Cancer Res *59*, 1869-1876.

Bonnelye, E., Merdad, L., Kung, V., and Aubin, J. E. (2001). The orphan nuclear estrogen receptor-related receptor α (ERRα) is expressed throughout osteoblast differentiation and regulates bone formation in vitro. J Cell Biol *153*, 971-984.

Bonnelye, E., Vanacker, J. M., Dittmar, T., Begue, A., Desbiens, X., Denhardt, D. T., Aubin, J. E., Laudet, V., and Fournier, B. (1997). The ERR-1 orphan receptor is a transcriptional activator expressed during bone development. Mol Endocrinol *11*, 905-916.

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.

Brass, L. M. (2004). Hormone replacement therapy and stroke: clinical trials review. Stroke *35*, 2644-2647.

Brown, A. M., Jeltsch, J. M., Roberts, M., and Chambon, P. (1984). Activation of pS2 gene transcription is a primary response to estrogen in the human breast cancer cell line MCF-7. Proc Natl Acad Sci U S A *81*, 6344-6348.

Byers, M., Kuiper, G. G. J. M., Gustafsson, J.-Å., and Park-Sarge, O.-K. (1997). Estrogen receptor-ß mRNA expression in rat ovary: down-regulation by gonadotropins. Mol Endocrinol *11*, 172-182.

Carrier, J. C., Deblois, G., Champigny, C., Levy, E., and Giguere, V. (2004). Estrogen related-receptor α (ERRα) is a transcriptional regulator of apolipoprotein A-IV and controls lipid handling in the intestine. J Biol Chem *279*, 52052-52058.

Carroll, J. S., and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. Mol Endocrinol *20*, 1707-1714.

Castro-Rivera, E., Samudio, I., and Safe, S. (2001). Estrogen regulation of cyclin D1 gene expression in ZR-75 breast cancer cells involves multiple enhancer elements. J Biol Chem *276*, 30853-30861.

Cavailles, V., Augereau, P., and Rochefort, H. (1993). Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate only TATA-dependent transcription in breast cancer cells. Proc Natl Acad Sci U S A *90*, 203-207.

Chen, H., Lin, R. J., Schlitz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatani, Y., and Evans, R. M. (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and form a multimeric activation complex with P/CAF and CBP/p300. Cell *90*, 569-580.

Choi, I., Gudas, L. J., and Katzenellenbogen, B. S. (2000). Regulation of keratin 19 gene expression by estrogen in human breast cancer cells and identification of the estrogen responsive gene region. Mol Cell Endocrinol *164*, 225-237.

Ciocca, D. R., and Elledge, R. (2000). Molecular markers for predicting response to tamoxifen in breast cancer patients. Endocrine *13*, 1-10.

Citarella, F., Misiti, S., Felici, A., Farsetti, A., Pontecorvi, A., and Fantoni, A. (1996). Estrogen induction and contact phase activation of human factor XII. Steroids *61*, 270-276.

Clarke, R. B., Howell, A., Potten, C. S., and Anderson, E. (1997). Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Res *57*, 4987-4991.

Couse, J. F., and Korach, K. S. (1999). Estrogen receptor null mice: what have we learned and where will they lead us? Endocr Rev *20*, 358-417.

Couse, J. F., Lindzey, J., Grandien, K., Gustafsson, J. A., and Korach, K. S. (1997). Tissue distribution and quantitative analysis of estrogen receptor-α (ER-α) and estrogen receptor-β (ER-β) messenger ribonucleic acid In the wild-type and ER-α-knockout mouse. Endocrinology *138*, 4613-4621.

Coward, P., Lee, D., Hull, M. V., and Lehmann, J. M. (2001). 4-Hydroxytamoxifen binds to and deactivates the estrogen-related receptor γ. Proc Natl Acad Sci U S A *98*, 8880-8884.

Cowley, S. M., Hoare, S., Mosselman, S., and Parker, M. G. (1997). Estrogen receptors α and β form heterodimers on DNA. J Biol Chem *272*, 19858-19862.

Currie, L. J., Harrison, M. B., Trugman, J. M., Bennett, J. P., and Wooten, G. F. (2004). Postmenopausal estrogen use affects risk for Parkinson disease. Arch Neurol *61*, 886-888.

Deblois, G., and Giguère, V. (2003). Ligand-independent coactivation of ERα AF-1 by steroid receptor RNA activator (SRA) via MAPK activation. J Steroid Biochem Mol Biol *85*, 123-131.

Dechering, K., Boersma, C., and Mosselman, S. (2000). Estrogen receptors alpha and beta: two receptors of a kind? Curr Med Chem *7*, 561-576.

Deroo, B. J., and Korach, K. S. (2006). Estrogen receptors and human disease. J Clin Invest *116*, 561-570.

Dong, L., Wang, W., Wang, F., Stoner, M., Reed, J. C., Harigai, M., Samudio, I., Kladde, M. P., Vyhlidal, C., and Safe, S. (1999). Mechanisms of transcriptional activation of bcl-2 gene expression by 17beta-estradiol in breast cancer cells. J Biol Chem *274*, 32099-32107.

Dotzlaw, H., Leygue, E., Watson, P. H., and Murphy, L. C. (1997). Expression of estrogen receptor-β in human breast tumors. J Clin Endocrinol Metab *82*, 2371-2374.

Dowsett, M., Nicholson, R. I., and Pietras, R. J. (2005). Biological characteristics of the pure antiestrogen fulvestrant: overcoming endocrine resistance. Breast Cancer Res Treat *93 Suppl 1*, S11-18.

Duan, R., Porter, W., and Safe, S. (1998). Estrogen-induced c-fos protooncogene expression in MCF-7 human breast cancer cells: role of estrogen receptor Sp1 complex formation. Endocrinology *139*, 1981-1990.

El-Ashry, D., Chrysogelos, S. A., Lippman, M. E., and Kern, F. G. (1996). Estrogen induction of TGF-alpha is mediated by an estrogen response element composed of two imperfect palindromes. J Steroid Biochem Mol Biol *59*, 261-269.

Elangovan, S., and Moulton, B. C. (1980). Progesterone and estrogen control of rates of synthesis of uterine cathepsin D. J Biol Chem *255*, 7474-7479.

Emens, L. A. (2005). Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. Am J Ther *12*, 243-253.

Endoh, H., Maruyama, K., Masuhiro, Y., Kobayashi, Y., Goto, M., Tai, H., Yanagisawa, J., Metzger, D., Hashimoto, S., and Kato, S. (1999). Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor α. Mol Cell Biol *19*, 5363-5372.

Enmark, E., Pelto-Huikko, M., Grandien, K., Lagercrantz, S., Lagercrantz, J., Fried, G., Nordenskjöld, M., and Gustafsson, J.-Å. (1997). Human estrogen receptor β-gene structure, chromosomal localization, and expression pattern. J Clin Endocrinol Metab *82*, 4258-4265.

Eudy, J. D., Yao, S., Weston, M. D., Ma-Edmonds, M., Talmage, C. B., Cheng, J. J., Kimberling, W. J., and Sumegi, J. (1998). Isolation of a gene encoding a novel member of the nuclear receptor superfamily from the critical region of Usher syndrome type IIa at 1q41. Genomics *50*, 382-384.

Fan, J. D., Wagner, B. L., and McDonnell, D. P. (1996). Identification of the sequences within the human complement 3 promoter required for estrogen responsiveness provides insight into the mechanism of tamoxifen mixed agonist activity. Mol Endocrinol *10*, 1605-1616.

Foekens, J. A., Rio, M. C., Seguin, P., van Putten, W. L., Fauque, J., Nap, M., Klijn, J. G., and Chambon, P. (1990). Prediction of relapse and survival in breast cancer patients by pS2 protein status. Cancer Res *50*, 3832-3837.

Fornander, T., Rutqvist, L. E., Cedermark, B., Glas, U., Mattsson, A., Silfversward, C., Skoog, L., Somell, A., Theve, T., Wilking, N., and et al. (1989). Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. Lancet *1*, 117-120.

Frasor, J., Danes, J. M., Komm, B., Chang, K. C., Lyttle, C. R., and Katzenellenbogen, B. S. (2003). 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 *144*, 4562-4574.

Frasor, J., Stossi, F., Danes, J. M., Komm, B., Lyttle, C. R., and Katzenellenbogen, B. S. (2004). Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. Cancer Res *64*, 1522-1533.

Ghosh, M. G., Thompson, D. A., and Weigel, R. J. (2000). PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res *60*, 6367-6375.

Giguere, V. (2002). To ERR in the estrogen pathway. Trends Endocrinol Metab *13*, 220-225.

Giguère, V. (1999). Orphan nuclear receptors: from gene to function. Endocr Rev *20*, 689-725.

Giguere, V., Yang, N., Segui, P., and Evans, R. M. (1988). Identification of a new class of steroid hormone receptors. Nature *331*, 91-94.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. Science *274*, 546, 563-547.

Green, S., Walter, P., Kumar, V., Krust, A., Bornet, J. M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erbA. Nature *320*, 134-139.

Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. (1986). Sequence and expression of human estrogen receptor complementary DNA. Science *231*, 1150-1154.

Greschik, H., Flaig, R., Renaud, J. P., and Moras, D. (2004). Structural basis for the deactivation of the estrogen-related receptor γ by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. J Biol Chem *279*, 33639-33646.

Greschik, H., Wurtz, J. M., Sanglier, S., Bourguet, W., van Dirsselaer, A., Moras, D., and Renaud, J. P. (2002). Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. Mol Cell *9*, 303-313.

Gronemeyer, H., Gustafsson, J. A., and Laudet, V. (2004). Principles for modulation of the nuclear receptor superfamily. Nat Rev Drug Discov *3*, 950-964.

Guan, X. Y., Meltzer, P. S., Dalton, W. S., and Trent, J. M. (1994). Identification of cryptic sites of DNA sequence amplification in human breast cancer by chromosome microdissection. Nat Genet *8*, 155-161.

Hahn, W. C., and Weinberg, R. A. (2002). Rules for making human tumor cells. N Engl J Med *347*, 1593-1603.

Hall, J. M., and McDonnell, D. P. (2005). Coregulators in nuclear estrogen receptor action: from concept to therapeutic targeting. Mol Interv *5*, 343-357.

Hall, J. M., McDonnell, D. P., and Korach, K. S. (2002). Allosteric regulation of estrogen receptor structure, function, and coactivator recruitment by different estrogen response elements. Mol Endocrinol *16*, 469-486.

Hermans-Borgmeyer, I., Susens, U., and Borgmeyer, U. (2000). Developmental expression of the estrogen receptor-related receptor γ in the nervous system during mouse embryogenesis. Mech Dev *97*, 197-199.

Herzig, S., Long, F., Jhala, U. S., Hedrick, S., Quinn, R., Bauer, A., Rudolph, D., Schutz, G., Yoon, C., Puigserver, P., *et al.* (2001). CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. Nature *413*, 179-183.

Hewitt, S. C., Harrell, J. C., and Korach, K. S. (2005). Lessons in estrogen biology from knockout and transgenic animals. Annu Rev Physiol *67*, 285-308.

Hewitt, S. C., and Korach, K. S. (2002). Estrogen receptors: structure, mechanisms and function. Rev Endocr Metab Disord *3*, 193-200.

Hiroi, H., Inoue, S., Watanabe, T., Goto, W., Orimo, A., Momoeda, M., Tsutsumi, O., Taketani, Y., and Muramatsu, M. (1999). Differential immunolocalization of estrogen receptor alpha and beta in rat ovary and uterus. J Mol Endocrinol *22*, 37-44.

Hoch, R. V., Thompson, D. A., Baker, R. J., and Weigel, R. J. (1999). GATA-3 is expressed in association with estrogen receptor in breast cancer. Int J Cancer *84*, 122-128.

Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G., and Evans, R. M. (1985). Primary structure and expression of a functional human glucocorticoid receptor cDNA. Nature *318*, 635-641.

Hong, H., Yang, L., and Stallcup, M. R. (1999). Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. J Biol Chem *274*, 22618-22626.

Ikeda, K., Orimo, A., Higashi, Y., Muramatsu, M., and Inoue, S. (2000a). Efp as a primary estrogen-responsive gene in human breast cancer. FEBS Lett *472*, 9-13.

Ikeda, K., Sato, M., Tsutsumi, O., Tsuchiya, F., Tsuneizumi, M., Emi, M., Imoto, I., Inazawa, J., Muramatsu, M., and Inoue, S. (2000b). Promoter analysis and chromosomal mapping of human EBAG9 gene. Biochem Biophys Res Commun *273*, 654-660.

Inadera, H., Hashimoto, S., Dong, H. Y., Suzuki, T., Nagai, S., Yamashita, T., Toyoda, N., and Matsushima, K. (2000). WISP-2 as a novel estrogen-responsive gene in human breast cancer cells. Biochem Biophys Res Commun *275*, 108-114.

Jakowlew, S. B., Breathnach, R., Jeltsch, J. M., Masiakowski, P., and Chambon, P. (1984). Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. Nucleic Acids Res *12*, 2861-2878.

Jensen, E. V. (2004). From chemical warfare to breast cancer management. Nat Med *10*, 1018-1021.

Jensen, E. V., and DeSombre, E. R. (1973). Estrogen-receptor interaction. Science *182*, 126-134.

Jensen, E. V., Desombre, E. R., Hurst, D. J., Kawashima, T., and Jungblut, P. W. (1967). Estrogen-receptor interactions in target tissues. Arch Anat Microsc Morphol Exp *56*, 547-569.

Jensen, E. V., Jacobson, H. I., Flesher, J. W., Saha, N. N., Gupta, G. N., Smith, S., Colucci, V., Shiplacoff, D., Neuman, H. G., Desombre, E. R., and Jungblut, P. W. (1966). Estrogen receptors in target tissues, In Steroid Dynamics, G. Pincus, T. Nakao, and J. F. Tait, eds. (New York: Academic Press), pp. 133-156.

Jensen, E. V., and Jordan, V. C. (2003). The estrogen receptor: a model for molecular medicine. Clin Cancer Res *9*, 1980-1989.

Jensen, E. V., Suzuki, T., Kawashima, T., Stumpf, W. E., Jungblut, P. W., and DeSombre, E. R. (1968). A two-step mechanism for the interaction of estradiol with rat uterus. Proc Natl Acad Sci U S A *59*, 632-638.

Joel, P. B., Traish, A. M., and Lannigan, D. A. (1998). Estradiol-induced phosphorylation of serine 118 in the estrogen receptor is independent of p42/p44 mitogen-activated protein kinase. J Biol Chem *273*, 13317-13323.

Johnston, S. D., Liu, X., Zuo, F., Eisenbraun, T. L., Wiley, S. R., Kraus, R. J., and Mertz, J. E. (1997). Estrogen-related receptor α1 functionally binds as a

monomer to extended half-site sequences including ones contained within estrogen-response elements. Mol Endocrinol *11*, 342-352.

Jones, C. D., Jevnikar, M. G., Pike, A. J., Peters, M. K., K., P. M., Black, L. J., Thompson, A. R., Falcone, J. F., and Clemens, J. A. (1984). Antiestrogens. 2. Structure-activity studies in a series of 3-aryl-2 arylbenzo[b]thiophene derivatives leading to [6-hydroxy-2-(4-hydroxyphenyl)benzo[b]thien-3-yl][4-[2-(1-piperidinyl0ethoxy]-phenyl]methanone hydroxyclhoride (LY156758), a remarkably effective estrogen antagonist with only minimal intrinsic estrogenicity. J Med Chem *27*, 1057-1066.

Jordan, V. C. (2006). Optimising endocrine approaches for the chemoprevention of breast cancer Beyond the Study of Tamoxifen and Raloxifene (STAR) Trial. Eur J Cancer *42*, 2909-2913.

Jordan, V. C., Phelps, E., and Lindgren, J. U. (1987). Effects of anti-estrogens on bone in castrated and intact female rats. Breast Cancer Res Treat *10*, 31-35.

Kamalakaran, S., Radhakrishnan, S. K., and Beck, W. T. (2005). Identification of estrogen-responsive genes using a genome-wide analysis of promoter elements for transcription factor binding sites. J Biol Chem *280*, 21491-21497.

Kamat, A., Hinshelwood, M. M., Murry, B. A., and Mendelson, C. R. (2002). Mechanisms in tissue-specific regulation of estrogen biosynthesis in humans. Trends Endocrinol Metab *13*, 122-128.

Karin, M., Liu, Z., and Zandi, E. (1997). AP-1 function and regulation. Curr Opin Cell Biol *9*, 240-246.

Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H., *et al.* (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. Science *270*, 1491-1494.

Kenealy, M. R., Flouriot, G., Sonntag-Buck, V., Dandekar, T., Brand, H., and Gannon, F. (2000). The 3'-untranslated region of the human estrogen receptor alpha gene mediates rapid messenger ribonucleic acid turnover. Endocrinology *141*, 2805-2813.

Kiang, D. T., and Kennedy, B. J. (1977). Tamoxifen (antiestrogen) therapy in advanced breast cancer. Ann Intern Med *87*, 687-690.

Kim, J. H., Li, H., and Stallcup, M. R. (2003). CoCoA, a nuclear receptor coactivator which acts through an N-terminal activation domain of p160 coactivators. Mol Cell *12*, 1537-1549.

Kim, M. Y., Woo, E. M., Chong, Y. T., Homenko, D. R., and Kraus, W. L. (2006). Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. Mol Endocrinol *20*, 1479-1493.

King, D. C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., and Hardison, R. C. (2005). Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res *15*, 1051-1060.

Klein-Hitpass, L., Schorpp, M., Wagner, U., and Ryffel, G. U. (1986). An estrogen-responsive element derived from the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells. Cell *46*, 1053-1061.

Kolbe, D., Taylor, J., Elnitski, L., Eswara, P., Li, J., Miller, W., Hardison, R., and Chiaromonte, F. (2004). Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. Genome Res *14*, 700-707.

Kraus, R. J., Ariazi, E. A., Farrell, M. L., and Mertz, J. E. (2002). Estrogen-related receptor α 1 actively antagonizes estrogen receptor-regulated transcription in MCF-7 mammary cells. J Biol Chem *277*, 24826-24834.

Krishnan, V., Wang, X., and Safe, S. (1994). Estrogen receptor-SP1 complexes mediate estrogen-induced cathepsin D gene expression in MCF-7 human breast cancer cells. J Biol Chem *269*, 15912-15917.

Kuiper, G. G. J. M., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., and Gustafsson, J. A. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. Endocrinology *138*, 863-870.

Kuiper, G. G. J. M., Enmark, E., Pelto-Huikko, M., Nilsson, S., and Gustafsson, J.-Å. (1996). Cloning of a novel estrogen receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A *93*, 5925-5930.

Kuo, M. H., and Allis, C. D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods *19*, 425-433.

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.

Kuukasjarvi, T., Kononen, J., Helin, H., Holli, K., and Isola, J. (1996). Loss of estrogen receptor in recurrent breast cancer is associated with poor response to endocrine therapy. J Clin Oncol *14*, 2584-2589.

Kyo, S., Takakura, M., Kanaya, T., Zhuo, W., Fujimoto, K., Nishio, Y., Orimo, A., and Inoue, M. (1999). Estrogen activates telomerase. Cancer Res *59*, 5917-5921.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell *97*, 17-27.

Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. J Mol Endocrinol *19*, 207-226.

Le, X. F., Pruefer, F., and Bast, R. C., Jr. (2005). HER2-targeting antibodies modulate the cyclin-dependent kinase inhibitor p27Kip1 via multiple signaling pathways. Cell Cycle *4*, 87-95.

Leclercq, G. (2002). Molecular forms of the estrogen receptor in breast cancer. J Steroid Biochem Mol Biol *80*, 259-272.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., *et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.

Lefstin, J. A., and Yamamoto, K. R. (1998). Allosteric effects of DNA on transcriptional regulators. Nature *392*, 885-888.

Lehman, J. J., Barger, P. M., Kovacs, A., Saffitz, J. E., Medeiros, D. M., and Kelly, D. P. (2000). Peroxisome proliferator-activated receptor γ coactivator-1 promotes cardiac mitochondrial biogenesis. J Clin Invest *106*, 847-856.

Lerner, L. J., and Jordan, V. C. (1990). Development of antiestrogens and their use in breast cancer: Eighth Cain memorial award lecture. Cancer Res *50*, 4177-4189.

Li, C., Briggs, M. R., Ahlborn, T. E., Kraemer, F. B., and Liu, J. (2001). Requirement of Sp1 and estrogen receptor alpha interaction in 17beta-estradiol-mediated transcriptional activation of the low density lipoprotein receptor gene expression. Endocrinology *142*, 1546-1553.

Liang, P., Averboukh, L., Keyomarsi, K., Sager, R., and Pardee, A. B. (1992). Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. Cancer Res *52*, 6966-6968.

Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science *257*, 967-971.

Liao, L., Kuang, S. Q., Yuan, Y., Gonzalez, S. M., O'Malley, B. W., and Xu, J. (2002). Molecular structure and biological function of the cancer-amplified nuclear receptor coactivator SRC-3/AIB1. J Steroid Biochem Mol Biol *83*, 3-14.

Lorke, D. E., Susens, U., Borgmeyer, U., and Hermans-Borgmeyer, I. (2000). Differential expression of the estrogen receptor-related receptor γ in the mouse brain. Brain Res Mol Brain Res *77*, 277-280.

Lowe, G. D. (2004). Hormone replacement therapy and cardiovascular disease: increased risks of venous thromboembolism and stroke, and no protection from coronary heart disease. J Intern Med *256*, 361-374.

Lu, D., Kiriyama, Y., Lee, K. Y., and Giguère, V. (2001). Transcriptional regulation of the estrogen-inducible pS2 beast cancer marker gene by the ERR family of orphan nuclear receptors. Cancer Res *61*, 6755-6761.

Lubahn, D. B., Moyer, J. S., Golding, T. S., Couse, J. F., Korach, K. S., and Smithies, O. (1993). Alteration of reproductive function but not prenatal sexual development after disruption of the mouse estrogen receptor. Proc Natl Acad Sci U S A *90*, 11162-11166.

Luo, J., Sladek, R., Bader, J. A., Matthyssen, A., Rossant, J., and Giguere, V. (1997). Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. Nature *388*, 778-782.

Luo, J., Sladek, R., Carrier, J., Bader, J.-A., Richard, D., and Giguère, V. (2003). Reduced fat mass in mice lacking orphan nuclear receptor estrogen-related receptor α. Mol Cell Biol *23*, 7947-7956.

MacGregor, J. I., and Jordan, V. C. (1998). Basic guide to the mechanisms of antiestrogen action. Pharmacol Rev *50*, 151-196.

Mahato, D., Goulding, E. H., Korach, K. S., and Eddy, E. M. (2000). Spermatogenic cells do not require estrogen receptor-alpha for development or function. Endocrinology *141*, 1273-1276.

Mahato, D., Goulding, E. H., Korach, K. S., and Eddy, E. M. (2001). Estrogen receptor-alpha is required by the supporting somatic cells for spermatogenesis. Mol Cell Endocrinol *178*, 57-63.

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

Martini, P. G., and Katzenellenbogen, B. S. (2001). Regulation of prothymosin alpha gene expression by estrogen in estrogen receptor-containing breast cancer cells via upstream half-palindromic estrogen response element motifs. Endocrinology *142*, 3493-3501.

Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A., and Chambon, P. (1982). Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line. Nucleic Acids Res *10*, 7895-7903.

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc Natl Acad Sci U S A *74*, 560-564.

McGhee, J. D., and von Hippel, P. H. (1975a). Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. Biochemistry *14*, 1281-1296.

McGhee, J. D., and von Hippel, P. H. (1975b). Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases. Biochemistry *14*, 1297-1303.

Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-α directs ordered, cyclical, and

combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751-763.

Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wilkstrom, A.-C., Gustafsson, J.-A., and Yamamoto, K. R. (1986). Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. Cell *46*, 389-399.

Miller, W. R. (2003). Aromatase inhibitors: mechanism of action and role in the treatment of breast cancer. Semin Oncol *30*, 3-11.

Mitsunaga, K., Araki, K., Mizusaki, H., Morohashi, K., Haruna, K., Nakagata, N., Giguere, V., Yamamura, K., and Abe, K. (2004). Loss of PGC-specific expression of the orphan nuclear receptor ERRβ results in reduction of germ cell number in mouse embryos. Mech Dev *121*, 237-246.

Mosselman, S., Polman, J., and Dijkema, R. (1996). ERβ: identification and characterization of a novel human estrogen receptor. Federation of European Biological Societies Letters *392*, 49-53.

Mueller, M. D., Vigne, J. L., Minchenko, A., Lebovic, D. I., Leitman, D. C., and Taylor, R. N. (2000). Regulation of vascular endothelial growth factor (VEGF) gene transcription by estrogen receptors alpha and beta. Proc Natl Acad Sci U S A *97*, 10972-10977.

Nagai, M. A., Ros, N., Bessa, S. A., Mourao Neto, M., Miracca, E. C., and Brentani, M. M. (2003). Differentially expressed genes and estrogen receptor status in breast cancer. Int J Oncol *23*, 1425-1430.

Nawaz, Z., Lonard, D. M., Dennis, A. P., Smith, C. L., and O'Malley, B. W. (1999). Proteasome-dependent degradation of the human estrogen receptor. Proc Natl Acad Sci U S A *96*, 1858-1862.

Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., and et al. (1987). The P450 gene superfamily: recommended nomenclature. DNA *6*, 1-11.

O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. (2004). Genomic targets of nuclear estrogen receptors. Mol Endocrinol *18*, 1859-1875.

Ogawa, S., Inoue, S., Watanabe, T., Hiroi, H., Orimo, A., Hosoi, T., Ouchi, Y., and Muramatsu, M. (1998). The complete primary structure of human estrogen

receptor β (hER-β) and its heterodimerization with ER α in vivo and in vitro. Biochem Biophys Res Commun *243*, 122-126.

Osborne, C. K. (1998). Tamoxifen in the treatment of breast cancer. N Engl J Med *339*, 1609-1618.

Osborne, C. K., Wakeling, A., and Nicholson, R. I. (2004). Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. Br J Cancer *90 Suppl 1*, S2-6.

Pethe, V., and Shekhar, P. V. (1999). Estrogen inducibility of c-Ha-ras transcription in breast cancer cells. Identification of functional estrogen-responsive transcriptional regulatory elements in exon 1/intron 1 of the c-Ha-ras gene. J Biol Chem *274*, 30969-30978.

Pettersson, K., Grandien, K., Kuiper, G. G. J. M., and Gustafsson, J.-Å. (1997). Mouse estrogen receptor β forms estrogen response element-binding heterodimers with estrogen receptor α. Mol Endocrinol *11*, 1486-1496.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R., and Berkenstam, A. (1996). Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. Mech Dev *54*, 211-223.

Petz, L. N., and Nardulli, A. M. (2000). Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor A promoter. Mol Endocrinol *14*, 972-985.

Pinkerton, J. V., and Henderson, V. W. (2005). Estrogen and cognition, with a focus on Alzheimer's disease. Semin Reprod Med *23*, 172-179.

Porter, W., Saville, B., Hoivik, D., and Safe, S. (1997). Functional synergy between the transcription factor Sp1 and the estrogen receptor. Mol Endocrinol *11*, 1569-1580.

Pratt, W. B., and Toft, D. O. (1997). Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocr Rev *18*, 306-360.

Preisler-Mashek, M. T., Solodin, N., Stark, B. L., Tyriver, M. K., and Alarid, E. T. (2002). Ligand-specific regulation of proteasome-mediated proteolysis of estrogen receptor-alpha. Am J Physiol Endocrinol Metab *282*, E891-898.

Preston-Martin, S., Pike, M. C., Ross, R. K., Jones, P. A., and Henderson, B. E. (1990). Increased cell division as a cause of human cancer. Cancer Res *50*, 7415-7421.

Puigserver, P., Wu, Z., Park, C. W., Graves, R., Wright, M., and Spiegelman, B. M. (1998). A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. Cell *92*, 829-839.

Qin, C., Singh, P., and Safe, S. (1999). Transcriptional activation of insulin-like growth factor-binding protein-4 by 17beta-estradiol in MCF-7 cells: role of estrogen receptor-Sp1 complexes. Endocrinology *140*, 2501-2508.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Richard, S., and Zingg, H. H. (1990). The human oxytocin gene promoter is regulated by estrogens. J Biol Chem *265*, 6098-6103.

Rio, M. C., and Chambon, P. (1990). The pS2 gene, mRNA, and protein: a potential marker for human breast cancer. Cancer Cells *2*, 269-274.

Rishi, A. K., Shao, Z. M., Baumann, R. G., Li, X. S., Sheikh, M. S., Kimura, S., Bashirelahi, N., and Fontana, J. A. (1995). Estradiol regulation of the human retinoic acid receptor alpha gene in human breast carcinoma cells is mediated via an imperfect half-palindromic estrogen response element and Sp1 motifs. Cancer Res *55*, 4999-5006.

Roberts, M., Wallace, J., Jeltsch, J. M., and Berry, M. (1988). The 5' flanking region of the human pS2 gene mediates its transcriptional activation by estrogen in MCF-7 cells. Biochem Biophys Res Commun *151*, 306-313.

Ross, J. S., and Fletcher, J. A. (1998). The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. Stem Cells *16*, 413-428.

Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A., Howard, B. V., Johnson, K. C., *et al.* (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. Jama *288*, 321-333.

Russo, J., Ao, X., Grill, C., and Russo, I. H. (1999). Pattern of distribution of cells positive for estrogen receptor α and progesterone receptor in relation to proliferating cells in the mammary gland. Breast Cancer Res Treat *53*, 217-227.

Sabbah, M., Courilleau, D., Mester, J., and Redeuilh, G. (1999). Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element. Proc Natl Acad Sci U S A *96*, 11217-11222.

Safe, S., and Kim, K. (2004). Nuclear receptor-mediated transactivation through interaction with Sp proteins. Prog Nucleic Acid Res Mol Biol *77*, 1-36.

Salvatori, L., Ravenna, L., Felli, M. P., Cardillo, M. R., Russo, M. A., Frati, L., Gulino, A., and Petrangeli, E. (2000). Identification of an estrogen-mediated deoxyribonucleic acid-binding independent transactivation pathway on the epidermal growth factor receptor gene promoter. Endocrinology *141*, 2266-2274.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*, 5463-5467.

Sar, M., and Welsch, F. (1999). Differential expression of estrogen receptor-beta and estrogen receptor-alpha in the rat ovary. Endocrinology *140*, 963-971.

Saunders-Pullman, R. (2003). Estrogens and Parkinson disease: neuroprotective, symptomatic, neither, or both? Endocrine *21*, 81-87.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T., and Kralli, A. (2003). The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor ERRα. J Biol Chem *278*, 9013-9018.

Schultz, J. R., Petz, L. N., and Nardulli, A. M. (2003). Estrogen receptor alpha and Sp1 regulate progesterone receptor gene expression. Mol Cell Endocrinol *201*, 165-175.

Schwabe, J. W. R., 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.

Seth, P., Krop, I., Porter, D., and Polyak, K. (2002a). Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene *21*, 836-843.

Seth, P., Porter, D., Lahti-Domenici, J., Geng, Y., Richardson, A., and Polyak, K. (2002b). Cellular and molecular targets of estrogen in normal human breast tissue. Cancer Res *62*, 4540-4544.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell *103*, 843-852.

Sherwin, B. B. (2003). Estrogen and cognitive functioning in women. Endocr Rev *24*, 133-151.

Shulman, L. M. (2002). Is there a connection between estrogen and Parkinson's disease? Parkinsonism Relat Disord *8*, 289-295.

Silberstein, G. B., van Horn, K., Shyamala, G., and Daniel, C. W. (1994). Essential role of endogenous estrogen in directly stimulating mammary growth demonstrated by implants containing pure antiestrogens. Endocrinology *134*, 84-90.

Simpson, E. R. (2002). Aromatization of androgens in women: current concepts and findings. Fertil Steril *77 Suppl 4*, S6-10.

Simpson, E. R., Misso, M., Hewitt, K. N., Hill, R. A., Boon, W. C., Jones, M. E., Kovacic, A., Zhou, J., and Clyne, C. D. (2005). Estrogen--the good, the bad, and the unexpected. Endocr Rev *26*, 322-330.

Sladek, R., Bader, J.-A., and Giguère, V. (1997a). The orphan nuclear receptor estrogen-related receptor α is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. Mol Cell Biol *17*, 5400-5409.

Sladek, R., Beatty, B., Squire, J., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Giguère, V. (1997b). Chromosomal mapping of the human and murine orphan nuclear receptor ERRα (ESRRA) and ERRβ (ESRRB) and identification of a novel human ERRα-related pseudogene. Genomics *45*, 320-326.

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science *235*, 177-182.

Smith, C. L., and O'Malley, B. W. (2004). Coregulator function: a key to understanding tissue specificity of selective receptor modulators. Endocr Rev *25*, 45-71.

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.

Spencer, T. E., Jenster, G., Burcin, M. M., Allis, C. D., Zhou, J., Mizzen, C. A., McKenna, N. J., Onate, S. A., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. Nature *389*, 194-198.

Stallcup, M. R., Chen, D., Koh, S. S., Ma, H., Lee, Y. H., Li, H., Schurter, B. T., and Aswad, D. W. (2000). Co-operation between protein-acetylating and protein-methylating co-activators in transcriptional activation. Biochem Soc Trans *28*, 415-418.

Stoner, M., Wang, F., Wormke, M., Nguyen, T., Samudio, I., Vyhlidal, C., Marme, D., Finkenzeller, G., and Safe, S. (2000). Inhibition of vascular endothelial growth factor expression in HEC1A endometrial cancer cells through interactions of estrogen receptor alpha and Sp3 proteins. J Biol Chem *275*, 22769-22779.

Suzuki, T., Miki, Y., Moriya, T., Shimada, N., Ishida, T., Hirakawa, H., Ohuchi, N., and Sasano, H. (2004). Estrogen-related receptor α in human breast carcinoma as a potent prognostic factor. Cancer Res *64*, 4670-4676.

Tanaka, N., Yonekura, H., Yamagishi, S., Fujimori, H., Yamamoto, Y., and Yamamoto, H. (2000). The receptor for advanced glycation end products is induced by the glycation products themselves and tumor necrosis factor-alpha through nuclear factor-kappa B, and by 17beta-estradiol through Sp-1 in human vascular endothelial cells. J Biol Chem *275*, 25781-25790.

Tang, S., Han, H., and Bajic, V. B. (2004). ERGDB: Estrogen Responsive Genes Database. Nucleic Acids Res *32*, D533-536.

Tavera-Mendoza, L. E., Mader, S., and White, J. H. (2006). Genome-wide approaches for identification of nuclear receptor target genes. Nucl Recept Signal *4*, e018.

Tcherepanova, I., Puigserver, P., Norris, J. D., Spiegelman, B. M., and McDonnell, D. P. (2000). Modulation of estrogen receptor-α transcriptional activity by the coactivator PGC-1. J Biol Chem *275*, 16302-16308.

Thompson, D. A., and Weigel, R. J. (1998). Characterization of a gene that is inversely correlated with estrogen receptor expression (ICERE-1) in breast carcinomas. Eur J Biochem *252*, 169-177.

Toft, D., and Gorski, J. (1966). A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. Proc Natl Acad Sci U S A *55*, 1574-1581.

Treilleux, Peloux, N., Brown, M., and Sergeant, A. (1997). Human estrogen receptor (ER) gene promoter-P1: estradiol-independent activity and estradiol inducibility in ER+ and ER- cells. Mol Endocrinol *11*, 1319-1331.

Tremblay, G. B., Bergeron, D., and Giguère, V. (2001a). 4-hydroxytamoxifen is an isoform-specific inhibitor of orphan estrogen-receptor-related (ERR) nuclear receptors β and γ. Endocrinology *142*, 4572-4575.

Tremblay, G. B., Kunath, T., Bergeron, D., Lapointe, L., Champigny, C., Bader, J.-A., Rossant, J., and Giguère, V. (2001b). Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERRβ. Genes Dev *15*, 833-838.

Tremblay, G. B., Tremblay, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Labrie, F., and Giguère, V. (1997). Cloning, chromosomal localization and functional analysis of the murine estrogen receptor β. Mol Endocrinol *11*, 353-365.

Turner, R. T., Wakley, G. K., Hannon, K. S., and Bell, N. H. (1987). Tamoxifen prevents the skeletal effects of ovarian hormone deficiency in rats. J Bone Miner Res *2*, 449-456.

Umayahara, Y., Kawamori, R., Watada, H., Imano, E., Iwama, N., Morishima, T., Yamasaki, Y., Kajimoto, Y., and Kamada, T. (1994). Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. J Biol Chem *269*, 16433-16442.

Vanacker, J.-M., Bonnelye, E., Chopin-Delannoy, S., Delmarre, C., Cavailles, V., and Laudet, V. (1999a). Transcriptional activities of the orphan nuclear receptor ERR α (estrogen receptor-related receptor-α). Mol Endocrinol *13*, 764-773.

Vanacker, J. M., Delmarre, C., Guo, X., and Laudet, V. (1998). Activation of the osteopontin promoter by the orphan nuclear receptor estrogen receptor related α. Cell Growth Differ *9*, 1007-1014.

Vanacker, J. M., Pettersson, K., Gustafsson, J. A., and Laudet, V. (1999b). Transcriptional targets shared by estrogen receptor- related receptors (ERRs) and estrogen receptor (ER) alpha, but not by ERbeta. Embo J *18*, 4270-4279.

Vega, R. B., Huss, J. M., and Kelly, D. P. (2000). The coactivator PGC-1 cooperates with peroxisome proliferator-activated receptor α in transcriptional control of nuclear genes encoding mitochondrial fatty acid oxidation enzymes. Mol Cell Biol *20*, 1868-1876.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. Science *270*, 484-487.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vogel, V. G., Costantino, J. P., Wickerham, D. L., Cronin, W. M., Cecchini, R. S., Atkins, J. N., Bevers, T. B., Fehrenbacher, L., Pajon, E. R., Jr., Wade, J. L., 3rd, *et al.* (2006). Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. Jama *295*, 2727-2741.

Vyhlidal, C., Samudio, I., Kladde, M. P., and Safe, S. (2000). Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. J Mol Endocrinol *24*, 329-338.

Wang, C., Fu, M., Angeletti, R. H., Siconolfi-Baez, L., Reutens, A. T., Albanese, C., Lisanti, M. P., Katzenellenbogen, B. S., Kato, S., Hopp, T., *et al.* (2001). Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity. J Biol Chem *276*, 18375-18383.

Wang, D. Y., Fulthorpe, R., Liss, S. N., and Edwards, E. A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. Mol Endocrinol *18*, 402-411.

Wang, F., Porter, W., Xing, W., Archer, T. K., and Safe, S. (1997). Identification of a functional imperfect estrogen-responsive element in the 5'-promoter region of the human cathepsin D gene. Biochemistry *36*, 7793-7801.

Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K., Ogawa, S., Arao, Y., Suzawa, M., Kobayashi, Y., Yano, T., Yoshikawa, H., *et al.* (2001). A

subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. Embo J *20*, 1341-1352.

Watanabe, T., Inoue, S., Hiroi, H., Orimo, A., Kawashima, H., and Muramatsu, M. (1998). Isolation of estrogen-responsive genes with a CpG island library. Mol Cell Biol *18*, 442-449.

Watanabe, T., Inoue, S., Hiroi, H., Orimo, A., and Muramatsu, M. (1999). NMDA receptor type 2D gene as target for estrogen receptor in the brain. Brain Res Mol Brain Res *63*, 375-379.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Watson, J. D., and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wolf, D. M., and Jordan, V. C. (1992). Gynecologic complications associated with long-term adjuvant tamoxifen therapy for breast cancer. Gynecol Oncol *45*, 118-128.

Wood, J. R., Greene, G. L., and Nardulli, A. M. (1998). Estrogen response elements function as allosteric modulators of estrogen receptor conformation. Mol Cell Biol *18*, 1927-1934.

Wu, Z., Puigserver, P., Andersson, U., Zhang, C., Adelmant, G., Mootha, V., Troy, A., Cinti, S., Lowell, B., Scarpulla, R. C., and Spiegelman, B. M. (1999). Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. Cell *98*, 115-124.

Xie, W., Duan, R., Chen, I., Samudio, I., and Safe, S. (2000). Transcriptional activation of thymidylate synthase by 17beta-estradiol in MCF-7 human breast cancer cells. Endocrinology *141*, 2439-2449.

Xie, W., Duan, R., and Safe, S. (1999a). Estrogen induces adenosine deaminase gene expression in MCF-7 human breast cancer cells: role of estrogen receptor-Sp1 interactions. Endocrinology *140*, 219-227.

Xie, W., Hong, H., Yang, N. N., Lin, R. J., Simon, C. M., Stallcup, M. R., and Evans, R. M. (1999b). Constitutive activation of transcription and binding of

coactivator by estrogen-related receptors 1 and 2. Mol Endocrinol *13*, 2151-2162.

Yang, N., Shigeta, H., Shi, H. P., and Teng, C. T. (1996). Estrogen-related receptor, hERR1, modulates estrogen receptor-mediated response of human lactoferrin gene promoter. J Biol Chem *271*, 5795-5804.

Ylikomi, T., Wurtz, J. M., Syvala, H., Passinen, S., Pekki, A., Haverinen, M., Blauer, M., Tuohimaa, P., and Gronemeyer, H. (1998). Reappraisal of the role of heat shock proteins as regulators of steroid receptor activity. Crit Rev Biochem Mol Biol *33*, 437-466.

Yoon, J. C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelmant, G., Stafford, J., Kahn, C. R., Granner, D. K., *et al.* (2001). Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. Nature *413*, 131-138.

Yue, W., Santen, R. J., Wang, J. P., Li, Y., Verderame, M. F., Bocchinfuso, W. P., Korach, K. S., Devanesan, P., Todorovic, R., Rogan, E. G., and Cavalieri, E. L. (2003). Genotoxic metabolites of estradiol in breast: potential mechanism of estradiol induced carcinogenesis. J Steroid Biochem Mol Biol *86*, 477-486.

Zhang, Z., and Teng, C. T. (2000). Estrogen receptor-related receptor α 1 interacts with coactivator and constitutively activates the estrogen response elements of the human lactoferrin gene. J Biol Chem *275*, 20837-20846.

Zhao, Y. Y., Zhou, J., Narayanan, C. S., Cui, Y., and Kumar, A. (1999). Role of C/A polymorphism at -20 on the expression of human angiotensinogen gene. Hypertension *33*, 108-115.

CHAPTER II: Functional Genomics Identifies a Mechanism For Estrogen Activation of the Retinoic Acid Receptor α1 Gene in Breast Cancer Cells

PREFACE

ERα is a major player in the initiation and progression of breast cancer. With the goal of identifying novel functional direct ERα genomic targets, I first developed the ChIP-cloning technology, a recent functional genomics approach, described in details in appendix 1 (Laganière et al., 2003). Following the optimization of the standard ChIP, I used the ChIP-cloning technique to identify novel genomic targets of endogenous ERα in MCF7 cells. Although we identified several new ERα binding sites, we characterized in more detail one intronic enhancer that controls the estrogen response of the *RARA* (RARα) gene, which is the main topic of this chapter. This study exemplified that regulation outside of promoter regions is a feature of transcriptional activation by ERα and proposed a novel mechanism for RARα upregulation by estrogens. The mechanism described here is different from those suggested in previous studies and illustrates the value of unbiased functional approaches for the identification of genuine transcriptional regulatory regions.

This chapter is a manuscript that has been published in the *Molecular Endocrinology* journal.

ABSTRACT

The identification of estrogen receptor (ERα) target genes is crucial to our understanding of its predominant role in breast cancer. In this study, we used a chromatin immunoprecipitation (ChIP)-cloning strategy to identify ERα regulatory modules and associated target genes in the human breast cancer cell line MCF-7. We isolated twelve transcriptionally active genomic modules that recruit ERα and the coactivator SRC-3 to different intensities *in vivo*. One of the ERα regulatory modules identified is located 3.7 kilobases downstream of the first transcriptional start site of the RARA locus which encodes retinoic acid receptor α1 (RARα1). This module, which includes an estrogen response element (ERE), is conserved between the human and mouse genomes. Direct binding of ERα to the ERE was shown using electromobility shift assays, and transient transfections in MCF-7 cells demonstrated that endogenous ERα can induce estrogen-dependent transcriptional activation from the module or the ERE linked to a heterologous promoter. Furthermore, ChIP assays showed that the coregulators SRC-1, SRC-3 and RIP140 are recruited to this intronic module in an estrogen-dependent manner. As expected from previous studies, the transcription factor SP1 can be detected at the RARA α1 promoter by ChIP. However, treatment with estradiol did not influence SP1 recruitment nor help recruit ERα to the promoter. Finally, ablation of the intronic ERE was sufficient to abrogate the up-regulation of RARA α1 promoter activity by estradiol. Thus, this study uncovered a mechanism by which ERα significantly activates RARα1 expression in breast cancer cells, and exemplifies the utility of functional genomics strategies in identifying long-distance regulatory modules for nuclear receptors.

INTRODUCTION

Nuclear receptors constitute a superfamily of transcription factors that control reproduction, embryonic development, homeostasis and play important roles in the initiation, progression and treatment of numerous diseases, including cancer. The cloning of the glucocorticoid and estrogen receptors 20 years ago (Green et al., 1986; Hollenberg et al., 1985) together with the development of a rapid assay for receptor activity in transfected cells (Giguère et al., 1986) set the stage for investigating the molecular mechanisms of small lipophilic ligand-regulated transcription. While early work focused on the receptors themselves, defining their functional domains (Giguère et al., 1986; Kumar et al., 1987) and their interaction with DNA (Umesono et al., 1991), it is now well understood that nuclear receptor-regulated gene expression requires the recruitment, in a multi-step fashion, of diverse sets of coregulatory proteins. To ensure modulated hormonal control of gene expression, these regulatory complexes can induce changes in chromatin structure, control the basal transcription machinery activity and regulate the degradation of the receptors and associated proteins (reviewed in Gamble and Freedman, 2002; Glass and Rosenfeld, 2000; McKenna et al., 1999). The ability of nuclear receptors to modulate transcription of target genes is achieved through recognition by the receptors of short sequences referred to as hormone response elements located in the promoters and enhancers of these genes (Yamamoto, 1985). In addition, molecular cross-talk allows nuclear receptors to regulate the expression of genes via association, either on DNA or in solution, with other transcription factors such as SP1 and AP-1 (reviewed in Herrlich, 2001; Pfahl, 1993; Safe and Kim, 2004). However, whereas nuclear receptors are expected to control a large number of genes given their expansive roles in development and physiology, relatively few genomic targets have been identified to date.

Although it has been known for decades that estrogens are potent stimulators of estrogen receptor-positive breast cancer cell proliferation (Lippman et al., 1976), the exact mechanisms underlying their growth

stimulating effects are still unknown. Delineation of estrogen action in these cells have been hindered by the paucity of *bona fide* ERa-regulated genes identified discovered thus far (O'Lone et al., 2004). Recently, gene expression profiling experiments have identified genes with altered expression upon estrogen treatment of human breast cancer cells, but very few have been confirmed as ERa primary targets (Charpentier et al., 2000; Coser et al., 2003; Cunliffe et al., 2003; Frasor et al., 2003; Inoue et al., 2002; Lin et al., 2004; Lobenhofer et al., 2002; Seth et al., 2002; Soulez and Parker, 2001; Wang et al., 2004; Weisz et al., 2004). Interested in isolating direct genomic targets of ERa, we recently developed a chromatin immunoprecipitation (ChIP)-derived approach to isolate genomic fragments bound to the receptor in MCF-7 cells (Laganière et al., 2003). Interestingly, among the cloned targets, endogenous ERα was found to bind *in vivo* to an estrogen response element (ERE) located in the first intron of the *RARA* locus, the gene encoding retinoic acid receptor a (RARa). In this report, we investigated the role played by this regulatory region in the control of *RARA* in response to estradiol. We found that in addition to recruiting coactivators, this intronic ERE provides the major estrogen response of the *RARA* gene in MCF-7 cells. This study highlights the predominant role that functional genomics will play in the near future in defining gene networks directly regulated by nuclear receptors and how, mechanistically, members of this superfamily of transcription factors exert this control.

RESULTS

Identification of ERα genomic targets in human breast cancer cells

To identify regulatory modules directly bound by ERα in human breast cancer cells, we performed ChIP using a specific antibody against ERα in MCF-7 cells followed by cloning and sequencing of the fragments obtained, as previously described (Laganière et al., 2003). The location of each fragment in the human genome was determined using the UCSC human genome database (http://genome.ucsc.edu). Twelve fragments containing either EREs or multiple half sites were selected for further analysis (see Table 2.1). First, the binding of ERα to these modules was re-evaluated using standard ChIPs and quantified by real-time PCR using primers specific for each fragment isolated. As shown in Fig. 2.1, we found that the genomic sequences examined were significantly bound by ERα *in vivo*, being enriched at least 2-fold over the control (no antibody) when retested by standard ChIP PCR with specific primers. This series of fragments included a genomic region located 220 bp upstream of the *TFF1* (pS2) start site (clone ER4282), a promoter region known to contain a well defined ERE and be estrogen responsive (Berry et al., 1989). Q-PCR quantification of the enrichment obtained by an independent ERα standard ChIP assay using primers specific for the *TFF1* promoter showed an enrichment of 108-fold over the control when cells were treated with estradiol (E_2) for 45 minutes prior to chromatin preparation (Fig. 2.1). Other fragments located near or at promoter regions were isolated and recognized *in vivo* by ERα: the promoters of *FLJ10618* that encodes a predicted protein, *NAP1L4* (nucleosome assembly protein 1-like 4) as well as a fragment located 2.2 kb upstream of *RNF14* that encodes a coactivator of the androgen receptor also known as ARA54 (Kang et al., 1999). Interestingly, in addition to binding to promoters, ERα also recognized modules distal from known transcriptional start sites. We found ERα regulatory modules located from 4 to 103 kb from the *DDEF2* (development and differentiation enhancing factor 2), *GPR81* (G protein-coupled receptor 81), *EDF1* (endothelial differentiation-related factor 1), *EDG1* (endothelial differentiation, sphingolipid G-protein-coupled receptor,

1), *FLJ16032*, *BCR* (breakpoint cluster region), *FLJ41849* and *RARA* genes. All modules except DER005 (Table 1) were well conserved between the mouse and human genomes (data not shown).

It is now well established that ERα recruits coactivators in response to estradiol to modulate gene expression (Hermanson et al., 2002; McKenna et al., 1999), and that recruitment of coactivators is a good indicator of the transcriptional activity of a transcription factor binding site. We thus investigated whether estradiol treatment would lead to SRC-3 recruitment to these modules. As shown in Fig. 2.1, SRC-3 was indeed recruited to the ERα-bound targets. We next investigated the estrogen responsiveness of a number of genes located nearby the ERα-bound modules. We performed quantitative RT-PCR for genes that were located close to an ERα target region, since regions located at very far distances are likely to control closer transcripts not yet annotated or annotated but not reviewed. As shown in Fig. 2.2, we observed, as expected, that TFF1 and RARA mRNA amounts were increased following estradiol treatment. In addition, we observed that FLJ10618 and RNF14 were also up-regulated after estradiol treatment, following distinct response profiles to the hormone. In contrast, GPR81 was downregulated by 2-fold while BCR was not significantly modulated by estradiol.

Our laboratory was particularly interested in the finding that enrichment of ERa by ChIP corresponded to a region, according to the UCSC genome database, located 8.36 kb upstream of the *RARA* gene that encodes a receptor for retinoic acid, RARα. In fact, it has been determined that transcription of the *RARA* gene is achieved through two distinct promoters, generating two different mRNAs encoding the RARα1 and RARα2 isoforms (Giguère, 1994). Close examination of the *RARA* gene shows that the ERE-containing module isolated by ChIP-cloning is indeed located within the first intron of *RARA*, 3.7 kb downstream of its first transcriptional start site (RARα1) and 8.36 kb upstream of the second (RARα2) promoter (Fig. 2.2A). In addition, recent studies have shown that RARα expression in MCF-7 cells solely depends on the *RARA* α1 promoter since the *RARA* α2 promoter is inactivated by methylation in these

cells (Farias et al., 2002). Therefore, we decided to focus our investigation on the significance of this ERE (herein referred to as ERE_{RARA}) in the control of the expression of the RARα1 isoform in MCF-7 cells.

The ERE_{RARA} is functional *in vitro* and *in vivo*

We were next interested in establishing the functionality of the novel ERE found in the first intron of the *RARA* gene. The ERE_{RARA} differs by only one base from the established ERE consensus (Fig. 2.2C) but an ERE with this sequence had not been previously reported to be functional *in vivo*. As mentioned above, both the ERE_{RARA} itself and the genomic fragment obtained by ChIP-cloning (130 bp in length) are well conserved from the mouse to the human genome (Fig. 2.2B and C). We first tested the ability of ERα to bind the ERE *in vitro*. As expected from the ChIP-cloning experiment and shown in Fig. 2.3A, an electromobility shift assay (EMSA) demonstrates that *in vitro* translated ERα recognized the ERE probe directly. In addition, endogenous ERα contained in total MCF-7 cell extract also affected the migration of the ERE probe in the gel. The presence of ERα in the retarded complex was confirmed by a supershift of the complex using an anti-ERα antibody (Fig. 2.3B). To determine if ERα could modulate gene transcription using this ERE, we performed transient transfections in MCF-7 cells of a TK-Luc reporter containing one copy of either ERE_{RARA} or the whole 130 bp fragment (DER001) isolated by ChIP-cloning. Our results show that treatment of MCF-7 cells with estradiol leads to a large induction of luciferase activity in cells transfected with either reporter plasmid, indicating that endogenous ERα can utilize the ERE_{RARA} to activate gene transcription (Fig. 2.3C).

The intronic ERE_{RARA} is transcriptionally active in MCF-7 cells

The recruitment of coregulators to the intronic ERE_{RARA} in the presence of estradiol would suggest that the ERE is actively engaged in transcriptional regulation of *RARA*. In addition, since it was previously suggested that ERα controls the expression of the RARα1 isoform through an ERα-SP1 interaction

occurring at the promoter (Sun et al., 1998; Zou et al., 1999), we also evaluated the binding of specific transcriptional modulators at the first *RARA* promoter. We thus performed ChIP experiments using primers specific for the *RARA* α1 promoter, the intronic ERE_{RARA}, and a generic control region located 4 kb upstream of the *TFF1* promoter. As previously shown in Fig. 2.1, ERα recognized the intronic region containing the ERE_{RARA} in the absence of estradiol, and its binding was significantly enhanced in the presence of the hormone (Fig. 2.4). Data presented in Fig. 2.4 also demonstrates that both SRC-1 and SRC-3 are recruited to the ERE_{RARA} region in the presence of estradiol, further supporting the *in vivo* activity of the ERα-bound ERE_{RARA}. Interestingly, the coregulator RIP140 was also strongly (>40-fold enrichment) recruited to the region containing the ERE_{RARA} (Fig. 2.4). RIP140 recruitment was also observed at the *TFF1* promoter and other regulatory modules isolated (data not shown). Modest binding of SP1 (6-fold enrichment) was also found at the ERE_{RARA} region. In sharp contrast, none of the coactivators tested, nor ERα, were significantly detected at the *RARA* (RARa1) promoter region. Similar results were also obtained following estradiol treatment at various time points taken between 5 minutes and 5 hours (data not shown). In addition, strengthening of protein-protein interactions with the crosslinking agent DTBP (dimethyl 3,3'-dithiobispropionimidate 2•HCl) prior to formaldehyde treatment of the prepared chromatin did not improve the detection of ERα at the promoter (data not shown). However, a significant presence of SP1 could be detected by ChIP at the promoter, but addition of estradiol had no significant effect on the recruitment of the protein.

The intronic ERE_{RARA} controls RARα1 response to estrogens

Finally, we sought to determine whether the newly discovered ERE_{RARA} plays a direct role in the control of the RARα1 gene expression. To this end, we cloned a 5 kb genomic DNA fragment containing the *RARA* α1 promoter and a segment of the first intron that includes the ERE_{RARA}. The same construct with a deleted ERE_{RARA} was also made (Fig. 2.5A). When the intact construct was

transfected in MCF-7 cells, we observed a strong induction of luciferase activity upon estradiol treatment, whereas less than 2-fold activation was obtained when the ERE_{RARA}-deleted construct was assayed in the same cells (Fig. 2.5B). Our data thus suggests that the main estrogenic response of the *RARA* gene is driven by the intronic ERE_{RARA} in MCF-7 cells. To confirm that ERα is responsible for this activity, we performed transient transfections of the same constructs in MDA-MB-231, an ER-negative breast cancer cell line, in the presence or absence of exogenous ERα. Figure 2.5C shows that these cells depend on both ERα and the ERE_{RARA} to provide estradiol-induced activation of the *RARA* locus, since the activation by estradiol is observed only in the presence of the receptor and the intronic ERE_{RARA}. We also wanted to determine if the induction of RARα1 through this ERE could be observed in other ER-positive cell lines. In Fig. 2.5D, we show that endogenous ERα from BT-474 breast cancer cells activates the reporter in a similar fashion.

DISCUSSION

In this report, we described the identification of direct genomic targets of endogenous ERα in the MCF-7 breast cancer cell line using a chromatin immunoprecipitation assay-derived approach. We found novel direct regulatory sequences directly bound *in vivo* by ERα, which were located upstream, downstream, or in annotated promoter regions. Among the twelve ERa binding modules described in this study, two were associated with genes, *TFF1* and *RARA*, previously known to be regulated by estradiol in human breast cancer cells. Although less high-throughput than promoter arrays (Odom et al., 2004), this functional approach has the advantage of identifying binding sites located outside of promoter regions. This is exemplified by the focus of our investigation on the ERE_{RARA} located in the first intron of the *RARA* gene, which was strongly bound by ERα *in vitro* and *in vivo*. In addition, the presence of the ERE_{RARA} was found to be essential for significant activation by estradiol of the *RARA* a1 promoter in transfected MCF-7 cells. Most importantly, we also showed that in addition to ERα, the SRC-1 and SRC-3 coactivators were recruited to the ERE_{RARA} in an estrogen-dependent manner, demonstrating that this intronic site becomes transcriptionally operational in MCF-7 cells upon estradiol treatment. Recent evidence indeed states the important regulatory role of the first intron in the regulation of gene expression. Based on chromosome 21 and 22 gene location analyses of SP1 and other transcription factors, a great proportion of all functional transcription factor binding sites are located a few hundred or thousand bases downstream of annotated promoters, strongly predicting an important regulatory role of the first intron for gene transcription in general (Cawley et al., 2004). Our study supports that, similar to other transcription factors such as c-Myc and SP1, ERα action is not restricted to promoters *in vivo*. Therefore, whole-genome strategies will be essential to reveal distal but functional EREs, indeed providing a much improved picture of ERα action in breast cancer cells and at other sites of estradiol action.

As mentioned above, *RARA* is a well known estrogen-regulated gene with potential as a target for cancer prevention and therapy since treatment with retinoids leads to cell growth arrest and apoptosis in ERα-positive breast cancer cells (Wang et al., 2000; Zhu et al., 1997). *RARA* can produce two different transcripts, encoding the RARα1 and RARα2 isoforms, originating from two distinct promoters (Giguère, 1994; Leroy et al., 1991). In MCF-7 cells, the RARα1 isoform is predominant since the *RARA* α2 promoter is methylated in these cells (Farias et al., 2002). It has also been demonstrated that RARα1 is the only estrogen-regulated isoform in breast cancer cells (van der Leede et al., 1995). Upregulation of RARα1 expression by estrogens was previously attributed to an ERα-SP1 interaction occurring at a GC-rich sequence in the first *RARA* promoter (Sun et al., 1998). Another laboratory had previously found an imperfect ERE half-site and SP1 motif in the promoter presumably responsible for the estrogen induction of RARα1 (Rishi et al., 1995). However, no direct binding of ERα was detected at these sites. In addition, the studies were performed with overexpressed exogenous proteins and restricted to proximal sections of the promoter. In order to have a more physiological representation of *RARA* α1 regulation, we transiently transfected a luciferase reporter construct of the genomic region containing the *RARA* α1 promoter and a portion of the first intron including the ERE_{RARA}, or its ERE-deleted mutant. Interestingly, the treatment of cells with estradiol lead to a strong increase in luciferase activity, indicating that endogenous ERα was sufficient to promote *RARA* α1 upregulation by the hormone. Moreover, in the ERE_{RARA} deletion mutant, the luciferase activity was nearly unchanged following estradiol treatment, hence the ERE_{RARA} confers the estrogenic response of RARα1 in MCF-7 cells.

Since the growth suppressive effect of retinoids is limited to ER-positive cells, it has been suggested that the action of retinoids is mainly antiestrogenic (Demirpence et al., 1994; Fontana et al., 1992; Muller et al., 2002; Rousseau et al., 2003). A recent study has shown that induction by retinoic acid of the coregulator RIP140, a known corepressor of ER activity (Chuang et al., 1997;

Eng et al., 1998; L'Horset et al., 1996), could mediate the antiestrogenic effects of retinoic acid in ER-positive human breast cancer cells (White et al., 2005). Interestingly, we showed that endogenous RIP140 was recruited to the ERE_{RARA}, an event that we also observed to occur at the estrogen-inducible *TFF1* promoter. This is, to our knowledge, the first demonstration that endogenous RIP140 is recruited to ERα target regulatory modules and promoters, a finding that further supports that RIP140 plays a central role in the estrogenic response of breast cancer cells. In a manner analogous to the switch between coactivators and the corepressor NCoR in apo-ERα-bound promoters recently shown by Métivier et al. (Metivier et al., 2004; Metivier et al., 2003), our finding suggests that a dynamic exchange may occur between the coactivators SRC-1 and SRC-3 and the corepressor RIP140 to provide subtle modulation of gene control in response to estradiol and retinoic acid in breast cancer cells.

In conclusion, this and other recent studies (Bourdeau et al., 2004; Odom et al., 2004) clearly demonstrate the importance that functional genomics will play in the near future in identifying, in a genome-wide and unbiased manner, hormone response elements and associated genes directly controlled by all members of the nuclear receptor superfamily. After twenty years of work dedicated to the receptors themselves and their interactions with DNA and other regulatory proteins, the tools are finally available to reveal the vast networks of genes regulated by nuclear receptors in a cell-specific manner.

MATERIALS AND METHODS

Plasmids.

To construct TK-ERE$_{RARA}$Luc, we used the synthetic oligonucleotides 5'-GATCCAGCGAGTGGGTCACGGTGACACTGCCTGCG-3' and 5'-TCGACGCAGGCAGTGTCACCGTGACCCACTCGCTG-3' that were annealed and subsequently ligated in TK-Luc, using *Sal*I and *Bam*HI restriction sites. For the pRL-RARa1 construct (Fig. 2.5), the *RARA* promoter/intron 1 locus was amplified from total MCF-7 genomic DNA with the following PCR primers: 5'-GTCTCTCAGATGGAGGGTGATTCAGATCC-3' and 5'-CGACGCGTCAGGAAGT-GACAGCCACGTGACAGGAAGAC-3'. The PCR product was digested with *Hind*III and *Mlu*I and ligated in corresponding pRL-Null vector (Promega, Madison, WI) restriction sites. The pRL-RARa1ΔERE construct was made using PCR products from 2 oligonucleotide sets, one designed for a PCR product upstream of ERE$_{RARA}$ and the other for the region downstream of ERE$_{RARA}$. PCRs were performed using Expand Long Template PCR System (Roche Applied Science, Germany). For the region upstream, the oligonucleotides used were 5'-GTCTCTCAGATGGAGGGTGATTCAGATCC-3' and 5'-GACTAGTGCCTGCGGGGTACAGTGACACATGGAGAC-3' followed by digestion with *Hind*III and *Spe*I. The region downstream of the ERE was amplified using oligos 5'-GACTAGTGTACACACCTACCTTGGAGTGGCTTTATCC-3' and 5'CGACGCGTCAGGAAGTGACAGCCACGTGACAGGAAGAC-3'. The digested PCR products were ligated into pRL-Null vector.

Cell culture and transient transfections.

MCF-7, MDA-MB 231 and BT-474 cells were cultured in Dulbecco's minimal essential medium (DMEM) containing penicillin (25 U/ml), streptomycin (25 U/ml), and 10% fetal calf serum at 37°C with 5% CO_2. At least 3 days prior to transfection, cells were cultured in phenol red-free DMEM supplemented with

10% charcoal-dextran-stripped fetal bovine serum. The cells were transfected with FuGENE 6 transfection reagent (Roche Applied Science), according to the protocol supplied by the manufacturer. Typically, 0.5 μg of reporter plasmid and 0.3 μg of pCMVßGal internal control were transfected per well. Twelve hours after transfection, fresh medium was added containing ethanol (vehicle) or estradiol (10^{-7} M). Cells were then harvested 24 hrs later and assayed for luciferase and ß-galactosidase. For pRL vector constructs, Renilla Luciferase was assayed using the Renilla Luciferase kit (Promega, Madison, WI).

Electromobility shift assay.

ERa proteins were synthesized by *in vitro* transcription-translation using rabbit reticulocyte lysates (Promega, Madison, WI). DNA-binding reactions were conducted as previously described (Tremblay et al., 1997) using 5 ml of programmed lysates in each binding reaction or 50 mg of protein obtained from a MCF-7 total cell extract.

RT-PCR

Total RNA was extracted from MCF-7 cells using RNeasy mini kit (Qiagen). Reverse transcription reactions were performed using Superscript II (Invitrogen) according to the manufacturer's recommendations. Quantitative RT-PCR was performed using Light Cycler (Roche) and QuantiTech SYBR Green PCR kit (Qiagen) according to the manufacturers' recommendations. Fold increase of transcription by estradiol were calculated using RPLP as an internal control. Primer sequences used for RT-PCR are available upon request.

Chromatin immunoprecipitation and ChIP-cloning.

The ChIP-cloning procedure has been described previously (Laganière et al., 2003). Briefly, the fragments obtained after ERα ChIP by double immunoprecipitation using an ERα antibody (HC20, Santa Cruz Biotechnology, Santa Cruz, CA) were repaired using T4 DNA polymerase, cloned in Ready-to-go pUC19-*Sma*1 BAP+Ligase (Pharmacia, NJ) and

sequenced. Quantitative PCR was performed using Light Cycler and SYBR Green Light cycler kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's recommendations. For standard ChIP, SRC-1, P/CIP and RIP140 antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA), and the SP1 antibody was obtained from Upstate (Lake Placid, NY). Primers used for standard ChIP at the *RARA* a1 promoter were 5'-TCTCCACCGAGCGCTATTTTCATTCTTTCC-3' and 5'-CTGACTGGTGATTGGTCGGTGGGCGGGCAG-3', the ERE_{RARA} region 5'-GAGGCTCAGGACAGGGCAAGAGTGGGGCAC-3' and 5'-GACAGAGGGAAGGA-GGGCTGAGGACCTGCG-3', the control region 5'-CTGGGCAATGCGAGGAGA-GTGAAGACTG-3' and 5'-GGGGAGGGAGGAGTTTGGGAGGAAGTGG-3'. All primers used for standard ChIP on novel regulatory modules are available upon request.

ACKNOWLEDGEMENTS

Financial support was provided by Genome Quebec/Canada and the Canadian Institutes for Health Research (CIHR). V. Giguère holds a CIHR senior scientist career award, and J. Laganière is a recipient of a U.S. Department of Defense Breast Cancer Research Program Predoctoral Traineeship Award (#W8IWXH-04-1-0399). We thank Yoshi Kiriyama for primers, Céline Lefebvre for expert technical assistance and Catherine Dufour for critical reading of the manuscript.

REFERENCES

Berry, M., Nunez, A.-M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci U S A *86*, 1218-1222.

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.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.

Charpentier, A. H., Bednarek, A. K., Daniel, R. L., Hawkins, K. A., Laflin, K. J., Gaddis, S., MacLeod, M. C., and Aldaz, C. M. (2000). Effects of estrogen on global gene expression: identification of novel targets of estrogen action. Cancer Res *60*, 5977-5983.

Chuang, F. M., West, B. L., Baxter, J. D., and Schaufele, F. (1997). Activities in Pit-1 determine whether receptor interacting protein 140 activates or inhibits Pit-1/nuclear receptor transcriptional synergy. Mol Endocrinol *11*, 1332-1341.

Coser, K. R., Chesnes, J., Hur, J., Ray, S., Isselbacher, K. J., and Shioda, T. (2003). Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in MCF7/BUS breast cancer cells by DNA microarray. Proc Natl Acad Sci U S A *100*, 13994-13999.

Cunliffe, H. E., Ringner, M., Bilke, S., Walker, R. L., Cheung, J. M., Chen, Y., and Meltzer, P. S. (2003). The gene expression response of breast cancer to growth regulators: patterns and correlation with tumor expression profiles. Cancer Res *63*, 7158-7166.

Demirpence, E., Balaguer, P., Trousse, F., Nicolas, J. C., Pons, M., and Gagne, D. (1994). Antiestrogenic effects of all-trans-retinoic acid and 1,25-dihydroxyvitamin D3 in breast cancer cells occur at the estrogen response element level but through different molecular mechanisms. Cancer Res *54*, 1458-1464.

Eng, F. C. S., Barsalou, A., Akutsu, N., Mercier, I., Zechel, C., Mader, S., and White, J. H. (1998). Different classes of coactivators recognize distinct but overlapping binding sites on the estrogen receptor ligand binding domain. J Biol Chem *273*, 28371-28377.

Farias, E. F., Arapshian, A., Bleiweiss, I. J., Waxman, S., Zelent, A., and Mira, Y. L. R. (2002). Retinoic acid receptor a2 is a growth suppressor epigenetically silenced in MCF-7 human breast cancer cells. Cell Growth Differ *13*, 335-341.

Fontana, J. A., Nervi, C., Shao, Z. M., and Jetten, A. M. (1992). Retinoid antagonism of estrogen-responsive transforming growth factor a and pS2 gene expression in breast carcinoma cells. Cancer Res *52*, 3938-3945.

Frasor, J., Danes, J. M., Komm, B., Chang, K. C., Lyttle, C. R., and Katzenellenbogen, B. S. (2003). 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 *144*, 4562-4574.

Gamble, M. J., and Freedman, L. P. (2002). A coactivator code for transcription. Trends Biochem Sci *27*, 165-167.

Giguère, V. (1994). Retinoic acid receptors and cellular retinoid binding proteins: complex interplay in retinoid signaling. Endocr Rev *15*, 61-79.

Giguère, V., Hollenberg, S. H., Rosenfeld, M. G., and Evans, R. M. (1986). Functional domains of the human glucocorticoid receptor. Cell *46*, 645-652.

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

Green, S., Walter, P., Kumar, V., Krust, A., Bornet, J. M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erbA. Nature *320*, 134-139.

Hermanson, O., Glass, C. K., and Rosenfeld, M. G. (2002). Nuclear receptor coregulators: multiple modes of modification. Trends in Endocrinology and Metabolism *13*, 55-60.

Herrlich, P. (2001). Cross-talk between glucocorticoid receptor and AP-1. Oncogene *20*, 2465-2475.

Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G., and Evans, R. M. (1985). Primary structure and expression of a functional human glucocorticoid receptor cDNA. Nature *318*, 635-641.

Inoue, A., Yoshida, N., Omoto, Y., Oguchi, S., Yamori, T., Kiyama, R., and Hayashi, S. (2002). Development of cDNA microarray for expression profiling of estrogen-responsive genes. J Mol Endocrinol *29*, 175-192.

Kang, H. Y., Yeh, S., Fujimoto, N., and Chang, C. (1999). Cloning and characterization of human prostate coactivator ARA54, a novel protein that associates with the androgen receptor. J Biol Chem *274*, 8570-8576.

Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R., and Chambon, P. (1987). Functional domains of the human estrogen receptor. Cell *51*, 941-951.
L'Horset, F., Dauvois, S., Heery, D. M., Cavaillès, V., and Parker, M. G. (1996). RIP-140 interacts with multiple nuclear receptors by means of two distinct sites. Mol Cell Biol *16*, 6029-6036.

Laganière, J., Deblois, G., and Giguère, V. (2003). Nuclear receptor target gene discovery using high throughput chromatin immunoprecipitation, In Methods in Enzymology, D. W. Russell, and D. J. Mangelsdorf, eds. (San Diego: Academic Press), pp. 339-350.

Leroy, P., Krust, A., Zelent, A., Mendelsohn, C., Garnier, J. M., Kastner, P., Dierich, A., and Chambon, P. (1991). Multiple isoforms of the mouse retinoic acid receptor α are generated by alternative splicing and differential induction by retinoic acid. European Molecular Biology Organization Journal *10*, 59-69.

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., Ramasamy, A., *et al.* (2004). Discovery of estrogen receptor a target genes and response elements in breast tumor cells. Genome Biology *5*, R66.

Lippman, M., Bolan, G., and Huff, K. (1976). The effects of estrogens and antiestrogens on hormone-responsive human breast cancer in long-term tissue culture. Cancer Res *36*, 4595-4601.

Lobenhofer, E. K., Bennett, L., Cable, P. L., Li, L., Bushel, P. R., and Afshari, C. A. (2002). Regulation of DNA replication fork genes by 17b-estradiol. Mol Endocrinol *16*, 1215-1229.

McKenna, N. J., Lanz, R. B., and O'Malley, B. W. (1999). Nuclear receptor coregulators: cellular and molecular biology. Endocrine Reviews *20*, 321-344.

Metivier, R., Penot, G., Carmouche, R. P., Hubner, M. R., Reid, G., Denger, S., Manu, D., Brand, H., Kos, M., Benes, V., and Gannon, F. (2004). Transcriptional complexes engaged by apo-estrogen receptor-a isoforms have divergent outcomes. European Molecular Biology Organization Journal *23*, 3653-3666.

Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-α directs ordered, cyclical, and

combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751-763.

Muller, P., Kietz, S., Gustafsson, J. A., and Strom, A. (2002). 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 *277*, 28376-28379.

O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. (2004). Genomic targets of nuclear estrogen receptors. Mol Endocrinol *18*, 1859-1875.

Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Pfahl, M. (1993). Nuclear receptor/AP-1 interaction. Endocr Rev *14*, 651-658.
Rishi, A. K., Shao, Z. M., Baumann, R. G., Li, X. S., Sheikh, M. S., Kimura, S., Bashirelahi, N., and Fontana, J. A. (1995). Estradiol regulation of the human retinoic acid receptor alpha gene in human breast carcinoma cells is mediated via an imperfect half-palindromic estrogen response element and SP1 motifs. Cancer Res *55*, 4999-5006.

Rousseau, C., Pettersson, F., Couture, M. C., Paquin, A., Galipeau, J., Mader, S., and Miller, W. H., Jr. (2003). The N-terminal of the estrogen receptor (ERa) mediates transcriptional cross-talk with the retinoic acid receptor in human breast cancer cells. J Steroid Biochem Mol Biol *86*, 1-14.

Safe, S., and Kim, K. (2004). Nuclear receptor-mediated transactivation through interaction with Sp proteins. Prog Nucleic Acid Res Mol Biol *77*, 1-36.

Seth, P., Krop, I., Porter, D., and Polyak, K. (2002). Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene *21*, 836-843.

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.

Sun, G., Porter, W., and Safe, S. (1998). Estrogen-induced retinoic acid receptor α 1 gene expression: role of estrogen receptor-Sp1 complex. Mol Endocrinol *12*, 882-890.

Tremblay, G. B., Tremblay, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Labrie, F., and Giguère, V. (1997). Cloning, chromosomal localization and functional analysis of the murine estrogen receptor β. Mol Endocrinol *11*, 353-365.

Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M. (1991). Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. Cell *65*, 1255-1266.

van der Leede, B. J., Folkers, G. E., van den Brink, C. E., van der Saag, P. T., and van der Burg, B. (1995). Retinoic acid receptor a1 isoform is induced by estradiol and confers retinoic acid sensitivity in human breast cancer cells. Mol Cell Endocrinol *109*, 77-86.

Wang, D. Y., Fulthorpe, R., Liss, S. N., and Edwards, E. A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. Mol Endocrinol *18*, 402-411.

Wang, Q., Yang, W., Uytingco, M. S., Christakos, S., and Wieder, R. (2000). 1,25-Dihydroxyvitamin D3 and all-trans-retinoic acid sensitize breast cancer cells to chemotherapy-induced cell death. Cancer Res *60*, 2040-2048.

Weisz, A., Basile, W., Scafoglio, C., Altucci, L., Bresciani, F., Facchiano, A., Sismondi, P., Cicatiello, L., and De Bortoli, M. (2004). Molecular identification of ERa-positive breast cancer cells by the expression profile of an intrinsic set of estrogen regulated genes. J Cell Physiol *200*, 440-450.

White, K. A., Yore, M. M., Deng, D., and Spinella, M. J. (2005). Limiting effects of RIP140 in estrogen signaling: potential mediation of anti-estrogenic effects of retinoic acid. J Biol Chem *280*, 7829-7835.

Yamamoto, K. R. (1985). Steroid receptor regulated transcription of specific gene networks. Annu Rev Genet *19*, 209-252.

Zhu, W. Y., Jones, C. S., Kiss, A., Matsukuma, K., Amin, S., and De Luca, L. M. (1997). Retinoic acid inhibition of cell cycle progression in MCF-7 human breast cancer cells. Exp Cell Res *234*, 293-299.

Zou, A., Marschke, K. B., Arnold, K. E., Berger, E. M., Fitzgerald, P., Mais, D. E., and Allegretto, E. A. (1999). Estrogen receptor β activates the human retinoic acid receptor α-1 promoter in response to tamoxifen and other estrogen receptor antagonists, but not in response to estrogen. Mol Endocrinol *13*, 418-430.

TABLES AND FIGURES

Table 2.1. Estrogen response elements and half-sites contained in the fragments obtained by ERα ChIP-cloning.

Estrogen response element and nuclear receptor half-site sequences contained in the MCF-7 genomic fragments isolated by ChIP-cloning using an ERα antibody are shown. For each clone, the EREs and half-sites are shown, allowing 1 or 2 mismatches, respectively. Capital letters indicate the bases corresponding to the ERE or half-site consensus.

Table 2.1.

Clone	EREs or A/GGGTCA-related half-sites
DER001	GGGACA, GGGTCAcggTGACAC, GTGTCA, AGGTCA
DER108	GGGTCAggcAGACCT, ATGTCA, AGTTCA, AGCTCA
ER4282	AGGTCA, AGGTCC, GGGTAA, GGCTCA, TGGCCAccgTGACCT, AGCTCA, GGCTCA, AGGCCA, GGATCA
DER7235	AGGCCA, AGGTCT, AGGTCA, GGATCA,
DER009	GGATCA, AGGTCT, AAGTCA
DERx15	GGGCCA, AGGTCA, TGGTCA, AGCTCA, GGGGCA
DER119	GGGTCA, GGGTCG, GGGTGA, AGGGCA, ATGTCA, GTGTCA, AGGTGA, AGGTCA
DER023	AGGGCA, GGGTGA, AGGTCA, GGGTCT, AGGTCA, AGGGCA
DER169	GGGTCA, GGGGCA, AGCTCA, GGGGCA, GGGACA, GGGTCA, GGGGCA, GGGTCA, AGGACA
DER005	AGATCA, AGGGCA
DER203	AGGTCA, TGGTCA, AGGTAA, GGGTCT, TGGTCA, GGCTCA, AAGTCA, GGGTTA, AGGGCA, GGTTCA, AAGTCA
DER148	GGCTCA, GGATCA, GGATCA, AGGTGA, AGGTCA, TGGTCA

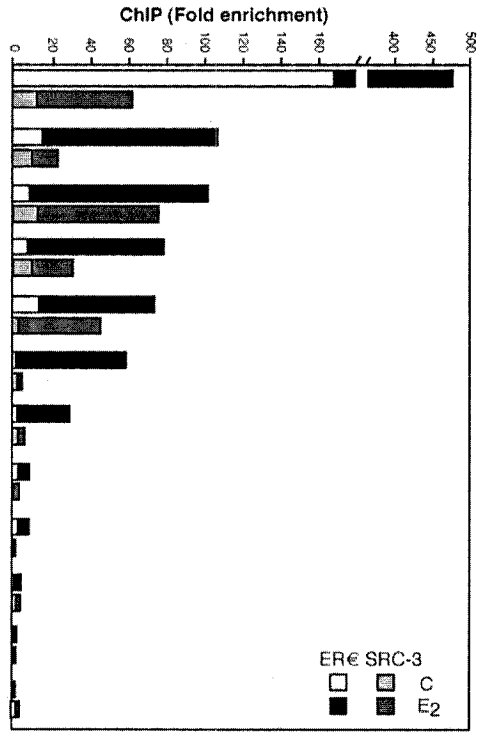
Figure 2.1. Identification of direct ERα target regulatory modules in MCF-7 cells.

A, *In vivo* binding of ERα and SRC-3 to novel regulatory modules identified by ChIP-cloning. Standard ERα and SRC-3 ChIP assays in MCF-7 cells, followed by quantitative PCR using primers specific for the region cloned. The graph shows enrichment in fold over the control (no antibody) by quantitative PCR. The results shown here are representative of 3 independent ChIP experiments. Clone: clone number of the fragments obtained by ChIP-cloning; Gene: closest annotated gene from the UCSC human genome database; Locus link: locus link number of the gene; Distance: distance from the closest annotated gene. B, Modulation of target gene expression by estradiol. MCF-7 cells were treated with estradiol for various time points. The total RNA was extracted and used for cDNA production followed by quantitative RT-PCR using specific primers.

Figure 2.1

A

Clone	Distance	Gene	Unigene #
DER001	- 8,62	RARA	Hs.361071
DER108	+ 13,6	FLJ41849	Hs.463024
ER4282	- 0,22	TFF1	Hs.350470
DER7235	- 0,14	FLJ10618	Hs.144130
DER009	+ 11,9	BCR	Hs.446394
DERx15	- 103	DDEF2	Hs.200666
DER119	+ 10	FLJ16032	Hs.211475
DER023	+ 21	EDG1	Hs.154210
DER169	+ 5,12	EDF1	Hs.174050
DER005	- 0,77	NAP1L4	Hs.269403
DER203	- 2,2	RNF14	Hs.170926
DER148	+ 2,6	GPR81	Hs.326712



B

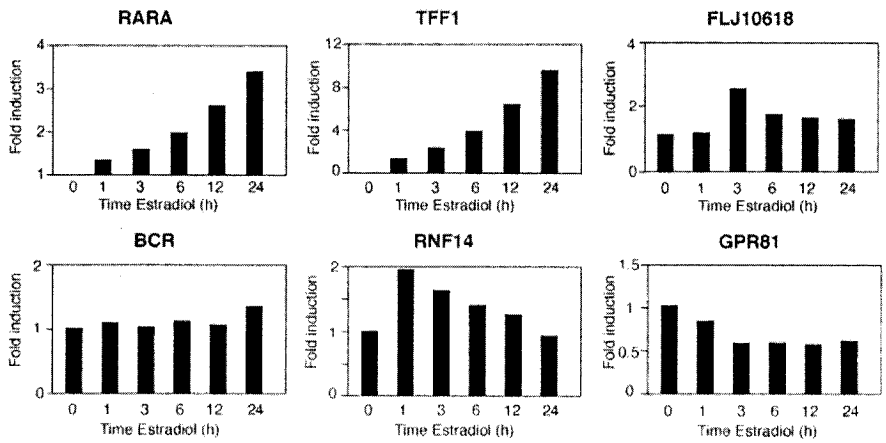


Figure 2.2. An ERE is located in the first intron of *RARA* gene.

A, The *RARA* transcripts are produced from two distinct promoters, α1 and α2. The ERE isolated by ChIP-cloning is located 3.7 kb downstream of the first *RARA* α1 promoter and 8.6 kb upstream of the second *RARA* α2 promoter (UCSC annotation). B, Complete sequence of the DER001 genomic fragment. The ERE_{RARA} is boxed. The fragment obtained by ERa ChIP-cloning is well conserved among the human, mouse and rat genomes. The * sign represents the bases that are conserved. Overall, 70 bases out of the 130 bp are conserved in the three genomes shown. C, The ERE_{RARA} is conserved among these genomes. The bases corresponding to the ERE consensus sequence are shown in capital letters.

Figure 2.2.

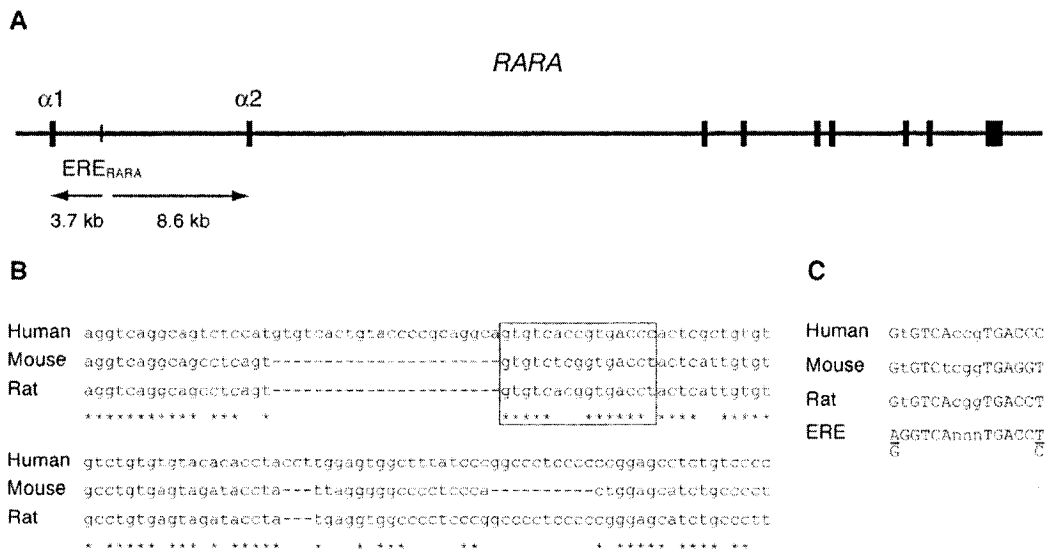

A
RARA
α1
α2
ERE$_{RARA}$
3.7 kb
8.6 kb
B
Human aggtcaggcagtctccatgtgtcactgtaccccgcaggcagtgtcaccgtgaccgactcgctgtgt
Mouse aggtcaggcagcctcagt----------gtgtctcggtgacctactcattgtgt
Rat aggtcaggcagcctcagt----------gtgtcacggtgacctactcattgtgt
Human gtctgtgtgtacacacctaccttggagtggctttatcccggccctccccccggagcctctgtcccc
Mouse gcctgtgagtagataccta---ttagggggccccctccca----------ctggagcatctgccccc
Rat gcctgtgagtagataccta---tgaggtggccccctcccgqccccctccccccgggagcatctgccctt
C
Human GtGTCAccgTGACCC
Mouse GtGTCtcggTGAGGT
Rat GtGTCAcggTGACCT
ERE AGGTCAnnnTGACCT
G C

Figure 2.3. The novel ERE_{RARA} isolated in the *RARA* first intron is functional.

A, EMSA using *in vitro* translated ERα and the ERE_{RARA} as a probe. Comp: vitallogenin ERE competitor probe. B, EMSA using endogenous ERα from MCF-7 total cell extracts. Ab: ERα antibody supershift. C, Transient transfection in MCF-7 cells. TK-Luc reporters containing either the whole DER001 fragment obtained by ChIP-cloning or ERE_{RARA} were transfected in MCF-7 cells, in the presence of estradiol (E2) or vehicle (C).

Figure 2.3.

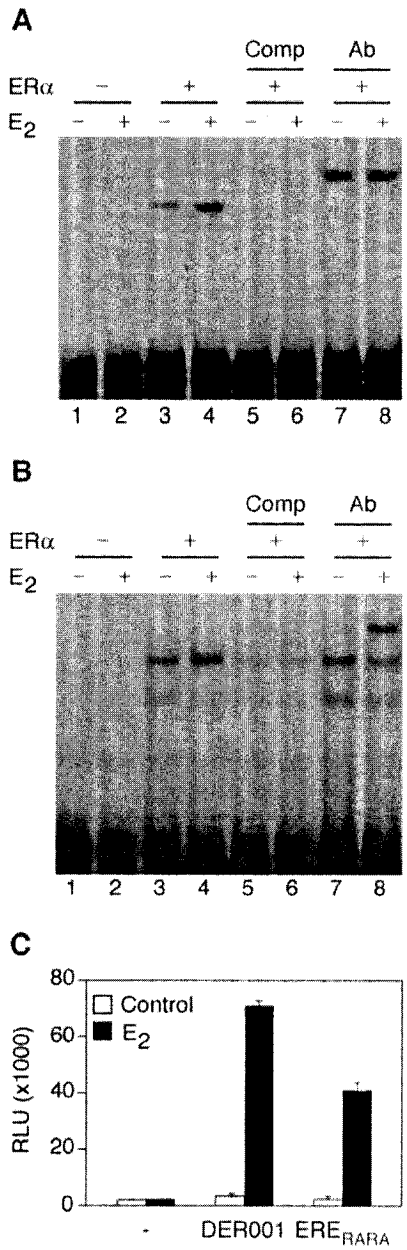

A
Comp
Ab
ERα
E2
1 2 3 4 5 6 7 8
B
Comp
Ab
ERα
E2
1 2 3 4 5 6 7 8
C
80
60
40
20
0
Control
E2
RLU (x1000)
DER001 ERE_RARA

Figure. 2.4. The ERE_{RARA} is active *in vivo*.

SP1, ERα, SRC-1, SRC-3 and RIP140 ChIP assays were performed using MCF-7 cells in the presence of estradiol (E2) or vehicle (C), followed by PCR using primers specific for the *RARA* a1 promoter, the intronic ERE (3.7 kb downstream the promoter) and a control region located 4 kb upstream of the *TFF1* promoter (control). No ab: no antibody control.

Figure 2.4

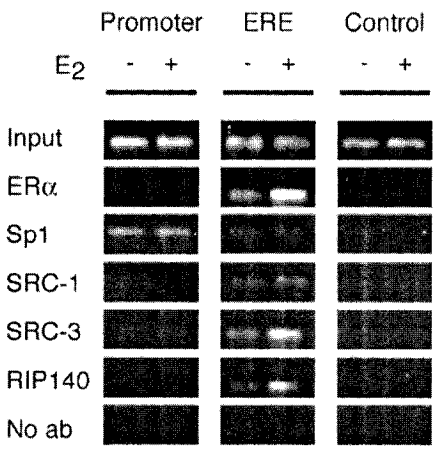
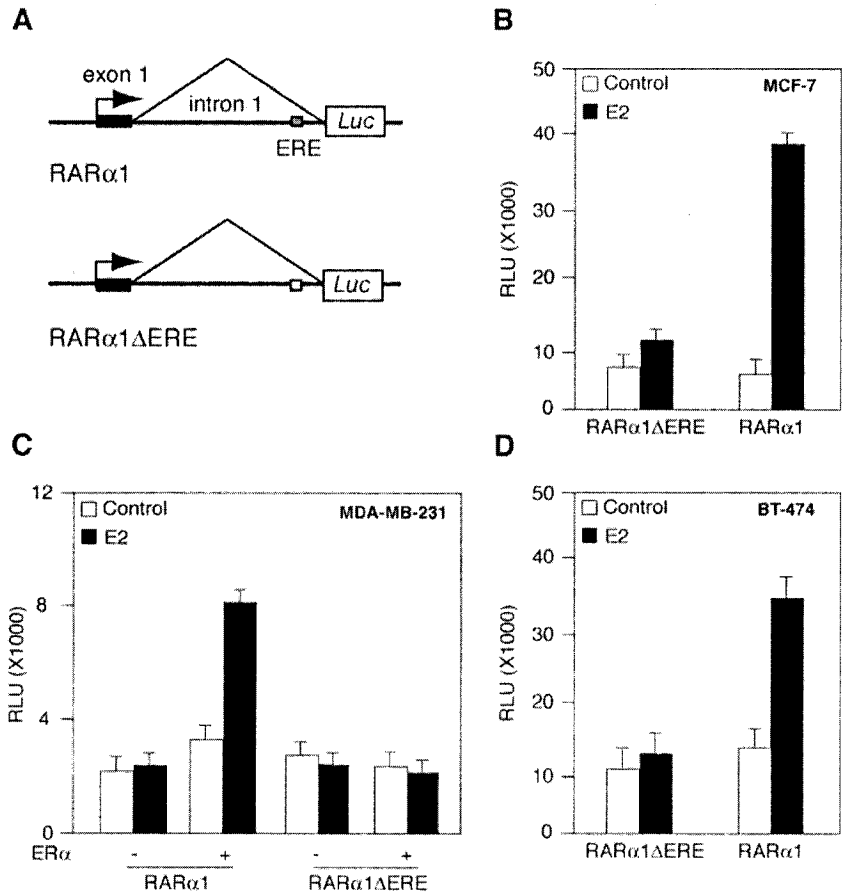


Figure. 2.5. The intronic ERE_{RARA} controls *RARA* α1 response to estradiol. A, Schematic representation of the region of the *RARA* locus inserted in the pRL-Null vector. RARα1ERE and RARα1ΔERE represents the locus with and without ERE_{RARA}, respectively. B, Transient transfections in MCF-7 cells in the presence of estradiol (E2) or vehicle (C). C, Transient co-transfections of the two reporter constructs in MDA-MB-231 cells together with ERα or empty vector (CMX), in the presence of estradiol or vehicle. D, Transient transfections of the reporters described in (A), in an ER-positive breast cancer cell line BT-474.Figure 2.5. The intronic ERE_{RARA} controls *RARA* α1 response to estradiol.

Figure 2.5.


A
exon 1
intron 1
Luc
ERE
RARα1
Luc
RARα1ΔERE
B
Control
E2
MCF-7
RLU (X1000)
50
40
30
20
10
0
RARα1ΔERE
RARα1
C
Control
E2
MDA-MB-231
RLU (X1000)
12
8
4
0
ERα
-
+
-
+
RARα1
RARα1ΔERE
D
Control
E2
BT-474
RLU (X1000)
50
40
30
20
10
0
RARα1ΔERE
RARα1

CHAPTER III: Location Analysis of Estrogen Receptor α Target Promoters Reveals that FOXA1 Defines a Domain of the Estrogen Response

PREFACE

In the previous chapter, I demonstrated the usefulness of ChIP-derived functional approaches in identifying genuine transcriptional regulatory regions. In the present chapter, I used the ChIP-on-chip technique to identify ERα genomic binding sites in a more high-throughput fashion, thereby allowing the analysis of its global action in response to its natural ligand in breast cancer cells. In this study, I provided the portrait of the gene promoters directly controlled by endogenous ERα in this context. In addition, a great finding was the identification of FOXA1 as a factor necessary for the recruitment of ERα to a subset of promoters, besides being a direct target of ER, indeed necessary for estrogen-induced cell cycle progression. This suggests that the breast cancer overexpressed FOXA1 can compartmentalize the estrogen response and control, with ERα, pathways monitoring cancer cell proliferation.

This chapter is an article that has been published in the *Proceedings of the National Academy of Sciences* journal. It was featured on the cover of the journal.

ABSTRACT

Nuclear receptors can activate diverse biological pathways within a target cell in response to their cognate ligands but how this compartmentalization is achieved at the level of gene regulation is poorly understood. We used a genome-wide analysis of promoter occupancy by the estrogen receptor (ERα) in MCF-7 cells to investigate the molecular mechanisms underlying the action of 17β-estradiol (E_2) in controlling the growth of breast cancer cells. We identified 153 promoters bound by ERα in the presence of E_2. Motif-finding algorithms demonstrated that the estrogen response element (ERE) is the most common motif present in these promoters while conventional chromatin immunoprecipitation assays showed E_2-modulated recruitment of co-activator AIB1 and RNA polymerase II at these loci. The promoters were linked to known ERα targets but also to many genes not directly associated with the estrogenic response, including the transcriptional factor FOXA1, whose expression correlates with the presence of ERα in breast tumors. We found that ablation of FOXA1 expression in MCF-7 cells suppressed ERα binding to the prototypic *TFF1* promoter that contains a FOXA1 binding site, hindered the induction of its expression by E_2 and prevented hormone-induced reentry into the cell cycle. Taken together, these results define a paradigm for estrogen action in breast cancer cells, and suggest that regulation of gene expression by nuclear receptors can be compartmentalized into unique transcriptional domains via licensing of their activity to cofactors such as FOXA1.

INTRODUCTION

Estradiol (E_2) is a potent growth factor of human breast cancer cells that exerts its action mainly through estrogen receptor α (NR3A1, ERα), a member of the superfamily of nuclear receptors (Osborne and Schiff, 2005). Despite significant advancement into our understanding of the molecular mechanisms of ERα action (Smith and O'Malley, 2004), little is known about mediators of the estrogen pathway that assist in the initiation, compartmentalization and propagation of its signal at the level of gene expression. Delineation of how ERα induces precise biological responses in breast cancer cells and other cell types has clearly been limited by the lack of data on the transcriptional regulatory regions of ERα direct target genes.

ERα regulates the expression of target genes by binding to specific sites in the chromatin, referred to as estrogen response elements (EREs) (O'Lone et al., 2004), or by interacting with other transcription factors and modulates gene transcription via their own specific binding sites (Herrlich, 2001; Pfahl, 1993; Safe and Kim, 2004). Determination of ERα target genes has recently been undertaken using DNA microarrays, identifying hundreds of genes with altered expression upon E_2 treatment of human breast cancer cells (Charpentier et al., 2000; Coser et al., 2003; Cunliffe et al., 2003; Frasor et al., 2003; Inoue et al., 2002; Lin et al., 2004; Lobenhofer et al., 2002; Seth et al., 2002; Soulez and Parker, 2001; Wang et al., 2004; Weisz et al., 2004). However, while providing information of the global action of E_2 in these cells, gene expression profiling can rarely discriminate between direct and indirect ERα targets. In addition, bioinformatic and comparative genomics have also been used successfully to identify high-affinity and physiologically relevant EREs encoded in the human genome (Bourdeau et al., 2004; Kamalakaran et al., 2005). These studies have also some constraints, including their limitation to consensus EREs and the general absence of large scale functional data linking these putative binding sites with gene expression in specific cell types.

Recently, chromatin immunoprecipitation (ChIP) has been used in combination with promoter or genomic DNA microarrays to identify loci

recognized by transcription factors in a genome-wide manner in mammalian cells (Cam et al., 2004; Cawley et al., 2004; Odom et al., 2004; Ren et al., 2002; Weinmann et al., 2002). This technology, termed ChIP-on-chip or location analysis, can therefore be used to determine the global gene expression program that characterize the action of a nuclear receptor in response to its natural ligand. For this study, we first constructed a human proximal promoter DNA microarray containing approximately 19,000 promoters and then monitored occupancy by ERα at these promoters in MCF-7 breast cancer cells in the presence of E_2. Our experiments identified genes that include known ERα targets, genes previously associated with the E_2 response but not characterized as direct targets, and several novel target genes. Among those genes, we identified the transcriptional factor FOXA1 whose expression correlates with the presence of ERα in breast tumors. We found that knock-down of FOXA1 expression in MCF-7 in cells using small interfering RNA (siRNA) depletion experiments diminished ERα binding to the prototypic *TFF1* promoter that containing a FOXA1 binding site, reduced the induction of its expression by E_2 and prevented hormone-induced reentry into the cell cycle. Our results demonstrate that FOXA1 licensing plays an unsuspected role in defining a sub-domain of the transcriptional response to E_2 in breast cancer cells, and suggest that more precise therapeutic approaches could be developed to target the wide-ranging action of E_2 in the normal and disease states.

MATERIALS AND METHODS

Human promoter microarray design. The strategy adopted to design our promoter microarray is similar to the one used by the Young group (Odom et al., 2004). Full length complementary DNAs were extracted from Refseq and mammalian gene collection databases and filtered to eliminate redundancy and incomplete cDNAs. Their transcription start sites were then located using the UCSC genome browser (Karolchik et al., 2003) and the sequence ranging from 800 base pairs (bp) upstream of the transcription start sites to 200 bp downstream of the transcription start sites was extracted using the UCSC database assemblage July 2003 (Karolchik et al., 2003). Primer pairs were designed using the Primer3 algorithm (Rozen and Skaletsky, 2000) and the specificity tested *in silico* using a virtual PCR algorithm (Lexa et al., 2001). When the primer pair gave no satisfactory virtual PCR results, a new primer pair was designed using Primer3 and tested again. The process was iterated 3 times to generate primer pairs predicted to be efficient to amplify promoter regions from human genomic DNA for almost all of our selected genes. This strategy was adopted after preliminary results showed that a more simple primer design approach did not generate good results when we tried to amplify promoter regions from human genomic DNA. This primer design pipeline allowed us to design primer pairs to amplify promoter regions from human genomic DNA with a success rate of ~80%, which is slightly better then that reported previously (Odom et al., 2004). At the date of the download (July 2004) 21,416 RefSeq and 16,521 MGC entries were retrieved. After the filtering process 18,741 of them were selected and submitted to primer design. Primers were obtains for 18,660 promoters and 188 controls were added (located in exons and far from any known genes).

Genome-wide location analysis and chromatin immunoprecipitation. After 72 h of steroid deprivation followed by 45 min of E_2 (100 nM) treatment, MCF-7 cells were fixed with 1% final concentration formaldehyde for 10 min at room temperature, harvested and rinsed with 1x PBS. The resultant cell pellet was sonicated, and protein-DNA complexes were enriched by immunoprecipitation

with the ERα specific antibody (Santa Cruz Biotechnology), beads were added and washed as previously described (Laganiere et al., 2005). After de-crosslinking, the enriched DNA was repaired with T4 DNA polymerase (New England Biolabs) and ligated with linkers, as described in (Odom et al., 2004). DNA was amplified using ligation-mediated PCR (LM-PCR), and then fluorescently labeled using BioPrime Array CGH genomic labeling kit and the Cy5 fluorophore (Invitrogen). A sample of DNA that has not been enriched by immunoprecipitation was subjected to LM-PCR and labeled with Cy3 fluorophore. Both IP-enriched and non-enriched pools of labeled DNA were hybridized to the human promoter array described above. The p-value threshold used to select target promoters for further analyses was determined empirically by testing randomly selected targets by standard ChIP/Q-PCR. Based on these experiments we used p=0.005 since our estimated false positive rate was less than 10% (genes tested=34, see Table 2 in supplementary information) using this threshold. FOXA1 ChIP assays were performed using two distinct antibodies from Chemicon and Santa Cruz Biotechnology. RNA polymerase II and AIB1 ChIP assays were performed using antibodies from Upstate Biotechnologies and Santa Cruz, respectively.

Promoter sequence analysis. We used a motif finding algorithm (MDScan) (Liu et al., 2002) to uncover motifs that are highly represented in our set of promoter sequences. The presence of EREs and FOXA1 binding sites was also determined using MacVector (Accelrys) and TRANSFAC (Matys et al., 2003). The logo pictured in Fig. 3.1A was generated using WebLogo (weblogo.berkeley.edu/logo.cgi).

Functional classification of target genes. Functional categories were assigned using both GO (www.fatigo.bioinfo.cnio.es) and manual inspection using PubMed (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed).

Cell culture, luciferase and cell cycle entry assays. MCF-7 cells were cultured as previously described (Laganiere et al., 2005). For the luciferase assay, cells were transfected with Lipofectamine 2000 (Invitrogen) with 0.4 μg of TFF1-Luc (Lu et al., 2001) and 0.2 μg of pCMVβGal internal control per

well, 0.1 μg of CMX-ERα, and 100 nM final concentration of FOXA1 or control siRNA (SMARTpool reagents, Dharmacon). Twelve hours after transfection, fresh medium was added, incubated for 12 h and then treated with ethanol (vehicle) or E_2 (10^{-7}M) for 20 h. Cells were then harvested and assayed for luciferase and β–Galactosidase activities. For FACS analysis, cells were cultured in steroid-deprived media for 48 h, transfected with *FOXA1* or control siRNAs and incubated for 36 h and treated with E_2 or vehicle for 20 h. Cells were then trypsinized, fixed in 70% EtOH and stored at -20 °C overnight. Before analysis, cells were washed in PBS, resuspended in a solution containing 0.5 mg/ml of RNase (Sigma) and 5 μg/ml of propidium iodide (Sigma) and analyzed on a FACScan (Becton Dickinson).

Western blot and RT-PCR.

Western blot was performed using FOXA1 and actin antibodies (Santa Cruz). RT-PCR was conducted as described in (Laganiere et al., 2005).

RESULTS AND DISCUSSION

ChIP-on-chip analysis of ERα binding. The MCF-7 cell line is a well-established model for the study of E_2-induced human breast cancer cell growth and was thus selected for this study (Lacroix and Leclercq, 2004). To identify targets of ERα in an unbiased genome-wide manner, we constructed a genomic DNA microarray containing the region spanning 800 base pairs upstream and 200 base pairs downstream of transcription start sites of 18,660 human genes. We identified a total of 153 promoters ($p<0.005$) bound by ERα in the presence of E_2 (Table 1, and Table 3 in supporting information). We confirmed binding by ERα to a subset of targets using conventional ChIP assays and quantitative PCR and determined that our rate of false positives was below 10% when previously established threshold criteria were used (see Materials and Methods). The results of the genome location experiment were further validated using a motif finding algorithm that examines the ChIP-on-chip selected sequences and searches for DNA sequence motifs representing the protein-DNA interaction sites (Liu et al., 2002). The consensus sequence derived from the most frequent motifs found in the ERα-bound promoters corresponds to a perfect estrogen response element (GGTCANNNTGACCT, Fig. 3.1*A*). If these genes are indeed regulated by E_2-bound ERα, co-regulator proteins and RNA polymerase II should also be recruited to the promoters in response to E_2. Examination of a subset of ERα-bound promoters using conventional ChIP demonstrated that a number of loci recruited the nuclear receptor co-activator AIB1 (also known as SRC-3, pCIP and ACTR) (Anzick et al., 1997; Chen et al., 1997; Torchia et al., 1997) in the presence of the hormone while the amount of RNA polymerase II was consistently increased above the basal level observed for each individual gene (Fig. 3.1*B*). One exception was for ABCC5, a gene previously found to be down regulated by E_2 (Vendrell et al., 2004), demonstrating that both up- and down-regulated genes can be identified using the promoter array.

FOXA1, a target of ERα co-expressed in breast tumors, is recruited to a subset of ERα targets. Although some known direct targets of ERα were

selectively enriched from the chromatin of MCF-7 cells (e.g. *CASP7*, *CYP1B1*, *GREB1*, *LY6E*, *SHP*, *SLC25A36/FLJ10618*, *TFF1*, *WISP2*), most of the genes identified represent novel primary targets of ERα. We used gene ontology (GO) (Al-Shahrour et al., 2004) to classify our ERα targets into functional categories and found that ERα regulates a wide array of cellular processes and molecular functions (Table 3.1 and Fig. 3.2*A*). Within these categories, we identified genes involved in Wnt signaling (*WNT16*, *WISP2*, *SEMA3B*, *CTNNBIP1*), steroid metabolism (*CYP1B1*, *STS*, *UGT2B15*, *UGT2B17*), multi-drug resistance (*ABCC5*, *ABCC11*) and cell cycle regulation (*CDK5* and *RBL2* also known as p130). Given the well-known property of E_2 to stimulate cell cycle progression of MCF-7 cells and other breast cancer cell lines (Prall et al., 1998), it was surprising that few key genes known to regulate the cell cycle were obtained in our location analysis. Although some ERα targets are likely to be regulated via enhancers located at a great distance form the transcription start sites and be missed by a promoter array, these results do suggest that ERα requires specific downstream effectors to regulate cell growth. These effectors are likely to be involved in transcriptional regulation, and this category was well represented among ERα targets (Fig. 3.2*A*). In addition to the known regulation by ERα of its own promoter (*ESR1*) and that of the orphan nuclear receptor SHP (*NR0B2*) (Lai et al., 2003), we identified the nuclear receptor coactivator PRC (*PPRC1*) and the forkhead transcription factor HNF3α/FOXA1 (*FOXA1*) as direct targets of ERα. Interestingly, the expression of *FOXA1*, a pioneer factor with the ability to initiate chromatin opening events (Cirillo et al., 2002) and previously shown to establish a promoter environment favorable to transcriptional activation by ERα (Robyr et al., 2000), correlates (Fig. 3.6 in supporting information, r^2=0.7987) with the presence of ERα in human breast tumors (Rhodes et al., 2004; van 't Veer et al., 2002), and is rapidly induced by E_2 in MCF-7 cells (Fig. 3.2*B*). In addition, motif-finding analysis using the consensus FOXA1 binding site WTGRTTNRTT revealed that a specific subset (~12%) of the ERα-bound promoters contained FOXA1 recognition sites. Conventional ChIP experiments

on selected promoter regions detected various levels of enrichment of these sequences with antibodies against FOXA1 in both the absence or the presence of E_2 (Fig. 3.2*C*). *TFF1*, a gene also referred to as pS2 and known to be strongly regulated by ERα (Berry et al., 1989), displayed the most robust enrichment of FOXA1 at its promoter, while control promoters without a FOXA1 binding site (*STS* and *HK1*) failed to recruit FOXA1. Taken together, these results suggest that FOXA1 could serve as a licensing factor to propagate a specific domain of the estrogenic response in breast cancer cells.

FOXA1 is required for ERα action on the *TFF1* promoter. We next examine whether FOXA1 plays a functional role in transcriptional activation of this subset of ERα target genes by transfecting siRNAs directed against *FOXA1* in MCF-7 cells. The presence of the siRNAs specifically knocked-down FOXA1 protein level (Fig. 3.3*A*) and reduced the ability of E_2 to stimulate the expression of a selected FOXA1/ERα target, *TFF1* (Fig. 3.3*B*), but not the control promoter *STS*. Similar results were obtained when the ability of ERα to stimulate the activity of the *TFF1* promoter was tested in a co-transfection assay in MCF-7 cells. As shown in Fig. 3.3*C*, introduction of siRNAs directed against FOXA1 considerably impaired the response of the *TFF1* promoter to E_2. The introduction of siRNA directed against FOXA1 did not affect the expression of ERα as monitored by Western blot (data not shown). Since FOXA1 binding to the *TFF1* promoter was not affected by treatment with E_2 (Fig. 3.2*B*), we next investigated whether the presence of FOXA1 is required for binding of ERα to the *TFF1* promoter as well as other ERα-bound promoters containing FOXA1 sites. As shown in Fig. 3.3*D*, knock-down of FOXA1 expression resulted in a marked reduction of the E_2-induced recruitment of ERα to the *TFF1* promoter, as well as to the *RPS6KL1*, *ABCC5* and *UGT2B17* promoters, while the recruitment of ERα to a control promoter (*STS*) was not affected. These results demonstrates that FOXA1 is playing an important role in ERα binding and transcriptional activity of a specific subset of FOXA1/ ERα target promoters in MCF-7 cells.

FOXA1 is required for E_2-induced reentry into the cell cycle. One hallmark of E_2 action is its ability to induce synchronous cell cycle reentry of steroid-deprived quiescent breast cancer cells (Musgrove and Sutherland, 1994). We thus tested the possibility that FOXA1 could serve as a mediator of ERα action in this process. MCF-7 cells synchronized in quiescence by depletion of steroid hormones for 48 h were released from quiescence by exposure to E_2 and harvested for cell cycle analysis by flow cytometry. As shown in Fig. 3.4, MCF-7 cells transfected with siRNAs directed against *FOXA1* failed to reenter the cell cycle upon stimulation with E_2.

Compartmentalization of the hormonal response. In this study, using a combination of genome-wide location, genetic analyses and functional assays, we identified FOXA1 as being essential for ERα binding to *TFF1*, a prototypic gene representing a subset of ERα target promoters, and required for E_2-induced reentry of quiescent breast cancer cells into the cell cycle. These results not only present a new paradigm in estrogen action but suggest a mechanism by which nuclear receptors can regulate specific subset of genes and biological responses with the cooperation of downstream effectors that are essential to both initiate and propagate the hormonal signal (Fig. 3.5). This study demonstrates that licensing factors, such as FOXA1, that are both under hormonal control and necessary for the hormonal response, can be used to compartmentalize the action of nuclear receptors at the level of the genome. These findings thus suggest the existence of new opportunities to target more precisely the action of nuclear receptors for the prevention and management of hormone-dependent diseases.

ACKNOWLEDGEMENTS

We thank Laurent Sansregret for expert advice. This work was supported by Genome Québec/Canada and the Canadian Institutes for Health Research (CIHR). V. Giguère holds a CIHR senior scientist career award, F. Robert holds a CIHR New Investigator Award and J. Laganière is a recipient of a U.S. Department of Defense Breast Cancer Research Program Predoctoral Traineeship Award (#W8IWXH-04-1-0399).

REFERENCES

Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics *20*, 578-580.

Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X.-Y., Sauter, G., Kallioniemi, O.-P., Trent, J. M., and Meltzer, P. S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science *277*, 965-968.

Berry, M., Nunez, A.-M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci U S A *86*, 1218-1222.

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.

Cam, H., Balciunaite, E., Blais, A., Spektor, A., Scarpulla, R. C., Young, R., Kluger, Y., and Dynlacht, B. D. (2004). A common set of gene regulatory networks links metabolism and growth inhibition. Mol Cell *16*, 399-411.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.

Charpentier, A. H., Bednarek, A. K., Daniel, R. L., Hawkins, K. A., Laflin, K. J., Gaddis, S., MacLeod, M. C., and Aldaz, C. M. (2000). Effects of estrogen on global gene expression: identification of novel targets of estrogen action. Cancer Res *60*, 5977-5983.

Chen, H., Lin, R. J., Schlitz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatani, Y., and Evans, R. M. (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and form a multimeric activation complex with P/CAF and CBP/p300. Cell *90*, 569-580.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell *9*, 279-289.

Coser, K. R., Chesnes, J., Hur, J., Ray, S., Isselbacher, K. J., and Shioda, T. (2003). Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in MCF7/BUS breast cancer cells by DNA microarray. Proc Natl Acad Sci U S A *100*, 13994-13999.

Cunliffe, H. E., Ringner, M., Bilke, S., Walker, R. L., Cheung, J. M., Chen, Y., and Meltzer, P. S. (2003). The gene expression response of breast cancer to growth regulators: patterns and correlation with tumor expression profiles. Cancer Res *63*, 7158-7166.

Frasor, J., Danes, J. M., Komm, B., Chang, K. C., Lyttle, C. R., and Katzenellenbogen, B. S. (2003). 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 *144*, 4562-4574.

Herrlich, P. (2001). Cross-talk between glucocorticoid receptor and AP-1. Oncogene *20*, 2465-2475.

Inoue, A., Yoshida, N., Omoto, Y., Oguchi, S., Yamori, T., Kiyama, R., and Hayashi, S. (2002). Development of cDNA microarray for expression profiling of estrogen-responsive genes. J Mol Endocrinol *29*, 175-192.

Kamalakaran, S., Radhakrishnan, S. K., and Beck, W. T. (2005). Identification of Estrogen-responsive Genes Using a Genome-wide Analysis of Promoter Elements for Transcription Factor Binding Sites. J Biol Chem *280*, 21491-21497.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., *et al.* (2003). The UCSC Genome Browser Database. Nucleic Acids Res *31*, 51-54.

Lacroix, M., and Leclercq, G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat *83*, 249-289.

Laganiere, J., Deblois, G., and Giguere, V. (2005). Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor alpha1 gene in breast cancer cells. Mol Endocrinol *19*, 1584-1592.

Lai, K., Harnish, D. C., and Evans, M. J. (2003). Estrogen receptor alpha regulates expression of the orphan receptor small heterodimer partner. J Biol Chem *278*, 36418-36429.

Lexa, M., Horak, J., and Brzobohaty, B. (2001). Virtual PCR. Bioinformatics *17*, 192-193.

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., Ramasamy, A., *et al.* (2004). Discovery of estrogen receptor a target genes and response elements in breast tumor cells. Genome Biology *5*, R66.

Liu, X. S., Brutlag, D. L., and Liu, J. S. (2002). An algorithm for finding protein DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol *20*, 835-839.

Lobenhofer, E. K., Bennett, L., Cable, P. L., Li, L., Bushel, P. R., and Afshari, C. A. (2002). Regulation of DNA replication fork genes by 17b-estradiol. Mol Endocrinol *16*, 1215-1229.

Lu, D., Kiriyama, Y., Lee, K. Y., and Giguère, V. (2001). Transcriptional regulation of the estrogen-inducible pS2 beast cancer marker gene by the ERR family of orphan nuclear receptors. Cancer Res *61*, 6755-6761.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., *et al.* (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res *31*, 374-378.

Musgrove, E. A., and Sutherland, R. L. (1994). Cell cycle control by steroid hormones. Semin Cancer Biol *5*, 381-389.

O'Lone, R., Frith, M. C., Karlsson, E. K., and Hansen, U. (2004). Genomic targets of nuclear estrogen receptors. Mol Endocrinol *18*, 1859-1875.

Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Osborne, C. K., and Schiff, R. (2005). Estrogen-receptor biology: continuing progress and therapeutic implications. J Clin Oncol *23*, 1616-1622.

Pfahl, M. (1993). Nuclear receptor/AP-1 interaction. Endocr Rev *14*, 651-658.

Prall, O. W., Rogan, E. M., and Sutherland, R. L. (1998). Estrogen regulation of cell cycle progression in breast cancer cells. J Steroid Biochem Mol Biol *65*, 169-174.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., and Dynlacht, B. D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev *16*, 245-256.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

Robyr, D., Gegonne, A., Wolffe, A. P., and Wahli, W. (2000). Determinants of vitellogenin B1 promoter architecture. HNF3 and estrogen responsive transcription within chromatin. J Biol Chem *275*, 28291-28300.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.

Safe, S., and Kim, K. (2004). Nuclear receptor-mediated transactivation through interaction with Sp proteins. Prog Nucleic Acid Res Mol Biol *77*, 1-36.

Seth, P., Krop, I., Porter, D., and Polyak, K. (2002). Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene *21*, 836-843.

Smith, C. L., and O'Malley, B. W. (2004). Coregulator function: a key to understanding tissue specificity of selective receptor modulators. Endocr Rev *25*, 45-71.

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.

Torchia, J., Rose, D. W., Inostroza, J., Kamei, Y., Westin, S., Glass, C. K., and Rosenfeld, M. G. (1997). The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. Nature *387*, 677-684.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

Vendrell, J. A., Magnino, F., Danis, E., Duchesne, M. J., Pinloche, S., Pons, M., Birnbaum, D., Nguyen, C., Theillet, C., and Cohen, P. A. (2004). Estrogen regulation in human breast cancer cells of new downstream gene targets involved in estrogen metabolism, cell proliferation and cell transformation. J Mol Endocrinol *32*, 397-414.

Wang, D. Y., Fulthorpe, R., Liss, S. N., and Edwards, E. A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. Mol Endocrinol *18*, 402-411.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H., and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev *16*, 235-244.

Weisz, A., Basile, W., Scafoglio, C., Altucci, L., Bresciani, F., Facchiano, A., Sismondi, P., Cicatiello, L., and De Bortoli, M. (2004). Molecular identification of ERa-positive breast cancer cells by the expression profile of an intrinsic set of estrogen regulated genes. J Cell Physiol *200*, 440-450.

Table 3.1. Functional classification of target genes bound by ERα in MCF-7 cells in the presence of estradiol

GENE	Description
	Apoptosis
CASP7	caspase 7
IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma
	Carbohydrate metabolism
GLT25D2	glycosyltransferase 25 domain containing 2
HK1	hexokinase 1
MDH1	malate dehydrogenase 1, NAD
	Cell adhesion
ANXA6	annexin A6
ANXA9	annexin A9
COL5A3	collagen, type V, alpha 3
NINJ2	ninjurin 2
	Cell-cell signaling
CTNNBIP1	catenin, beta interacting protein 1
SEMA3B	sema domain, Ig domain, short basic domain, secreted, (semaphorin)
WISP2	WNT1 inducible signaling pathway protein 2
WNT16	wingless-type MMTV integration site family, member 16
	Cell growth/maintenance
CHPT1	choline phosphotransferase 1
EPS8	epidermal growth factor receptor pathway substrate 8
PRCC	papillary renal cell carcinoma
SEL1L	sel-1 suppressor of lin-12-like (C. elegans)
TBC1D3	TBC1 domain family, member 3
	Cell motility
CRKL	v-crk sarcoma virus CT10 oncogene homolog
	Cell cycle
ARKRD15	ankyrin repeat domain 15
BANP	BTG3 associated nuclear protein
CDK5	cyclin-dependent kinase 5
RBL2	retinoblastoma-like 2 (p130)
TUSC4	tumor suppressor candidate 4
	Chromosome biogenesis
SMYD3	SET and MYND domain containing 3
	Co-enzyme metabolism
COQ4	coenzyme Q4 homolog (yeast)
MOCS2	molybdenum cofactor synthesis 2
	Cytoskeleton
FGD3	FYVE, RhoGEF and PH domain containing 3
KRT13	Keratin 13
SPTBN4	spectrin, beta, non-erythrocytic 4
TTID	titin immunoglobulin domain protein (myotilin)
TNS	Tensin
	Defense response
LY6E	lymphocyte antigen 6 complex, locus E
PGLYRP2	peptidoglycan recognition protein 2
TFF1	trefoil factor 1
TFF3	trefoil factor 3
	DNA repair
RECQL4	RecQ protein-like 4
	Immune response
IL20	interleukin 20
	Lipid metabolism
ALDH3B2	aldehyde dehydrogenase 3 family, member B2
PAFAH2	platelet-activating factor acetylhydrolase 2, 40kDa
PCYT1A	phosphate cytidylyltransferase 1, choline, alpha isoform
	Protein metabolism and modification
AHSA1	HA1, activator of heat shock 90kDa protein ATPase homolog
B3Gn-T6	beta-1,3-N-acetylglucosaminyl transferase protein
CST5	cystatin D
FBXO33	F-box protein 33
H11	protein kinase H11
HSPH1	heat shock 105kDa/110kDa protein 1
PKIB	protein kinase (cAMP-dependent, catalytic) inhibitor beta
RPS6KL1	ribosomal protein S6 kinase-like 1
TIPARP	TCDD-inducible poly(ADP-ribose) polymerase
TMPRSS3	transmembrane protease, serine 3
	RNA processing
DDX23	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23
PRPF31	pre-mRNA processing factor 31 homolog (yeast)
QTRTD1	queuine tRNA-ribosyltransferase domain containing 1
THOC3	THO complex 3
	Signal transduction
P2RY6	pyrimidinergic receptor P2Y, G-protein coupled, 6
	Steroid and drug metabolism
BAAT	bile acid CoA:amino acid N-acyltransferase (glycine N-choloyltransferase)
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3
CYP4F11	cytochrome P450, family 4, subfamily F, polypeptide 11
STS	steroid sulfatase, arylsulfatase C, isozyme S
UGT2B15	UDP glycosyltransferase 2 family, polypeptide B15
UGT2B17	UDP glycosyltransferase 2 family, polypeptide B17
	Transcriptional regulator
CARP	cardiac ankyrin repeat protein
ESR1	estrogen receptor 1
FLJ20097	hypothetical protein FLJ20097
FOXA1	forkhead box A1
NR0B2	nuclear receptor subfamily 0, group B, member 2
PHF15	PHD finger protein 15
PPRC1	PPAR, gamma, coactivator-related 1
PROP1	prophet of Pit1, paired-like Hox transcription factor
TRIM16	tripartite motif-containing 16
ZNF140	zinc finger protein 140
ZNF302	zinc finger protein 302
ZNF485	zinc finger protein 485
	Transport
ABCA3	ATP-binding cassette, sub-family A (ABC1), member 3
ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5
ABCC11	ATP-binding cassette, sub-family C (CFTR/MRP), member 11
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
DSCR3	Down syndrome critical region gene 3
NDUFA2	NADH dehydrogenase 1 alpha subcomplex, 2
NDUFB9	NADH dehydrogenase 1 beta subcomplex, 9
P2RX7	purinergic receptor P2X, ligand-gated ion channel, 7
PDZK1	PDZ domain containing 1
PKD2l2	polycystic kidney disease 2-like 2
RAB7L1	RAB7, member RAS oncogene family-like 1
SLC7A3	solute carrier family 7, member 3
SLC9A8	solute carrier family 9 (sodium/hydrogen exchanger), isoform 8
SLC25A36	solute carrier family 25, member 36
SLC27A2	solute carrier family 27, member 2
SYT12	synaptotagmin XII
UCRC	ubiquinol-cytochrome c reductase complex (7.2 kD)
ZFYVE1	zinc finger, FYVE domain containing

Genes without an assigned function at this level of analysis: C9orf11, C14orf61, C14orf133, C20orf172, CBWD2, CHD1L, CYB561D2, DKFZp434B1272, DKFZp547E1912, DKFZP564I122, DKFZP566J2046, DNCI2, DOC1, Eny2, FAIID1, FAM3C, FEM1A, FLJ10871, FLJ11267, FLJ13710, FLJ20094, FLJ20772, FLJ31882, FLJ33761, FLJ33868, GREB1, HAGH, HSPC138, IGSF3, INVS, KIAA1536, KSP37, LOC90668, LOC114926, MDH1, MDS025, MGC8902, MGC10200, MGC11242, MGC26694, MGC35361, MGC47799, MR-1, MSMB, NALP6, NAV3, NUDCD1, PRUNE, RGN, S100A10, SCGB1D2, SMAP, SMILE, TFPT, TRIM51, TSNAXIP1, TSSC4, VEPH1, YIF1B, ZMAT5. In the case that one locus could be assigned to two distinct genes, both genes were included in the analysis.

Figure 3.1. Genome-wide location analysis of direct ERα transcriptional targets in MCF-7 breast cancer cells

Genome-wide location analysis of direct ERα transcriptional targets in MCF-7 breast cancer cells. (*A*) Motif-finding algorithms identifies the consensus ERE (GGTCANNNTGACCT) as the most common transcription factor binding motif present in the promoters bound by ERα in the promoter array. The motif was present in 60% of the promoters used for the analysis. (*B*) E_2-modulated recruitment of ERα, the co-activator AIB1 and RNA polymerase II at selected ERα targets in MCF-7 cells as assayed by conventional ChIP.

Figure 3.1

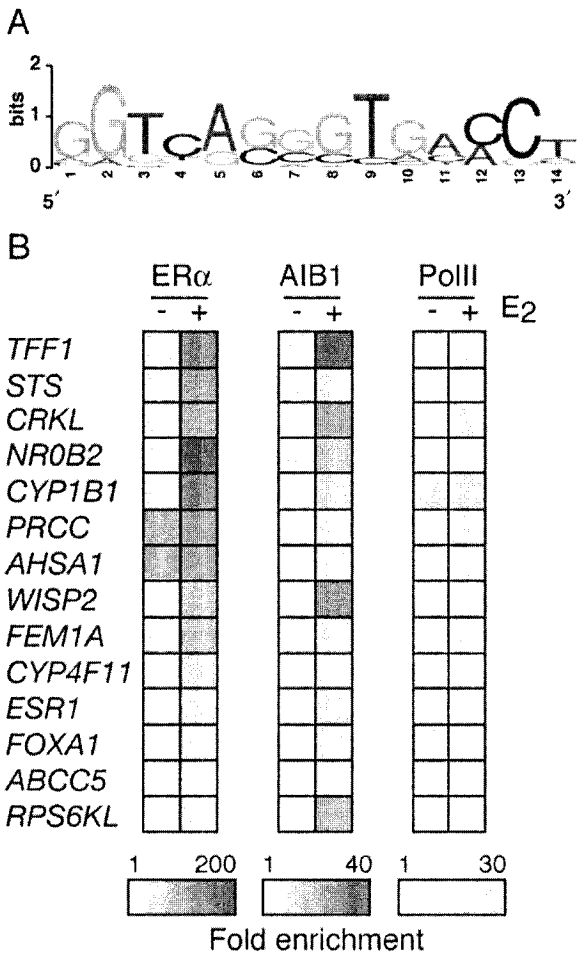

A
bits
2
1
0
5′
3′
B
ERα
AIB1
PolII
- +
- +
- +
E2
TFF1
STS
CRKL
NR0B2
CYP1B1
PRCC
AHSA1
WISP2
FEM1A
CYP4F11
ESR1
FOXA1
ABCC5
RPS6KL
1 200
1 40
1 30
Fold enrichment

Figure 3.2. FOXA1, a target of ERα, is recruited to a subset of ERα targets.

FOXA1, a target of ERα, is recruited to a subset of ERα targets. (*A*) Pie chart representing major biological functions and processes associated with ERα targets (153) enriched in E_2-treated MCF-7 cells. (*B*) Induction of FOXA1 expression by E_2 as monitored by Western blot. (*C*) FOXA1 recruitment to a subset of ERα-bound promoters containing FOXA1 binding sites as assayed by conventional ChIP. The *STS* and *HK1* promoters serve as a ERα-bound control promoter without a FOXA1 binding site. The results presented are from a single experiment representative of three independent experiments.

Figure 3.2

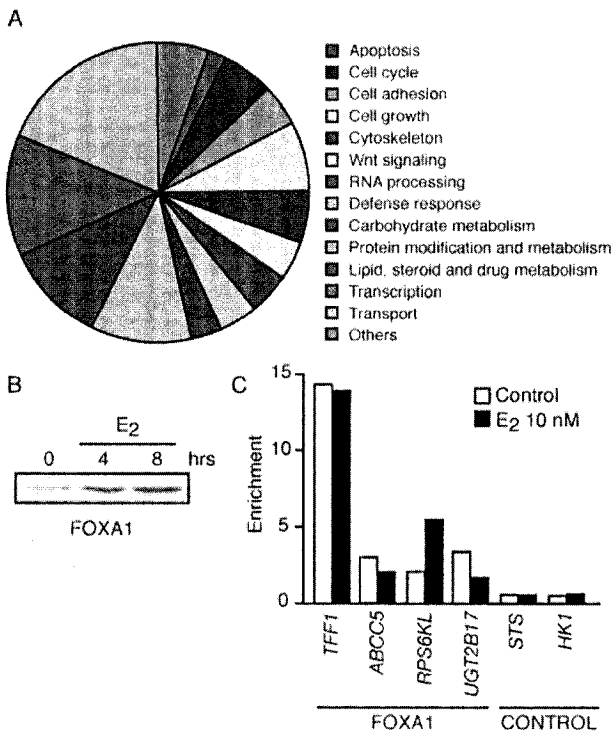

A
Apoptosis
Cell cycle
Cell adhesion
Cell growth
Cytoskeleton
Wnt signaling
RNA processing
Defense response
Carbohydrate metabolism
Protein modification and metabolism
Lipid, steroid and drug metabolism
Transcription
Transport
Others
B
E2
0 4 8 hrs
FOXA1
C
Control
E2 10 nM
Enrichment
15
10
5
0
TFF1
ABCC5
RPS6KL
UGT2B17
STS
HK1
FOXA1
CONTROL

Figure 3.3. FOXA1 is required for ERα activity on a subset of target promoters.

FOXA1 is required for ERα activity on a subset of target promoters. (*A*) FOXA1 expression in MCF-7 cells transfected with control (siC) and FOXA1 (siF) siRNAs. Actin levels serve as a control for specificity and gel loading. (*B*) FOXA1 is required for the E_2 regulation of *TFF1* expression in MCF-7 cells. RT-PCR analysis of *TFF1* expression was performed with extracts obtained from cells transfected with control (siC) and FOXA1 (siF) siRNAs in the presence or absence of E_2. The *STS* promoter serves as a ERα-bound control promoter without a FOXA1 binding site. (*C*) Knock-down of FOXA1 expression decreases the ability of ERα to stimulate transcription from the *TFF1* promoter. MCF-7 cells were co-transfected with ERα, the TFF1-Luc reporter and control (siC) or FOXA1 (siF) siRNAs in the presence or absence of E_2. (*D*) FOXA1 is required for E_2-induced recruitment of ERα to the *TFF1*, *RPS6KL1*, *ABBC5* and *UGT2B17* promoters as assayed by conventional ChIP. The *STS* promoter acts as a control as described in (*B*). The cells were treated with vehicle (C) or 100 nM E_2. Results are expressed as % of maximal ERα binding observed in the presence of E_2. For panel *A*, *C* and *C*, the results presented are from a single experiment representative of at least two independent experiments.

Figure 3.3

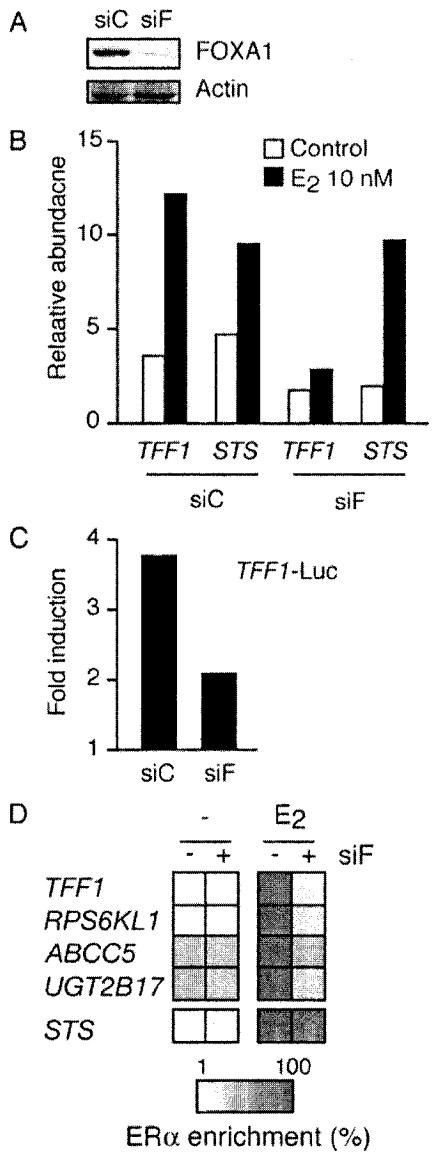

A
siC siF
FOXA1
Actin
B
Relaative abundacne
15
10
5
0
Control
E2 10 nM
TFF1 STS TFF1 STS
siC
siF
C
Fold induction
4
3
2
1
TFF1-Luc
siC siF
D
-
E2
- + - + siF
TFF1
RPS6KL1
ABCC5
UGT2B17
STS
1
100
ERα enrichment (%)

Figure 3.4. Effect of FOXA1 knock-down on cell cycle entry in response to E_2.

Effect of FOXA1 knock-down on cell cycle entry in response to E_2. Results shown represent the % change in cells in S, G2 and M phases stimulated by E_2 relative to untreated cells in the presence of control (siC) or FOXA1 (siF) siRNAs. The results presented are from a single experiment representative of two independent experiment.

Figure 3.4

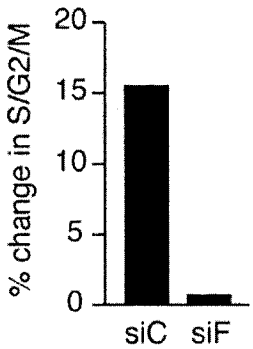


Figure 3.5. Model illustrating how FOXA1 licensing defines sub-domains of E_2 action in breast cancer cells.

Model illustrating how FOXA1 licensing defines sub-domains of E_2 action in breast cancer cells. Green arrows represent direct transcriptional activity of ERα and FOXA1, and the dashed blue arrow indicates the action of FOXA1 as a modulator of ERα binding to a subset of promoters. The presence of FOXA1 thus grants permission to ERα to regulate a subset of the hormonal response which can be further amplified by positive regulation of *FOXA1* expression by E_2-bound ERα.

Figure 3.5

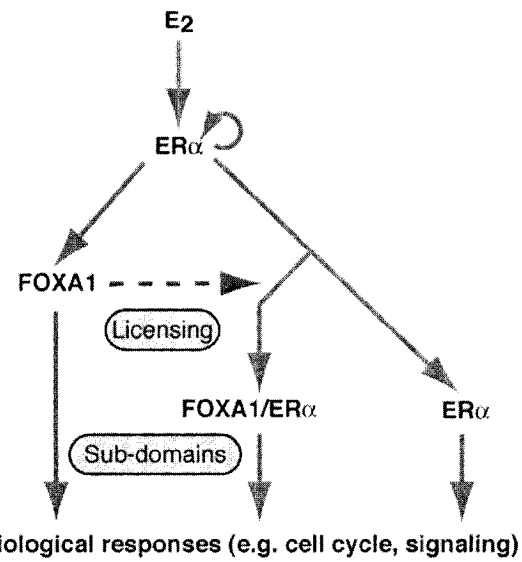
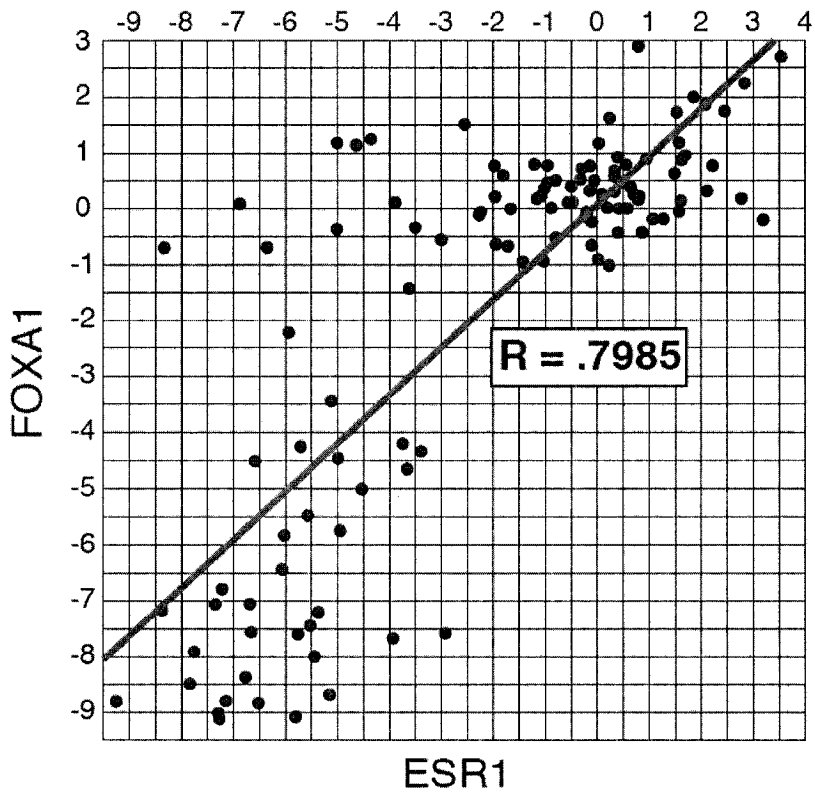

E2
ERα
FOXA1
Licensing
FOXA1/ERα
ERα
Sub-domains
Biological responses (e.g. cell cycle, signaling)

Figure 3.6. suppl. Scatter plot of the correlation of the expression of ESR1 and FOXA1 in breast tumors.

The *x* and *y* axes represent fold change in expression. The plot was generated using the Cancer Microarray Database Oncomine (www.oncomine.org) from RNA expression data originally published in (van 't Veer et al., 2002).

Figure 3.6 Suppl.



CHAPTER IV: A Polymorphic Autoregulatory Hormone Response Element in the Human ERRα Promoter Dictates PGC-1α Control of ERRα Expression

PREFACE

Variations in the genome sequence among individuals play a role in determining predisposition to diseases and response to therapies. This chapter describes a regulatory polymorphism found within the promoter of *ESRRA*, the gene encoding ERRα. We show that a variable copy number of a 23 bp sequence influences ERRα transcriptional response to PGC1, which is indeed dependent on the presence of ERRα on its own promoter. This section thus describes a polymorphic autoregulatory mechanism for ERRα transcriptional response and provides a direct genomic target of ERRα which will be useful for subsequent studies.

This chapter is a manuscript that has been published in the *Journal of Biological Chemistry.*

ABSTRACT

The orphan nuclear estrogen-related receptor α (ERRα) and transcriptional cofactor PGC-1α are both involved in the regulation of energy metabolism. Recently, extensive cross-talk between PGC-1α and ERRα has been demonstrated: the presence of PGC-1α is associated with an elevated expression of ERRα, while the two proteins can influence each other's transcriptional activities. Using a candidate gene approach to detect regulatory variants within genes encoding nuclear receptors, we have identified a 23 base-pair sequence (ESRRA23) containing two nuclear receptor recognition half-site motifs that is present in 1 to 4 copies within the promoter of the human *ESRRA* gene encoding ERRα. The ESRRA23 sequence contains a functional ERR response element that is specifically bound by ERRα, and chromatin immunoprecipitation shows that endogenous ERRα occupies its own promoter *in vivo.* However, ERRα transcriptional activity on the *ESRRA* promoter or on the ESRRA23 element linked to the thymidine kinase (TK) promoter is low when assayed by transient transfections in HeLa cells. Strikingly, introduction of PGC-1α induces the activity of the *ESRRA* promoter in a manner that is dependent on the presence of the ESRRA23 element and of its dosage. Co-expression of ERRα and PGC-1α results in a synergistic activation of the *ESRRA* promoter, and this effect is dependent on the presence of an intact ERRα DNA binding domain. In experiments using ERRα null fibroblasts, the ability of PGC-1α to stimulate the *ESRRA* promoter is considerably reduced but can be restored by addition of ERRα. Taken together, these results demonstrate that an interdependent ERRα/PGC-1α-based transcriptional pathway targets the ESRRA23 element to dictate the level of ERRα expression. This study further suggests that this regulatory polymorphism may provide differential responses to ERRα/PGC-1α-mediated metabolic cues in the human population.

INTRODUCTION

Nuclear hormone receptors are transcription factors that control essential developmental and physiological pathways (Mangelsdorf et al., 1995). Although the transcriptional activity of nuclear receptors is primarily regulated by specific ligands, several members of the superfamily of nuclear receptors have no known natural ligands and are therefore referred to as orphan receptors (Giguère, 1999). Estrogen-related receptor α (ERRα; NR3B1) was the first orphan nuclear receptor to be identified on the basis of its similarity with estrogen receptor α (ERα; NR3A1) (Giguere et al., 1988). Phylogenic tree reconstruction confirmed that ERRα belongs to the subgroup of receptors for steroid hormones (Laudet, 1997), and ERRα was subsequently shown to share both structural and functional attributes with the ERs, including binding to synthetic estrogenic ligands (Giguere, 2002). ERRα also recognizes estrogen response elements (EREs) but characterization of its DNA binding properties demonstrated a preference for sites composed of a single half-site preceded by three nucleotides with the consensus sequence TNAAGGTCA, referred to as an ERRE (Johnston et al., 1997; Sladek et al., 1997; Vega and Kelly, 1997; Wiley et al., 1993; Yang et al., 1996). The transcriptional activity of ERRα is independent of exogenously added ligand and its relative potency as a transcriptional activator appears to be cell context- and promoter-dependent (Bonnelye et al., 1997; Chen et al., 2001; Giguere et al., 1988; Lu et al., 2001; Sladek et al., 1997; Vanacker et al., 1999; Zhang and Teng, 2000). ERRα has also been described as a potent repressor of the SV40 late promoter (Wiley et al., 1993) and to interfere with the functions of glucocorticoid, retinoic acid and peroxisome proliferator-activated receptors (Maehara et al., 2003; Sladek et al., 1997; Trapp and Holsboer, 1996). While the exact physiological role of ERRα has not been precisely defined, increasing evidence suggest that ERRα plays an important role in regulating cellular energy balance. ERRα is predominantly expressed in tissues demonstrating a high capacity for fatty acid β-oxidation (Giguere et al., 1988; Ichida et al., 2002; Sladek et al., 1997), and has been shown to regulate the medium-chain acyl coenzyme A dehydrogenase gene

(MCAD) (Sladek et al., 1997; Vega and Kelly, 1997). More recently, ERRα null mice have been shown to display reduced body weight and peripheral fat deposition, and be resistant to high-fat diet-induced obesity (Luo et al., 2003). In agreement with the observed phenotype, gene microarray experiments demonstrated alteration in the expression of genes involved in adipogenesis, mitochondrial biogenesis and energy metabolism, including cytochrome c, MCAD, acetyl coenzyme A synthetase 2 and fatty acid synthase (Luo et al., 2003).

The transcriptional activity of nuclear receptors is dependent on specific interactions with coregulatory proteins (McKenna et al., 1999). The recent identification and functional characterization of PPARγ coactivator-1α (PGC-1α), PGC-1β and PGC-1 related protein revealed the existence of a family of coactivators that possess the unique characteristic of relaying diverse physiological signals to transcription factors that regulate gene networks controlling energy balance (reviewed in Knutti and Kralli, 2001; Puigserver and Spiegelman, 2003). In particular, PGC-1α has been shown to regulate thermogenesis in brown fat (Puigserver et al., 1998), mitochondrial biogenesis and respiration in skeletal muscle (Lin et al., 2002) and glucogenesis in the liver (Herzig et al., 2001; Puigserver et al., 2003; Yoon et al., 2001). PGC-1β may also contribute to the control of energy metabolism as over-expression of this gene in transgenic mice induces a high-energy expenditure and antagonizes obesity (Kamei et al., 2003). While PGC-1α was originally identified as a transcriptional coactivator specific for PPARγ (Puigserver et al., 1998), subsequent studies have demonstrated that PGC-1α influences the activity of numerous transcription factors, including a wide array of nuclear receptors (Knutti and Kralli, 2001). Recently, two groups, using similar yeast two-hybrid approaches, independently identified ERRα as a novel PGC-1α-binding protein (Huss et al., 2002; Ichida et al., 2002). Huss et al. (Huss et al., 2002) demonstrated that PGC-1α enhances ERRα transcriptional activity on the MCAD promoter. In contrast, Ichida et al. (Ichida et al., 2002) described ERRα as a repressor of PGC-1α activity on the PEPCK promoter. ERRα

transcriptional activity has also been shown to be stimulated by PGC-1α and PGC-1β in transient transfection assays using synthetic promoters (Schreiber et al., 2003). Interestingly, ERRα and PGC-1α show similar expression profiles in adult tissues, including induction of expression of both genes by exposure to cold (Ichida et al., 2002; Schreiber et al., 2003). Consistent with this observation, PGC-1α has been shown to induce the expression of ERRα (Schreiber et al., 2003). However, the molecular mechanisms underlying this phenomenon remain to be elucidated.

In this study, we first used a candidate gene approach to detect variants within genes encoding nuclear hormone receptors likely to play a role in physiology and be associated with disease. We first performed a family-wide screen of the genes encoding nuclear receptors in French Canadian women for novel frequent variants. This screen included exons, splice consensus sites and approximately 1 kb of sequence located upstream of the first exon presumably containing promoter regulatory sequences. This led us to identify a 23 bp sequence referred to as ESRRA23, located at position -682 in the *ESRRA* gene promoter that can be found in 1 to 4 copies in human chromosomes. Remarkably, this sequence includes a functional ERRE that is also responsive to the presence of PGC-1α. In this report, we describe the properties of this novel polymorphic regulatory element and functional consequences of its dosage in the *ESRRA* promoter responses to ERRα and PGC-1α. Our results demonstrate the existence of an autoregulatory mechanism by which ERRα can control its own expression, and further suggest the existence of an interdependent PGC-1α/ERRα pathway involved in the control of energy balance.

EXPERIMENTAL PROCEDURES

Identification of a polymorphic repeat in the ESRRA promoter. The primers 5' CCTTGGTGTGGCCTCGACTG 3' and 5' GCACTCGCGAGCCAAGAGA 3' were used to produce a 1054 bp fragment upstream of *ESRRA* exon 1. PCR was performed according to standard protocols with TAQ polymerase from QIAGEN. PCR products were purified with a QIAquick PCR purification kit (QIAGEN) and quantified by gel electrophoresis with standardized amounts of DNA. Automated sequencing of PCR products was performed with fluorescently labeled dideoxy terminators using the Big Dye terminator cycle sequencing kit on a ABI 377 DNA-Sequencer (Applied Biosystems).

Genotyping. Genomic DNA was purified from 200 μl of peripheral blood leukocytes with QIAGEN's whole blood DNA purification kits for 96-well plates. The collected DNA was further diluted 5 times with a solution 10 mM TRIS-HCl pH 7.5 containing 56.3 μM of an inert fluorescent dye, ROX (Molecular Probes, C-1309). The final concentration of ROX was 45 μM in each DNA sample and the mean DNA concentration was 5 ng/μl. For each PCR, 5 μl was used regardless of the DNA concentration. Genotyping for the ESRRA23 minisatellite in the *ESRRA* promoter was performed using the following primers; 5' CGTGGCCCCGCCCTTCC 3' and 5' GTAGACCCAGTAGCCCCACAG 3'. PCR was performed in a 96-well microplate (Axygen) with 5 μl (25 ng) of genomic DNA and 20 μl PCR premix containing 2.5 μl 10X buffer (QIAGEN), 200 μM of each dNTP, 7.5 pmoles of each primer, 1X Q-solution (QIAGEN), 2% DMSO and 1 unit HotStart Taq DNA polymerase (QIAGEN). PCR setup in the microplates was performed with a Qiagen Biorobot 3000 (QIAGEN) or manually with a multichannel (8- or 12-channel) pipette. The plates were then covered with a silicone mat (Axymat from Axygen) properly sealed using a roller (MJ Research). PCR was performed on MJ PTC-200 (MJ Research), 95^{0}C for 15 min, 30 cycles of 45 sec at 95^{0}C, 45 sec at 58^{0}C, 45 sec at 72^{0}C and a final extension at 72^{0}C for 7

minutes. PCR products were run on a 2% agarose gel in TBE 1X buffer for 2 hrs at 10 volts/cm. The reaction produces amplicons of 198 bp with 2 repeats, of 221 bp with 3 repeats, of 244 bp with 4 repeats and 175 bp with only one repeat. The gels were photographed and the genotype assigned by two independent readers. 180 samples were run in duplicate with a concordance rate of 99%.

Cloning of the ESRRA promoter. Genomic DNA from blood samples or lymphoblastoid cell lines containing two or three repeats was prepared with the QIAamp kit (QIAGEN) and used as template in a PCR reaction to amplify the promoter region of the human *ESRRA* gene. The primers used had the following sequences, 5' GCGGTACCTGAGTGCCCTGCGCTAC 3' (forward) and 5' CCCAAGATTCCTACTCCGCTTCCTC 3' (reverse) and produced a product of 1.2 kb. This fragment was digested with *Kpn*I and *Hind*III and subcloned into the luciferase reporter plasmid pGL3 (Promega, Madison, WI). All the selected clones were sequenced with fluorescently labeled dideoxy terminators using the Big Dye terminator cycle sequencing kit on a ABI 377 DNA-Sequencer (Applied Biosystems).

Plasmids and cell transfections. The ERRα cDNA was cloned into the expression vector pCMX. Plasmids expressing the ERRα-VP16 fusion protein was constructed by subcloning PCR amplified ERRα cDNA into pCMX-VP16 downstream of the VP16 activation domain. The DNA binding null mutant ERRα^{DBDm} was generated by substituting the glutamic acid and alanine residues of the ERRα P box for glycine residues. ERRα^{DBDm} does not bind DNA as examined by EMSA *in vitro* but locates to the nucleus when transfected in mammalian cells (Y. Kiriyama and V. Giguère, unpublished results). Expression vector for human ERα has been described (Tremblay et al., 1998). The pCDNA3.1 HA-hPGC-1α vector was described previously (Kressler et al., 2002) and obtained from A. Kralli (La Jolla, CA). The luciferase reporter plasmid ESRRA23-TKLuc and ESRRA23(3)-TKLuc contained one and three copies respectively, of the ESRRA23 response element (Fig. 4.1A) cloned into pTKLuc. A fragment containing *ESRRA* promoter sequence 1.2 kb upstream of

the transcriptional start site was subcloned into the luciferase reporter plasmid pGL3 (Promega, Madison, WI) to give pGL3ESRRA. To construct the ΔESRRA promoter luciferase reporter gene, sequences 5' and 3' adjacent to the ESRRA23 elements and putative ERRα binding site were PCR amplified and subcloned sequentially into pGL3. HeLa cells were obtained form American Type Culture Collection and maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum. Mouse embryonic fibroblasts (MEFs) were isolated from 13.5 day-old wild-type and ERRα null embryos (Luo et al., 2003). The embryos were minced with a razor blade and the cells were dissociated by trypsin. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) and supplemented with 10% heat-inactivated fetal bovine serum, 100 units/ml penicillin and 100 μg/ml streptomycin. Twenty-four hrs prior to transfection, cells were seeded in 12-well plates and grown in phenol red-free DMEM containing charcoal-treated fetal bovine serum. One hundred ng of expression vector for nuclear receptors, 0.4 μg for the PGC-1α expression vector, 0.4 μg luciferase reporter and 0.3 μg CMXβgal plasmids were introduced into cells using Lipofectamine (Invitrogene Life Technologies) or FuGENE 6 Transfection Reagent (Roche Applied Science). When using Lipofectamine, cells were maintained in the presence of liposomes for 16 hrs and cultured for an additional 24 hrs. Cells were harvested in potassium phosphate buffer containing 1% Triton X-100. Luciferase activity was determined using Steady-Glo (Promega, Madison, WI) and values were read with the Victor2 in the luminescence mode. The transfection was normalized to the β-galactosidase activity of each sample. All results represent experiments conducted in duplicate at least three times.

Electromobility shift assay. ERRα and ERα proteins were synthesized by *in vitro* transcription-translation using rabbit reticulocyte lysates (Promega, Madison, WI). DNA-binding reactions were conducted as previously described (Tremblay et al., 1997) using 5 μl of programmed lysates in each binding reaction. The entire reaction was loaded onto a 5% polyacrylamide gel and electrophoresed at 150 V at room temperature. The gel was dried and exposed

overnight at –85^{0}C. The oligonucleotides and their complements that were used as probes and competitors are shown in Fig. 4.2A.

Chromatin Immunoprecipitation Assay. ChIP assays were performed as previously described (Laganière et al., 2003). Briefly, MCF-7 cells were crosslinked for 10 min at room temperature. Cells were then washed twice with ice-cold PBS, collected and centrifuged. Pellets were incubated in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1, supplemented with protease inhibitor cocktail (Roche)) for 15 min, and sonicated 4 times for 10 sec using a VirSonic 100 (Virtis) sonicator. After centrifugation at 12,000 rpm to remove the debris, soluble chromatin was diluted 10-fold in ChIP dilution buffer (1% TritonX-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1) and immunoprecipitated using a specific anti-hERRα polyclonal antibody developed in our laboratory against the non-conserved amino terminal region of the protein. This antibody does not recognize ERRβ, ERRγ, ERα or ERβ proteins. Following immunoprecipitation, 40 µl of salmon sperm DNA/protein A agarose (Upstate) was added and incubated for 2 hrs. The precipitates were washed sequentially for 10 min each with buffer I (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris HCl pH 8.1, 150 mM NaCl), with buffer II (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HC1 pH 8.1, 150 mM HC1 pH 8.1, 500 mM NaCl) and buffer III (0.25 mM LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1). Precipitates were then washed three times with TE buffer and eluted with 1% SDS, 0.1 M $NaHCO_3$ buffer. The eluates were pooled and incubated at 65^{0}C for at least 6 hrs. The isolated DNA fragments were purified according to QIAquick Spin Kit (QIAGEN). Real time quantitative PCR was performed using LightCycler and Fast Start DNA Master SYBR Green1 (Roche) and primers 5' CCATCCGAGTGGAATTTGAGTCCTAAAG and 5'-GAACCGTAGACCCAGTAGCCC-CACAGAG for *ESRRA* promoter region containing the ESRRA23 elements. For the negative control region located 4 kb upstream of the initiation start site, the primers were: 5' GTGGCCCACAGGTGTCGCTCAAGTCTTC 3' and 5'

GGATGCAGTGTCCTTCTCCCCCAGATTG 3'. Enrichment of the *ESRRA* promoter region was normalized against the upstream control region. PCR products were visualized after migration on a 2% agarose gel.

RESULTS

Identification of a polymorphic hormone response element in the ESRRA promoter. Our search for functional coding and regulatory polymorphisms in genes encoding members of the nuclear receptor superfamily led us to identify a 23 bp element (Fig. 4.1A) located at position -682 in the *ESRRA* promoter that is present in 1 to 4 copies in human chromosomes. Sequence analysis of the 23 bp element, herein referred to as ESRRA23, revealed the presence of two nuclear receptor half-site recognition motifs (Fig. 4.1A). The upstream half-site is preceded by the three nucleotides TGA thus generating a consensus ERRα binding site, also referred to as an ERRE (Sladek et al., 1997). An additional putative ERRE, TCAAGGTCA, can also be found in the promoter region 1 bp downstream of the ESRRA23 sequence (Fig. 4.1B). The ESRRA23 element and few base pairs flanking it, including the putative ERRE downstream of the repeated element, are absolutely conserved between human and mouse genomic sequences (Fig. 4.1C). However, the ESRRA23 element is present in a single copy in the mouse genome. The observed *ESRRA* allelic frequencies amongst 5,490 human chromosomes (2,745 individuals) were "1" = 0.06, "2" = 93.15, "3"= 6.3 and "4" = 0.36 (Table 1).

Functional characterization of the polymorphic ESRRA23 element. Given the observation that the polymorphic sequence contained a putative ERRE, we first tested whether the ESRRA23 motif could serve as an autoregulatory element for ERRα. Electromobility shift assays using *in vitro* translated proteins and a set of oligonucleotide probes derived from the ESRRA23 element (Fig. 4.2A) showed that ERRα binds ESRRA23 with high specificity (Fig. 4.2B). Nucleotide changes within the upstream AGGTCA motif (ESRRA23m1) abolished ERRα binding while similar mutations in the downstream CGGTCA half-site (ESRRA23m2) had little effect on recognition of the element by ERRα (Fig. 4.2B). The ERRβ and γ isoforms also bound the ESRRA23 element with similar affinity (data not shown). It has recently been shown that treatment with the synthetic estrogen diethylstilbestrol can enhance the expression of ERRα (Liu et al., 2003). However, the related ERα did not

significantly bind the ESRRA23 element or the downstream half-site in this assay, but did recognize to a control ERE (Fig. 4.2B and data not shown). Previous studies have shown that the ERRs can bind to their response element either as monomers or homodimers (Johnston et al., 1997; Lu et al., 2001; Pettersson et al., 1996; Sladek et al., 1997). The presence of an intermediate retarded complex in a binding experiment using a mixture of wild-type and amino-terminal truncated receptors clearly demonstrates that ERRα preferentially binds ESRRA23 as a homodimer (see arrow in Fig. 4.2C). We next examined whether the ESRRA23 element could act as a functional ERRE *in vivo* when linked to a heterologous promoter. As expected, the generally transcriptionally silent ERRα failed to generate a significant response when assayed in HeLa cells (Fig. 4.2D). In order to test whether ERRα recognizes the ESRRA23 element *in vivo*, we used a mammalian one-hybrid system in which ERRα is linked to the potent transcriptional activation domain of the viral VP16 protein. Indeed, the constitutively active ERRα-VP16 chimera induced strong transcriptional responses (10- to 57-fold) in an element-dosage manner (Fig. 4.2E), demonstrating that ERRα can recognize the ESRRA23 element *in vivo*. We next studied the functional consequence of the ESRRA23 regulatory variant on the *ESRRA* promoter itself. Human *ESRRA* promoters containing either two or three copies of the ESRRA23 element representing the most common observed genotypes were cloned upstream of the luciferase reporter gene (Fig. 4.3A). We also engineered a mutant *ESRRA* promoter construct in which all copies of the ESRRA23 element as well as the non-polymorphic putative downstream ERRE were removed (ΔESRRA, Fig. 4.3A). ERRα induced a small but significant transcriptional response of 1.5- or 2-fold on the ESRRA promoter containing 2 or 3 copies of ESSRA23, respectively (Fig. 4.3B). The effect is specific as removal of the elements abolishes the ERRα-induced transcriptional response. As observed with the synthetic ESRRA23-TK promoters, the ERRα-VP16 chimera elicited strong responses from the reporter gene driven by the *ESSRA* promoters containing either 2 or 3 copies of ESSRA23, being 7- and 14-fold respectively (Fig. 4.3C). The mammalian one-

hybrid assay thus confirms that ERRα can directly interact with the *ESRRA* promoter. The specificity of the transcriptional effect was demonstrated in that the strength of the response was directly related to the copy number of ESRRA23, and that ablation of the ESRRA23 completely abolished ERRα-VP16-induced luciferase activity. Lastly, we used a ChIP assay to test whether endogenous ERRα interacts with the *ESRRA* promoter in the context of the native chromatin. The human breast cancer cell line MCF-7 was previously shown to express endogenous ERRα and therefore used for this assay (Lu et al., 2001). As shown in Fig. 4.4D, an antibody raised against human ERRα immunoprecipitates a DNA fragment that includes the ESRRA23 elements. Quantitative PCR showed a 25-fold enrichment of the promoter fragment over the control fragment located 4 kb upstream of the ESRRA23 element. Taken together, these experiments clearly show that ERRα recognizes its own promoter via the polymorphic ESRRA23 element.

The ESRRA23 element dictates PGC-1α control of ERRα expression. As introduced above, recent studies have shown that the coactivator PGC-1α can regulate both the expression and transcriptional activity of ERRα (Huss et al., 2002; Ichida et al., 2002; Schreiber et al., 2003). However, the molecular mechanism underlying the action of PGC-1α on ERRα expression has not yet been elucidated. As shown in Fig. 4.4A, introduction of PGC-1α alone by transient transfection in HeLa cells has a significant effect on *ESRRA* promoter activity, leading to a 4- and 6.5-fold induction in luciferase activity generated by *ESRRA* promoter reporter constructs containing 2 and 3 copies of the ESRRA23 element, respectively. The increased *ESRRA* promoter activity induced by PGC-1α is not only ESRRA23 dosage-dependent but mediated through ESRRA23 and possibly with a contribution by the flanking ERRE given that deletion of the region encoding these elements from the *ESRRA* promoter resulted in a complete loss of the stimulatory activity (Fig. 4.4A). The direct involvement of the ESRRA23 element in the PGC-1α response was further demonstrated by the observation that PGC-1α can activate TK-luciferase reporter genes containing 3 copies of the ESRRA23 element but not the parent

vector (Fig. 4.4B). The presence of a single copy of the ESRRA23 element was not sufficient to confer PGC-1α responsiveness to the TK promoter in HeLa cells. However, PGC-1α activity is much more potent in Cos-1 cells on both the *ESRRA* and ESRRA23-TK promoters suggesting that cell context may be important for the PGC-1α response (data not shown).

ERRα-dependent PGC-1α activity. PGC-1α has recently been described as a potent coactivator of ERRα and the related ERRγ isoform (Huss et al., 2002; Schreiber et al., 2003). We therefore investigated the interaction between ERRα and PGC-lα on the polymorphic *ESSRA* promoter in HeLa cells. As observed in Fig. 4.5, co-expression of ERRα and PGC-lα in HeLa cells results in a synergistic activation of the *ESRRA* promoter. This set of experiments also demonstrates that both the independent and combined transcriptional activities of ERRα and PGC-1α are not observed in the absence of either the region containing the ESRRA23 elements or a functional ERRα DNA binding domain. The response to ERRα and PGC-lα is also element dosage-dependent as the promoter containing 3 copies of the ESRRA23 element displays higher activity than the promoter containing 2 copies of the element in the presence of these regulatory factors (Fig. 4.5).

We next tested whether the presence of ERRα was absolutely essential to the activity of PGC-1α on the ESRRA23 element. Mouse embryonic fibroblasts (MEFs) were derived from both wild-type and ERRα null mice (Luo et al., 2003) and transfected with PGC-1α and ERRα, alone or in combination with the *ESRRA* promoter reporter construct containing 3 copies of the ESRRA23 element. ERRα is not transcriptionally active in MEFs, but the constitutively active ERRα-VP16 chimera displays identical activity in MEFs derived from both strains indicating that ERRα recognizes the *ESRRA* promoter in a similar manner in both cell types (Fig. 4.6). As previously observed in HeLa cells, introduction of PGC-1α in wild-type MEFs leads to a significant (3-fold) induction of *ESRRA* promoter activity. However, PGC-1α transcriptional activity is considerably reduced in ERRα null MEFs, but this activity can be completely restored by introduction of exogenous ERRα (Fig. 4.6). The

response to PGC-1α was not observed when the Δ*ESRRA* construct was used as a reporter in this assay (data not shown). Taken together, these results demonstrate a central role for ERRα in PGC-1α-induced activation of the *ESRRA* promoter via the ESRRA23 element.

DISCUSSION

In this report, we described the identification and functional characterization of a new polymorphic hormone response element (ESRRA23) present in the *ESRRA* gene promoter, the gene encoding the orphan nuclear receptor ERRα. Functional analysis of the polymorphic ESRRA23 sequence showed it to act as an autoregulatory element for ERRα whose activity is dependent on the presence of PGC-1α. Conversely, our results also demonstrate that PGC-1α activity on the *ESRRA* promoter is dependent on the presence of ERRα. Our study thus delineates the molecular mechanisms by which PGC-1α can up-regulate ERRα expression (Schreiber et al., 2003) and by which ERRα can control its own expression in a positive fashion. Our results also clearly establish a direct correlation between the number of ESRRA23 repeat elements present and the response of the *ESRRA* promoter to ERRα and PGC-lα proteins, alone or in combination. To our knowledge, this is the first example of natural regulatory polymorphism consisting of a sequence, present in one copy or in tandem repeated elements two, three or four times, containing a functional hormone response element for a nuclear receptor/coactivator complex.

The expression patterns of ERRα and PGC-lα and their response to specific physiological stimuli such as cold and starvation are nearly identical (Ichida et al., 2002; Schreiber et al., 2003). Furthermore, in agreement with these observations, PGC-lα has been shown to induce ERRα expression (Schreiber et al., 2003). Our functional characterization of the ESSRA23 element clearly demonstrates it to be the direct target of PGC-lα action. However, PGC-lα is a coactivator protein that does not bind ESRRA23 (J. Barry and V. Giguère, unpublished results) and thus requires interaction with a transcription factor that has the requisite docking site on the target promoter. One clear candidate is ERRα itself. ERRα binds to the ESRRA23 element and activates transcription from it in the presence of PGC-lα and other coactivators such as GRIP-1 (data not shown). The role of ERRα as a PGC-1α DNA binding partner is further corroborated by the transient transfection experiments

performed in HeLa cells and MEFs. A strong activation of the *ESRRA* promoter in HeLa cells is only observed in the presence of both factors, and transcriptional activation is not detected with an ERRα mutant unable to bind DNA. Using MEFs obtained from ERRα null mice, we have also shown that PGC-1α activity is considerably reduced in those cells, indicating an important and direct role for ERRα in PGC-1α action at the *ESRRA* promoter. We have also observed that PGC-1α retains some transcriptional activity in the ERRα null MEFs, suggesting that factors other than ERRα can transduce PGC-1α activity in these cells. Indeed, PGC-1α has been shown to interact with and stimulate the activity of a large number of transcription factors, including many nuclear receptors (Puigserver and Spiegelman, 2003). The likely candidates are nuclear receptors since the region encoding the ESRRA23 elements contains several nuclear receptor binding sites but no recognizable binding sites for other transcription factors. It has recently been shown that treatment with the estrogen agonist diethylstilbestrol stimulates the expression of ERRα (Liu et al., 2003). However, our studies do not support a role for ERα in the ERRα-independent PGC-1α activity observed in ERRα null MEFs since this activity can be detected with the use of steroid-deprived serum and that ERα does not significantly bind to the ESRRA23 element or adjacent half-site (Fig. 4.2B and data not shown). We are currently investigating the role of the ERRγ isoform, a known partner of PGC-1α that recognizes the ESRRA23 element, in the control of ERRα expression.

Increasing evidence point to a role for ERRα in regulating energy homeostasis. ERRα is expressed in tissues that demonstrate a high capacity for fatty acid β-oxidation such as kidneys, heart and brown fat (Sladek et al., 1997; Vanacker et al., 1998). Its expression has been shown to be up-regulated by physiological stimuli such as cold and starvation (Ichida et al., 2002; Schreiber et al., 2003), and ablation of the gene in mice results in reduced fat mass and resistance to high fat diet-induced obesity (Luo et al., 2003). Similarly, PGC-1α is a transcriptional coactivator for many transcription factors that control biological programs linked to energy needs (Puigserver and Spiegelman, 2003).

The results presented in this study reaffirm the existence of strong physiological and functional links between ERRα and PGC-1α action and also demonstrate the convergence of PGC-1α-based transcriptional pathways on a polymorphic hormone response element controlling ERRα expression.

In conclusion, this study clearly demonstrates that identification of regulatory variants in the human genome can reveal physiologically relevant interactions between distinct components of complex transcriptional pathways. It would be of interest to pursue further genetic studies to investigate whether the ESRRA23 polymorphism is linked to a particular phenotype or susceptibility to a metabolic disease in the human population, including obesity and diabetes.

ACKNOWLEDGEMENTS

Financial support was provided by the Canadian Institutes for Health Research (CIHR), the Canadian Genetic Diseases Network (CGDN) and Genome Quebec/Canada. V. Giguère holds a CIHR senior scientist career award, J. Laganière is a recipient of a U.S. Army Breast Cancer Research Program Predoctoral Traineeship Award, C. Dufour is supported by the Canderel Foundation and F. Rousseau holds a Fonds de la Recherche en Santé du Québec senior scientist award. We thank Anatasia Kralli for the PGC-1α expression vectors, Yoshimitsu Kiriyama and Janelle Barry for the ERRα DBD null mutant and Guy Cardinal, Latifa Elfassihi, Elaine Gravel, Denis Bergeron, Line Lapointe and Céline Champigny for expert technical work.

TABLES AND FIGURES

Table 4.1 Observed ESRRA allelic frequencies among 5490 human chromosomes

Genotype:	1,2	1,3	2,2	2,3	2,4	3,3	3,4
#	2	1	2381	334	16	8	3
%	0.074	0.037	86.74	12.17	0.58	0.29	0.11

Figure 4.1. Identification and structural organization of the polymorphic ESRRA23 element.

A, Sequence of the ESRRA23 element. The nuclear receptor half-site consensus sequences are indicated by arrows. B, The ESRRA23 element and neighboring sequence of the ESRRA promoter. The displayed allelic sequence of *ESRRA* contains two copies of the ESRRA23 element (boxed). The nuclear receptor half-site consensus sequences and the ERRα binding sites (ERRE) are indicated by arrows and brackets, respectively. The ERRE is independent of the ESRRA23 repeat element. C, Comparison between the human and mouse *ESRRA* promoter genomic sequences encompassing the ESRRA23 element. Vertical bars indicate identity between nucleotide residues.

Figure 4.1.

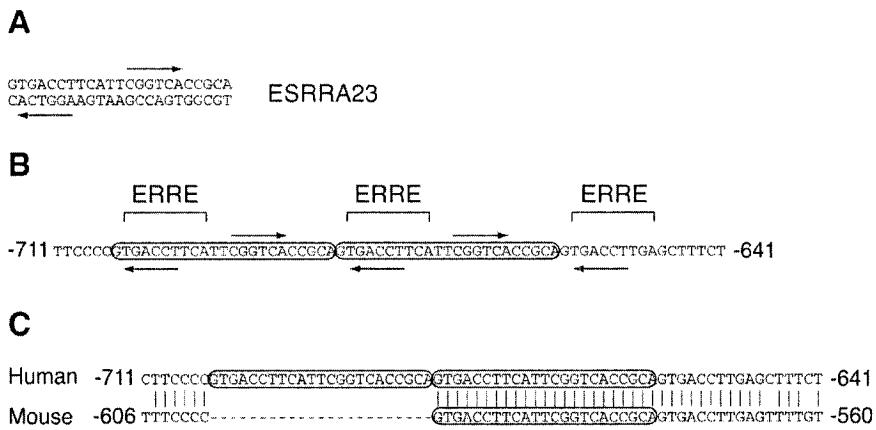

A
GTGACCTTCATTCGGTCACCGCA
CACTGGAAGTAAGCCAGTGGCGT
ESRRA23
B
ERRE
ERRE
ERRE
-711 TTCCCCGTGACCTTCATTCGGTCACCGCAGTGACCTTCATTCGGTCACCGCAGTGACCTTGAGCTTTCT -641
C
Human -711 CTTCCCCGTGACCTTCATTCGGTCACCGCAGTGACCTTCATTCGGTCACCGCAGTGACCTTGAGCTTTCT -641
Mouse -606 TTTCCCC----------GTGACCTTCATTCGGTCACCGCAGTGACCTTGAGTTTTGT -560

Figure 4.2. Functional characterization of the ESRRA23 element.

A, Sequences of the DNA probes used in the EMSA experiments and schematic representation of the ESRRA23-TK reporter constructs. Mutated nucleotides are shown in lower case letters. B, ERRα binds to the 5'-half-site of the ESRRA23 element. Mutation of the 5'-half-site (ESRRA23m1) abolishes binding while similar changes in the 3'-half site (ESRRA23m2) have no effect on binding. ERα, a closely related nuclear receptor, does not recognize the ESRRA23 element. An EMSA performed with an ERE probe is shown as a positive control. C, ERRα binds as a homodimer to the ESRRA23 element. EMSA was performed with wild-type and an N-terminal truncated ERRα, individually (lanes 2 and 3) or in combination (lane 4). The intermediate band (arrow) indicates the formation of a dimeric complex. D and E, The ESRRA23 element confers ERR responsiveness to a heterologous promoter. The ESRRA23 element was cloned in 1 (1X) or 3 (3X) copies upstream of the herpes simplex TK promoter and cotransfected in HeLa cells together with wild-type ERRα or the ERRα-VP16 chimera to assess responsiveness. Results are expressed as fold induction over control vector in the absence of receptor.

Figure 4.2.

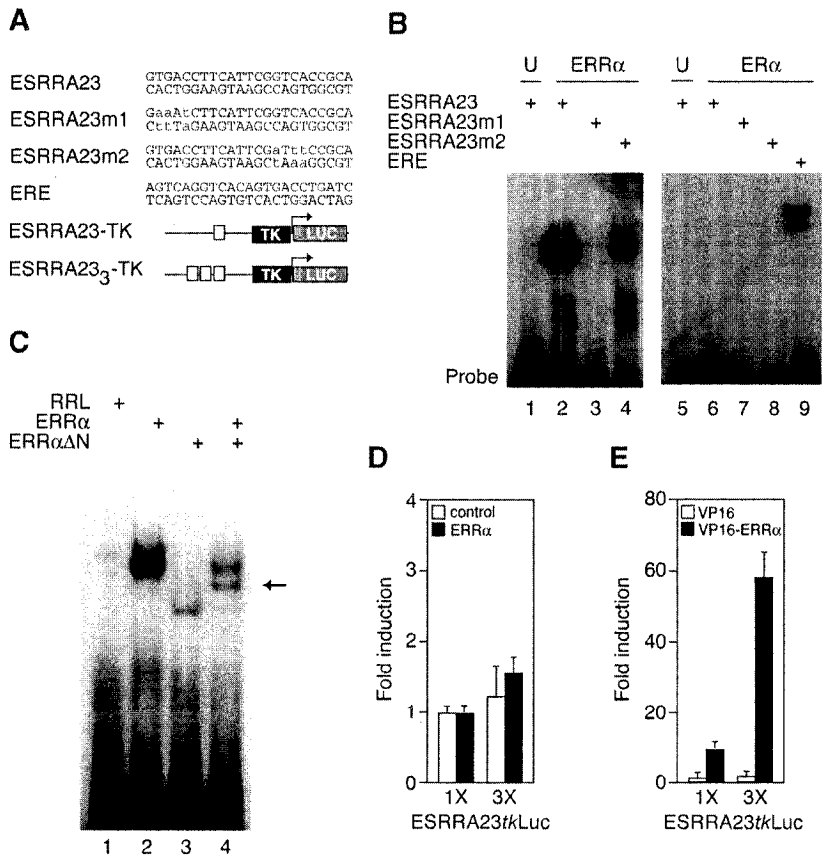

A
ESRRA23
GTGACCTTCATTCGGTCACCGCA
CACTGGAAGTAAGCCAGTGGCGT
ESRRA23m1
GaaAtCTTCATTCGGTCACCGCA
CttTaGAAGTAAGCCAGTGGCGT
ESRRA23m2
GTGACCTTCATTCGaTttCCGCA
CACTGGAAGTAAGCtAaaGGCGT
ERE
AGTCAGGTCACAGTGACCTGATC
TCAGTCCAGTGTCACTGGACTAG
ESRRA23-TK
TK
LUC
ESRRA23$_3$-TK
TK
LUC
B
U
ERRα
U
ERα
ESRRA23
ESRRA23m1
ESRRA23m2
ERE
Probe
1 2 3 4 5 6 7 8 9
C
RRL +
ERRα + +
ERRαΔN + +
1 2 3 4
D
control
ERRα
Fold induction
0 1 2 3 4
1X 3X
ESRRA23tkLuc
E
VP16
VP16-ERRα
Fold induction
0 20 40 60 80
1X 3X
ESRRA23tkLuc

Figure 4.3. The polymorphic ESRRA23 element is a functional ERRα response element in the context of the *ESRRA* promoter.

A, Schematic representation of the allelic *ESRRA* promoter luciferase reporter constructs containing 2 (2X) or 3 (3X) copies of the ESRRA23 element or a synthetic mutant in which ESRRA23 is absent (A). B and C, Effect of the presence of ERRα on *ESRRA* promoter activity. HeLa cells were co-transfected with the three ESRRA-based reporter constructs and wild-type ERRα (B) or the chimeric ERRα-VP16 construct (C). Results are expressed as fold induction over control vector in the absence of receptor. D, Binding of ERRα to the *ESRRA* promoter as determined by chromatin immunoprecipitation assay. PCRs containing primer pairs amplifying a region of the *ESRRA* promoter containing the polymorphic ESRRA23 element (ESRRA23) or the non-specific region (-4 kb control) were performed following immunoprecipitation of DNA/protein complexes with the ERRα antibody (α-ERRα).

Figure 4.3.

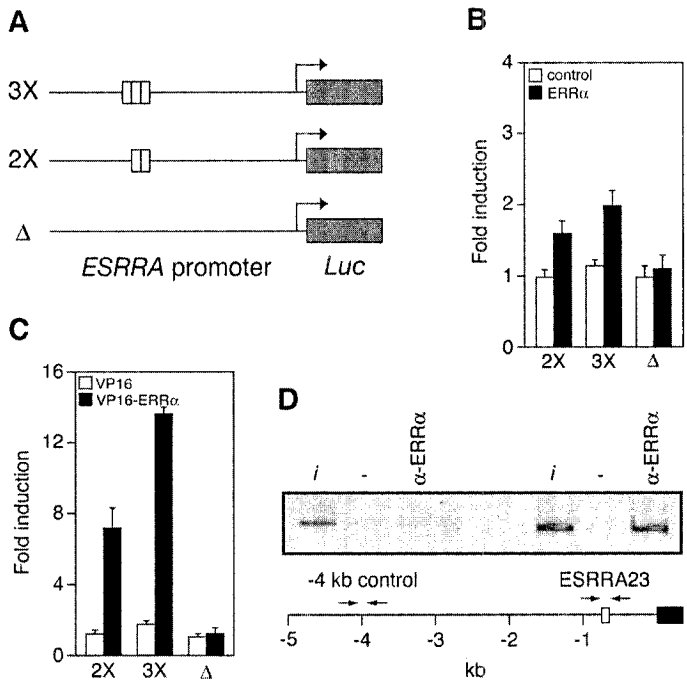

A
3X
2X
Δ
ESRRA promoter
Luc
B
control
ERRα
Fold induction
4
3
2
1
0
2X
3X
Δ
C
VP16
VP16-ERRα
Fold induction
16
12
8
4
0
2X
3X
Δ
D
i
-
α-ERRα
i
-
α-ERRα
-4 kb control
ESRRA23
-5
-4
-3
-2
-1
kb

Figure 4.4. PGC-1α induces *ESRRA* promoter activity through the polymorphic ESRRA23 element.

A, The allelic *ESRRA* promoter luciferase reporter constructs containing 2 (2X) or 3 (3X) copies of the ESRRA23 element or a synthetic mutant in which ESRRA23 is absent (A) were cotransfected in HeLa cells and cell extracts were assayed for luciferase activity. Results are expressed as fold induction over control vector in the absence of receptor. B, The ESRRA23 element was cloned in 1 (1X) or 3 (3X) copies upstream of the herpes simplex TK promoter and cotransfected in HeLa cells with a PGC-1α expression vector.

Figure 4.4.

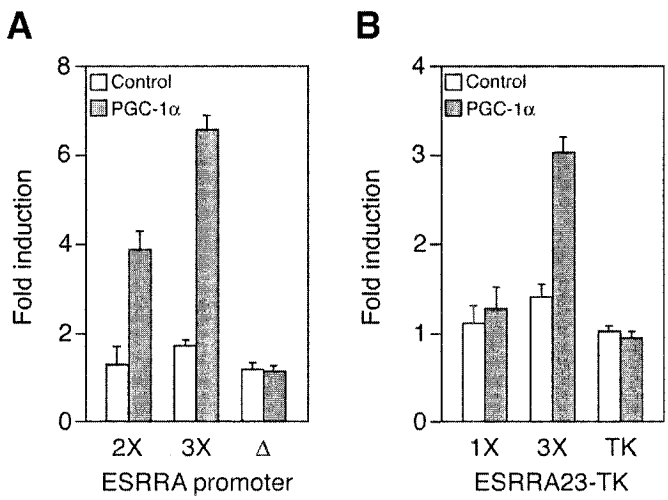

A
B
Control
PGC-1α
Fold induction
8
6
4
2
0
2X
3X
Δ
ESRRA promoter
4
3
2
1
0
1X
3X
TK
ESRRA23-TK

Figure 4.5. ERRα/PGC-lα interaction on the *ESRRA* promoter.

Synergistic activity of ERRα and PGC-1α on the polymorphic *ESRRA* promoters. HeLa cells were cotransfected with *ESRRA* promoter luciferase reporter constructs containing 2 (2X) or 3 (3X) copies of the ESRRA23 element or a synthetic mutant in which the region containing the ESRRA23 element is absent, and wild-type ERRα or an ERRα^{DBDm} mutant and PGC-1α. Results are expressed as fold induction over control vector in the absence of receptor.

Figure 4.5.

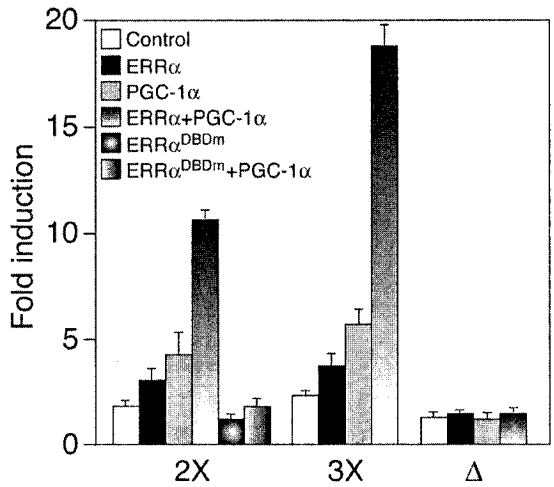
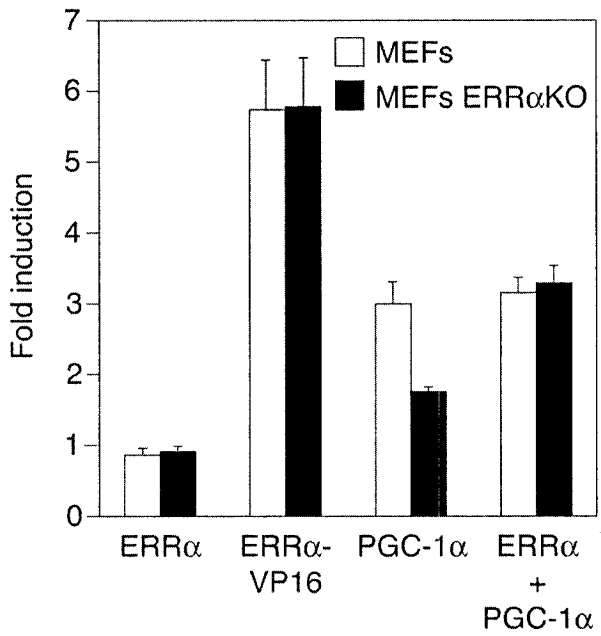

20
15
10
5
0
Fold induction
Control
ERRα
PGC-1α
ERRα+PGC-1α
ERRα^{DBDm}
ERRα^{DBDm}+PGC-1α
2X
3X
Δ

Figure 4.6. ERRα and ERR-like dependent activity of PGC-1α.

MEFs obtained from wild-type or ERRα null mice were cotransfected with an *ESRRA* promoter luciferase reporter construct containing 3 copies of the ESRRA23 element and expression vectors for ERRα, ERRα-VP16 and PGC-1α, alone or in combination. Results are expressed as fold induction over control vector in the absence of receptor or PGC-1α.

Figure 4.6.


Fold induction
0
1
2
3
4
5
6
7
MEFs
MEFs ERRαKO
ERRα
ERRα-VP16
PGC-1α
ERRα + PGC-1α

REFERENCES

Bonnelye, E., Vanacker, J. M., Dittmar, T., Begue, A., Desbiens, X., Denhardt, D. T., Aubin, J. E., Laudet, V., and Fournier, B. (1997). The ERR-1 orphan receptor is a transcriptional activator expressed during bone development. Mol Endocrinol *11*, 905-916.

Chen, S., Zhou, D., Yang, C., and Sherman, M. (2001). Molecular basis for the constitutive activity of estrogen-related receptorα-1. J Biol Chem *276*, 28465-28470.

Giguere, V. (2002). To ERR in the estrogen pathway. Trends Endocrinol Metab *13*, 220-225.

Giguère, V. (1999). Orphan nuclear receptors: from gene to function. Endocr Rev *20*, 689-725.

Giguere, V., Yang, N., Segui, P., and Evans, R. M. (1988). Identification of a new class of steroid hormone receptors. Nature *331*, 91-94.

Herzig, S., Long, F., Jhala, U. S., Hedrick, S., Quinn, R., Bauer, A., Rudolph, D., Schutz, G., Yoon, C., Puigserver, P., *et al.* (2001). CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. Nature *413*, 179-183.

Huss, J. M., Kopp, R. P., and Kelly, D. P. (2002). Peroxisome proliferator-activated receptor coactivator-1α (PGC-1α) coactivates the cardiac-enriched nuclear receptors estrogen-related receptor-α and -γ. Identification of novel leucine-rich interaction motif within PGC-1α. J Biol Chem *277*, 40265-40274.

Ichida, M., Nemoto, S., and Finkel, T. (2002). Identification of a specific molecular repressor of the peroxisome proliferator-activated peceptor γ coactivator-1α (PGC-α). J Biol Chem *277*, 50991-50995.

Johnston, S. D., Liu, X., Zuo, F., Eisenbraun, T. L., Wiley, S. R., Kraus, R. J., and Mertz, J. E. (1997). Estrogen-related receptor α1 functionally binds as a monomer to extended half-site sequences including ones contained within estrogen-response elements. Mol Endocrinol *11*, 342-352.

Kamei, Y., Ohizumi, H., Fujitani, Y., Nemoto, T., Tanaka, T., Takahashi, N., Kawada, T., Miyoshi, M., Ezaki, O., and Kakizuka, A. (2003). PPARγ coactivator 1β/ERR ligand 1 is an ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. Proc Natl Acad Sci U S A *100*, 12378-12383.

Knutti, D., and Kralli, A. (2001). PGC-1, a versatile coactivator. Trends in Endocrinology and Metabolism *12*, 360-365.

Kressler, D., Schreiber, S. N., Knutti, D., and Kralli, A. (2002). The PGC-1-related protein PERC is a selective coactivator of estrogen receptor α. J Biol Chem *277*, 13918-13925.

Laganière, J., Deblois, G., and Giguère, V. (2003). Nuclear receptor target gene discovery using high throughput chromatin immunoprecipitation, In Methods in Enzymology, D. W. Russell, and D. J. Mangelsdorf, eds. (San Diego: Academic Press), pp. 339-350.

Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. J Mol Endocrinol *19*, 207-226.

Lin, J., Wu, H., Tarr, P. T., Zhang, C. Y., Wu, Z., Boss, O., Michael, L. F., Puigserver, P., Isotani, E., Olson, E. N., *et al.* (2002). Transcriptional co-activator PGC-1 alpha drives the formation of slow-twitch muscle fibres. Nature *418*, 797-801.

Liu, D., Zhang, Z., Gladwell, W., and Teng, C. T. (2003). Estrogen stimulates estrogen-related receptor α gene expression through conserved hormone response elements. Endocrinology *144*, 4894-4904.

Lu, D., Kiriyama, Y., Lee, K. Y., and Giguère, V. (2001). Transcriptional regulation of the estrogen-inducible pS2 beast cancer marker gene by the ERR family of orphan nuclear receptors. Cancer Res *61*, 6755-6761.

Luo, J., Sladek, R., Carrier, J., Bader, J.-A., Richard, D., and Giguère, V. (2003). Reduced fat mass in mice lacking orphan nuclear receptor estrogen-related receptor α. Mol Cell Biol *23*, 7947-7956.

Maehara, K., Hida, T., Abe, Y., Koga, A., Ota, K., and Kutoh, E. (2003). Functional interference between estrogen-related receptor alpha and peroxisome proliferator-activated receptor alpha/9-cis-retinoic acid receptor alpha heterodimer complex in the nuclear receptor response element-1 of the medium chain acyl-coenzyme A dehydrogenase gene. J Mol Endocrinol *31*, 47-60.

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

McKenna, N. J., Lanz, R. B., and O'Malley, B. W. (1999). Nuclear receptor coregulators: cellular and molecular biology. Endocr Rev *20*, 321-344.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R., and Berkenstam, A. (1996). Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. Mech Dev *54*, 211-223.

Puigserver, P., Rhee, J., Donovan, J., Walkey, C. J., Yoon, J. C., Oriente, F., Kitamura, Y., Altomonte, J., Dong, H., Accili, D., and Spiegelman, B. M. (2003). Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1α interaction. Nature *423*, 550-555.

Puigserver, P., and Spiegelman, B. M. (2003). Peroxisome Proliferator-Activated Receptor-γ Coactivator 1α (PGC-1α): Transcriptional Coactivator and Metabolic Regulator. Endocr Rev *24*, 78-90.

Puigserver, P., Wu, Z., Park, C. W., Graves, R., Wright, M., and Spiegelman, B. M. (1998). A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. Cell *92*, 829-839.

Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T., and Kralli, A. (2003). The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor ERRα. J Biol Chem *278*, 9013-9018.

Sladek, R., Bader, J.-A., and Giguère, V. (1997). The orphan nuclear receptor estrogen-related receptor α is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. Mol Cell Biol *17*, 5400-5409.

Trapp, T., and Holsboer, F. (1996). Nuclear orphan receptor as a repressor of glucocorticoid receptor transcriptional activity. J Biol Chem *271*, 9879-9882.

Tremblay, G. B., Tremblay, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Labrie, F., and Giguère, V. (1997). Cloning, chromosomal localization and functional analysis of the murine estrogen receptor β. Mol Endocrinol *11*, 353-365.

Tremblay, G. B., Tremblay, A., Labrie, F., and Giguère, V. (1998). Ligand-independent activation of the estrogen receptors α and β by mutations of a conserved tyrosine can be abolished by antiestrogens. Cancer Res *58*, 877-881.

Vanacker, J.-M., Bonnelye, E., Chopin-Delannoy, S., Delmarre, C., Cavailles, V., and Laudet, V. (1999). Transcriptional activities of the orphan nuclear

receptor ERR α (estrogen receptor-related receptor-α). Mol Endocrinol *13*, 764-773.

Vanacker, J. M., Bonnelye, E., Delmarre, C., and Laudet, V. (1998). Activation of the thyroid receptor α gene promoter by the orphan nuclear receptor ERRα. Oncogene *17*, 2429-2435.

Vega, R. B., and Kelly, D. P. (1997). A role for estrogen-related receptor α in the control of mitochondrial fatty acid β-oxidation during brown adipocyte differentiation. J Biol Chem *272*, 31693-31699.

Wiley, S. R., Kraus, R. J., Zuo, F., Murray, E. E., Loritz, K., and Mertz, J. E. (1993). SV40 early-to-late switch involves titration of cellular transcriptional repressors. Genes Dev *7*, 2206-2219.

Yang, N., Shigeta, H., Shi, H. P., and Teng, C. T. (1996). Estrogen-related receptor, hERR1, modulates estrogen receptor-mediated response of human lactoferrin gene promoter. J Biol Chem *271*, 5795-5804.

Yoon, J. C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelmant, G., Stafford, J., Kahn, C. R., Granner, D. K., *et al.* (2001). Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. Nature *413*, 131-138.

Zhang, Z., and Teng, C. T. (2000). Estrogen receptor-related receptor α 1 interacts with coactivator and constitutively activates the estrogen response elements of the human lactoferrin gene. J Biol Chem *275*, 20837-20846.

CHAPTER V: ERRα directly controls genes involved in mitochondrial function

PREFACE

This chapter is a description of my contribution to a collaborative project between the laboratories of Drs Evans and Giguère. In this work, we demonstrate the role of ERRα in IFN-γ induced host defense though activation of mitochondrial activity. We contributed to this work by showing that a great proportion of IFN-γ induced genes are directly controlled by ERRα in macrophages, using a 19K mouse promoter array for ERRα ChIP-on-chip. This chapter contains this data set that is part of the paper:

Nuclear receptor ERRα is an essential effector of IFN-γ induced host defense
Junichiro Sonoda, Josée Laganière*, Isaac Mehl*, Grant Barish, Ling-Wa Chong, Immo E. Scheffler, Alain R. Bataille, François Robert, Chih-Hao Lee, Vincent Giguère and Ronald M. Evans.

*These authors contributed equally.

ABSTRACT

ERRα (estrogen-related receptor α, NR3B1) expression has recently been shown to be induced by the pro-inflammatory cytokine interferon-γ (IFN-γ) in mouse macrophages, a critical component of the host innate response to bacterial pathogenesis. Here we show using genome-wide chromatin binding profiling that genes encoding mitochondrial respiratory chain machinery are directly controlled by ERRα. These findings provide valuable information on the genomic action of this orphan nuclear receptor in coordinating mitochondrial function and suggest a role in IFN-γ-induced host defense.

INTRODUCTION

ERRα was the first orphan nuclear receptor to be identified (Giguère et al., 1988). ERRα has very few known target genes, but has been shown to share some ERα-target promoters and to control the promoter of medium-chain acyl coenzyme A dehydrogenase (MCAD), the enzyme catalyzing the initial step of the mitochondrial fatty-acid oxidation (FAO) pathway (Giguère, 2002; Sladek et al., 1997; Vega and Kelly, 1997). More recently, ERRα was shown to act as a target of the inducible coactivators PGC-1α and PGC-1β to control broad aspects of mitochondria biology including mitochondria biogenesis, FAO and oxidative respiration (Kamei et al., 2003; Mootha et al., 2004; Schreiber et al., 2004). However, whether ERRα directly controls the expression of the genes involved in those processes has not specifically been shown.

Macrophages are the principal phagocytic cell in the immune system and thus play a key defense role during bacterial infection. Upon infection, macrophages ingest bacterial pathogens through phagocytosis, which eventually matures into a phago-lysosome within which pathogens are destroyed. The anti-bacterial activities of macrophages are triggered by IFN-γ, a pro-inflammatory cytokine which exerts its effects through activation of the JAK/STAT pathway and transcriptional induction of anti-bacterial genes including nitric oxide synthase (iNOS) and GTP-binding protein LRG-47 (Darnell et al., 1994; Gutierrez et al., 2004; MacMicking et al., 1995; MacMicking et al., 2003). In addition, IFN-γ alters the expression of as much as 10% of the genome in macrophages, although contribution of most of IFN-γ inducible genes in IFN-γ-induced host defense is largely unknown (Ehrt et al., 2001).

One of the genes whose expression in macrophage was found inducible by IFN-γ is nuclear receptor ERRα (estrogen related receptor α, NR3B1) (Barish et al., 2005). However, whether the IFN-γ dependent induction of

ERRα and subsequent changes in global gene expression have physiological consequences in macrophages is unknown. To decipher the roles of ERRα in activated macrophages, we performed genome-wide chromatin binding profiling in these conditions.

METHODS

Cell culture

Bone marrow derived-macrophages were isolated and differentiated in the presence of M-CSF as previously described (Lee, 2003) and subsequently cultured in RPMI medium containing 10% FBS and 0.2% glucose or galactose.

Mouse promoter microarray design

The strategy adopted to design our mouse promoter microarray is similar to the one used for our human promoter array (Laganiere et al., 2005). Briefly, full length cDNAs were extracted from Refseq and mammalian gene collection (MGC) databases and filtered to eliminate redundancy and incomplete cDNAs. Their transcription start sites (TSS) were then located using the University of California at Santa Cruz (UCSC) genome browser (Karolchik et al., 2003) and the sequence ranging from 800 base pairs (bp) upstream to 200 bp downstream of the TSS was extracted using the UCSC database assemblage May 2004 (Karolchik et al., 2003). Primer pairs were designed using the Primer3 algorithm (Rozen and Skaletsky, 2000) and the specificity tested *in silico* using the vPCR algorithm (Lexa et al., 2001). When the primer pair gave no satisfactory vPCR results, a new primer pair was designed using primer3 and tested again. The process was iterated 3 times to generate primer pairs predicted to be efficient to amplify promoter regions from mouse genomic DNA for almost all of our selected genes. This strategy was adopted after preliminary results showed that a more simple primer design approach did not generate good results when we tried to amplify promoter regions from mouse genomic DNA. This primer design pipeline allows us to design primer pairs to amplify promoter regions from mouse genomic DNA with a success rate of about 80%. At the date of the download (November 2004) 17947 RefSeq and 16,390 MGC entries were retrieved, after the filtering process 18,892 of them were selected

and submitted to primer design. Finally primers were obtained for 18,655 promoters.

Genome-wide location analysis and chromatin immunoprecipitation

After IFN-γ treatment (10mg/ml for 12 hours), macrophages were fixed with 1% final concentration formaldehyde for 10 min at room temperature, rinsed and harvested with 1x PBS. The resultant cell pellet was sonicated, and protein-DNA complexes were enriched by immunoprecipitation with the ERRα specific antibody, beads were added and washed as previously described (Laganière et al., 2004). After de-crosslinking, the enriched DNA was repaired with T4 DNA polymerase (New England Biolabs) and ligated with linkers, as described. DNA was amplified using ligation-mediated PCR (LM-PCR), and then fluorescently labeled using BioPrime Array CGH genomic labeling kit and the Cy5 fluorophore (Invitrogen). A sample of DNA that has not been enriched by immunoprecipitation was subjected to LM-PCR and labeled with Cy3 fluorophore. Both IP-enriched and non-enriched pools of labeled DNA were hybridized to the mouse promoter array described above. The p-value threshold used to select target promoters for further analyses was determined empirically by testing randomly selected targets by standard ChIP/Q-PCR. Based on these experiments we used $p<0.01$ since our estimated false positive rate was less than 10% using this threshold (Laganière et al., 2005). Functional categories were assigned using both GO (www.fatigo.org) and manual inspection using PubMed (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed).

RESULTS AND DISCUSSION

To examine the role of ERRα, macrophages were prepared from mouse bone marrow and resulting cells were activated with the IFN-γ cytokine to undertake a chromatin-binding profiling experiment. ChIP was performed followed by hybridization to a genomic DNA microarray containing -800 to + 200 segments of approximately 19,000 promoters. We identified a total of 215 promoters ($P < 0.01$) bound by ERRα in the presence of IFN-γ, spanning various functional groups (Figure 1 and Table 1).

We previously demonstrated by ChIP assay using the human breast cancer cell line MCF7 that ERRα recognized its own promoter *in vivo* (Laganière et al., 2004). ERRα could do so through an ERRE conserved in the mouse genome. Interestingly, we found that ERRα also regulates itself in mouse macrophages, indicating a conserved mechanism of autoregulation in other organisms and tissues.

ERRα is known to be expressed predominantly in tissues with high mitochondrial oxidative capability such as heart and muscle (Giguère et al., 1988; Sladek et al., 1997). In addition, previous studies have demonstrated that PGC-1α can induce mitochondrial proliferation, stimulate mitochondrial gene expression and increase levels of fatty acid oxidation and oxidative phosphorylation in these tissues with the help of ERRα (Huss et al., 2004; Lehman et al., 2000; Schreiber et al., 2004). Consistently, a great proportion of ERRα direct target genes identified in the present study in macrophages encode mitochondrial proteins, some of which being involved in cellular respiration (Fig. 5.1, 5.2 and Table 1). Mitochondrial complex I comprises the NDUF proteins (NADH deshydrogenase (ubiquinone)) that transfers electrons from NADH to the respiratory chain. In our experiments, ERRα was shown to directly regulate several genes of this complex (*Ndufa5, Ndufa9, Ndufb4, Ndufb5, Ndufs1 and Ndufs7*) through direct promoter binding. Complex II of

the respiratory chain, which is involved in the oxidation of succinate, carries electrons from FADH to CoQ (of which *Coq9* gene is regulated by ERRα). Again, ERRα regulates genes of this complex (succinate dehydrogenase *Sdha, Sdhb* and *Sdhd*). Genes representing complex III (*1110020p15rik*), complex IV (*Cox5b Cox7a2* and *Cox8a*) and complex V (*Atp5k, Atp5c1, Atp5d, Atp5g3* and *Atp5b*) responsible for electron transfer (III and IV) and Atp synthesis (V) respectively are also well represented among ERRα targets. ERRα also controls other genes encoding mitochondrial proteins such as *Cycs* (Cytochrome c, somatic), a previously identified ERRα/PGC-1α target (Schreiber et al., 2004).

ERRα has previously been demonstrated to mediate PGC-1α activation of the nuclear respiratory factor (NRF) pathway through regulation of the *Gabpa* gene, which encodes for the NRF-2 subunit (Mootha et al., 2004). Interestingly, our chromatin binding profiling demonstrated that ERRα directly control *Gabpa* through its promoter region.

This work performed in macrophages is in agreement with others suggesting a key role for ERRα as a regulator of mitochondrial genes. Interestingly, PGC-1β expression was also induced following IFN-γ treatment in macrophages (Barish et al., 2005). It will be interesting to determine whether ERRα and PGC1-β, known to strongly coactivate ERRα, act together in the control of energy production and mitochondrial biogenesis in macrophages, similar to ERRα and PGC-1α in other organs (Kamei et al., 2003; Schreiber et al., 2004).

REFERENCE

Barish, G. D., Downes, M., Alaynick, W. A., Yu, R. T., Ocampo, C. B., Bookout, A. L., Mangelsdorf, D. J., and Evans, R. M. (2005). A Nuclear Receptor Atlas: macrophage activation. Mol Endocrinol *19*, 2466-2477.

Darnell, J. E., Jr., Kerr, I. M., and Stark, G. R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science *264*, 1415-1421.

Ehrt, S., Schnappinger, D., Bekiranov, S., Drenkow, J., Shi, S., Gingeras, T. R., Gaasterland, T., Schoolnik, G., and Nathan, C. (2001). Reprogramming of the macrophage transcriptome in response to interferon-gamma and Mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase. J Exp Med *194*, 1123-1140.

Giguère, V. (2002). To ERR in the estrogen pathway. Trends in Endocrinology and Metabolism *13*, 220-225.

Giguère, V., Yang, N., Segui, P., and Evans, R. M. (1988). Identification of a new class of steroid hormone receptors. Nature *331*, 91-94.

Gutierrez, M. G., Master, S. S., Singh, S. B., Taylor, G. A., Colombo, M. I., and Deretic, V. (2004). Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. Cell *119*, 753-766.

Huss, J. M., Pineda Torra, I., Staels, B., Giguere, V., and Kelly, D. P. (2004). Estrogen-related receptor a directs peroxisome proliferator-activated receptor a signaling in the transcriptional control of energy metabolism in cardiac and skeletal muscle. Mol Cell Biol *24*, 9079-9091.

Kamei, Y., Ohizumi, H., Fujitani, Y., Nemoto, T., Tanaka, T., Takahashi, N., Kawada, T., Miyoshi, M., Ezaki, O., and Kakizuka, A. (2003). PPARγ coactivator 1β/ERR ligand 1 is an ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. Proc Natl Acad Sci U S A *100*, 12378-12383.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., *et al.* (2003). The UCSC Genome Browser Database. Nucleic Acids Res *31*, 51-54.

Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., and Giguere, V. (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A *102*, 11651-11656.

Laganière, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., and Giguère, V. (2005). Location analysis of estrogen receptor a target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A *102*, 11651-11656.

Laganière, J., Tremblay, G. B., Dufour, C. R., Giroux, S., Rousseau, F., and Giguere, V. (2004). A polymorphic autoregulatory hormone response element in the human estrogen related receptor α (ERRα) promoter dictates PGC-1α control of ERRα expression. J Biol Chem *279*, 18504-18510.

Lehman, J. J., Barger, P. M., Kovacs, A., Saffitz, J. E., Medeiros, D. M., and Kelly, D. P. (2000). Peroxisome proliferator-activated receptor γ coactivator-1 promotes cardiac mitochondrial biogenesis. J Clin Invest *106*, 847-856.

Lexa, M., Horak, J., and Brzobohaty, B. (2001). Virtual PCR. Bioinformatics *17*, 192-193.

MacMicking, J. D., Nathan, C., Hom, G., Chartrain, N., Fletcher, D. S., Trumbauer, M., Stevens, K., Xie, Q. W., Sokol, K., Hutchinson, N., and et al. (1995). Altered responses to bacterial infection and endotoxic shock in mice lacking inducible nitric oxide synthase. Cell *81*, 641-650.

MacMicking, J. D., Taylor, G. A., and McKinney, J. D. (2003). Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science *302*, 654-659.

Mootha, V. K., Handschin, C., Arlow, D., Xie, X., St Pierre, J., Sihag, S., Yang, W., Altshuler, D., Puigserver, P., Patterson, N., *et al.* (2004). Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. Proc Natl Acad Sci U S A *101*, 6570-6575.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.

Schreiber, S. N., Emter, R., Hock, M. B., Knutti, D., Cardenas, J., Podvinec, M., Oakeley, E. J., and Kralli, A. (2004). The estrogen-related receptor alpha (ERRα) functions in PPARγ coactivator 1α (PGC-1α)-induced mitochondrial biogenesis. Proc Natl Acad Sci U S A *101*, 6472-6477.

Sladek, R., Bader, J.-A., and Giguère, V. (1997). The orphan nuclear receptor estrogen-related receptor α is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. Mol Cell Biol *17*, 5400-5409.

Vega, R. B., and Kelly, D. P. (1997). A role for estrogen-related receptor α in the control of mitochondrial fatty acid β-oxidation during brown adipocyte differentiation. J Biol Chem *272*, 31693-31699.

Table 1. Functional classification of target genes bound by ERRα in mouse macrophages

GENE	Description
	Amino acid metabolism
Got2	glutamate oxaloacetate transaminase 2, mitochondrial
Hmgcl	3-hydroxy-3-methylglutaryl-Coenzyme A lyase
	Carbohydrate metabolism
Man2c1	mannosidase, alpha, class 2C, member 1
	Cell growth/migration/adhesion
Islr	immunoglobulin superfamily containing leucine-rich repeat
Itgax	integrin alpha X
Rtn4	reticulon 4
Stoml2	stomatin (Epb7.2)-like 2
	Chromosome biogenesis
Hist1h2bc	histone 1, H2bc
Hist1h4b	histone 1, H4b
Hist1h4d	histone 1, H4d
Hist1h4m	histone 1, H4m
Tacc2	transforming, acidic coiled-coil containing protein 2
Tacc3	transforming, acidic coiled-coil containing protein 3
	DNA metabolism
Kin	antigenic determinant of rec-A protein
Mcm3ap	minichromosome maintenance deficient 3 (S. cerevisiae) associated protein
Pole3	polymerase (DNA directed), epsilon 3 (p17 subunit)
Polq	polymerase (DNA directed), theta
Supv3l1	suppressor of var1, 3-like 1 (S. cerevisiae)
	Glycolysis
Gpi1	glucose phosphate isomerase 1
	Heme metabolism
Cox10	COX10 homolog, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast)
Ncb5or	cytochrome b5 reductase 4
	Inflammatory response
Alox5ap	arachidonate 5-lipoxygenase activating protein
Mpal2	macrophage activation 2 like
Prdx5	peroxiredoxin 5
	Immune response
Ccl12	chemokine (C-C motif) ligand 12
Cd6	CD6 antigen
Csf2rb2	colony stimulating factor 2 receptor, beta 2, low-affinity (granulocyte-macrophage)
Mx1	myxovirus (influenza virus) resistance 1
Tlr1	toll-like receptor 1
Tlr5	toll-like receptor 5
	Lipid metabolism
Pafah1b3	platelet-activating factor acetylhydrolase, isoform 1b, alpha1 subunit
	Oxidative phosphorylation/TCA cycle
Atp5b	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit
Atp5c1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1
Atp5g2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 2
Atp5g3	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 3
Atp5k	ATP synthase, H+ transporting, mitochondrial F1F0 complex, subunit e
Cox5b	cytochrome c oxidase, subunit Vb
Cox7a2	cytochrome c oxidase, subunit VIIa 2
Cox8a	cytochrome c oxidase, subunit VIIIa
Cycs	cytochrome c, somatic
Ndufa5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5
Ndufa9	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9
Ndufb4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4
Ndufb5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5
Ndufs1	NADH dehydrogenase (ubiquinone) Fe-S protein 1
Ndufs7	NADH dehydrogenase (ubiquinone) Fe-S protein 7
Sdha	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
Sdhd	succinate dehydrogenase complex, subunit D, integral membrane protein
Suclg1	succinate-CoA ligase, GDP-forming, alpha subunit
	Protein metabolism and modification
Ahsa1	AHA1, activator of heat shock 90kDa protein ATPase homolog 1 (yeast)
Birc6	baculoviral IAP repeat-containing 6

Bms1l	BMS1-like, ribosome assembly protein (yeast)
Cct7	chaperonin subunit 7 (eta)
Clgn	calmegin
Comtd1	catechol-O-methyltransferase domain containing 1
Coq7	demethyl-Q 7
Dnajc17	DnaJ (Hsp40) homolog, subfamily C, member 17
Eef1b2	eukaryotic translation elongation factor 1 beta 2
Eif3s6	eukaryotic translation initiation factor 3, subunit 6
Galnt12	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 12
Gfm2	G elongation factor, mitochondrial 2
Hspa9a	heat shock protein 9A
Lnpep	leucyl/cystinyl aminopeptidase
Mrpl11	mitochondrial ribosomal protein L11
Mrpl19	mitochondrial ribosomal protein L19
Mrpl47	mitochondrial ribosomal protein L47
Mprs18b	mitochondrial ribosomal protein S18B
Nek4	NIMA (never in mitosis gene a)-related expressed kinase 4
Pfdn5	prefoldin 5
Pja2	praja 2, RING-H2 motif containing
Ppil3	peptidylprolyl isomerase (cyclophilin)-like 3
Ppp3cc	protein phosphatase 3, catalytic subunit, gamma isoform
Prcp	prolylcarboxypeptidase (angiotensinase C)
Prepl	prolyl endopeptidase-like
Psma6	proteasome (prosome, macropain) subunit, alpha type 6
Psmb1	proteasome (prosome, macropain) subunit, beta type 1
Rpl7l1	ribosomal protein L7-like 1
Rps14	ribosomal protein S14
Sec11l1	Sec11-like 1 (S. cerevisiae)
Trim30	tripartite motif protein 30
Trp53rk	Trp53 regulating kinase
Zfyve19	zinc finger, FYVE domain containing 19
	RNA metabolism and processing
Dhx38	DEAH (Asp-Glu-Ala-His) box polypeptide 38
Exosc4	exosome component 4
Grsf1	G-rich RNA sequence binding factor 1
Phf5a	PHD finger protein 5A
Snrpe	small nuclear ribonucleoprotein E
Ttc14	tetratricopeptide repeat domain 14
Txnl4b	thioredoxin-like 4B
Usp52	ubiquitin specific peptidase 52
Xrn1	5'-3' exoribonuclease 1
Zfp36	zinc finger protein 36
	Signal transduction
Akt1	thymoma viral proto-oncogene 1
Arhgef1	Rho guanine nucleotide exchange factor (GEF) 1
Bmx	BMX non-receptor tyrosine kinase
Casr	calcium-sensing receptor
Cerk	ceramide kinase
Edg8	endothelial differentiation, sphingolipid G-protein-coupled receptor, 8
Gnb2	guanine nucleotide binding protein, beta 2
Itgb3bp	integrin beta 3 binding protein (beta3-endonexin)
Mgst3	microsomal glutathione S-transferase 3
Olfr56	olfactory receptor 56
Olfr315	olfactory receptor 315
Olfr1015	olfactory receptor 1015
Plxna2	plexin A2
Sphk2	sphingosine kinase 2
	Steroid and drug metabolism
Sult4a1	sulfotransferase family 4A, member 1
Ugt1a6	UDP glucuronosyltransferase 1 family, polypeptide A6
	Transcriptional regulator
Asb5	ankyrin repeat and SOCs box-containing protein 5
Asb15	ankyrin repeat and SOCS box-containing protein 15
Atf2	activating transcription factor 2
Crsp3	cofactor required for Sp1 transcriptional activation, subunit 3
Esrra	estrogen related receptor, alpha
Gabpa	GA repeat binding protein, alpha
Gabpb2	GA repeat binding protein, beta 2
Med4	mediator of RNA polymerase II transcription, subunit 4 homolog (yeast)
Polrj2	polymerase (RNA) II (DNA directed) polypeptide J
Ppp1r10	protein phosphatase 1, regulatory subunit 10
Rnpc2	RNA-binding region (RNP1, RRM) containing 2
Rxrb	retinoid X receptor beta
Smarca2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
Smarca5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5

Supt7l	suppressor of Ty 7 (S. cerevisiae)-like
Taf6	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor
Tbp	TATA box binding protein
Tead3	TEA domain family member 3
Tfb2m	transcription factor B2, mitochondrial
Whsc2	Wolf-Hirschhorn syndrome candidate 2 (human)
Zpf142	zinc finger protein 142
Zfp143	zinc finger protein 143
	Transport
Abcb8	ATP-binding cassette, sub-family B (MDR/TAP), member 8
Accn5	amiloride-sensitive cation channel 5, intestinal
Ap1g2	adaptor protein complex AP-1, gamma 2 subunit
Atp2b2	ATPase, Ca++ transporting, plasma membrane 2
Exoc3	exocyst complex component 3
Gosr1	golgi SNAP receptor complex member 1
Nup54	nucleoporin 54
Rab2b	RAB2B, member RAS oncogene family
Rab21	RAB21, member RAS oncogene family
Rhcg	Rhesus blood group-associated C glycoprotein
Sec22l2	SEC22 vesicle trafficking protein-like 2 (S. cerevisiae)
Slc4a1ap	solute carrier family 4 (anion exchanger), member 1, adaptor protein
Slc41a3	solute carrier family 41, member 3
Slc9a1	solute carrier family 9 (sodium/hydrogen exchanger), member 1
Slc35b1	solute carrier family 35, member B1
Slc39a7	solute carrier family 39 (zinc transporter), member 7
Timm8b	translocase of inner mitochondrial membrane 8 homolog b (yeast)
Timm10	translocase of inner mitochondrial membrane 10 homolog (yeast)
Trappc3	trafficking protein particle complex 3
Vps28	vacuolar protein sorting 28 (yeast)

Genes without an assigned function at this level of analysis: 0610011N22Rik; 0610038D11Rik; 0610039G24Rik; 1110002N22Rik; 1110020P15Rik; 1500034E06Rik; 1810003N24Rik; 1810026J23Rik; 2310015N07Rik; 2610019P18Rik; 2610044O15Rik; 3100002L24Rik; 3110005O21Rik; 4632434I11Rik; 4930506M07Rik; 4930583K01Rik; 4933414I15Rik; 4933430I17; 4933434G05Rik; 5730427N09Rik; 5730589K01Rik; 9430023L20Rik; 9630058J23Rik; AA407526; A130042E20Rik; Acp2; AI413782; AI553587; Atxn10; BC019806; BC020077; BC051227; C530044N13Rik; Cd300lf; Chordc1; Cln8; Dock7; Dock11; D930001I22Rik; Dscr5; Gdpd5; Grwd1; Hbld2; Hdhd2; Hdhd3; Mbd6; Nsfl1c; Rab11fip4; Rfn187; Sacm1l; Sepm; Sh2d4b; Sh3rf2; Spata11; Theg; Tpr; Trim31; Ttc5; Unc93b; Yipf2; Zcchc14; Zcsl2; Zfp353; Zfp655; Zmat5.

In the case that one locus could be assigned to two distinct genes, both genes were included in the analysis.

Figure 5.1. Promoter binding of ERRα in IFN-γ-treated macrophages.
Chromatin/protein complex were precipitated from IFN-γ-treated macrophages using ERRα antibody, and hybridized to a mouse promoter DNA chip. The wheel represents all the 215 genes (with GO functions) that were identified. The genes identified in both the genome-wide expression in IFN-γ treated cells and promoter-binding experiments are listed around the wheel (red: mitochondrial genes; black: other location). Two genes had no function assigned to them (listed below the categories). The genes that were also identified by the IFN-γ/IL-4 comparison are shown with bold letter. Asterisks indicate genes that were not identified by the promoter DNA chip, but were positive by the conventional chromatin immunoprecipitation.

Figure 5.1.

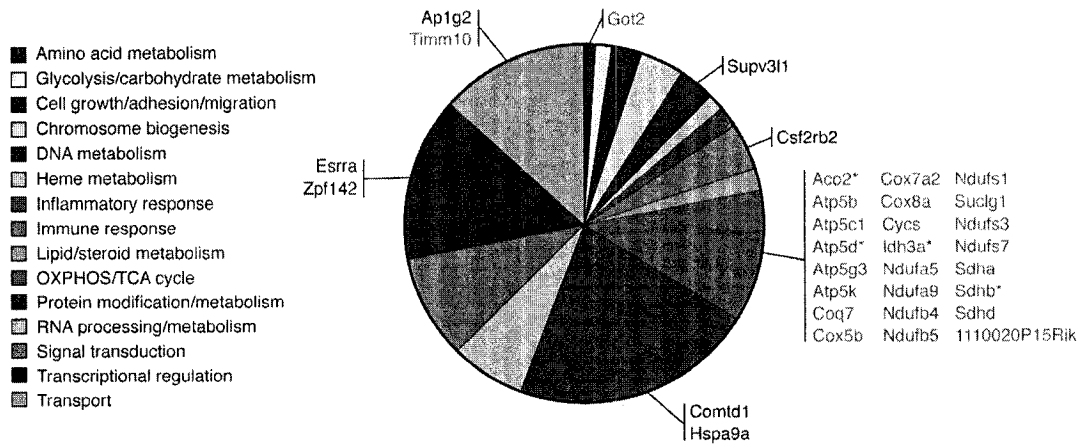

Amino acid metabolism
Glycolysis/carbohydrate metabolism
Cell growth/adhesion/migration
Chromosome biogenesis
DNA metabolism
Heme metabolism
Inflammatory response
Immune response
Lipid/steroid metabolism
OXPHOS/TCA cycle
Protein modification/metabolism
RNA processing/metabolism
Signal transduction
Transcriptional regulation
Transport
Ap1g2
Timm10
Got2
Supv3l1
Csf2rb2
Esrra
Zpf142
Acc2* Cox7a2 Ndufs1
Atp5b Cox8a Suclg1
Atp5c1 Cycs Ndufs3
Atp5d* Idh3a* Ndufs7
Atp5g3 Ndufa5 Sdha
Atp5k Ndufa9 Sdhb*
Coq7 Ndufb4 Sdhd
Cox5b Ndufb5 1110020P15Rik
Comtd1
Hspa9a

Figure 5.2. ERRα targets identified by genome-wide expression and chromatin binding profilings.

Functional locations (nucleus, plasma membrane, cytoplasm or mitochondria) of the identified gene products are schematically shown. Blue: common between the expression profiling of IFN-γ- and IL-4-treated cells, Green: common between expression profiling and chromatin-binding profiling of IFN-γ-treated cells, and Red: common for both of these two criteria. Abbreviations: MRP (mitochondrial ribosomal proteins); PDH (pyruvate dehydrogenase complex); TIMM (translocase of inner mitochondria membrane).

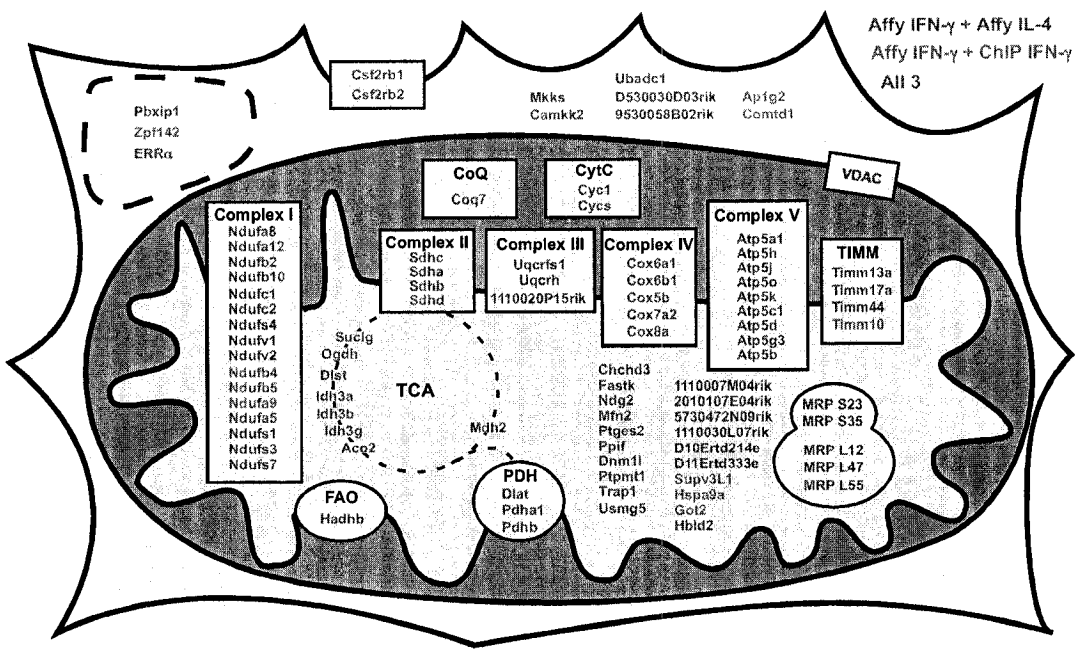

Affy IFN-γ + Affy IL-4
Affy IFN-γ + ChIP IFN-γ
All 3
Csf2rb1
Csf2rb2
Ubadc1
Mkks
D530030D03rik
Ap1g2
Camkk2
9530058B02rik
Comtd1
Pbxip1
Zpf142
ERRα
CoQ
Coq7
CytC
Cyc1
Cycs
VDAC
Complex I
Ndufa8
Ndufa12
Ndufb2
Ndufb10
Ndufc1
Ndufc2
Ndufs4
Ndufv1
Ndufv2
Ndufb4
Ndufb5
Ndufa9
Ndufa5
Ndufs1
Ndufs3
Ndufs7
Complex II
Sdhc
Sdha
Sdhb
Sdhd
Complex III
Uqcrfs1
Uqcrh
1110020P15rik
Complex IV
Cox6a1
Cox6b1
Cox5b
Cox7a2
Cox8a
Complex V
Atp5a1
Atp5h
Atp5j
Atp5o
Atp5k
Atp5c1
Atp5d
Atp5g3
Atp5b
TIMM
Timm13a
Timm17a
Timm44
Timm10
Sucl g
Ogdh
Dlst
Idh3a
Idh3b
Idh3g
Acq2
TCA
Mdh2
Chchd3
Fastk
Ndg2
Mfn2
Ptges2
Ppif
Dnm1l
Ptpmt1
Trap1
Usmg5
1110007M04rik
2010107E04rik
5730472N09rik
1110030L07rik
D10Ertd214e
D11Ertd333e
Supv3L1
Hspa9a
Got2
Hbld2
MRP S23
MRP S35
MRP L12
MRP L47
MRP L55
FAO
Hadhb
PDH
Dlat
Pdha1
Pdhb

Figure 2

CHAPTER VI GENERAL DISCUSSION

ERα and ERRα are two related nuclear receptors from the steroid receptor subfamily. ERα generally modulates transcription in the presence of its ligand estrogen, while ERRα is an orphan receptor that displays constitutive activity. These two receptors are particularly related in their DBD, which indeed allows the recognition of similar DNA response elements *in vitro* and in reporter assays (Giguère, 1999; Giguère, 2002). In addition, they have both been linked to the breast cancer disease in several independent reports.

The goal of this thesis was to develop chromatin immunoprecipitation approaches for the identification and/or characterization of *bona fide* ERα and ERRα direct target genes and regulatory modules. The identification of ERα and ERRα target genes is expected to significantly increase our knowledge of the function and specific action of these receptors at the genome level especially that very few have been discovered so far.

6.1 ERα ChIP-cloning

Although the molecular mechanisms of ERα transcriptional activation are well understood, very little is known about the direct genomic regulatory sequences it controls which postpones our full comprehension of the role of this receptor in physiologic and pathologic states. As a first step in the goal of characterizing novel *in vivo* ERα target functional in breast cancer cells, we developed a ChIP-derived approach that we called ChIP-cloning which consists in the cloning of the DNA fragments obtained after the ERα ChIP, their sequencing and subsequent identification using human genome databases recently made

available (Laganière et al., 2003). This approach had also been applied for the discovery of E2F target genes (Weinmann et al., 2001; Weinmann et al., 2002). This first involved the optimization of ERα ChIP on the previously characterized prototypic estrogen target *pS2* (*TFF1*) promoter (Metivier et al., 2003; Shang et al., 2000). In opposition to many other methods used to characterize target genes, ChIP-based strategies have the advantage of leading to the direct identification of elements functional *in vivo*, in the conditions used for the experiment. In chapter II, we developed and used the ChIP-cloning strategy to successfully identify 12 ERα-bound sequences (Laganière et al., 2005a). The *TFF1* promoter sequence was actually found in our ChIP-cloning experiment, our first indication of the validity of the approach to identify ERα genomic targets. Another promoter region was identified as an ERα target, that of *FLJ10618*, a gene from the solute carrier family with no clear assigned function to date (UCSC database: genome.ucsc.edu). As envisioned, our initial ChIP-cloning results markedly showed that ERα controls genomic sequences located distally from known annotated genes. For example, ERα recognized a genomic region located 12 kb downstream of the Breakpoint Cluster Region (BCR) gene. Although the BCR-ABL fusion protein produced by the translocation from both BCR and ABL genes has been extensively studied for its crucial role in chronic myeloid leukemia (Elrick et al., 2005), the function of the normal BCR protein is not yet clear. However, the recent finding that BCR is a negative regulator of the Wnt signaling pathway could be of particular interest in the context of hormone-dependent breast cancers (Ress and Moelling, 2005).

6.1.1 An intronic ERE responsible for RAR regulation by estrogens

RARα and ERα genes are coordinately expressed in breast cancer cells (Lu et al., 2005). Although the role of RARα in breast cancer is not entirely clear, it is worth noting that its expression provides retinoid-induced growth inhibition of breast cancer cells (Afonja et al., 2002; Fitzgerald et al., 1997).

Interestingly, estrogen treatments have previously been shown to increase RARα gene expression (Elgort et al., 1996; van der Leede et al., 1995). After the analysis of the *RARA* (RARα gene) proximal promoter region, this estrogen up-regulation was attributed to an indirect action of ERα through the SP1 transcription factor (Safe, 2001; Safe and Kim, 2004; Sun et al., 2001). With our unbiased ChIP-cloning method for the isolation of ERα-regulated genomic elements, we identified a functional ERE located within the first intron of RARα1, 3.7 kb downstream the RARα1 promoter. We have shown using standard ChIP that ERα, as well as coregulators SRC-1, SRC-3 and RIP140 were recruited in an estrogen-dependent manner, showing that this site is hormone-responsive. In contrast, neither ERα nor coactivator recruitment was observed at the promoter region, in contrast to SP1 and RNA polymerase II. Moreover, we demonstrated that this intronic ERE provided the major estrogenic response of the RARα gene with the transfection of the RARA locus containing a 4 kb fragment including the promoter and the downstream ERE. Our results suggests that the functional intronic ERE identified acts as an intronic enhancer for the control of RARα up-regulation by estrogens.

Therefore, our ChIP-cloning approach was successful for the identification of novel ERα targets and confirms the value of ChIP-based strategies for the characterization of transcriptionally functional regions of the human genome, including the ones located distally from annotated promoters. Another important aspect revealed by this study concerns RIP140, one of the first receptor coregulator to be identified and isolated through its recruitment by ERα-AF2 in the presence of ligand (Cavaillès et al., 1995). We provided the first *in vivo* evidence that ERα recruits endogenous RIP140/NRIP to regulatory regions involved in gene activation like the *pS2/TFF1* promoter and the *RARA* downstream enhancer, although it is believed to possess repressive properties. We thus proposed that the corepressor RIP140, which functions on agonist-

bound receptors, is recruited for the fine-tuning of ERα-transcriptional activation.

ChIP-cloning has not been used widely for genome-wide location analyses due to the extensive sequencing step and mapping to the genome that has to be executed, perhaps explaining why ChIP-on-chip became an appealing alternative option.

6.2 ERα ChIP-on-chip.

The ChIP-on-chip technology was first developed for studies in *S. cerevisiae* for the identification of complete sets of genomic targets (Iyer et al., 2001; Ren et al., 2000; Simon et al., 2001; Wyrick et al., 2001). Ideally, these ChIP-on-chip studies would also be performed in complex organisms, using an array that represents the entire genome. Unfortunately, because of the length of mammalian genomes, human for instance, that are composed of long intergenic regions that are likely to be transcription factor targets, it was until very recently technically impossible to represent whole genomes on a single or a reasonable amount of arrays for experimentation. Consequently, compromises had to be made in laboratories regarding the sequences chosen to compose an array. The first arrays were thus covering a number of human promoters, as for the identification of E2F targets an array was made with PCR products of 1200 promoter regions (Ren et al., 2002). Similarly, a human CpG microarray was constructed by another group for the identification of E2F4 targets (Weinmann et al., 2002). The binding profile of GATA-1 was performed on an array representing the β-globin locus (Horak et al., 2002). The challenge soon became to render the studies more and more complete through the production of arrays with increased capacities. In 2004, the binding profiles of 3 HNF transcription factors in pancreas and liver were revealed using a 13,000 promoter array (Odom et al., 2004). Likewise, E2F4 and p130 were shown to cooperatively repress a common set of genes under growth arrest condition using a 13,000 promoter array (Cam et al., 2004). Alternatively, high-density

oligonucleotide arrays covering all non-repetitive sequences of the entire chromosome 22 or 21 and 22 were also developed to perform ChIP-on-chip of NFκB, CREB, SP1, c-Myc and p53 (Cawley et al., 2004; Euskirchen et al., 2004; Martone et al., 2003). These tiling arrays have the advantage of allowing the profiling of transcription factors without bias for promoters or other regions, which are useful to our in-depth understanding the transcriptional regulation mechanisms and the annotation of the functional regulatory sequences along a whole chromosome. Another advantage is the possibility of performing gene expression profiling on the same chips. However, they have the inconvenience of representing a very small number of genes and the association of distal regulatory region to a specific gene is not simple due to similar proximities of several genes; chromosome 22 comprises 756 protein-coding genes, making it impossible to decipher pathways controlled by a transcription factor when using only this type of array.

In an attempt to identify ERα genomic targets in breast cancer cells in a high-throughput manner, we constructed a 19,000 promoter array and performed ChIP-on-chip using MCF7 breast cancer cells treated with estrogens (Laganière et al., 2005b). Valuably, the evaluation of the relative binding of ERα to thousands of promoters in a single experiment gives the opportunity to reveal pathways directly controlled by estrogens and provide information regarding the genomic sequences used *in vivo* by this nuclear receptor for transcriptional activation of target genes, which will be considerably helpful in the delineation of biological responses controlled by estrogen-bound ERα in breast cancer cells. Although our choice of a promoter array does not allow the identification of distal regulatory sequences, we hypothesize based on the looping model for transcriptional regulation that some functional enhancers will connect at some point with promoter regions and will therefore be enriched in our ChIP (since protein-protein interactions are fixed with formaldehyde) and consequently identified as an ERα target gene, even if the regulatory sequence is not directly within the promoter.

6.2.1 ERα promoter-binding profiling in breast cancer cells

Our genome-wide analysis of promoter occupancy was a great accomplishment in the isolation of functional ERα transcriptional regulatory regions. While only about a few of direct estrogen-responsive genes were defined in the literature to date, we identified 153 promoters controlled by ERα in MCF7 cells in the presence of estradiol in a single experiment (Laganière et al., 2005b; Laganiere and Giguere, 2006). Concomitant with our publication, a chromosome-wide mapping of ERα binding was also issued (Carroll et al., 2005). In their study, Carroll et al. used a chromosome 21 and 22 tiling array made by Affymetrix for ChIP-on-chip of estradiol-treated MCF7 cells. They identified 57 ER binding sites, most of them located distally from known genes and transcriptional start sites.

Since the chromatin shearing step carried out for the ChIP procedure produces DNA fragments of about 500 to 2000 bp, ChIP and ChIP-on-chip allow the isolation of direct target loci of a transcription factor with a 1-2 kb resolution. Therefore, informatics tools have been developed to locate binding sites that are overrepresented within the enriched sequences. One such motif-finding algorithm used for analyzing ChIP-on-chip data is called MDscan (http://ai.stanford.edu/~xsliu/MDscan/). Interestingly, an MDscan search for overrepresented motifs within ERα-bound promoter sequences revealed an ERE present in 60% of all promoters, confirming that the technique was successful in finding ERα targets, in addition to underscore that ERα really uses a sequence close to the actual ERE consensus *in vivo* in the majority of the cases. For promoters where no ERE was found, it is possible that ERα acts through binding to other transcription factors present on promoters. Another possibility is that it binds an ERE just off the sequence entered in MDscan program for the search and that the enriched fragment could overlap and hybridize to the sequence on the array, since ChIP has a resolution of 2kb approximately. In addition, as mentioned above, since formaldehyde

crosslinking stabilizes protein-protein in addition to protein-DNA interactions and functional enhancers can potentially interact with promoter regions, these can also be detected as positive promoters without containing an ERE, but it is possible that the distal ERα-bound sequences possess one and that the promoter represents an active promoter.

In ChIP-chip, the threshold p-value is set experimentally, through standard ChIP and confirmation of targets by quantitative PCR (Ren et al., 2000). The cut-off is established by reconfirmation of the enriched promoters by standard ChIP, at the p-value giving less than 10% false-positive results. This procedure assures that targets within that threshold are very limited in false-positives. To further validate that the promoters were truly estrogen-responsive, we assessed the recruitment of coactivators and RNA polymerase II in addition to that of ERα and thus demonstrated the reactivity of the promoters following hormonal treatment.

Some of the bound promoters identified corresponded to known direct ERα targets. This is the case for *TFF1/pS2*, already known as an ERα responsive direct target (Berry et al., 1989; Metivier et al., 2003; Roberts et al., 1988; Shang et al., 2000). In addition, we identified the *FLJ10618/SLC25A36* promoter that we had previously identified in our ERα ChIP-cloning study (Laganière et al., 2005a). Caspase 7 (*CASP7*), *GREB1* and *Ly6e* promoters all contain EREs and were shown to bind ERα in MCF7 cells prior to our work (Bourdeau et al., 2004; Lin et al., 2004). *GREB1* is a gene of unknown function whose expression is induced by estrogens in breast cancer cells (Ghosh et al., 2000). Remarkably, a recent study suggests that knocking down GREB1 expression by siRNA suppresses cell proliferation of estrogen dependent MCF-7 cells, but the function of the protein remains to be elucidated (Rae et al., 2005). *Ly6e* (lymphocyte antigen 6 complex e) was also identified as estrogen-responsive at a 24h time point in the breast cancer cell line ZR-75-1 (Soulez and Parker, 2001b). The nuclear receptor SHP (small heterodimer partner,

NR0B2) was also characterized for its estrogen-inducibility in HepG2 cells occurring through an ERE located in its promoter (Lai et al., 2003). Similarly, WISP2 (Wnt-1 induced signaling pathway protein 2) was inducible by estrogens in an MCF7 cell experiment through a promoter ERE (Frasor et al., 2003; Inadera et al., 2000). *CYP1B1* (cytochrome P450, family 1, subfamily B, polypeptide 1) is another gene that was known to be up-regulated by estrogens in T-47D and MCF7 cells through EREs near the start site (Finlin et al., 2001; Lin et al., 2004; Tsuchiya et al., 2004; Tsuchiya et al., 2005).

Other genes whose expression were augmented in the presence of estradiol with no previous evidence of direct regulation by ERα were found in our ChIP-on-chip experiment and include: *HK1* (hexokinase 1) in rat uteri (Andrade et al., 2002), ANXA9 (annexin A9) in MCF7 cells (Cicatiello et al., 2004), the secreted SEMA3B protein (sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B) in MCF-7 cells (Terasaka et al., 2004), *COQ4* in T47D cells (Lin et al., 2004), *ABCA3* (ATP-binding cassette, sub-family A (ABC1), member 3) and *TFF3* (trefoil factor 3, Intestinal) in breast cancer cells ZR-75-1 (Soulez and Parker, 2001a), *H11* (protein kinase H11, HSPB8 heat shock 22kDa protein 8) in MCF7 (Charpentier et al., 2000), *PKIB* (protein kinase (cAMP-dependent, catalytic) inhibitor beta) and *FLJ13710* in MCF7 (Finlin et al., 2001) *P2RY6* (pyrimidinergic receptor P2Y, G-protein coupled, 6) in human umbilical vein endothelial cells (Pedram et al., 2002), *PDZK1* (PDZ domain containing 1) in MCF7 cells (Inoue et al., 2002), *DOC1* or *CDK2AP1* (CDK2 (cyclin-dependent kinase 2)-associated protein 1) in mouse pregnant uteri (Reese et al., 2001), *ENY2* (enhancer of yellow 2 homolog (Drosophila)) in ZR-75-1 breast cancer cells (Seth et al., 2002), and recently, *UGT2B15* (UDP (uridine diphosphate)-glucuronosyltransferase (UGT) 2B15) enzyme in MCF7 cells (Harrington et al., 2006).

Our list of ERα targets also includes some genes that were downregulated by estrogens: *RBL2* (retinoblastoma-like 2 (p130)) was downregulated by estrogen in MCF7 cells (Frasor et al., 2003), *ESR1* (estrogen receptor α) in several conditions (Inoue et al., 2000; Lin et al., 2004; Prange-Kiel et al., 2001; Seo et al., 1998; Terasaka et al., 2004), *ABCC5* (ATP-binding cassette, sub-family C (CFTR/MRP), member 5) (Finlin et al., 2001; Lin et al., 2004) in MCF7 and T-47D breast cancer cells, *S100A10* (S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11)) in ZR-75-1 breast cancer cells (Soulez and Parker, 2001a), *SMAP* or *KIFAP3* (kinesin-associated protein 3) in T47D breast cancer cells (Lin et al., 2004). These observations suggest a direct gene repression exerted by ERα, a mechanism of action for which not much information is available yet. We thus provide model promoters for the study of this ER-driven mechanism, which should greatly help refine our knowledge. Interestingly, ERα seems to downregulate its own expression through a direct mechanism of action by promoter repression. As mentioned in the present work, estrogen mediate the protein stability of ERα, in addition to the 3'UTR important for mRNA stability, implicating a complex regulation of ERα at several levels (Reid et al., 2002).

Most of the promoters identified in our ChIP-on-chip experiment represent novel direct ERα target genes. Several genes and pathways isolated could provide insight into the role that ERα plays in breast cancer. Since the identification of WNT-1 as a mammary oncogene in MMTV (mouse mammary tumor virus) infected mice, the Wnt signaling pathway is the object of studies evaluating its importance for human breast cancer (Nusse and Varmus, 1982). Although no equivalent of MMTV exists for humans, increasing evidence suggest that Wnt proteins and/or constituent of their pathways are important in human breast cancer. Interestingly, ERα was shown to bind to 4 promoters of genes part of the Wnt signaling pathway: *WNT16*, *WISP2*, *SEMA3B* and *CTNNB1P1*. However, the way these proteins act in ER-positive breast cancer remains to be elucidated. Additionally, ERα controlled genes involved in the

metabolism of steroids (*CYP1B1*, *STS*, *UGT2B15*, *UGT2B17*). STS (steryl-sulfatase precursor) catalyzes the conversion of sulfated steroid precursors to estrogens in tissues. Interestingly, STS inhibitors are currently being developed as a breast cancer therapy (Nussbaumer and Billich, 2004). The CYP450 are oxidative enzymes with an important role in the metabolism of a diverse range of xenobiotics (Gonzalez and Gelboin, 1994). Estrogens can be metabolized by CYPs through hydroxylation at different sites and produce estrogenic, non-estrogenic or even carcinogenic molecules. Combined polymorphisms in *CYP* genes are thought to enhance cancer risk (Murray, 2000). CYP1B1 is expressed is breast tumors but not in normal breast tissue, appearing as a tumor-specific *CYP*, and is associated with the increased 4-Hydroxy-17β-estradiol production in breast cancer, a potentially carcinogenic metabolite (Liehr et al., 1986). Its colocalization with CYP19, the enzyme producing estrogen, could lead to high local production of carcinogenic estrogen metabolites. In addition, polymorphisms have been identified in the *CYP1B1* gene and are thought to have divergent outcomes in various ethnicities.

The UGT (UDP-glucuronosyltransferase) family is comprised of enzymes that catalyze the formation of water-soluble metabolites of many biologically active substrates through the transfer of glucuronic acid from the cofactor UDP-glucuronic acid (UDPGA) to substrates. Localized to the endoplasmic reticulum, they convert their targets (endogenous biomolecules and xenobiotics including steroids, bile acids, bilirubin, dietary constituents, drugs, environmental toxicants, and carcinogens), to more hydrophilic, acidic and less biologically active components, facilitating their excretion through bile and urine (Tukey and Strassburg, 2000). *UGT2B15* is highly expressed in liver, prostate, kidney, testis, mammary gland, placenta, adipose, and uterus and is known to glucuronidate a wide range of endogenous estrogens and androgens as well as tamoxifen, and variants might be associated with breast cancer survival and recurrence (Nowell et al., 2005). UGT2B17 is identical to UGT2B15 at 95% but has only androgens as substrates, which might also have

non-negligible impact in breast cancer patients since androgens are thought to be antiproliferative in this context (Harrington et al., 2006; Tukey and Strassburg, 2000). We also identified two ERα targets in the *ABC* (ATP-binding cassette) gene family (*ABCC5*, *ABCC11*) which are involved in multidrug resistance and could influence response to chemotherapy (Park et al., 2006). ERα was also shown to control several other transcription factors that could be important in mediating secondary effects of estrogen. Among them was NR0B2 (Small heterodimer partner, Shp), an atypical orphan nuclear receptor nuclear receptor lacking a conventional DNA binding domain actually able to repress ligand-bound ERα activity (Johansson et al., 1999). *PPRC1* (PGC-1-related coactivator) was also a target of ERα, and was previously described as an ER-specific coactivator (Kressler et al., 2002). Together, these data show how ERα controls multiple pathways at the transcription level that hold the potential of influencing the response to estrogens and antiestrogens through multiple direct and feedback mechanisms.

ChIP validations measuring RNA pol II occupancy also gave us interesting insights about various mechanisms of action by ERα. Not all promoters showed similar RNA pol II levels or recruitment after estrogen treatments. For example, *CYP1B1* and *FOXA1*, whose gene expression are known to be upregulated by estrogens, showed high levels of RNA pol II at their promoters, but not much, if any, increased recruitment following estrogen treatments. This could suggest that RNA pol II could be present but not highly active without the addition of estrogens. It would be interesting to test whether antibodies against activated RNA pol II (phosphorylated C-terminal domain (CTD)) could detect an increased promoter activity. Also, promoters whose genes are downregulated by estrogens were isolated in our ChIP-on-chip study, showing that ERα potentially acts as a direct repressor of gene expression on some targets, although the clear mechanism for such activity remains to be elucidated.

Further examination of their roles and expression levels in breast cancer subtypes might provide important insights into the understanding of breast cancer and susceptibility to the disease, in addition to presenting possible novel biomarkers.

6.2.2 ERα-FOXA1 cooperation defines a domain of the estrogen response

Forkhead box transcription factors (FOX) were given their name based on the discovery of mutations in *Drosophila* causing a characteristic spiked head appearance due to defects in head fold involution (Friedman and Kaestner, 2006; Weigel et al., 1989). The *FOXA* subfamily comprises 3 subtypes that are the most similar to the Drosophila gene among the 100 *Fox* genes; *Foxa1, Foxa2 and Foxa3*. *Foxa* genes were found to have important roles in liver, pancreas, lung, prostate and kidney function (Friedman and Kaestner, 2006). A the molecular level, the *Foxa forkhead* box or winged helix is composed of 3 α-helices arranged in a helix-turn-helix core and allows DNA binding (Clark et al., 1993). In addition, it was noted that this *Forkhead* box structure resembles that of linker histone H1 (Clark et al., 1993). Moreover, the C-terminal domain of Foxa proteins has the ability to interact with histones H3 and H4 (Cirillo et al., 2002). Accordingly, together with the observation that they can open highly compacted chromatin *in vitro* in an ATP-independent manner, FOXA were proposed as "pioneer" transcription factors (Holmqvist et al., 2005). Thus, FOXA binding would render promoters competent for recruitment of other transcription factors needed for responding to signals induced in particular physiological or developmental stages. This competence phenomenon was indeed observed for activity of some hormone receptors. Transcriptional activation mediated by GR was shown to be dependent on FOXA2 pioneering action during fasting (Zhang et al., 2005). AR and Foxa2 were also shown to act cooperatively on prostate-specific gene regulatory regions (Gao et al., 2003). In addition, it was shown that ERα and FOXA1 co-operatively activate transcription of the liver-specific vitellogenin B1 gene (Robyr et al., 2000).

Interestingly, FOXA1 expression levels are strongly associated to that of ERα in breast cancer cells and tumors (Bertucci et al., 2000; Gruvberger et al., 2001; Lacroix and Leclercq, 2004; Perou et al., 2000; Ross and Perou, 2001; Sorlie et al., 2001; Sorlie et al., 2003; Sotiriou et al., 2003; van 't Veer et al., 2002; West et al., 2001). Sequence analysis of ERα-bound promoters found in our chromatin binding profiling experiment revealed enrichment in FOXA1 consensus sites. This observation prompted us to verify by ChIP whether FOXA1 could bind these promoters *in vivo*, which we could clearly show. In contrast, on promoters containing no FOXA1 sites we could not detect FOXA1 binding, showing the specificity of binding to a subset of ERα target promoters. Interestingly, in agreement with the assumption that FOXA1 can act as a pioneer factor, we showed by ChIP that its knock down using siRNA inhibited ERα recruitment to target promoters that are usually shared with FOXA1. Similar observations were simultaneously made by Carroll et al. who showed impaired recruitment of ERα on distal binding sites following FOXA1 knock down (Carroll et al., 2005). In addition, we showed that FOXA1 is a direct target of ERα whose protein levels are augmented by estrogen treatment. Moreover, we demonstrated that FOXA1 was necessary for estrogen-induced cell proliferation as FOXA1 knock down inhibited cell cycle entry by blocking cells in G1. Together, our data demonstrate that ERα and Foxa1 cooperate on a defined set of genes and form a domain of estrogen response. By being inducible by estrogens and necessary for hormonal response, we suggest that Foxa1 compartimentalizes estrogen signaling in breast cancer cells. Whether this intriguing cooperation between Foxa1 and ERα occurs only in pathologic stages remains to be elucidated but we suggest that the targeting of Foxa1 or ER-Foxa1 pathways might provide more specific therapies for efficient diminution of breast cancer cell proliferation.

6.2.3 Genome-wide computational prediction of transcriptional regulatory modules and regulation by ERα

Our ChIP-cloning study and the work performed by Carroll et al. on the chromosome 21/22-wide tiling array illustrated well that ERα recruitment is not restricted to promoter regions *in vivo*. Collaboration with bioinformaticians gave us the chance to study ER action on regulatory regions with no bias for promoter regions.

Blanchette et al. described a genome-wide computational prediction of human transcriptional regulatory modules based on the fact that transcription factors rarely work on their own and rather bind to DNA in cooperation with others (Blanchette et al., 2006). While the prediction of individual TFBS is a laborious problem, the DNA footprints of sets of TFBS, called cis-regulatory modules or CRM believed to be constituents of most transcriptional regulatory processes in mammals, represent a novel tactic for identification of regulatory regions (Howard and Davidson, 2004). With the goal of constructing a global map of regulatory modules, Blanchette et al. built an algorithm based on the detection of phylogenetically conserved clusters of TFBS. The algorithm involved the identification and scoring of putative human-mouse-rat-conserved TFBSs from the Transfac position-weight matrix (PWM), followed by the detection of clusters enriched in putative binding sites.

We contributed to the experimental validation of predicted CRM (pCRM) for ERα binding using ChIP-microarray. Among the sequences chosen to be represented by the array, 758 modules were predicted to be bound by ERα. Our ChIP-on-chip validation from MCF7 cells confirmed binding of ERα to about 3% of pCRMs. Since these CRM can be used by a transcription factor in a cell-specific manner, more pCRM are expected to be functional *in vivo*, especially for ERα that plays important roles in many other tissues. However, this fairly low efficiency of *in vivo* confirmation can also illustrate how difficult

it is to perform *in silico* predictions even using complex algorithms reflecting our current knowledge on the constitution of gene regulatory regions.

Isolation of 4 modules was of particular interest in the context of breast cancer cells: 2 modules around the PgR or PR (progesterone receptor) locus, one for NRIP/RIP140 and one around the CALCOCO1 gene. PR, like ER, is an important predictive marker of response to endocrine therapy and estrogens are known to control the expression of PR in breast cancer (Horwitz et al., 1978). We showed that this estrogen response of PR is likely to be regulated through distal regulatory modules discovered in this study and bound by ERα in MCF-7 cells. NRIP (nuclear receptor interacting protein) is a coregulator whose expression is believed to repress ERα activity (Cavaillès et al., 1995). We showed earlier that ERα recruits NRIP on genomic targets, and we provide evidence supporting the regulation of NRIP by ERα, consistent with the observation that short estrogen treatments (30 minutes) increases NRIP mRNA levels in MCF-7 and human ovary cells (Thénot et al., 1999). Interestingly, one module controlled by ERα was near another nuclear receptor coregulator transcriptional start site, CoCoA (coil-coil coactivator) or CALCOCO1. CoCoA is a coregulator that binds the basic-helix-loop-helix/Per-Arnt-Sim (bHLH-PAS) of p160 coactivators thereby enhancing ERα transcriptional activity (Kim et al., 2003). Again, ERα seems to directly regulate, at the genome level, regulators of its own activity.

An important technical breakthrough has recently been achieved in two industries: Agilent and Affymetrix were both able to represent the whole genome on chips thereby providing the tools for binding and mRNA profilings. Dr. Young's group published the first complete genome-wide location analysis using an array constructed by Agilent covering all non-repetitive sequences of the human genome on 115 slides (Lee et al., 2006). Interestingly, the Affymetrix technology achieved complete human genome covering on 14 slides, which allowed the genome-wide location analysis of ER binding sites,

providing a detailed description of its action (Carroll et al., 2006). We can expect that ChIP-on-chip combined with gene expression profiling studies using these whole-genome chips will become more accessible and reveal impressive amount of information regarding general and specific mechanisms of gene expression regulation.

6.3 A functional autoregulatory variant responsive to ERRα-PGC1 complexes

Recent studies have stated that ERRα and PGC1α expression profiles are nearly identical in response to physiological stimuli like exposure to cold and starvation (Ichida et al., 2002). Compatible with these observations, one group has shown that PGC-1α can induce ERRα expression (Schreiber et al., 2003). In chapter IV, we showed that PGC-1α is a potent activator of the ERRα promoter (Laganière et al., 2004). Furthermore, this PGC1α activation of the ERRα promoter is dependent on both the presence of ERRα itself, as well as an ERRE located within the ERRα promoter. With these findings, we proposed a model explaining how ERRα and PGC-1α coexpress in tissues.

While most of the studies for genetic variants affecting an individual's phenotype like the susceptibility to certain diseases or response to treatments yet implicate the coding sequence of genes, it is acknowledged that variations in the gene regulatory regions also contribute to the complexity of human traits (Kleinjan and van Heyningen, 2005). Interestingly, the autoregulatory mechanism of ERRα expression and induction by PGC1α relies on an ERRE present within a 23 bp polymorphic repeat (ESRRA23). Our human genetic screen revealed that the ESRRA23 element was found from 1 to 4 copies on each allele in the human population. While the most common genotype was the presence of 2 copies of the ESRRA23 on each allele, representing more than 85% of the population, other frequent variants were observed, for instance 12% exhibit a 2.3. genotype (2 copies on one allele and 3 copies on the other one).

Other genotype combinations were found at lower frequencies (less than 0.15% each). Interestingly, increasing copy number of the ESRRA23 was associated to a higher activity of PGC1α through ERRα on its own promoter. Therefore, the ESRRA23 functional regulatory polymorphism may provide differential responses to ERRα-PGC1α-controlled physiological pathways. Several studies support a role for ERRα-PGC1α pathways in the regulation of energy metabolism through the induction of mitochondrial biogenesis and the regulation of the oxidative phosphorylation, fatty acid β-oxidation as well as glucose metabolism pathways (Carrier et al., 2004; Huss et al., 2004; Mootha et al., 2004; Mootha et al., 2003; Schreiber et al., 2004; Sladek et al., 1997; Vega et al., 2000; Vega and Kelly, 1997; Wende et al., 2005; Willy et al., 2004). Accordingly, high levels of ERRα-PGC1α expression of were previously associated with an increased energy expanditure that antagonize obesity (Kamei et al., 2003). The ERRα-PGC1α-driven pathways have also been linked to related human diseases such as obesity and type II diabetes (Mootha et al., 2004). Interestingly, the role of the *ESRRA23* polymorphic element was evaluated in the human population and was significantly associated to higher body mass index (BMI) in a study of 703 Japanese people (Kamei et al., 2005). ERRα is expressed at high levels during all stages of bone development and overexpression or inhibition of its expression have effect on the formation of bone nodules in rat calvaria cell culture (Bonnelye et al., 2001). In agreement with these findings, a recent study conducted in humans suggests an association between the *ESRRA23* regulatory polymorphism with bone density: in premenopausal women, an increased copy number of the repeated element correlated with a higher bone mineral density (BMD), a measure indicating the susceptibility to osteoporosis and bone fractures (Laflamme et al., 2005). Therefore, the regulatory polymorphism described in our study may have important roles in the determination of human traits implicating ERRα and ERRα-PGC1α directed pathways.

6.3.1 Autoregulation of ERRα gene expression

We described a human genetic screen that allowed us to discover a polymorphic 23 bp sequence (ESRRA23) in the *ESRRA* gene promoter containing a hormone response element (HRE). The HRE identified actually corresponded to a perfect ERRE consensus sequence that is indeed nicely recognized by ERRα in conventional *in vitro* and reporter assays (Laganière et al., 2004). In this study, ChIP was instrumental to certify that endogenous ERRα from MCF7 breast cancer cells could truly recognize its own promoter *in vivo*, in the region containing the polymorphic element. Considering that one important limitation in the development of ChIPs and the discovery of direct target genes is often hindered by the absence of a known functional target of the factor of interest, publishing the first ERRα ChIP will make more accessible subsequent ChIP studies for our and other laboratories. Since we described the autoregulation of ERRα, we think that ERRα is likely to modulate its expression through this mechanism in most systems and thus be a valid ERRα target gene in the majority of cells and tissues where it is expressed. In addition, the 23 bp sequence containing the responsive element is present in the mouse genome. Therefore, we presume that the discovery of this direct target gene will be of great use for the development of the mouse and human ERRα ChIP-based approaches in various cell lines and tissues.

6.4 ERRα directly controls mitochondrial function

Recent studies revealed that ERRα levels were increased following treatments with cytokines in macrophages (Barish et al., 2005). This observation prompted us, in collaboration with Dr. Evans laboratory, to evaluate its importance in these cells in addition to providing us with a relevant and appealing physiological context for the identification of ERRα direct target genes.

We used a 19,000 promoters array constructed by Dr. Robert laboratory and performed ERRα ChIP-on-chip in mouse macrophages stimulated by the cytokine IFN-γ.

An impressive amount of ERRα targets (11%) were promoters of genes coding for proteins involved in all steps controlling the mitochondrial respiratory chain. Therefore, our genome-wide promoter binding profiling is in agreement with other reports suggesting a role for ERRα in the control of energy homeostasis through mitochondrial functions but reveals an extensive cataloguing of its direct target genes, which had not been previously described in the literature. It also seems that ERRα controls mitochondrial function regardless of the tissue type, a finding that is further supported by other recent experiments in Dr. Giguère's laboratory showing that these pathways are also targeted by ERRα in the mouse heart.

6.5 ERRα in cancer

Recent studies have demonstrated that ERRα expression is a negative prognostic factor for disease-free survival in breast cancer. It was demonstrated that ERRα expression in greater than 10% of malignant cells was associated with a 20% decrease in overall disease-free survival at 13 years, in accordance with the observation that ERRα and HER2 expression positively associate in advanced, tamoxifen-resistant, and/or ERα-negative tumors (Ariazi et al., 2002; Suzuki et al., 2004). ERRα has also been suggested to play a role in ovarian, colorectal and prostate cancers (Cavallini et al., 2005; Cheung et al., 2005; Sun et al., 2005). Although the role of ERRα in these cancers remains to be determined, its expression and association with a negative prognosis suggests that it may be a useful target for therapy.

Current suggested roles for ERRα in breast cancer mostly have to do with its resemblance to ERα. It has been put forward that ERRα could somehow replace ERα action in an estrogen-independent manner, through the

control of the same regulatory regions due to high degree of similarity of their DBDs. This way, ERα target genes would remain expressed and tumors highly proliferative in a hormone-independent manner, i.e. even in the absence of functional ERα, in addition to be resistant to endocrine therapies. However, the first part of ChIP studies performed with the MCF-7 breast cancer model contrasts with this hypothesis. In fact, from the ERα-bound promoters identified, in addition to distal regulatory modules isolated by Carroll et al. on chromosome 21 and 22 and targets isolated by ChIP-cloning, we tested ERRα binding by standard ChIP to over 20 sites. Surprisingly, none were bound by ERRα neither in absence or presence of estrogen (data not shown). The reverse was also true since no ERRα target was bound by ERα in these cells (about 10 tested). Thus, although no genome-wide location analysis have been performed for ERα and ERRα in the same cell type for the exact comparison of their targets, it seems that ERRα functions are less related to that of ERα than anticipated. Despite their similar DBD that allows the recognition of similar response elements *in vitro*, ERα and ERRα seem to have distinct requirements for DNA binding *in vivo*. These observations do not exclude a role for ERRα in breast and other cancers. The fact that ERRα can control mitochondrial function in several systems namely macrophages, heart, and breast cancer cells suggests that attention should be paid to the roles played by these pathways in cancer. It is tempting to suggest an association between ERRα and the production of reactive oxygen species (ROS) in cancer cells. ROS are oxygen-containing chemical species with reactive chemical properties, such as superoxide (O2−), hydroxyl radicals (HO−) and hydrogen peroxide (H2O2). They are generated through a variety of pathways, most importantly during oxidative phosphorylation occurring in mitochondria, which are considered as the major source of ROS in a cell (Richter et al., 1995). Increased ROS generation in cancer cells has been reported (Konstantinov et al., 1987; Szatrowski and Nathan, 1991; Zhou et al., 2003). The augmented amounts of ROS in cancer cells may have important consequences such as stimulation of cellular proliferation, promotion of mutations and genetic instability and

alterations in cellular sensitivity to anticancer agents (Pelicano et al., 2004). Therefore, increased ERRα expression leading to augmentation of oxidative phosphorylation could cause higher ROS production and play a role in breast cancer. Another possibility concerns the observation that ERRα controls cellular movement in Zebrafish. This could be another interesting avenue for studies, since ERRα might regulate other related events in human pathology such as metastasis (Bardet et al., 2005; Bardet et al., 2006). Therefore, albeit ERRα might not influence estrogenic pathways as first expected, ligands currently under development might offer promising avenues in the treatments of some cancers where ERRα is overexpressed.

CONCLUSION

We have successfully developed and used ChIP-derived approaches and significantly contributed to the knowledge of nuclear-receptor driven mechanisms of transcriptional regulation at the genome level. We portrayed previously unexplored regions of the genome that are crucial regulators of the estrogen dependence of breast cancer and ERRα's predominant action in mitochondrial function. We have elucidated novel aspects of ER and ERR functions notably the genomic pathways they control, the complexity of their regulatory feedback mechanisms, their collaboration with transcriptional partners as well as providing insights into their action in physiology and pathology.

REFERENCES

Afonja, O., Raaka, B. M., Huang, A., Das, S., Zhao, X., Helmer, E., Juste, D., and Samuels, H. H. (2002). RAR agonists stimulate SOX9 gene expression in breast cancer cell lines: evidence for a role in retinoid-mediated growth inhibition. Oncogene *21*, 7850-7860.

Andrade, P. M., Silva, I. D., Borra, R. C., de Lima, G. R., and Baracat, E. C. (2002). Estrogen regulation of uterine genes in vivo detected by complementary DNA array. Horm Metab Res *34*, 238-244.

Ariazi, E. A., Clark, G. M., and Mertz, J. E. (2002). Estrogen-related receptor α and estrogen-related receptor γ associate with unfavorable and favorable biomarkers, respectively, in human breast cancer. Cancer Res *62*, 6510-6518.

Bardet, P. L., Horard, B., Laudet, V., and Vanacker, J. M. (2005). The ERRalpha orphan nuclear receptor controls morphogenetic movements during zebrafish gastrulation. Dev Biol *281*, 102-111.

Bardet, P. L., Laudet, V., and Vanacker, J. M. (2006). Studying non-mammalian models? Not a fool's ERRand! Trends Endocrinol Metab *17*, 166-171.

Barish, G. D., Downes, M., Alaynick, W. A., Yu, R. T., Ocampo, C. B., Bookout, A. L., Mangelsdorf, D. J., and Evans, R. M. (2005). A Nuclear Receptor Atlas: macrophage activation. Mol Endocrinol *19*, 2466-2477.

Berry, M., Nunez, A.-M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci U S A *86*, 1218-1222.

Bertucci, F., Houlgatte, R., Benziane, A., Granjeaud, S., Adelaide, J., Tagett, R., Loriod, B., Jacquemier, J., Viens, P., Jordan, B., *et al.* (2000). Gene expression profiling of primary breast carcinomas using arrays of candidate genes. Hum Mol Genet *9*, 2981-2991.

Blanchette, M., Bataille, A. R., Chen, X., Poitras, C., Laganiere, J., Lefebvre, C., Deblois, G., Giguere, V., Ferretti, V., Bergeron, D., *et al.* (2006). Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res 16, 656-668.

Bonnelye, E., Merdad, L., Kung, V., and Aubin, J. E. (2001). The orphan nuclear estrogen receptor-related receptor α (ERRα) is expressed throughout osteoblast differentiation and regulates bone formation in vitro. J Cell Biol *153*, 971-984.

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.

Cam, H., Balciunaite, E., Blais, A., Spektor, A., Scarpulla, R. C., Young, R., Kluger, Y., and Dynlacht, B. D. (2004). A common set of gene regulatory networks links metabolism and growth inhibition. Mol Cell *16*, 399-411.

Carrier, J. C., Deblois, G., Champigny, C., Levy, E., and Giguere, V. (2004). Estrogen related-receptor α (ERRα) is a transcriptional regulator of apolipoprotein A-IV and controls lipid handling in the intestine. J Biol Chem *279*, 52052-52058.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., *et al.* (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33-43.

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., *et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. Nat Genet *38*, 1289-1297.

Cavaillès, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P. J., and Parker, M. G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. European Molecular Biology Organization Journal *14*, 3741-3751.

Cavallini, A., Notarnicola, M., Giannini, R., Montemurro, S., Lorusso, D., Visconti, A., Minervini, F., and Caruso, M. G. (2005). Oestrogen receptor-related receptor alpha (ERRalpha) and oestrogen receptors (ERalpha and ERbeta) exhibit different gene expression in human colorectal tumour progression. Eur J Cancer *41*, 1487-1494.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., *et al.* (2004). Unbiased mapping of transcription factor binding sites along human

chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.

Charpentier, A. H., Bednarek, A. K., Daniel, R. L., Hawkins, K. A., Laflin, K. J., Gaddis, S., MacLeod, M. C., and Aldaz, C. M. (2000). Effects of estrogen on global gene expression: identification of novel targets of estrogen action. Cancer Res *60*, 5977-5983.

Cheung, C. P., Yu, S., Wong, K. B., Chan, L. W., Lai, F. M., Wang, X., Suetsugi, M., Chen, S., and Chan, F. L. (2005). Expression and functional study of estrogen receptor-related receptors in human prostatic cells and tissues. J Clin Endocrinol Metab *90*, 1830-1844.

Cicatiello, L., Scafoglio, C., Altucci, L., Cancemi, M., Natoli, G., Facchiano, A., Iazzetti, G., Calogero, R., Biglia, N., De Bortoli, M., *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.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell *9*, 279-289.

Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature *364*, 412-420.

Elgort, M. G., Zou, A., Marschke, K. B., and Allegretto, E. A. (1996). Estrogen and estrogen receptor antagonists stimulate transcription from the human retinoic acid receptor-α 1 promoter via a novel sequence. Mol Endocrinol *10*, 477-487.

Elrick, L. J., Jorgensen, H. G., Mountford, J. C., and Holyoake, T. L. (2005). Punish the parent not the progeny. Blood *105*, 1862-1866.

Euskirchen, G., Royce, T. E., Bertone, P., Martone, R., Rinn, J. L., Nelson, F. K., Sayward, F., Luscombe, N. M., Miller, P., Gerstein, M., *et al.* (2004). CREB binds to multiple loci on human chromosome 22. Mol Cell Biol *24*, 3804-3814.

Finlin, B. S., Gau, C. L., Murphy, G. A., Shao, H., Kimel, T., Seitz, R. S., Chiu, Y. F., Botstein, D., Brown, P. O., Der, C. J., *et al.* (2001). RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. J Biol Chem *276*, 42259-42267.

Fitzgerald, P., Teng, M., Chandraratna, R. A. S., Heyman, R. A., and Allegretto, E. A. (1997). Retinoic acid receptor a expression correlates with retinoid-induced growth inhibition of human breast cancer cells regardless of estrogen receptor status. Cancer Res *57*, 2642-2650.

Frasor, J., Danes, J. M., Komm, B., Chang, K. C., Lyttle, C. R., and Katzenellenbogen, B. S. (2003). 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 *144*, 4562-4574.

Friedman, J. R., and Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci *63*, 2317-2328.

Gao, N., Zhang, J., Rao, M. A., Case, T. C., Mirosevich, J., Wang, Y., Jin, R., Gupta, A., Rennie, P. S., and Matusik, R. J. (2003). The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. Mol Endocrinol *17*, 1484-1507.

Ghosh, M. G., Thompson, D. A., and Weigel, R. J. (2000). PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res *60*, 6367-6375.

Giguère, V. (1999). Orphan nuclear receptors: from gene to function. Endocr Rev *20*, 689-725.

Giguère, V. (2002). To ERR in the estrogen pathway. Trends in Endocrinology and Metabolism *13*, 220-225.

Gonzalez, F. J., and Gelboin, H. V. (1994). Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. Drug Metab Rev *26*, 165-183.

Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Ferno, M., Peterson, C., and Meltzer, P. S. (2001). Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res *61*, 5979-5984.

Harrington, W. R., Sengupta, S., and Katzenellenbogen, B. S. (2006). Estrogen regulation of the glucuronidation enzyme UGT2B15 in estrogen receptor-positive breast cancer cells. Endocrinology *147*, 3843-3850.

Holmqvist, P. H., Belikov, S., Zaret, K. S., and Wrange, O. (2005). FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. Exp Cell Res *304*, 593-603.

Horak, C. E., Mahajan, M. C., Luscombe, N. M., Gerstein, M., Weissman, S. M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci U S A *99*, 2924-2929.

Horwitz, K. B., Koseki, Y., and McGuire, W. L. (1978). Estrogen control of progesterone receptor in human breast cancer: role of estradiol and antiestrogen. Endocrinology *103*, 1742-1751.

Howard, M. L., and Davidson, E. H. (2004). cis-Regulatory control circuits in development. Dev Biol *271*, 109-118.

Huss, J. M., Pineda Torra, I., Staels, B., Giguere, V., and Kelly, D. P. (2004). Estrogen-related receptor a directs peroxisome proliferator-activated receptor a signaling in the transcriptional control of energy metabolism in cardiac and skeletal muscle. Mol Cell Biol *24*, 9079-9091.

Ichida, M., Nemoto, S., and Finkel, T. (2002). Identification of a specific molecular repressor of the peroxisome proliferator-activated peceptor γ coactivator-1α (PGC-α). J Biol Chem *277*, 50991-50995.

Inadera, H., Hashimoto, S., Dong, H. Y., Suzuki, T., Nagai, S., Yamashita, T., Toyoda, N., and Matsushima, K. (2000). WISP-2 as a novel estrogen-responsive gene in human breast cancer cells. Biochem Biophys Res Commun *275*, 108-114.

Inoue, A., Yoshida, N., Omoto, Y., Oguchi, S., Yamori, T., Kiyama, R., and Hayashi, S. (2002). Development of cDNA microarray for expression profiling of estrogen-responsive genes. J Mol Endocrinol *29*, 175-192.

Inoue, H., Tanabe, T., and Umesono, K. (2000). Feedback control of cyclooxygenase-2 expression through PPARγ. J Biol Chem *275*, 28028-28032.

Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Johansson, L., Thomsen, J. S., Damdimopoulos, A. E., Spyrou, G., Gustafsson, J., and Treuter, E. (1999). The orphan nuclear receptor SHP inhibits agonist-

dependent transcriptional activity of estrogen receptors ERα and ERβ. J Biol Chem *274*, 345-353.

Kamei, Y., Lwin, H., Saito, K., Yokoyama, T., Yoshiike, N., Ezaki, O., and Tanaka, H. (2005). The 2.3 genotype of ESRRA23 of the ERR alpha gene is associated with a higher BMI than the 2.2 genotype. Obes Res *13*, 1843-1844.

Kamei, Y., Ohizumi, H., Fujitani, Y., Nemoto, T., Tanaka, T., Takahashi, N., Kawada, T., Miyoshi, M., Ezaki, O., and Kakizuka, A. (2003). PPARγ coactivator 1β/ERR ligand 1 is an ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. Proc Natl Acad Sci U S A *100*, 12378-12383.

Kim, J. H., Li, H., and Stallcup, M. R. (2003). CoCoA, a nuclear receptor coactivator which acts through an N-terminal activation domain of p160 coactivators. Mol Cell *12*, 1537-1549.

Kleinjan, D. A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet *76*, 8-32.

Konstantinov, A. A., Peskin, A. V., Popova, E., Khomutov, G. B., and Ruuge, E. K. (1987). Superoxide generation by the respiratory chain of tumor mitochondria. Biochim Biophys Acta *894*, 1-10.

Kressler, D., Schreiber, S. N., Knutti, D., and Kralli, A. (2002). The PGC-1-related protein PERC is a selective coactivator of estrogen receptor α. J Biol Chem *277*, 13918-13925.

Lacroix, M., and Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. Mol Cell Endocrinol *219*, 1-7.

Laflamme, N., Giroux, S., Loredo-Osti, J. C., Elfassihi, L., Dodin, S., Blanchet, C., Morgan, K., Giguere, V., and Rousseau, F. (2005). A frequent regulatory variant of the estrogen-related receptor α gene associated with BMD in French-Canadian premenopausal women. J Bone Miner Res *20*, 938-944.

Laganière, J., Deblois, G., and Giguère, V. (2003). Nuclear receptor target gene discovery using high throughput chromatin immunoprecipitation, In Methods in Enzymology, D. W. Russell, and D. J. Mangelsdorf, eds. (San Diego: Academic Press), pp. 339-350.

Laganière, J., Deblois, G., and Giguère, V. (2005a). Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor a1 gene in breast cancer cells. Mol Endocrinol *19*, 1584-1592.

Laganière, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F., and Giguère, V. (2005b). Location analysis of estrogen receptor a target promoters reveals that FOXA1 defines a domain of the estrogen response. Proc Natl Acad Sci U S A *102*, 11651-11656.

Laganiere, J., and Giguere, V. (2006). [Decoding the mode of action of the estrogen receptor through functional genomics]. Bull Cancer *93*, 883-887.

Laganière, J., Tremblay, G. B., Dufour, C. R., Giroux, S., Rousseau, F., and Giguere, V. (2004). A polymorphic autoregulatory hormone response element in the human estrogen related receptor α (ERRα) promoter dictates PGC-1α control of ERRα expression. J Biol Chem *279*, 18504-18510.

Lai, K., Harnish, D. C., and Evans, M. J. (2003). Estrogen receptor alpha regulates expression of the orphan receptor small heterodimer partner. J Biol Chem *278*, 36418-36429.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Liehr, J. G., Fang, W. F., Sirbasku, D. A., and Ari-Ulubelen, A. (1986). Carcinogenicity of catechol estrogens in Syrian hamsters. J Steroid Biochem *24*, 353-356.

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., Ramasamy, A., *et al.* (2004). Discovery of estrogen receptor a target genes and response elements in breast tumor cells. Genome Biology *5*, R66.

Lu, M., Mira-y-Lopez, R., Nakajo, S., Nakaya, K., and Jing, Y. (2005). Expression of estrogen receptor alpha, retinoic acid receptor alpha and cellular retinoic acid binding protein II genes is coordinately regulated in human breast cancer cells. Oncogene *24*, 4362-4369.

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T. E., Luscombe, N. M., Rinn, J. L., Nelson, F. K., Miller, P., Gerstein, M., *et al.* (2003).

Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A *100*, 12247-12252.

Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751-763.

Mootha, V. K., Handschin, C., Arlow, D., Xie, X., St Pierre, J., Sihag, S., Yang, W., Altshuler, D., Puigserver, P., Patterson, N., *et al.* (2004). Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. Proc Natl Acad Sci U S A *101*, 6570-6575.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 267-273.

Murray, G. I. (2000). The role of cytochrome P450 in tumour development and progression and its potential in therapy. J Pathol *192*, 419-426.

Nowell, S. A., Ahn, J., Rae, J. M., Scheys, J. O., Trovato, A., Sweeney, C., MacLeod, S. L., Kadlubar, F. F., and Ambrosone, C. B. (2005). Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients. Breast Cancer Res Treat *91*, 249-258.

Nussbaumer, P., and Billich, A. (2004). Steroid sulfatase inhibitors. Med Res Rev *24*, 529-576.

Nusse, R., and Varmus, H. E. (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell *31*, 99-109.

Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., *et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Park, S., Shimizu, C., Shimoyama, T., Takeda, M., Ando, M., Kohno, T., Katsumata, N., Kang, Y. K., Nishio, K., and Fujiwara, Y. (2006). Gene expression profiling of ATP-binding cassette (ABC) transporters as a predictor

of the pathologic response to neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Res Treat *99*, 9-17.

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.

Pelicano, H., Carney, D., and Huang, P. (2004). ROS stress in cancer cells and therapeutic implications. Drug Resist Updat *7*, 97-110.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Prange-Kiel, J., Rune, G. M., Zwirner, M., Wallwiener, D., and Kiesel, L. (2001). Regulation of estrogen receptor alpha and progesterone receptor (isoform A and B) expression in cultured human endometrial cells. Exp Clin Endocrinol Diabetes *109*, 231-237.

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.

Reese, J., Das, S. K., Paria, B. C., Lim, H., Song, H., Matsumoto, H., Knudtson, K. L., DuBois, R. N., and Dey, S. K. (2001). Global gene expression analysis to identify molecular markers of uterine receptivity and embryo implantation. J Biol Chem *276*, 44137-44145.

Reid, G., Denger, S., Kos, M., and Gannon, F. (2002). Human estrogen receptor-alpha: regulation by synthesis, modification and degradation. Cell Mol Life Sci *59*, 821-831.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., and Dynlacht, B. D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev *16*, 245-256.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Ress, A., and Moelling, K. (2005). Bcr is a negative regulator of the Wnt signalling pathway. EMBO Rep *6*, 1095-1100.

Richter, C., Gogvadze, V., Laffranchi, R., Schlapbach, R., Schweizer, M., Suter, M., Walter, P., and Yaffee, M. (1995). Oxidants in mitochondria: from physiology to diseases. Biochim Biophys Acta *1271*, 67-74.

Roberts, M., Wallace, J., Jeltsch, J. M., and Berry, M. (1988). The 5' flanking region of the human pS2 gene mediates its transcriptional activation by estrogen in MCF-7 cells. Biochemical Biophysical Research Communication *151*, 306-313.

Robyr, D., Gegonne, A., Wolffe, A. P., and Wahli, W. (2000). Determinants of vitellogenin B1 promoter architecture. HNF3 and estrogen responsive transcription within chromatin. J Biol Chem *275*, 28291-28300.

Ross, D. T., and Perou, C. M. (2001). A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. Dis Markers *17*, 99-109.

Safe, S. (2001). Transcriptional activation of genes by 17 beta-estradiol through estrogen receptor-Sp1 interactions. Vitam Horm *62*, 231-252.

Safe, S., and Kim, K. (2004). Nuclear receptor-mediated transactivation through interaction with Sp proteins. Prog Nucleic Acid Res Mol Biol *77*, 1-36.

Schreiber, S. N., Emter, R., Hock, M. B., Knutti, D., Cardenas, J., Podvinec, M., Oakeley, E. J., and Kralli, A. (2004). The estrogen-related receptor alpha (ERRα) functions in PPARγ coactivator 1α (PGC-1α)-induced mitochondrial biogenesis. Proc Natl Acad Sci U S A *101*, 6472-6477.

Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T., and Kralli, A. (2003). The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor ERRα. J Biol Chem *278*, 9013-9018.

Seo, H. S., Larsimont, D., Querton, G., El Khissiin, A., Laios, I., Legros, N., and Leclercq, G. (1998). Estrogenic and anti-estrogenic regulation of estrogen receptor in MCF-7 breast-cancer cells: comparison of immunocytochemical data with biochemical measurements. Int J Cancer *78*, 760-765.

Seth, P., Krop, I., Porter, D., and Polyak, K. (2002). Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene *21*, 836-843.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell *103*, 843-852.

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N. J., Volkert, T. L., Wyrick, J. J., Zeitlinger, J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. Cell *106*, 697-708.

Sladek, R., Bader, J.-A., and Giguère, V. (1997). The orphan nuclear receptor estrogen-related receptor α is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. Mol Cell Biol *17*, 5400-5409.

Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A *98*, 10869-10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., *et al.* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A *100*, 8418-8423.

Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L., and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A *100*, 10393-10398.

Soulez, M., and Parker, M. G. (2001a). Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. J Mol Endocrinol *27*, 259-274.

Soulez, M., and Parker, M. G. (2001b). Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. J Mol Endocrinol *27*, 259-274.

Sun, J. M., Chen, H. Y., and Davie, J. R. (2001). Effect of estradiol on histone acetylation dynamics in human breast cancer cells. J Biol Chem *276*, 49435-49442.

Sun, P., Sehouli, J., Denkert, C., Mustea, A., Konsgen, D., Koch, I., Wei, L., and Lichtenegger, W. (2005). Expression of estrogen receptor-related receptors, a subfamily of orphan nuclear receptors, as new tumor biomarkers in ovarian cancer cells. J Mol Med *83*, 457-467.

Suzuki, T., Miki, Y., Moriya, T., Shimada, N., Ishida, T., Hirakawa, H., Ohuchi, N., and Sasano, H. (2004). Estrogen-related receptor α in human breast carcinoma as a potent prognostic factor. Cancer Res *64*, 4670-4676.

Szatrowski, T. P., and Nathan, C. F. (1991). Production of large amounts of hydrogen peroxide by human tumor cells. Cancer Res *51*, 794-798.

Terasaka, S., Aita, Y., Inoue, A., Hayashi, S., Nishigaki, M., Aoyagi, K., Sasaki, H., Wada-Kiyama, Y., Sakuma, Y., Akaba, S., *et al.* (2004). Using a customized DNA microarray for expression profiling of the estrogen-responsive genes to evaluate estrogen activity among natural estrogens and industrial chemicals. Environ Health Perspect *112*, 773-781.

Thénot, S., Charpin, M., Bonnet, S., and Cavaillès, V. (1999). Estrogen receptor cofactors expression in breast and endometrial human cancer cells. Mol Cell Endocrinol *156*, 85-93.

Tsuchiya, Y., Nakajima, M., Kyo, S., Kanaya, T., Inoue, M., and Yokoi, T. (2004). Human CYP1B1 is regulated by estradiol via estrogen receptor. Cancer Res *64*, 3119-3125.

Tsuchiya, Y., Nakajima, M., and Yokoi, T. (2005). Cytochrome P450-mediated metabolism of estrogens and its regulation in human. Cancer Lett *227*, 115-124.

Tukey, R. H., and Strassburg, C. P. (2000). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. Annu Rev Pharmacol Toxicol *40*, 581-616.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

van der Leede, B. J., Folkers, G. E., van den Brink, C. E., van der Saag, P. T., and van der Burg, B. (1995). Retinoic acid receptor a1 isoform is induced by estradiol and confers retinoic acid sensitivity in human breast cancer cells. Mol Cell Endocrinol *109*, 77-86.

Vega, R. B., Huss, J. M., and Kelly, D. P. (2000). The coactivator PGC-1 cooperates with peroxisome proliferator-activated receptor α in transcriptional control of nuclear genes encoding mitochondrial fatty acid oxidation enzymes. Mol Cell Biol *20*, 1868-1876.

Vega, R. B., and Kelly, D. P. (1997). A role for estrogen-related receptor α in the control of mitochondrial fatty acid β-oxidation during brown adipocyte differentiation. J Biol Chem *272*, 31693-31699.

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E., and Jackle, H. (1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. Cell *57*, 645-658.

Weinmann, A. S., Bartley, S. M., Zhang, T., Zhang, M. Q., and Farnham, P. J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol *21*, 6820-6832.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H., and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev *16*, 235-244.

Wende, A. R., Huss, J. M., Schaeffer, P. J., Giguere, V., and Kelly, D. P. (2005). PGC-1alpha coactivates PDK4 gene expression via the orphan nuclear receptor ERRalpha: a mechanism for transcriptional control of muscle glucose metabolism. Mol Cell Biol *25*, 10684-10694.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Jr., Marks, J. R., and Nevins, J. R. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A *98*, 11462-11467.

Willy, P. J., Murray, I. R., Qian, J., Busch, B. B., Stevens, W. C., Jr., Martin, R., Mohan, R., Zhou, S., Ordentlich, P., Wei, P., *et al.* (2004). Regulation of PPARgamma coactivator 1α (PGC-1α) signaling by an estrogen-related receptor α (ERRα) ligand. Proc Natl Acad Sci U S A *101*, 8912-8917.

Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., and Aparicio, O. M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E., and Kaestner, K. H. (2005). Foxa2 integrates the transcriptional response of the hepatocyte to fasting. Cell Metab *2*, 141-148.

Zhou, Y., Hileman, E. O., Plunkett, W., Keating, M. J., and Huang, P. (2003). Free radical stress in chronic lymphocytic leukemia cells and its role in cellular sensitivity to ROS-generating anticancer agents. Blood *101*, 4098-4104.

CHAPTER VII: Contribution to Original Research

Chapter II:

- I have defined 11 novel functional binding sites of ERα in breast cancer cells.

- I have shown that an intronic ERE controls RARα expression, providing a new mechanism for its regulation by estrogens, as well as highlighting that ERα action is not restricted to promoter regions.

Chapter III:

- I have identified over 150 direct target promoters bound by ERα *in vivo* in breast cancer cells, which:

    1. significantly expands our comprehension of the primary pathways monitored by estrogens in breast cancer.
    2. provides a great number of *bona fide* target regulatory regions for the study of ERα various mechanisms of action on specific genes.

- I have shown an essential role of FOXA1 for the recruitment of ERα to a subset of promoters, in addition to being regulated by estrogens, which implies that:

    1. FOXA1 cooperates with ERα in breast cancer cells.
    2. FOXA1 is a licensing factor that compartmentalizes the hormonal response.

Chapter IV:

- I have described an autoregulatory polymorphism in the *ESRRA* promoter.

- I have shown that ERRα controls its own regulation through direct binding to its own promoter, through the polymorphic element.

Chapter V:

- I have described over 100 genuine targets for ERRα in mouse macrophages

- I have provided data supporting an important role for ERRα in mitochondrial function

APPENDIX 1

[19] Nuclear Receptor Target Gene Discovery Using High-Throughput Chromatin Immunoprecipitation

By JOSÉE LAGANIÈRE, GENEVIÈVE DEBLOIS, and VINCENT GIGUÈRE

Introduction

Nuclear receptors are master transcription factors that regulate the development, physiology, and homeostasis of whole organisms through direct control of gene expression in response to diverse ligands and hormonal stimuli.[1] Nuclear receptors regulate the expression of their target genes through association with specific DNA regulatory elements.[2] While a

[1] D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and R. M. Evans, *Cell* **83**, 835 (1995).
[2] C. K. Glass, *Endocr. Rev.* **15**, 391 (1994).

0076-6879/2003 $35.00

significant number of nuclear receptor target genes have been identified to date, it is believed that these genes represent only a small fraction of the regulatory units likely to be under the control of nuclear receptors. Most nuclear receptor target genes identified so far were characterized through "gene oriented" approaches that study the regulation of one candidate gene at a time, and these studies are usually limited to the promoter region. However, in order to understand the complex nuclear receptor-driven transcriptional networks that operate in a living organism, a whole genome approach is required and now feasible. Here, we describe a powerful "nuclear receptor/whole genome-oriented" approach to identify and more accurately study nuclear receptor regulatory networks.

Strategy

We took advantage of the recent advances in chromatin immunoprecipitation (ChIP)[3,4] and associated cloning procedures[5–7] to develop a high-throughput ChIP technique to identify primary nuclear receptor target genes. Nuclear receptors associate with the regulatory elements of their potential target genes leading to transcriptionally active chromatin. Formaldehyde can be used to cross-link the receptors and associated cofactors bound to DNA in living cells. Following isolation and fragmentation, the chromatin is immunoprecipitated using a specific antibody raised against the nuclear receptor of interest. Cross-link reversal and DNA purification is then performed, and the isolated fragments are cloned in a suitable vector for sequencing. Bioinformatic analysis of the fragments obtained is performed using the recently available human genome databases to localize the isolated regulatory elements and thus, identify associated target genes. The strategy is outlined in Fig. 1.

We have adapted a high throughput ChIP technique to clone target sequences bound by the estrogen receptor α (NR3A1, ERα) in the MCF-7 cell line upon estradiol (E_2) stimulation. This technique, used for the isolation of regulatory elements bound by the ERα *in vivo*, can be applied to other sources of materials (cell lines or dissociated primary cells) and other classic or orphan nuclear receptors.

[3] V. Orlando, *Trends Biochem. Sci.* **25**, 99 (2000).
[4] Y. Shang, X. Hu, J. DiRenzo, M. A. Lazar, and M. Brown, *Cell* **103**, 843 (2000).
[5] J. Wells and P. J. Farnham, *Methods* **26**, 48 (2002).
[6] A. S. Weinmann, S. M. Bartley, T. Zhang, M. Q. Zhang, and P. J. Farnham, *Mol. Cell. Biol.* **21**, 6820 (2001).
[7] A. S. Weinmann, P. S. Yan, M. J. Oberley, T. H. Huang, and P. J. Farnham, *Genes Dev.* **16**, 235 (2002).

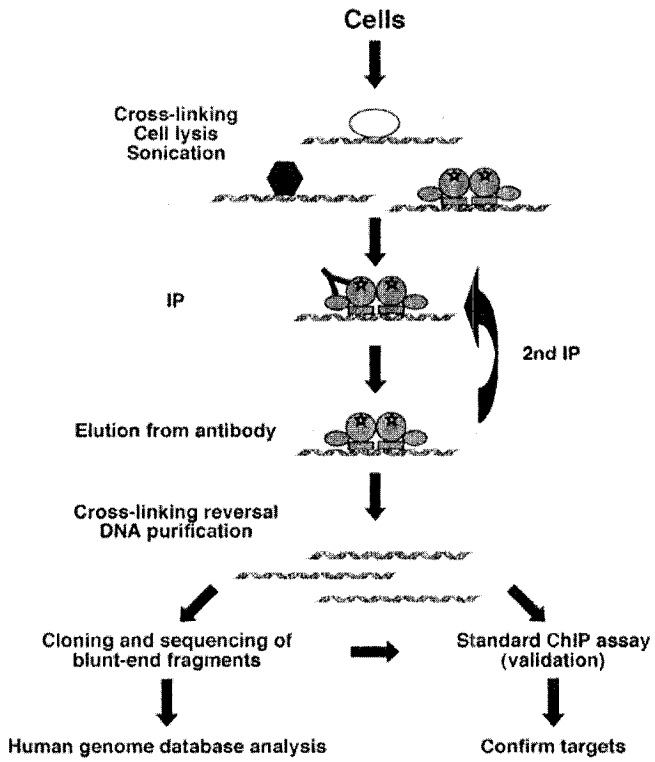


FIG. 1. Use of ChIP for the isolation of nuclear receptors DNA regulatory sequences. Following a cross-linking step, the cells are lysed and the chromatin is fragmented by sonication. A specific antibody directed against the nuclear receptor of interest is used in two sequential IP experiments to improve the purity of the complexes. After cross-link reversal and DNA purification, the isolated fragments are repaired, cloned, sequenced, and localized in the human genome database. For new target validation, an independent standard ChIP assay with PCR primers specific for the newly isolated sequence is performed to assess the actual binding efficiency of this sequence to the receptor. Isolated DNA fragments are subcloned into a reporter plasmid and assayed in transient transfection to confirm their regulatory properties. Quantitative RT-PCR is performed in the same cell line used for the ChIP assay, and analyzed for changes in levels of expression of a candidate target gene to confirm the transcriptional regulatory role of the receptor.

Cell Culture and Chromatin Preparation

To study ERα association with regulatory DNA sequences upon E_2 stimulation, we used the ER-positive human breast carcinoma cell line MCF-7. Approximately, 1.5×10^7 cells were used for a single ChIP on endogenous ERα. Cells were routinely cultured in phenol-red-free Dulbecco's

Minimal Essential Medium (DMEM) as previously described.[8] Seventy-two hours prior to chromatin extraction, the media containing 10% complete fetal bovine serum (FBS) was replaced by phenol red-free DMEM supplemented with 10% charcoal–dextran treated FBS (steroid deprived). Prior to chromatin extraction, the cells are treated with either 10^{-7} *M* E_2 (Sigma, St. Louis, MO) or vehicule (ethanol) for 30–45 min.[4]

When studying the association between orphan nuclear receptors and their target regulatory sequences, no addition of ligand is required prior to performing the ChIP, and the cells can be routinely cultured in DMEM supplemented with 10% complete serum. However, in the event that cross-talk in gene regulation between an orphan nuclear receptor and a ligand-inducible nuclear receptor is suspected, the cells should be maintained in a manner not to interfere with specific nuclear receptor signaling, such as in a steroid-deprived serum.

Following E_2 treatment, bound proteins are immediately cross-linked to DNA upon addition of formaldehyde directly into the medium to a final concentration of 1%. After 10 min of incubation at room temperature on a shaking platform, the cross-linking reaction is stopped by addition of glycine to a final concentration of 0.125 *M*, and incubated for 5 more minutes at room temperature as previously described.[7] The duration of the cross-linking reaction needs to be accurately monitored since extensive exposure of the cells to formaldehyde can lead to a decreased yield in chromatin isolation and to poor immunoprecipitation (IP) efficiency. Following the cross-linking step, the cells are washed twice with ice-cold phosphate-buffered saline (PBS) and scraped in ice-cold PBS.

Chromatin preparation is carried out as follows. The PBS-harvested cross-linked cells are centrifuged at 1500 rpm for 10 min at 4°C, the pellet is resuspended in a suitable volume (200 μl per 1.5×10^7) of lysis buffer [1% SDS, 10 m*M* EDTA, 50 m*M* Tris–HCl, pH 8.1, supplemented with complete, mini, EDTA-free protease inhibitor cocktail (Roche, Molecular Biochemicals, Indianapolis, IN)] and incubated on ice for 10 min. The lysates are sonicated five times at setting 10 with the sonicator (model Virsonic 60, Virtis, NY) for 7–8 sec each time. This step needs to be optimized according to the sonicator and the cell line used. At this point, it is important to verify the length of the chromatin fragments resulting from the sonication step. An aliquot of the sonicated lysates is removed, incubated at 65°C for at least 6 hr to reverse the formaldehyde cross-linking,

[8] G. B. Tremblay, A. Tremblay, N. G. Copeland, D. J. Gilbert, N. A. Jenkins, F. Labrie, and V. Giguère, *Mol. Endocrinol.* **11**, 353 (1997).

purified using QIAquick spin PCR purification kit (Qiagen, CA), and analyzed on a 1% agarose gel. For a standard ChIP assay and for cloning ChIP-obtained DNA fragments, the average size of the DNA fragments should be approximately 1000 bp.[7] While this verification step is in progress, the remaining sonicated lysates can be frozen at −80°C. Further sonication steps can be carried out to obtain the desired fragment length. At this point, the sonicated lysates are centrifuged for 10 min, the supernatants collected and diluted 10 × in ChIP dilution buffer (0.5% Triton X-100, 2 m*M* EDTA, 150 m*M* NaCl, 20 m*M* HEPES, pH 8) to achieve a final SDS concentration of 0.1%.

Chromatin Immunoprecipitation

Prior to performing the actual ChIP, it is important to remove and freeze an aliquot of the diluted fragmented chromatin corresponding to 10% of the total amount used for one IP. The antibodies (Ab) used to perform ChIP on endogenous ERα in MCF-7 cells were αERα Ab-1 (Neomarker, Fremont, CA) or αERα HC-20 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA).[4] The Abs must have previously been shown to be suitable for IP. The ChIP procedure is essentially carried out as previously described.[4] Immunoclearing of the diluted chromatin is achieved by incubating for 1 hr at 4°C with 40 μl of salmon sperm DNA–protein A agarose (Upstate Biotechnology, Inc., Lake Placid, NY; provided as 50% slurry in 10 m*M* Tris–HCl, pH 8, 1 m*M* EDTA, 0.05% sodium azide). IP is performed overnight at 4°C using a specific antibody. Following IP, 40 μl of salmon sperm DNA–protein A agarose is added and the incubation is pursued for 2 more hours. The precipitates are washed sequentially for 10 min each with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 m*M* EDTA, 20 m*M* Tris–HCl, pH 8.1, 150 m*M* NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 m*M* EDTA, 20 m*M* Tris–HCl, pH 8.1, 500 m*M* NaCl), and LiCl wash buffer (0.25 m*M* LiCl, 1% NP-40, 1% deoxycholate, 1 m*M* EDTA, 10 m*M* Tris–HCl, pH 8.1). Precipitates are then washed three times with TE buffer and eluted twice for 15–30 min each time on vortex set at 3 with 75 μl of elution buffer (1% SDS, 0.1 *M* $NaHCO_3$).[7] The pooled eluates and the input are incubated at 65°C for at least 6 hr to reverse the formaldehyde cross-linking. The isolated DNA fragments are then purified according to the QIAquick Spin Kit protocol (Qiagen, CA) and used for quantitative PCR analysis of the ChIP assay. It is advisable to validate that the first ChIP was successful and specific by standard ChIP assay of a known target sequence (refer to section "Validation of the ChIP Experiment (Positive control)" and Fig. 2), and

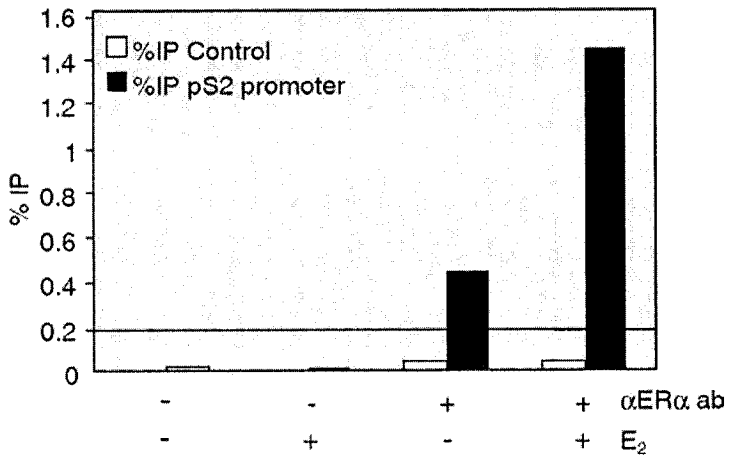


FIG. 2. Standard ChIP assay for ERα in MCF-7 cells on the pS2 promoter. MCF-7 cells were treated with estradiol or vehicle for 45 min and submitted to standard chromatin IP with ERα antibody. The purified DNA was used as a template for quantitative PCR amplification of the pS2 promoter and for a 5-kb upstream region (Control). The percentage of IP (%IP) was determined by comparing with a 2% input sample. Any %IP below 0.2% is considered nonsignificant. The pS2 promoter can be efficiently immunoprecipitated by an ERα antibody in E_2-treated cells, and to a lower extent in nontreated cells.

that the criteria for standard ChIP discussed in this section are achieved before beginning the cloning procedure.

Cloning of Fragments Isolated by ChIP

For cloning the fragments, we suggest that a second purification step be performed; cloning subsequently to a single ChIP may lead to the isolation of a considerable amount of clones containing nonspecific sequences, such as repeat DNA. The first ChIP is carried out as described earlier, with the exception that the eluates are pooled, rediluted 10 × in ChIP dilution buffer, and reimmunoprecipitated using the same Ab for ERα prior to the reversal of the cross-link. The use of a different Ab (if available) for the second IP is a good way to increase the stringency of the technique, further reducing the isolation of nonspecific sequences. The subsequent steps for the second ChIP are carried out as described for the first ChIP.

The ends of the isolated fragments from the double ChIP procedure are then repaired with T4 DNA polymerase (MBI Fermentas Inc., ON) and repurified with the QIAquick Spin Kit (Qiagen, CA). The blunt fragments are then cloned into a suitable vector [such as the Ready-to-go pUC18 *Sma*I/BAP + Ligase kit (Pharmacia, NJ)] for further sequencing analysis.

Following transformation of the ligation reaction and DNA purification, inserts are sequenced using an automated sequencer and appropriate primers.

Criteria for the Analysis of the Isolated Sequences

The cloned sequences are analyzed using BLASTn search on the private CELERA human genome database (http://www.celeradiscoverysystem.com) or the public database (http://www.ncbi.nlm.nih.gov/BLAST/). Several criteria are taken into consideration for the analysis and validation of the sequences obtained.

1. The sequence isolated has to be long enough (> 300 bp) to facilitate following sequence analysis.
2. A perfect match has to be obtained between the isolated sequence and the genome sequence to ensure the localization specified by the search.
3. The BLASTn result has to be unique among the genome, since a sequence leading to multiple blast results could represent a repeat that was isolated in a nonspecific manner.
4. Localization to the closest annotated gene from the isolated sequence is the next criteria to consider. Immediate proximity ($< \sim 1$ kb) of the isolated sequence to the transcriptional start site of a gene indicates a possible promoter function of the sequence isolated. Sequences situated further 5′ or 3′ from a gene (we used 50 kb as an arbitrary cutting point) indicates a possible role as an enhancer.
5. The sequence is examined for the presence of consensus half-sites or complete hormone response elements (HRE).[9] However, analysis of the sequences should be extended to sites for other transcription factors, as nuclear receptors can activate the transcription of target genes via interaction with other complexes that associate with DNA (such as Sp1).[10] In addition, as transcription factors are known to bind next to each other in regulatory regions, the presence of other known transcription factor consensus binding sites should also be taken into consideration as it can give insights about the potential regulatory properties of the isolated sequences.

[9] K. Umesono, K. K. Murakami, C. C. Thompson, and R. M. Evans, *Cell* **65**, 1255 (1991).
[10] V. Krishnan, X. Wang, and S. Safe, *J. Biol. Chem.* **269**, 15912 (1994).

Validation and Analysis of the Isolated Sequences

Standard ChIP Assay

As discussed earlier, a standard ChIP assay defined as a PCR on the DNA fragments isolated using a single IP should first be performed to:

1. Validate each ChIP experiment using a known target sequence to which the receptor binds (positive control).
2. Validate the IP of a newly isolated and cloned sequence (standard ChIP).

For convenience and accuracy, we used quantitative PCR for the standard ChIP assay.

Validation of the ChIP Experiment (Positive Control)

Primers are designed to amplify a positive control sequence to which the receptor of interest is known to associate to upon ligand binding. This is done to validate the actual ChIP experiment. Negative control primers situated form 3 to 5 kb upstream or downstream of the region analyzed for nuclear receptor binding are also designed to ensure the specificity of the immunoprecipitated fragment. The PCR product length should be around 200–400 bp as suggested by the manufacturer of the Light Cycler apparatus. For standard ChIP assay, a control IP performed in the absence of Ab (basal control) is carried out for each treatment and used to compare with the actual IP reaction. Samples corresponding to 0.5, 1, and 2% of input are kept for quantification on the Light Cycler to calculate the ChIP efficiency of the target sequence compared to the negative control, for both the IP and the control.

The percentage of sequence immunoprecipitated reflects the affinity of association of the nuclear receptor to the target sequence and can be calculated using the formula:

$$\%\mathrm{IP} = 2^{1+n_i-n} \tag{1}$$

where n_i and n are the number of cycles at which the exponential portion of the curves for product formation is most efficient for the 2% input and the sample, respectively. This formula is derived from the assumption that an efficiency of two is achieved for the Light Cycler PCR reactions. In a standard ChIP assay, the sequence has to be immunoprecipitated at higher percentages than the basal control (no Ab); a significant IP percentage (%IP) has to reach at least 0.2%.[6] The %IP of the 5-kb upstream negative control must be insignificant (lower than 0.2% IP) in all the samples. Moreover, the difference in the %IP between the basal control

(no Ab) and the sample (with Ab) must be significantly higher in the target sequence compared to the negative control (~5 kb upstream). The difference in the IP percentages between the treated and the non-treated purified DNA fragments from the ChIP experiment indicates at which extent the addition of the ligand recruits the receptor to the regulatory element.

Since it had already been shown that the human TFF1 (pS2) gene promoter is occupied by ERα upon E_2 treatment,[11] it was used as a positive control for the validation of our ChIP experiment (see Fig. 2).[4] The pS2 promoter primers were used with an annealing temperature of 62°C on the quantitative Light Cycler, using SYBR Green Light Cycler kit 1 (Roche Molecular Biochemicals, CA) (Fig. 2). To further confirm the specificity of the binding, another set of primers was designed for a negative control sequence situated approximately 4-kb upstream of the studied regulatory sequence. As shown, all negative controls are under the 0.2% IP baseline, and ChIP targets reach higher %IP.

Validation of a ChIP experiment with a known target sequence as positive control is an important step to ensure the accuracy and the specificity of the ChIP technique, before cloning the purified DNA fragments. However, we are aware that it may not always be possible to have access to a known target sequence for some specific nuclear receptors in some particular cell lines. The use of other cell lines with known target sequences for the receptor of interest could also be used (when available) to develop the experimental conditions of the ChIP.

Validation of a Newly Isolated and Cloned Sequence

The standard ChIP assay is similarly used to validate the potential association between the nuclear receptor and the newly cloned sequence. Specific ChIP primers are designed for the cloned sequence of interest, as well as for a negative control sequence (~5 kb upstream or downstream), and are used with quantitative Light Cycler to assess the enrichment of this potential regulatory sequence in the ChIP samples. This step allows to discriminate between true and false positives that have been cloned using the double ChIP-cloning technique.

It is interesting to note that one of the clones isolated upon ChIP cloning following a single ER IP was a 389-bp fragment encompassing the pS2 promoter. The isolation of this promoter in the cloning step, thus served as a positive control for our technique.

[11] C. Giamarchi, M. Solanas, C. Chailleux, P. Augereau, F. Vignon, H. Rochefort, and H. Richard-Foy, *Oncogene* **18**, 533 (1999).

Changes in Gene Expression Levels Induced by Nuclear Receptors

Once the association of the receptor with the cloned regulatory sequence has been firmly established using standard ChIP assay, the influence of the nuclear receptor on mRNA expression of a specific target gene close to the sequence of interest needs to be assessed. Quantitative RT-PCR experiment is intended to quantify the levels of specific mRNA corresponding to the expression of this specific gene in the cell line over several time points, following E_2 or other ligand treatment. For example, pS2 RT-PCR primers were designed to perform quantitative RT-PCR on cDNA from MCF-7 cells that were either treated or not with E_2 (10^{-7} *M*) for a specific period of time. Quantitative RT-PCR using Light Cycler instrumentation and the SYBR Green detection kit 1 was carried out, and the upregulation of the target gene was obvious upon treatment of the cells with E_2, at all the time course tested with respect to the samples obtained from untreated cells (Fig. 3).

This assay should be applied to all the potential target genes, and can also be adapted for orphan nuclear receptors by introducing siRNA (or a vector-based expression of siRNA) directed against the specific orphan nuclear receptor in order to knockdown its expression and analyze changes in the levels of expression of putative target genes.[12]

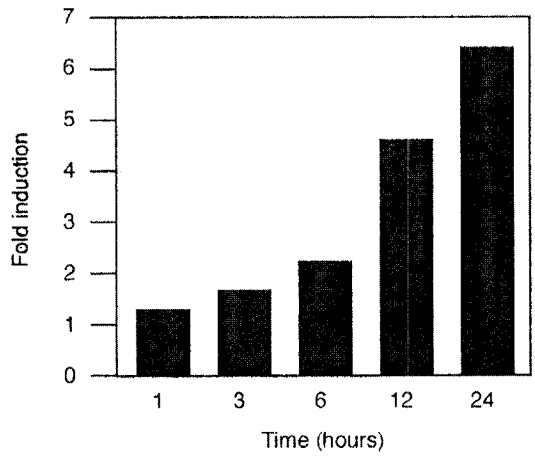


FIG. 3. Quantitative RT-PCR showing the E_2 response of the pS2 gene in MCF-7 cells. Total cDNA from MCF-7 cells treated with E_2 for 1, 3, 6, 12, and 24 hr is used for quantitative RT-PCR of a specific gene, and compared with control cells.

[12] T. R. Brummelkamp, R. Bernards, and R. Agami, *Science* **296**, 550 (2002).

Transient Transfections with Reporter Luciferase Assay

Once it has been established that the cloned sequence is enriched in standard ChIP, and thus associates with the receptor or a complex containing the receptor, it is suggested to test the ability of the cloned sequence to confer nuclear receptor responsiveness to a basal reporter gene. The DNA sequence is subcloned upstream of the TK-Luciferase (TK-Luc) reporter gene (or other suitable basal expression vectors) to be used in transient transfection assays in responsive cells using Fugene 6 reagent (Roche Molecular Biochemicals, CA). The cell line used for transfections should preferentially be the same as the one used as the starting material for the double ChIP cloning reactions. In the case of ligand-induced nuclear receptors, addition of ligand is necessary to observe the transcriptional activity of the receptor on the TK-Luc gene via the sequence of interest. On the other hand, it would be difficult to study the level of transcriptional activation induced by orphan nuclear receptors via its target regulatory sequences, since the absence of ligand for these receptors imply the lack of a basal transcriptional level to compare with. This problem can be overcome by using a different cell line that does not express the orphan receptor to perform the luciferase assay. However, this alternative might as well become problematic, since the cell line used might lack other factors necessary for transcriptional activation by the receptor on the studied regulatory sequence. The use of siRNA as described earlier constitutes a viable alternative.

The pS2 promoter isolated was cloned upstream of a luciferase reporter gene and transfected in MCF-7 cells that were either treated or not with E_2 (10^{-8} *M*). It was found that the luciferase gene was transcriptionally activated upon E_2 treatment, confirming the regulatory effect of the isolated pS2 promoter on transcriptional activation of the target gene as previously observed.[13]

At this point, it may be suitable to determine which sequence(s) present in the cloned fragment is the site of action of the nuclear receptor. A panoply of well-developed techniques can be used to characterize the regulatory element.[14]

Conclusion

The development of high-throughput ChIP technology, now permits a whole-genome analysis of gene regulation by nuclear receptors. The efficient

[13] D. Lu, Y. Kiriyama, K. Y. Lee, and V. Giguère, *Cancer Res.* **61**, 6755 (2001).
[14] M. Shago and V. Giguère, *Mol. Cell. Biol.* **16**, 4337 (1996).

cloning of new regulatory elements harboring high-affinity-binding sites for specific nuclear receptors can now be efficiently achieved, and allows for the identification of new target genes within a particular cell context. These studies will increase our knowledge of gene regulation via their promoters and enhancers. Since these procedures can be adapted to all nuclear receptors and other transcription factors, it should also be possible to discover transcriptional cross-talks between distinct regulatory networks. For this need, ChIP cloning can be used as a basis for the generation of regulatory regions arrays, allowing ChIP-microarray studies to analyze the binding of different nuclear receptors and transcription factors to these regulatory modules in different cell contexts, such as in normal versus cancer cells. In turn, these studies should lead to a better understanding of many diseases and to the development of better drugs for their treatments.

Acknowledgments

This work was supported by the Canadian Institutes for Health Research (CIHR), the National Cancer Institute of Canada, and Genome Québec/Canada. V. G. is a Senior Scientist of the CIHR.

APPENDIX 2

GENOME RESEARCH

Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression

Mathieu Blanchette, Alain R. Bataille, Xiaoyu Chen, Christian Poitras, Josée Laganière, Céline Lefèbvre, Geneviève Deblois, Vincent Giguère, Vincent Ferretti, Dominique Bergeron, Benoit Coulombe and François Robert

Genome Res. 2006 16: 656-668; originally published online Apr 10, 2006;
Access the most recent version at doi:10.1101/gr.4866006

Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression

Mathieu Blanchette,[1,5] Alain R. Bataille,[2] Xiaoyu Chen,[1] Christian Poitras,[2] Josée Laganière,[3] Céline Lefèbvre,[3] Geneviève Deblois,[3] Vincent Giguère,[3] Vincent Ferretti,[4] Dominique Bergeron,[2] Benoit Coulombe,[2] and François Robert[2,5]

[1]McGill Centre for Bioinformatics, Montreal, Quebec, Canada, H3A 2B4; [2]Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada H2W 1R7; [3]Molecular Oncology Group Department of Medicine, Oncology and Biochemistry, McGill University, Montreal, Quebec, Canada H3A 1A1; [4]McGill University and Genome Quebec Innovation Center, Montreal, Quebec, Canada H3A 1A4

The identification of regulatory regions is one of the most important and challenging problems toward the functional annotation of the human genome. In higher eukaryotes, transcription-factor (TF) binding sites are often organized in clusters called *cis*-regulatory modules (CRM). While the prediction of individual TF-binding sites is a notoriously difficult problem, CRM prediction has proven to be somewhat more reliable. Starting from a set of predicted binding sites for more than 200 TF families documented in Transfac, we describe an algorithm relying on the principle that CRMs generally contain several phylogenetically conserved binding sites for a few different TFs. The method allows the prediction of more than 118,000 CRMs within the human genome. A subset of these is shown to be bound in vivo by TFs using ChIP-chip. Their analysis reveals, among other things, that CRM density varies widely across the genome, with CRM-rich regions often being located near genes encoding transcription factors involved in development. Predicted CRMs show a surprising enrichment near the 3′ end of genes and in regions far from genes. We document the tendency for certain TFs to bind modules located in specific regions with respect to their target genes and identify TFs likely to be involved in tissue-specific regulation. The set of predicted CRMs, which is made available as a public database called PReMod (http://genomequebec.mcgill.ca/PReMod), will help analyze regulatory mechanisms in specific biological systems.

[Supplemental material is available online at www.genome.org.]

The regulation of gene expression is at the core of many important biological processes such as cell growth, division, differentiation, and adaptation to the extracellular environment. Gene expression is regulated in large part at the transcription level, with transcription factors (TFs) binding their specific DNA regulatory elements and activating or repressing transcription. The identification and characterization of these DNA regulatory elements are among the most important and challenging tasks for molecular biologists in the post-genome era.

TFs typically have an affinity for short, 5–15 bp, degenerate DNA sequences. Decades of work in many laboratories have led to the identification of consensus-binding motifs for hundreds of these TFs. These binding motifs are generally represented by position-weighted matrices (PWM). In principle, examination of the human genome with these PWM should allow for the identification of TF-binding sites (TFBSs), and hence, regulatory regions; but the size of the genome, combined with the fact that TF-binding motifs are short and degenerate, complicates this task enormously. Indeed, these motifs can be found everywhere in the genome and experiments have shown that only an extremely small proportion represent bona fide TFBSs. The binding of a TF is thus not simply a function of the theoretical affinity for a DNA site, but also of a number of other factors like the chromatin environment and the cooperation or competition with other DNA-binding proteins. In higher eukaryotes, TFs rarely operate by themselves, but rather bind to DNA in cooperation with other DNA-binding proteins. The DNA footprint of this set of factors is called a *cis*-regulatory module (CRM), which consists of a set of TFBSs located in a DNA region of up to a few hundred bases located in the vicinity of the gene being regulated. These modules have been the focus of much work recently (Davidson 2001), particularly in the context of the gene regulation during development (Howard and Davidson 2004), and are believed to be key features of most transcriptional regulatory processes in mammals.

Several features of known CRMs can be used to recognize new modules as follows: (1) CRMs are generally composed of several binding sites for a few different TFs; (2) CRMs, and in particular the binding sites they contain, are generally more evolutionarily conserved than their flanking intergenic regions, and (3) genes regulated by a common set of TFs tend to be coexpressed. Different combinations of those characteristics have been used, often in conjunction with PWM information, to predict regulatory elements for specific TFs. However, very few existing methods are designed to be applied on a genome-wide

[5]Corresponding authors.
E-mail blanchem@mcb.mcgill.ca; fax (514) 398-3387.
E-mail francois.Robert@ircm.qc.ca; fax (514) 987-5743.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.4866006.

16:656–668 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org

scale without prior knowledge about sets of interacting TFs or sets of coregulated genes (the main exception being the regulatory potential analysis of Kolbe et al. [2004] and King et al. [2005]). To date, the general properties of human nonpromoter regulatory regions indeed remain largely unexplored.

Here, we describe an algorithm that allows the identification of about 118,000 putative CRMs, based on predicted sites of 229 families of human TFs (represented by 481 PWMs). We refer to these regions as "predicted *cis*-regulatory modules" (pCRMs). Together with the regions predicted for regulatory potential by Kolbe et al. (2004), this constitutes the first genome-wide, nonpromoter centric set of human *cis*-regulatory modules, although related studies have been reported for yeast (Segal et al. 2003) and for human promoters (Bajic et al. 2004; Segal and Sharan 2005; Robertson et al. 2006). More importantly, in the analysis our set of pCRMs yields a number of novel insights into the mechanisms of gene regulation. After experimental validation of some of our predictions using a combination of chromatin immunoprecipitation and DNA microarrays (ChIP-chip), we used these predictions to explore the regulatory potential of the human genome. We show that, despite the fact that our pCRMs undoubtedly contain a significant number of false positives, the whole-genome approach provides sufficient statistical power to formulate specific biological hypotheses. For example, (1) the CRM density is unexpectedly high downstream of the 3′ end of genes, hinting at a possible involvement in regulating antisense transcription; (2) the regions that are the densest in CRMs are associated with developmental TFs; (3) different TF families have binding sites that are enriched in different regions relative to their target genes; (4) certain TFs or combination of TFs are associated with tissue-specific regulation. The Web-accessible database that accompanies this study will prove useful to experimental biologists interested in the regulation of specific genes, and will allow further bioinformatics and data-mining efforts.

Results and Discussion

Existing methods for *cis*-regulatory module prediction

The problem of computationally predicting *cis*-regulatory modules has been extensively studied in the last few years. Most predictive methods are either based exclusively on sequence data (see below), but some attempt to take advantage of gene expression data (Segal et al. 2003; Ihmels et al. 2004; Kloster et al. 2005; Wang et al. 2005) or DNaseI hypersensitivity data (Noble et al. 2005). Sequence-based algorithms have been developed along several lines. In the most studied case, the promoters of a set of (presumably) coregulated genes obtained from some prior experiments is analyzed to identify overrepresented motif combinations likely to be responsible for the gene's coregulation (Wasserman and Fickett 1998; Krivan and Wasserman 2001; Aerts et al. 2003, 2004; Sharan et al. 2004; Thompson et al. 2004; Zhou and Wong 2004; Gupta and Liu 2005; Segal and Sharan 2005). Other approaches assume that the user provides a small set of transcription-factor PWMs that are expected to co-occur in modules, and identifies genomic regions densely populated in putative sites for these TFs (Bailey and Noble 2003; Frith et al. 2003; Johansson et al. 2003; Sinha et al. 2003, 2004; Alkema et al. 2004). None of these two families of approaches are applicable in our setting, where we do not have sets of coregulated genes to train from, and where we have little prior knowledge about combinations of factors that are likely to co-occur to form modules.

To our knowledge, the only computational approach that has been used for de novo, genome-wide prediction of regulatory regions is the method of regulatory potential estimation from Hardison's group (Kolbe et al. 2004; King et al. 2005). This method is trained to recognize sequence features and interspecies conservation patterns that allow us to distinguish between known regulatory regions and nonfunctional sequences. A comparison of the results obtained by this approach and ours is given below.

A new algorithm for prediction of *cis*-regulatory modules

We designed a computational method with the goal of (1) identifying the DNA regions within the human genome that are likely to be important for regulating gene expression and (2) predicting what TFs are likely to bind these regions. Because our interest does not lie on any specific TF or specific system, but rather on having a global map of the regulatory elements of the entire genome, we exploited the fact that PWMs representing binding sites for a few hundreds of TFs have been described in databases such as Transfac (Matys et al. 2003) and JASPAR (Sandelin et al. 2004). Our algorithm takes advantage of the fact that regulatory regions often consist of clusters of binding sites for a few different TFs and that they are more conserved than their flanking intergenic DNA (Davidson 2001; Bulyk 2003; Levine and Tjian 2003). Our approach, based on the detection of statistically significant clusters of phylogenetically conserved TFBSs, shares some of the features of algorithms previously proposed by Sharan et al. (2004) and Aerts et al. (2004), but differs in that it allows the detection of modules without prior knowledge regarding which TFs are likely to be involved together in modules of interest. Our method also shares some similarities with the word-based approach of Philipakis et al. (2005), but uses a very different approach to module scoring.

Our algorithm involves two steps (see Fig. 1 and Methods for more details) as follows:

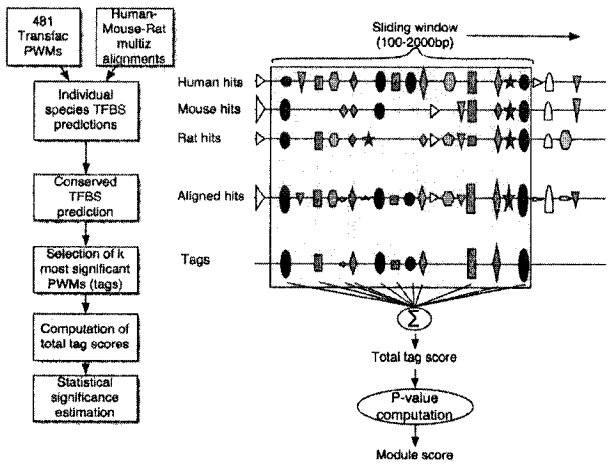


Figure 1. Overview of the CRM prediction algorithm. TFBS predictions for different PWMs are shown with different geometric shapes and their size indicates the score of the hit. Hits from individual species are combined using a weighted average method to compute the "Aligned hits." The most significant (up to five) aligned hits are considered as "Tags" for the corresponding region. The sum of the Tags scores is used to calculate a "Module score" using a statistical significance estimation. This operation is performed for each position of the human genome, for sliding windows of size 100, 200, 500, 1000, and 2000 bp.

1. Identification and scoring of putative TFBSs using 481 Transfac PWMs for vertebrate TFs (representing a total of 229 TF families). To this end, each noncoding, nonrepetitive position of the human genome within a human–mouse–rat alignment block (based on MULTIZ genome-wide alignments [Blanchette et al. 2004]) was evaluated for its similarity to each PWM using a log-likelihood ratio score with a third-order Markov background model parameterized based on the local GC content. Corresponding orthologous positions in mouse and rat genomes were evaluated similarly and a weighted average of the human, mouse, and rat log-likelihood scores at aligned positions was used to define a "hit score" for each human genomic position and each PWM. The scoring method favors simultaneous matches in all three species, which greatly reduces the false-positive rate of predictions. Notice, however, that the sites predicted need not be located within large phylogenetically conserved regions, nor do they need to be perfectly conserved across species.
2. Detection of clustered putative binding sites. Regulatory modules are often characterized by the presence of several binding sites for each of a small number of TFs (Howard and Davidson 2004). We identified regions of, at most, 2 kb that are significantly enriched in binding sites for one to up to five different TFs. To assign a "module score" to a given region, the five TFs with the highest total nonoverlapping scoring hits are chosen as tags for the putative module, and a *P*-value is assigned to the total score observed for the top one, two, three, four, or five tags. The number of tags for a given module is chosen so as to maximize the statistical significance of the hit density, so a short region that would be dense in sites for one TF would score well, as would a larger region with a few binding sites for each of a handful of factors. The *P*-value computation takes into consideration the number of factors involved (1–5), their total hit scores, the overall genome-wide frequency of their predicted hits, and the length and GC content of the region under evaluation (see Methods).

Our algorithm was used to scan the regions of the human genome that were alignable to the mouse and rat genome using the MULTIZ program (Blanchette et al. 2004; these regions cover 34% of the human genome). This resulted in the identification of 118,402 predicted modules, covering 2.88% of the human genome. Taken as a whole, this set of pCRMs, although likely to contain a non-negligible fraction of false positives, reveals a number of properties of human gene regions.

Although we considered putative modules of size up to 2000 bp, 58% of the pCRMs are less than 500 bp long, with an overall average length of 635 bp per CRM (see Supplemental Fig. S1A for a size histogram). This size distribution is quite close to that of the experimentally verified modules contained in the TRRD database (Kolchanov et al. 2002). However, we cannot exclude the possibility that some of the larger pCRMs are in fact made of more than one biological CRM. Modules have, on average, 3.1 tags (see Supplemental Fig. S1B), with shorter modules usually built from fewer tags than larger ones.

While the total number of individual sites predicted in phase (1) of our algorithm varies significantly from one PWM to another (see Supplemental Table S1), our procedure for correcting for low-specificity matrices ensures that no PWM is chosen as a tag too frequently. Supplemental Table S2 shows that tags are not seriously biased toward particular matrices, a sign that our algorithm for tag selection is sufficiently robust to avoid PWMs with low specificity. The PWM chosen as a tag the most often (5401 times, of 118,402 modules) is that for E2F, while the median PWM is selected as a tag in 704 modules. The PWMs that are the most often chosen for tags fall under two categories. The first is that of general promoter-associated factors, like E2F, ZF5, and TBP, which are indeed expected to bind a large number of regulatory regions. The second set of common tags consists of homeobox TFs (e.g., NKX family, POU family, etc.).

In silico validation of predicted modules

We evaluated the biological relevance of the pCRMs by measuring the extent to which they overlap known regulatory elements such as those compiled in the TRRD (Kolchanov et al. 2002), Transfac (Matys et al. 2003), and GALA (Giardine et al. 2003) databases. We also measured the overlap between the pCRMs and other putative regulatory elements, such as "promoter" regions (defined as the 1-kb region upstream of the transcription start sites [TSS] of all known genes), CpG islands (based on the UCSC Genome Browser annotation [Karolchik et al. 2003]), and DNaseI hypersensitive sites (Dorschner et al. 2004; Sabo et al. 2004;) from the Encode regions (Thomas et al. 2003). Figure 2A shows that despite the fact that only about 2.88% of the genome belongs to pCRMs, our predictions contain about 40% of the bases within modules annotated in GALA, 34% of the bases within Transfac binding sites, and 20% of the bases within the TRRD database. Our pCRMs are highly enriched within promoter regions, especially those containing CpG islands. Indeed, when considering the overlap between pCRMs and nonproximal (>1 kb upstream) annotated regulatory regions, our sensitivity (Fig. 2C) drops for all indicators except for the modules from the GALA database, though all remain severalfold higher than expected by chance (Fig. 2B,D). The significant enrichment for DNaseI hypersensitive sites is particularly interesting, as those represent an unbiased probing of chromatin structure. Although the function of these hypersensitive sites remains in most cases undetermined, many are likely to be CRMs.

By definition, the sensitivity of our method for detecting annotated regulatory regions increases with the number of modules that are predicted. This increase is very rapid for the first ~20,000 modules predicted, but the sensitivity for most indicators then increases more slowly. This observation is likely due to the fact that the modules that are the easiest to detect are those located in promoter regions. These also turn out to be the regions where most regulatory modules have been studied. However, the fact that our most reliable indicators of performance (TRRD modules, GALA modules, and, to a lesser extent, hypersensitive sites) continue to grow steadily after the first 20,000 pCRMs indicates that nonproximal modules can still be identified, and justifies considering a much larger set of modules.

Comparison to other genome-wide predictions

The ability of our algorithm to take advantage of interspecies TFBS conservation contributes in good part to the accuracy of the predictions. Indeed, the 34% of the human genome that lies within an alignment block with the mouse and rat genome contains 90% of bases within Transfac sites, 67% of those within TRRD modules, and 87% of those within GALA regulatory regions. Nonetheless, the sensitivity obtained by our pCRMs on these indicators remains three to five times higher than what would be obtained if modules were randomly predicted within the alignment blocks. To measure more accurately the extent to

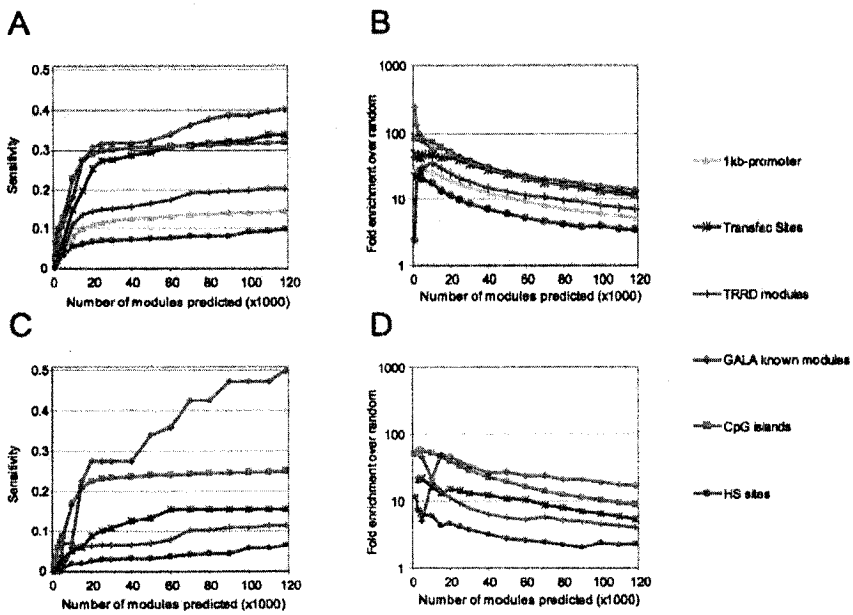


Figure 2. Sensitivity and enrichment of pCRMs for various regions of interest. (*A*) Sensitivity of the module predictions at varying score threshold, with respect to likely regulatory regions. Along the *y*-axis is the fraction of the bases within known regulatory regions that are predicted to belong to a pCRM. Along the *x*-axis is the number of predicted modules above a given threshold. Regions of interest are: 1 kb upstream: regions upstream of the TSS of Known Genes (based on the UCSC Genome Browser); Transfac sites: a set of 1209 experimentally verified binding sites from Transfac 7.2, mapped onto the human genome; TRRD modules: a set of 601 experimentally verified regulatory modules from the TRRD database; GALA modules: a set of 93 modules for the GALA database; CpG islands (based on the UCSC Genome Browser annotation); 1 kb upstream: regions upstream of the TSS of Known Genes that are not annotated as CpG islands; HS sites: a set of DNaseI hypersensitive sites from the Encode regions. (*B*) The fold enrichment is computed as the ratio between the size of the intersection between modules and regions of interest and the expected intersection size if modules were randomly positioned in the genome. (*C*,*D*) The analogous data, but restricting our attention to non proximal regulatory regions, i.e., those located more than 1 kb away from the TSS of the closest gene.

which sequence conservation alone can be used to predict known regulatory modules, sensitivity curves were computed based on the noncoding interspecies conserved regions identified by the PhastCons program (Siepel et al. 2005) (See Supplemental Fig. S2). The sensitivity of pCRMs is consistently 30%–70% higher than that of PhastCons elements for 1-kb "promoter" regions and TRRD and GALA modules, while it is comparable for Transfac and DNaseI hypersensitive sites. The advantage of pCRMs over PhastCons is most marked when only the highest-scoring half of each set of predictions is considered, in which case, the pCRMs sensitivity is at least twice that of PhastCons for all indicators.[6] Overall, 41% of the bases within pCRMs lie within a PhastCons region (and 31% of PhastCons bases are within a pCRM), an 11-fold enrichment over what would be expected by chance.

Kolbe et al. (2004) and King et al. (2005) have developed a method called "regulatory potential," which has been applied to the complete human genome to yield a set of CRM predictions. The method is trained to identify sequence features and interspecies conservation patterns that allow one to distinguish between a set of known regulatory regions and a set of nonfunctional regions. The overlap between the regulatory regions predicted by King et al. and our pCRMs is very significant—choosing a score threshold that results in about the same number of predicted bases as we get in our pCRMs (2.88% of the genome); more than 25% of the bases in pCRMs are also in King's regions (nine times more than would be expected by chance). The accuracy of the two sets of predictions was compared based on the set of known regulatory regions used above, and none of the two methods appears significantly better than the other (see Supplemental Fig. S2), despite the fact that King's method was trained on some of the specific regulatory regions used here for validation.

Experimental validation of predicted modules

In order to further validate our pCRMs, we took advantage of a technique called genome-wide location analysis (or ChIP-chip) (Ren et al. 2000; Iyer et al. 2001). This method allows for the large-scale identification of protein–DNA interactions as they occur in vivo. Briefly, proteins are cross-linked to DNA by treating live cells with formaldehyde and specific protein–DNA complexes are enriched by immunoprecipitation of fragmented chromatin using antibodies directed against a protein of interest. After reversal of the cross-links, the enriched DNA fragments are identified by hybridization onto DNA microarrays. We selected modules predicted to be bound by the estrogen receptor (ER), the E2F transcription factor 4 (E2F4), the signal transducer and activator of transcription 3 (STAT3), and the hypoxia-inducible factor 1 (HIF1) to print a DNA microarray. The microarray contains 758, 1370, 860, and 1882 modules predicted to be bound by ER, E2F4, STAT3, and HIF1, respectively. In the current study, the microarray was then probed by ChIP-chip for ER and E2F4 (see Methods for experimental details). After statistical analysis and experimental validation of the data (see Methods and Supplemental Table S3), we have identified 55 and 433 modules bound by ER and E2F4, respectively (see Supplemental Tables S4 and S5, respectively, and Table S6 for full ChIP-chip results). Approximately 3% of the 758 ER-predicted pCRMs on the microarray actually proved to be bound by ER, while 17% of the 1370 E2F4-predicted pCRMs on the microarray were bound by E2F4.

These numbers need to be considered as an underestimation of the actual specificity of the algorithm, since the protein–DNA interactions were tested in a single cell type, while TFs are known to regulate different sets of genes in different cell types, physiological conditions, and time in development (Zeitlinger et al. 2003; Hartman et al. 2005). For example, ER was tested in MCF-7, a breast cancer-derived cell line, due to its importance in breast cancer. ER, however, also plays important roles in many tissues such as ovaries, bone, brain, liver, and more. It is very likely that ER binds many pCRMs in some of these tissues, but not in MCF-

[6]Since PhastCons was designed to detect any type of region under selective pressure, many of its noncoding predictions are likely to have other nonregulatory functions.

7. In addition, the experiment was conducted under a single set of conditions (concentration of estradiol, time of treatment, etc.). For all of these reasons, it is difficult to determine the real accuracy of the algorithm.

Because our microarray contains predicted modules for four different TFs, the data can be used to assess the specificity of our TFBS predictions, e.g., to evaluate whether our prediction of which TFs should bind to each module is accurate. Among the 55 modules bound by ER, 44% (24/55, whereas 8/55 would be expected by chance) had indeed been selected for their ER-binding sites, and among the 433 modules bound by E2F4, 54% (236/433, whereas 147/433 would be expected by chance) had been selected for that factor. In addition to false-positive ChIP-chip signals or the failure of the algorithm to detect some binding sites, it is likely that binding of TFs through alternative mechanisms such as protein–protein interactions contributes to this result. For example, ER has been shown to be recruited to DNA by interaction with AHR to repress AHR-dependent gene regulation in an ER-responsive element-independent manner (Beischlag and Perdew 2005). It is important to note that our algorithm can only predict the binding of TF through direct DNA-binding interactions. It is likely that other TFs, in addition to those predicted here, may play roles in these modules. Of note, while 87% of the validated pCRMs for E2F4 were located in promoter regions, only 20% of those for ER were in these regions, confirming that our nonproximal pCRMs are also highly enriched for functional CRMs. Finally, Carroll et al. (2005) have used ChIP-chip on a tiling array to identify ER-binding sites on human chromosomes 21 and 22. Of the 57 regions they found to be bound by ER in MCF-7 cells, 14 overlap our predicted modules (five times more than expected by chance).

Despite the fact that the goal of this study is not to discuss specific interactions, we would like to highlight an interesting result that came out of the ChIP-chip experiments. While it is well known that the expression of the progesterone receptor gene *PGR* is up-regulated in breast cancer cells in response to estradiol, the absence of consensus estrogen response elements (ERE) in the two promoters driving its expression led to the suggestion that ER binds via other TFBSs (Petz et al. 2004). However, our data show that ER binds pCRMs present both ~35 kb upstream of the TSS and ~5 kb downstream of the 3′ end. Functional characterization of these pCRMs may reveal important clues about the molecular mechanisms implicated in long-range regulation by ER and other nuclear receptors (Carroll et al. 2005; Laganière et al. 2005).

A global view of the gene regulatory landscape

Having validated our predictions, we went on using them to study different global aspects of gene regulation. The genome-wide distribution of predicted modules is exemplified by Figure 3, which shows the pCRMs in a typical genomic region of human chromosome 11 containing the progesterone receptor gene *PGR*. The module density varies widely across the genome, with an average of four modules per 100 kb and a maximum of 44 modules per 100-kb window, covering from 0% to 55% of such a region. The presence of pCRMs is significantly correlated with the presence of a gene's TSSs (correlation coefficient = 0.17, P-value < 10^{-308}) on a local scale (10-kb window), but on a larger scale (1-Mb windows), no such correlation is observed. This indicates that the correlation between TSSs and pCRMs only extends to a few kilobases (Fig. 3B), and that distal pCRMs do not have strong location preferences relative to TSSs.

As illustrated in Figure 3, some regions are rich in modules, but relatively poor in genes. In some cases, this could reflect the presence of many unknown protein-coding genes, or at least of many alternative TSSs. Another possible explanation is that some of these modules may be regulating the transcription of noncoding transcripts. Cumulating evidence indeed shows that much more transcription happens in the genome than what can be accounted for by traditional genes (Cawley et al. 2004; Cheng et al. 2005; The FANTOM Consortium 2005). Finally, this observation may be due to the presence of long-range enhancers, which may affect transcription of genes up to several hundreds of kilobases away (Bejerano et al. 2004; Baroukh et al. 2005; Woolfe et al. 2005). Clearly, a sizeable fraction of the module predictions is likely to be false positives, but there are no a priori reasons to expect false-positive predictions to cluster in any particular regions of the genome.

The genomic locations that are the densest in predicted modules (measured over 100-kb windows) are listed in Table 1. Most of these are located upstream, in the introns, or downstream of genes that are themselves TFs often involved in development. Among the 15 densest regions, we find parts of all four HOX clusters that operate differential genetic programs along the anterior–posterior axis of animal bodies (Alonso 2002), and regions near the *EBF3, ZFHX1B, NR2F2, BCOR, MEIS2,* and *DLX5-6* genes, all of which are characterized TFs. The pCRMs in these regions have the unusual property of often being significantly conserved back to zebrafish and *fugu*, an indication that they may be part of the core regulatory mechanism of vertebrate development. There are 137 100-kb regions covered at least at 20% of CRMs, and these regions contain the TSSs of 115 genes with GO annotations (Harris et al. 2004). These genes are very strongly enriched for involvement in the regulation of transcription (79 genes, P-value 10^{-89}), morphogenesis (24 genes, P-value 10^{-13}), organogenesis (17 genes, P-value 3×10^{-5}), and neurogenesis (10 genes, P-value 4×10^{-4}), based on the Gostat program (Beissbarth and Speed 2004). We conjecture that genes involved in these processes often require very tight regulation, which in turn requires an elaborate set of regulatory modules. Notably, the presence in that group of *ZBTB20*, a poorly characterized gene encoding a predicted zinc finger TF, suggests the intriguing possibility that this TF may have a critical biological role, perhaps in regulating development.

There also exist regions that are very sparsely populated in predicted modules. One of the most striking examples is a 4-Mb region of chromosome 2 (chr2:123,000,001–127,000,000), of which <0.1% is covered by predicted modules. The region is somewhat of a gene desert, containing only one large gene annotated, hypothetical gene *CNTNAP5*. Other gene deserts are the opposite, quite rich in pCRMs. Many of those appear to be located in the vicinity of developmental TFs. For example, the homeobox gene *MEIS1* is surrounded by a 1-Mb region devoid of any other TSS, but contains >130 kb of pCRMs.

Regulatory modules are preferentially located in specific regions relative to genes

We studied the position of pCRMs with respect to their closest gene. The genome was divided into several types of noncoding regions, i.e., upstream of a gene, 5′ UTR, 1st intron, internal introns, last intron, 3′ UTR, and downstream region. Within

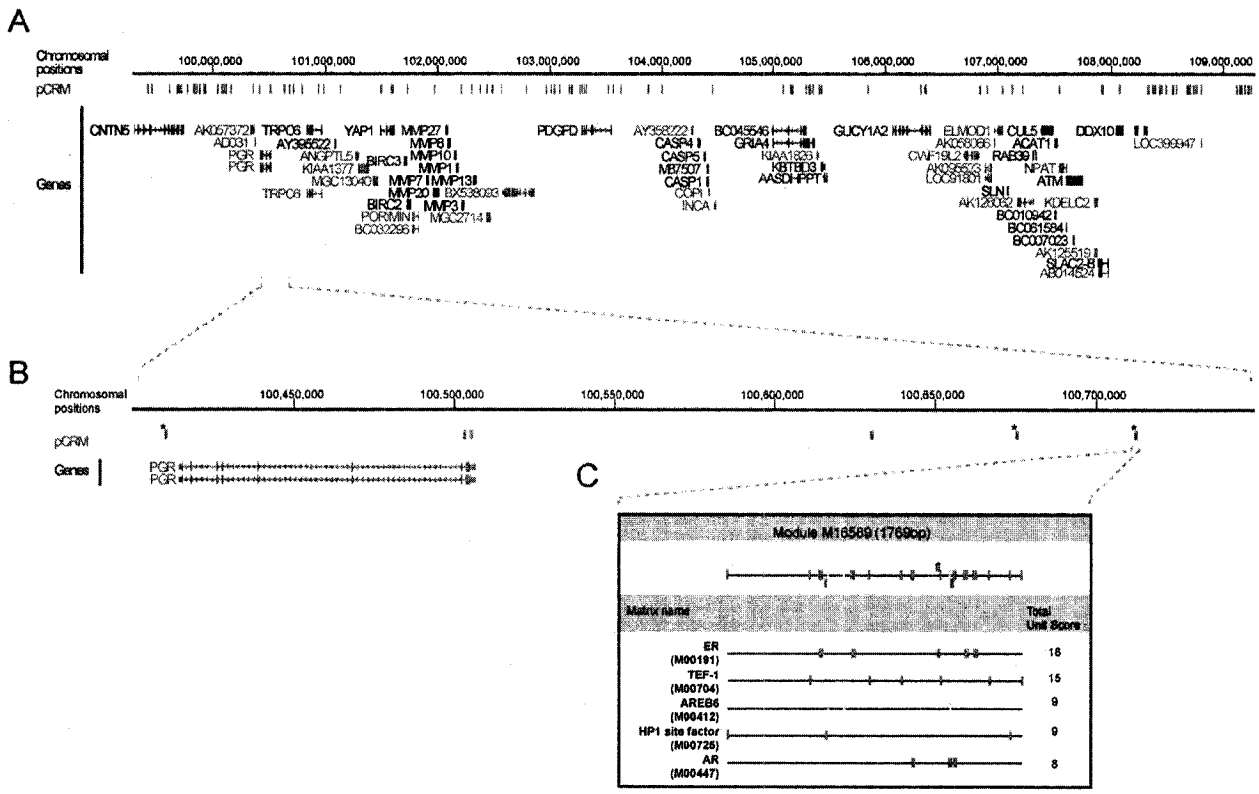


Figure 3. Distribution of pCRMs along a region of chromosome 11. (*A*) A 10-megabase region from chromosome 11 is shown (coordinates 99, 308, 463–109, 308, 463). The position of the pCRMs (red) and the known genes (blue, from the UCSC Genome Browser) is shown. (*B*) A zoom in a 350-kilobase region containing the progesterone receptor gene (*PGR*) (coordinate 100, 400,000–100,750,000). The pCRM marked with an asterisk are those printed on our DNA microarray. (*C*) The composition of the Module M16589 is depicted as can be found in the PReMod database accompanying this study (http://genomequebec.mcgill.ca/PReMod). The position of the hits for five TRANSFAC matrices chosen as tags for this module is shown together with their individual scores.

each type of region, we computed the fraction of bases included in a pCRM as a function of the distance to a reference point for each type of region (e.g., for upstream regions and 5′ UTR, the reference point is the TSS; see legend of Figure 4 for more details). This positional distribution was also compared with the positional distribution of a set of interspecies-conserved regions identified by the phastCons program (Siepel et al. 2005) on a set of aligned vertebrate genomes, using a conservation score threshold that results in a total number of noncoding bases predicted to be the same as the number of bases within pCRMs.

From Figure 4, a number of striking observations are possible as follows:

1. Regions immediately surrounding TSSs are highly enriched for predicted modules. This was to be expected as this region often contains the promoter of the gene. More surprising is the presence of modules immediately downstream of the TSSs (either in the 5′ UTR or the first few kilobases of the first intron). These may represent alternative promoters for initiation downstream from the annotated transcripts. Alternatively, they may represent a yet underappreciated mode of activation that would take place from downstream proximal binding sites.
2. Regions surrounding the sites of termination of transcription are also highly enriched for modules. 3′ UTRs are essentially as enriched as 5′ UTRs for pCRMs, and module enrichment continues several kilobases past the end of the transcript, though to a lesser degree than in the upstream regions. At least two reasons may explain the presence of regulatory elements in the 3′ region of genes. First, these may represent enhancer type of regulatory elements that activate the upstream gene via a DNA-looping mechanism. Second, these may represent promoter elements driving noncoding transcript, antisense relative to the coding gene. Such antisense transcripts may regulate gene expression by a post-transcriptional mechanism (Cawley et al. 2004). Alternatively, these transcripts (or this transcription) may have biological roles of their own, independently of the coding transcript itself. For example, recent work in yeast showed that intergenic transcription could regulate gene expression by interfering with activation of a neighboring gene (Martens et al. 2004). It is possible that these TFBSs in the 3′ region of genes could give rise to antisense transcription that would interfere with sense transcription (Katayama et al. 2005). Recent analysis of the transcriptome of mammalian genomes revealed that a large proportion of all transcripts detected represent noncoding transcription (Kapranov et al. 2002; Cheng et al. 2005; The FANTOM Consortium 2005). Many of these noncoding transcripts map to the 3′ UTR of coding transcripts. ChIP-chip experiments performed on chromosome 21 and 22 (Cawley et al. 2004) have revealed that TFs can indeed bind these regions with a fre-

Table 1. Human genomic region densest in predicted CRMs

Region[a]	#CRMs[b]	Genomic location	Gene annotation[c]	Main gene function[c]
chr12:52600000–52700000	44 (55%)	HOXC cluster	Homeobox TFs	Anterior-posterior differentiation during development
chr7:26900000–27000000	44 (54%)	HOXA cluster	Homeobox TFs	Idem
chr10:131500000–131600000	43 (44%)	Up., intron, and down. of EBF3	COE-type TF	Regulation of development
chr17:44000000–44100000	37 (42%)	HOXB cluster	Homeobox TFs	Anterior-posterior differentiation during development
chr7:96200000–96300000	35 (34%)	DLX5-DLX6 intergenic region	Homeobox TFs	Central role in development of several structures
chrX:39700000–39800000	35 (43%)	Up. and 1st intron of BCOR	Transcription corepressor	BCL6 repressor
chr2:176800000–176900000	34 (47%)	HOXD cluster	Homeobox TFs	Idem
chr3:115600000–115700000	34 (36%)	Introns of ZBTB20	Zing-finger BTB/POZ TF	Possibly involved in hematopoiesis, oncogenesis, and immune responses
chr2:145000000–145100000	33 (41%)	Up. and introns of ZFHX1B	Zinc-finger BTB/POZ TF	Transcription inhibitor, interacting with SMAD proteins
chr15:94600000–94700000	33 (38%)	Up. intron, and down. of NR2F2 (COUP-TFII)	Nuclear hormone receptor, zinc-finger TF	Regulation of Notch signaling and vein identity
chr11:114600000–114700000	32 (36%)	Introns of IGSF4	Immunoglobulin-like domain	Intercellular adhesion molecule; Involved in human oncogenesis
chr11:114800000–114900000	30 (34%)	Up. and intron of IGSF4	Immunoglobulin-like domain	Intercellular adhesion molecule; Involved in human oncogenesis
chr15:35100000–35200000	29 (37%)	Up. and introns of MEIS2	Homeobox TF	Essential contributor to developmental programs
chr12:52700000–52800000	28 (34%)	Beginning of HOXC cluster	Homeobox TFs	Anterior-posterior differentiation during development

Human regions with the highest concentration of predicted regulatory modules, computed over windows of 100 kb.
[a]Human genome coordinates (build 34).
[b]Number of pCRMs predicted and percentage of the region they cover.
[c]Based on the UCSC Genome Browser Known Gene track information and PubMed literature searches.

quency higher than expected. These experimental data on chromosomes 21 and 22 are in agreement with our genome-wide predictions and likely reflect a yet understudied aspect of gene expression regulation.

3. Another surprising observation is that the density of modules is the lowest in regions located 10–50 kb upstream of the TSS and, symmetrically, 10–30 kb downstream of the end of transcription. This is unexpected, as one would expect that these regions (at least those upstream of the TSS) would be prime estate for transcriptional regulation. However, this is confirmed by the density of interspecies conserved elements, which is also at its lowest in those regions. We believe that this can be explained as follows: Thanks to their relative proximity to the TSS, regulatory elements in these regions may be allowed to contain fewer binding sites (or binding sites with less affinity), making them difficult to detect using our method. Alternatively, these regions may actually be depleted for regulatory elements. This could be due to constraints imposed by the chromatin structure of the nuclear architecture, making it more difficult for the DNA of these regions to come in physical proximity to the TSS. After the first 50 kb upstream of the TSS, the density of modules (and, to a lesser extent, of conserved regions) starts increasing with the distance to the TSS, with regions located >200 kb upstream of the TSS, being about 50% more densely populated in modules than the −50 to −10 kb region. We believe that this may be explained by the fact that regulatory modules that are located very far from the gene they regulate would often require many strong binding sites, making their computational detection easier. The symmetric effect is observed in regions downstream of genes, although at these large distances it is unclear whether these modules would regulate the sense or antisense transcription.
4. The density of predicted modules in intronic regions is very low in the close vicinity of exons (except the first and last ones), but increases with the distance to the closest exon. Although some of the intronic pCRMs may turn out to be splicing regulatory regions, this is unlikely to be the case for a large fraction of them, as intronic splicing elements usually cluster near exon boundaries (Sorek and Ast 2003). Instead, we speculate that CRMs within these very large introns may be located in genes that require tighter transcriptional regulation, resulting in a higher module density in these regions.
5. Although the module density usually follows closely the interspecies conservation density, a few notable exceptions indicate that our module predictions are doing more than merely detecting conserved regions. First, the regions surrounding the TSS (on either side) are much richer in modules than in conserved regions. Second, the 1-kb regions immediately flanking internal exons tend to be highly conserved (Sorek and Ast 2003) and they are believed to be involved in splicing regulation. However, these regions are depleted from pCRMs, as indeed these regions are not involved in transcriptional regulation and lack the signature sought by our algorithm. As a side note, pCRMs are also twofold depleted in known RNA genes, although these too tend to be well conserved evolutionarily.

Specific TFs target different regions relative to their target genes

As described above, our predictions, when taken altogether, are enriched in the 5′ and 3′ region of known genes. When broken down into predictions for individual TFs, however, a great variability in observed. For example, our predictions of ER modules

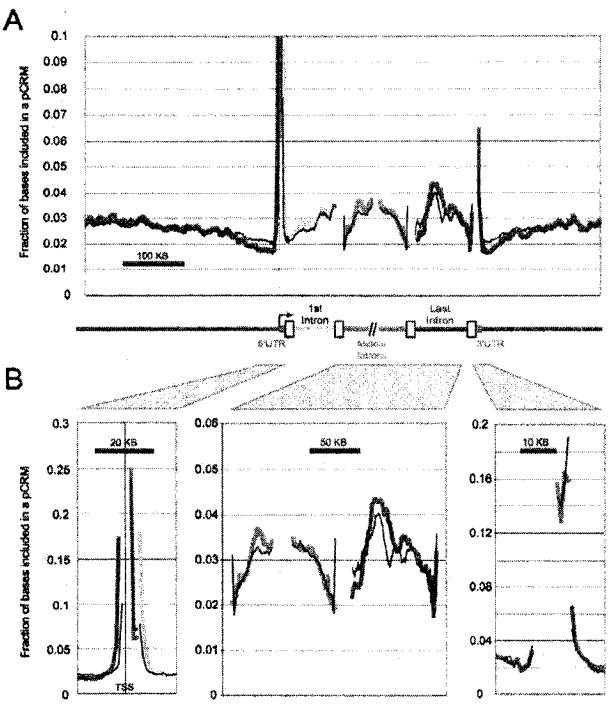


Figure 4. Distribution of pCRMs relative to specific regions of genes. The genome was divided into several types of noncoding regions: upstream of a gene (dark blue), 5′ UTR (pink), 1st intron (yellow), internal introns (light blue), last intron (brown), 3′ UTR (aqua), and downstream region (dark blue). (*A*) For each type of region, the fraction of bases included in a pCRM is graphed as a function of the distance to a reference point. For upstream regions, 5′ UTR, and first intron the reference point is the gene's TSS. For middle introns the closest 5′ or 3′ intron boundary is used. For the last intron, the 3′ UTR and the region 3′ of the last exon, the 3′ end of the mRNA is used. Note that the 3′ UTR is off the scale in *A*. (*B*) Same as in *A*, but different scales are used for the *x*- and *y*-axes in order to better show the characteristics of all regions.

(e.g., modules predicted to contain at least one high-scoring ER-binding site) are enriched in regions located more than 10 kb upstream of known genes, while our predictions for E2F4 are enriched in the proximal 5′ region of known genes. This suggests that ER functions mainly through distal, enhancer-like elements, while E2F4 regulates gene transcription via promoter-proximal elements. Notably, evidence in the literature supports this hypothesis (see Blais and Dynlacht 2005; Carroll et al. 2005). Importantly, our ChIP-chip data also supports this model. Indeed, despite the fact that pCRMs printed on the array were uniformly distributed with respect to genes, only 20% of the pCRMs bound by ER in our ChIP-chip experiments were within 1 kb on either side of the TSS, while the proportion is of 87% for the pCRM bound by E2F4. Based on this observation, we have computed the location preferences of each of the 229 TF families represented by the PWMs used in our predictions (see Figure 5 and Supplemental Table S7). Figure 5 shows that more than 70 of the 229 TFs families considered exhibit a significant enrichment for one or more types of genomic regions (see Methods). These TFs separate quite clearly into two groups with very little overlap. A number of TFs show preference for distal positions, mostly those located more than 100 kb upstream of the TSS, and are also often enriched within introns. This set of TFs is enriched for factors containing homeo domains or basic helix-loop-helix domains and are often involved in regulating development. Some of these factors have indeed been shown to bind distal modules and activate transcription during early development (Bejerano et al. 2004; Woolfe et al. 2005). Notably, we find no TFs enriched for introns only (except within 1 kb downstream of the TSS), which indicates that regulatory modules located in introns are of the same type as those located far away from genes. In fact, it is likely that certain intronic modules do not regulate the gene in which they are located, but rather another gene located nearby, as reported recently for sonic hedgehog (Sagai et al. 2005)

A second set of TFs preferentially binds within 1 kb of the TSSs. This set is enriched for leucine zipper TF and factors from the Ets family. Notably, most of these factors, contrary to what is observed for those binding distal sites, are involved in basic cellular functions. Among the best-known examples we found NF-Y, E2F, CREB, ATF, and others. Interestingly, and much to our surprise, most of these TFs show a clear preference for either the 1 kb upstream or the 1 kb downstream of the TSS, but not both. The most striking example is Nuclear Factor Y (NF-Y), which is highly enriched 1 kb upstream, but highly depleted 1 kb downstream of the TSS. This preference may reflect a mechanistic characteristic of these TFs. Finally, note that when we computed enrichment statistics based on all genome-wide predicted TFBSs instead of based only on those located in modules, much fewer TFs obtained significant enrichment in any given type of region, indicating that our pCRMs are effective at reducing the false-positive rate in TFBS predictions.

Long-range correlation of TFBS predictions

We observe that the closer together two modules are on the genome, the more likely they are to contain predicted binding sites for the same factors. Part of this is simply due to isochors, those broad variations of GC content along the genome (International Human Genome Sequencing Consortium 2001). However, even after correcting for this factor (see Methods), a number of TFs show significant long-range correlation between their predicted sites (Supplemental Fig. S3; Supplemental Table S8). This is likely to be due to the fact that if several regulatory modules regulate a gene, they are likely to be bound by a similar set of TFs. Not surprisingly, most of the TFs that exhibit long-range correlation are those that show preferences for binding sites located more than 10 kb upstream of the TSS. The set of nearby pCRMs that contain binding sites for similar TFs tends to be located in large intergenic or intronic regions and they tend to be located near genes encoding TFs.

Predicted TFBSs induce correlated tissue-specific gene expression

Comparison of TF-binding data with gene expression data in yeast showed that genes bound by a common set of TFs tend to be coregulated (Lee et al. 2002). Such a correlation is expected to occur in mammalian cells as well, but was never thoroughly tested because of the lack of genome-wide data for TF binding. Our predicted module data allows us to investigate this question. For each TF family in our study, a set of putatively regulated genes was identified as those with at least one predicted high-scoring site in a pCRM located within 10 kb upstream of the TSS. We computed the average pairwise Pearson correlation coefficient between tissue-specific expression levels of the genes of the set using expression data from 79 human cell types or tissues from the GNF Atlas 2 (Su et al. 2004). A total of 27 of the 229 TF

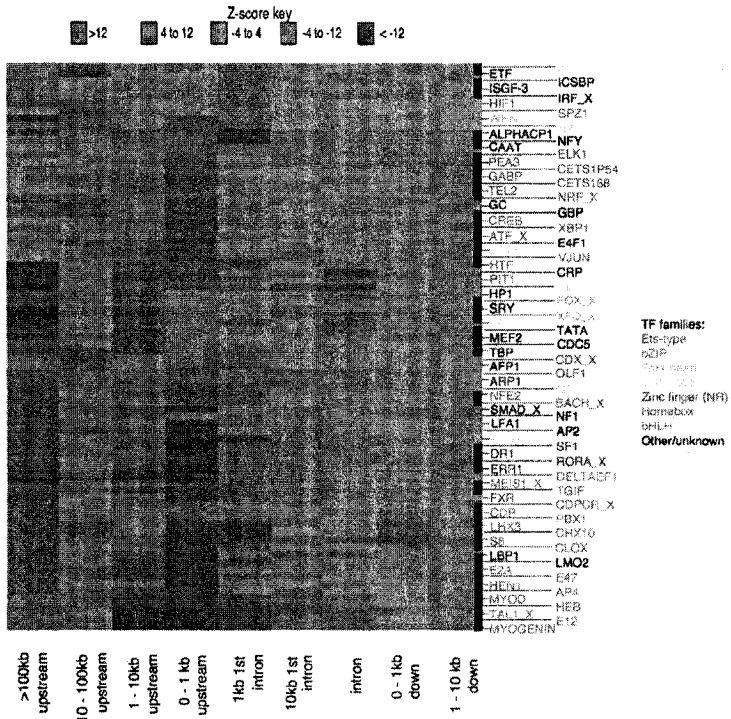


Figure 5. Many TFs preferentially bind to specific regions relative to the TSS of their target genes. A heat map of the enrichment (represented as a Z-score) of a TF for different regions relative to TSSs is shown. Regions in red are highly enriched for binding sites for the given TF, while those in blue are depleted. The regions shown on the *x*-axis are as follows: *>100kb upstream*, pCRMs located more than 100 kb upstream from a TSS; *10–100kb upstream*, pCRMs located >10 kb, but <100 kb upstream from a TSS; *1–10kb upstream*, pCRMs located >1 kb but <10 kb upstream from a TSS; *0–1kb upstream*, pCRMs located within 1 kb upstream of a TSS; *1kb 1st intron*, intronic pCRMs located within 1 kb downstream of the TSS of a gene; *10kb 1st intron*, intronic pCRMs located within 10 kb downstream of a TSS; *intron*, intronic pCRM located >10 kb from the TSS; *0–1kb down*, pCRM located within 1 kb from the 3′ end of a gene; *1–10kb down*, pCRM located >1 kb but <10 kb downstream from the 3′ end of a gene. See Methods for details on the computation of Z-scores.

families are associated to a significant expression correlation (*P*-value < 0.01, false-discovery rate (FDR) = 8%; see Supplemental Table S9). We repeated our correlation analysis, this time measuring the expression correlation for genes sharing binding sites for pairs of TFs. Of the 26,106 pairs of TF families considered, 595 are associated to a significant expression correlation (*P*-value < 0.01, FDR = 43%) (See Supplemental Table S10 for a complete list). For example, most of the 20 genes that have a pCRM containing OCT-1 and BACH1-binding sites are highly expressed in various brain tissues, excluding the cerebellum and the olfactory bulb, and in the pituary gland. While the role of OCT-1 in brain cells has already been characterized (Givens et al. 2004), its association with BACH1 has not been reported before.

Since most TFs are only expressed in a subset of the 79 cell types considered, they are unlikely to induce significant coexpression when measured over all 79 cell types. In order to identify transcription factors regulating expression in specific cell types, we analyzed each pair of TF and cell type. For each pair, the average expression level of the genes associated with predicted binding sites for the TF was computed and its significance assessed by a permutation test. Of the 229 × 79 = 18,091 possible (TF-cell type) pairs, we found 119 where genes are overexpressed (*P*-value < 0.001, FDR = 15%), and 78 where genes are underexpressed (*P*-value < 0.001, FDR = 23%). Table 2 lists the pairs with the most significant associations (see Supplemental Table S11 for the complete list). For example, the genes associated with pCRMs for MyoD tend to be highly expressed in skeletal muscle cells, while those associated to Ets are highly expressed in white blood cells. Both the role of MyoD in skeletal muscles and that of Ets in blood cells are very well characterized, thereby validating the approach.

We also discovered associations that are not well character-

Table 2. Tissue-specific expression for genes predicted to be regulated by various types of transcription factors

TRANSFAC matrices	Tissues with high expression[a]	Tissues with low expression[b]	Evidence from the literature
ETS, NRF2, ELK1, PEA3, PU1	White blood cells (Dentritic, NK, B, and T cells)	Most brain tissues	Reviewed in Sharrocks (2001)
MyoD	Skeletal muscle	Lung	Reviewed in Tapscott (2005)
NF-Y, CCAAT-box	Thymus, leukemia lymphoblastic, B lymphoblasts	Ciliary and superior cervical ganglions	Reviewed in Mach et al. (1996). See also Mantovani (1999)
AP-4	Various brain tissues	Leukemia lymphoblastic	No evidence found
Ahr/Arnt	Most brain tissues		Pravettoni et al. (2005)
Areb6	Fetal thyroid, salivary gland, trachea		No evidence found
NERF-1A	Subthalamic nucleus	Bone marrow, heart, lung, kidney, liver	No evidence found
NF-kappaB	Tonsil, lymphoblasts, Burkitts lymphoma, smooth muscle,	Thalamus	Reviewed in Viatour et al. (2005)
COUP-TF/DR1	Kidney, liver, tongue		Kerber et al. (1998)
SREBP	Fetal brain		Reviewed in Medina and Tabernero (2002)
MZF1		Kidney, liver	Lantinga-van Leeuwen et al. (2005)

[a]Tissues expressing high level of putative target for the given TF.
[b]Tissues expressing low level of putative target genes for the given TF. See Methods for details.

ized. For instance, we found that genes around pCRMs for NF-Y tend to have low expression in the ciliary and superior cervical ganglia and high expression in thymus and lymphoblasts. NF-Y binds an element called the CCAAT box, which has been reported to be present within promoters of genes activated during peptide presentation in antigen presenting cells (APC) (Mach et al. 1996) and within the promoters of housekeeping genes such as those regulated during the cell cycle (Mantovani 1999). From this literature, one would not have predicted a role for NF-Y in the brain and the thymus, but the fact that ciliary and ganglia cells are not (or only slowly) dividing and that some APC originate from thymus (Choi et al. 2005) is however consistent with our findings.

The average expression levels were also computed for the set of genes associated with each pair of TFs. Of the roughly 2 million triplets (TF_1, TF_2, cell-type) tested, 5242 triplets show significant overexpression (P-value < 0.001, FDR <39%), while 6407 triplets show significant underexpression (P-value < 0.001, FDR <31%; see Supplemental Table S12).

A searchable public database of predicted regulatory modules

The modules predicted by the algorithm were stored in a database with a Web-based interface (http://genomequebec.mcgill.ca/PReMod). The database supports a variety of queries and contains hyperlinks pointing to the NCBI Entrez of the closest gene. The module information includes its genomic position as well as its TFBS content. A graphical view of the TFBS distribution of the highest scoring matrices is also provided (see, for example, Fig. 3C). Queries can reveal relationships such as the set of modules associated with a specific matrix, the set of modules located in the vicinity of a gene of interest, the set of the modules located within a specific distance from any gene, the set of modules associated with CpG islands, etc. Output from queries can be viewed as html or Excel tables. Genomic sequence of the whole set of modules can also be downloaded in fasta format from the Web site.

Conclusions

Using the literature as a guideline, we have identified a set of rules describing the architecture of DNA regulatory elements and used them to build an algorithm allowing us to explore the regulatory potential of the human genome. Although the error rate in CRM predictions is likely to be relatively high, the statistical power obtained through a large-scale, genome-wide approach revealed new insights into the biology of transcriptional regulation. Among other things, we observe a strong enrichment for pCRMs in regions at the 3′ end of genes. By concentrating on predicted TF-binding sites within pCRMs, we are able to improve the specificity of individual TFBS predictions, which allows the detection of signals that could not be seen otherwise. For example, we noted that a significant number of TFs have a strong bias for regulating genes either from a great distance or from promoter-proximal binding sites. Noteworthy is the fact that most TFs that preferentially work from a large distance are involved in development, while those predicted to work from promoter-proximal sites tend to regulate genes involved in basic cellular processes. We have identified a set a TFs that are predicted to play important roles in specific tissues, including cells and tissues issued from tumors and metastases. Finally, our data provides a starting point for the elaboration of human gene networks.

In a bootstrap-like fashion, several of the features derived from our pCRMs could be used to design improved CRM prediction algorithms. For example, the fact that specific TFs prefer binding at specific locations with respect to genes and that CRMs tend to organize in larger and looser clusters often containing binding sites for similar sets of factors could allow improved predictions.

We expect that the database containing the modules predicted in this study may speed up the discovery and experimental validation of CRMs. Finally, deeper data-mining approaches are likely to yield a plethora of specific testable biological hypotheses.

Methods

Transfac position weight matrices

A set of 481 vertebrate PWMs from Transfac 7.2 (Matys et al. 2003) was used for the analysis. Pseudocounts were introduced to regularize matrices based on few known sites (Durbin et al. 1998). Many PWMs represent the same or very similar factors. This does not cause any problem to our CRM prediction algorithm (since it excludes overlapping sites), but it is undesirable for downstream analyses of individual TF properties, e.g., localization with respect to the genes and tissue-specific expression. For these sections of the study, PWMs were grouped into 229 families based on the following rule: If many related TFs had individual PWMs, but Transfac also contained a generic PWM for the family, then only that generic matrix was used.

Module prediction algorithm

The outline of our module prediction algorithm is provided in Figure 1. We used a genome-wide multiple alignment of the human, mouse, and rat genomes (versions hg16, mm3, and rn2) produced by the MULTIZ program (Blanchette et al. 2004) and available from the UCSC Genome Browser (Karolchik et al. 2003). Only regions within MULTIZ alignment blocks are considered in what follows. These regions cover 34% of the human genome. For each of the 481 PWMs, individual binding sites are first predicted as follows. The human, mouse, and rat genomic regions are first scanned separately, on both strands, and a log-likelihood ratio score is computed in the standard way (Durbin et al. 1998). The only improvement is that we use a set of 3rd-order Markov models for background, and the choice of model depends on the local GC-content of the 1-kb region surrounding the position. Twenty different Markov models have been trained, based on nonrepetitive, noncoding human genomic regions with 0%–5% GC, 5%–10% GC, 95%–100% GC, and at every position the most appropriate background model is used.

Species-specific scores are then mapped onto the alignment and for each alignment column p and PWM m, we compute: $\mathrm{hitScore}_{\mathrm{aln}}(m,p) = \mathrm{hitScore}_{\mathrm{Hum}}(m,p) + 1/2 \max(0, \mathrm{hitScore}_{\mathrm{Mou}}(m,p) + \mathrm{hitScore}_{\mathrm{Rat}}(m,p))$. Thus, $\mathrm{hitScore}_{\mathrm{aln}}(m,p)$ will be high if all three species have a high-scoring site at position p. Notice that if the hit score of human is very high, the resulting $\mathrm{hitScore}_{\mathrm{aln}}$ may be relatively good even if mouse and/or rat do not have high-scoring hits at that position. This allows us to predict human-specific binding sites, provided that they are very good matches to the PWM considered. Once the alignment scan is completed, only positions with $\mathrm{hitScore}_{\mathrm{aln}}(m,p) > 10$ are retained to construct modules. This results in a total number of predicted sites that varies from 1.5 million for E2F (M00103) to about 8000 for Hogness (M00316), many of which are expected to be false positives (see Supplemental Table S1).

We now discuss how to compute moduleScore(p_1,...p_2) for

the alignment region going from position p_1 to p_2 of human. We first define TotalScore(m, $p_1...p_2$) to be the sum of the hitScores$_{aln}$ of all nonoverlapping hits for m in the region $p_1...p_2$. Formally, letting H_m be the set of all hits for matrix m in region $p_1...p_2$, we have TotalScore(m,$p_1...p_2$) = max $_{\{H \subseteq Hm \text{ s.t. hits in } H \text{ do not overlap}\}}$ $\Sigma_{h \in H}$ hitScore(m,p).

The optimization problem of choosing the best set of nonoverlapping hits is solved heuristically, using a greedy algorithm that iteratively selects the hit with the maximal score that does not overlap with the other hits previously chosen. For each matrix and each region, a P-value is assigned to the TotalScore observed, measuring the probability that a random region of the human–mouse–rat alignment would have a total score that would exceed the observed one. This P-value takes into consideration the length and GC-content of the region considered, as well as the overall frequency and score distribution of hits predicted for that matrix in the genome. This allows for a region dense in hits for a rare matrix (i.e., one with few hits in the genome) to obtain a higher score than a region equally dense in hits for a more common matrix. Matrices that tend to have a large number of hits throughout the genome are thus penalized. More precisely, for each matrix m, GC-content g and window length l, the distribution of TotalScore is estimated empirically through simulation, repeating 10 million times the following procedure: (1) choose l random positions from alignment regions with GC-content g and (2) compute the TotalScore of the set of positions selected, assuming that the l positions chosen form a contiguous region.

The score of a candidate module is computed based on one to five PWMs called tags. The first tag for region $p_1...p_2$ is the matrix with the most significant TotalScore, i.e., tag_1 = argmin$_{m \in PWMs}$ pValue(TotalScore(m,$p_1...p_2$)). The regions belonging to the hits selected for tag_1 are then masked out and the TotalScores for each matrix are recomputed, excluding hits overlapping those of tag_1. The second tag is then the matrix that achieves the most significant totalScore, and its occurrences are masked out. The process is repeated until five tags are selected, if possible. Finally, we define moduleScore($p_1...p_2$) = max$_{\{k=1...5\}}$ −log (pValueMaxUnif (k, 481, $\Pi_{l=1...k}$ pValue(totalScore(tag_k, $p_1...p_2$)))), where pValueMaxUnif(k, *481*, a) is the probability that the product of k random variables, each defined as the maximum of 481 uniform(0,1) random variables, is smaller than a.[7] A module can thus consist of one to five tags, depending on which number of tags yields the highest statistical significance.

The above procedure was used to search for modules of maximal length 100, 200, 500, 1000, and 2000bp.[8] For each window size, regions with moduleScore > 10 (i.e., P-value < e^{-10}) were identified. This choice of threshold is somewhat arbitrary, but results in a total number of bases predicted in pCRMs to be ~2.88% of the genome, a reasonable upper bound for the fractions of bases in regulatory regions. To address the fact that many of these modules overlap each other, a greedy algorithm was used to repeatedly select the highest-scoring module not overlapping any of the previously selected higher-scoring modules. This resulted in the set of 118,402 nonoverlapping modules studied in this work. Predictions were then mapped onto the latest human assembly (hg17) using the liftOver program (Karolchik et al. 2003; <0.1% of modules could not be mapped onto the new assembly and were discarded).

[7]Note that the formula for moduleScore is actually an approximation of the true P-value, for the following reasons: (1) Since competition for space between different tags is not modeled, the computed P-value of the total score of the 2nd, 3rd, 4th, and 5th tags are slightly conservative; (2) since the totalScores are discrete variables (but with a very large number of possible values), the approximation with a continuous uniform distribution introduces a small error; (3) since the moduleScore is obtained by selecting the best of five P-values, a multiple hypothesis testing correction should be applied. However, since we are mostly interested in the ranking of modules, this correction would make no difference.

[8]Only a small number of maximal lengths could be tried, as the calculation of the TotalScore P-values are computationally expensive and depend on that length.

Microarray design and production

A subset of the pCRMs was selected to build a microarray to be used for ChIP-chip validation experiments. For each TFs among ER, HIF1, STAT3, and E2F4, at most 50 pCRMs were randomly selected for each combination of the following categories: (1) module score: High vs. non-high; (2) totalScore for the given TF: High vs. non-high; (3) genomic location with respect to closest TSS: 10–100 kb upstream, 800 bp–10 kb upstream, -800 to +200 bp, +200 bp to +1000 bp, +1 kb to +10 kb, 0–10 kb downstream of 3′ UTR, or other. Most combinations could be not filled up to their quota. Each pCRM selected was extended symmetrically to a size of 1 kb, excluding repetitive regions. Primer pairs were designed for each region, using the Primer3 algorithm (Rozen and Skaletsky 2000), and the specificity was tested in silico by using a virtual PCR algorithm (Lexa et al. 2001). When the primer pair gave no satisfactory virtual PCR results, a new primer pair was designed by using Primer3 and tested again. The process was iterated three times to generate primer pairs predicted to be efficient to amplify regions from human genomic DNA for almost all of our selected pCRMs. This primer design pipeline allowed us to design primer pairs to amplify pCRMs from human genomic DNA with a success rate of ≈85%.

ChIP-chip assay and data analysis

ER ChIP-chip experiments were performed as described previously (Laganière et al. 2005). E2F4 ChIP-chip experiments were performed as follows: T98G cells (ATCG) were grown in DMEM containing 10% FBS and arrested through contact inhibition by allowing cells to reach confluence. Medium was changed after the second day of confluence and cells harvested on the third day. Confluent T98G cells were fixed with 1% formaldehyde, rinsed twice with PBS, and harvested. The cell pellet was lysed and sonicated to obtain DNA fragments of 600 bp on average. ChIP was performed using anti-E2F4 antibody (sc-1082, Santa-Cruz) and Dynabeads (Dyna). ChIP samples and nonimmunoprecipitated fragments were blunted with T4 DNA polymerase and ligated to unidirectional linkers. The DNA was then amplified by LM–PCR and labeling carried out post PCR by incorporation of Cy5 or Cy3-dUTP using Klenow polymerase reaction. Detail protocol can be found at http://www.ircm.qc.ca/microsites/francoisrobert/en.

Data were normalized and triplicates were combined using a weighted average method as described previously (Ren et al. 2000). The P-value threshold used for the analysis was established by testing the enrichment of 10 targets for each of the following P-value intervals for both ER and E2F4 ChIPs using quantitative PCR with SYBR Green: <0.001, 0.001–0.005, 0.005–0.01, 0.01–0.05, 0.05–0.1, 0.1–0.5, 0.5–1. The results of this validation process are shown in Supplemental Table S1. Using $P < 0.01$ (ER) and $P < 0.1$ (E2F4), virtually all targets are bona fide binding sites (see Supplemental Tables S2 and S3). All microarray data will be deposited to ArrayExpress.

Statistical significance of TF location preferences and spatial correlation

We used a permutation test to estimate the statistical significance of the observed number of binding sites predicted in each type of

region of the genome. Given the set of all predicted sites for all TFs, we first removed from consideration all but one of the hits of a TF within a given module. Each module thus contains at most one binding site for a given TF. To perform our permutation test, we repeatedly randomly chose two sites for two different factors, and exchanged their labels (but kept the original positions), provided they both lie in regions of the same GC-content (within 1% difference, measured over 1 kb). The scrambling procedure was sufficiently repeated often to reach a random distribution, at which point the number of sites in each region was counted. The experiment was repeated 100 times, from which the expectation and variance of the count of each TF in each region was estimated and the Z-score calculated. Notice that this procedure preserves the varying density of binding sites across the genome (since only labels, but not positions, are modified), as well as the local GC-content preferences of each TF. To estimate the significance of the long-range spatial correlations observed between sites of a given TF, a similar permutation test was applied and the observed number of co-occurrence within a given distance was compared with those obtained in the permuted data sets, allowing to compute a Z-score for each TF and distance interval.

Correlation between predicted TFBS and tissue-specific gene expression

For each TF, a set of putative target genes was defined as the genes with at least one high-scoring predicted site for that TF within a pCRM and within 10 kb of the TSS. The average expression level of these genes in each of 79 tissues (GNF Atlas II) was calculated and its significance was estimated using a permutation test. Tissues showing overexpression or underexpression with Z-score > 5 are reported in Table 2.

Acknowledgments

This work was funded by grants from Génome Québec and Génome Canada (M.B., V.G., B.C., and F.R.) and by the Canadian Institutes for Health Research (V.G). A.R.B. is a recipient of a doctoral fellowship from the IRCM/CIHR Cancer Research Program. X.C. is a recipient of a Génome Québec Comparative and Integrative Genomics Program. J.L. is a recipient of a U.S. Department of Defense Breast Cancer Research Program Predoctoral Traineeship Award (#W8IWXH-04-1-0399). F.R. holds a new investigator award from the CIHR. We thank Adam Siepel for his PhastCons data, UCSC Genome Browser group for their support, and John Stamatoyannopoulos for the DNaseI hypersensitive regions data.

References

Aerts, S., Loo, P.V., Thijs, G., Moreau, Y., and Moor, B.D. 2003. Computational detection of *cis*-regulatory modules. *Bioinformatics* (Suppl 2) **19:** II5–II14.

Aerts, S., Loo, P.V., Moreau, Y., and Moor, B.D. 2004. A genetic algorithm for the detection of new *cis*-regulatory modules in sets of coregulated genes. *Bioinformatics* **20:** 1974–1976.

Alkema, W.B.L., Johansson, O., Lagergren, J., and Wasserman, W.W. 2004. MSCAN: Identification of functional clusters of transcription factor binding sites. *Nucleic Acids Res.* **32:** W195–W198.

Alonso, C.R. 2002. Hox proteins: Sculpting body parts by activating localized cell death. *Curr. Biol.* **12:** R776–R778.

Bailey, T.L. and Noble, W.S. 2003. Searching for statistically significant regulatory modules. *Bioinformatics* **19:** II16–II25.

Bajic, V.B., Choudhary, V., and Hock, C.K. 2004. Content analysis of the core promoter region of human genes. *In Silico Biol.* **4:** 109–125.

Baroukh, N., Ahituv, N., Chang, J., Shoukry, M., Afzal, V., Rubin, E.M., and Pennacchio, L.A. 2005. Comparative genomic analysis reveals a distant liver enhancer upstream of the COUP-TFII gene. *Mamm. Genome* **16:** 91–95.

Beischlag, T.V. and Perdew, G.H. 2005. ER α-AHR-ARNT protein-protein interactions mediate estradiol-dependent transrepression of dioxin-inducible gene transcription. *J. Biol. Chem.* **280:** 21607–21611.

Beissbarth, T. and Speed, T.P. 2004. GOstat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20:** 1464–1465.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, J., Mattick, J., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304:** 1321–1325.

Blais, A. and Dynlacht, B.D. 2005. Constructing transcriptional regulatory networks. *Genes & Dev.* **19:** 1499–1511.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14:** 708–715.

Bulyk, M.L. 2003. Computational prediction of transcription-factor binding site locations. *Genome Biol.* **5:** 201.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122:** 33–43.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116:** 499–509.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308:** 1149–1154.

Choi, E.Y., Jung, K.C., Park, H.J., Chung, D.H., Song, J.S., Yang, S.D., Simpson, E., and Park, S.H. 2005. Thymocyte-thymocyte interaction for the efficient positive selection and maturation of CD4 T cells. *Immunity* **23:** 387–396.

Davidson, E. 2001. *Genomic regulatory systems: Development and evolution*, Academic Press, NY.

Dorschner, M.O., Hawrylycz, M., Humbert, R., Wallace, J.C., Shafer, A., Kawamoto, J., Mack, J., Hall, R., Goldy, J., Sabo, P.J., et al. 2004. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat. Methods* **1:** 219–225.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. 1998. *Biological sequence analysis*. Cambridge University Press, Cambridge, UK.

The FANTOM Consortium. 2005. The transcriptional landscape of the mammalian genome. *Science* **309:** 1559–1563.

Frith, M.C., Li, M.C., and Weng, Z. 2003. Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* **31:** 3666–3668.

Giardine, B., Elnitski, L., Riemer, C., Makalowska, I., Schwartz, S., Miller, W., and Hardison, R.C. 2003. GALA, a database for genomic sequence alignments and annotations. *Genome Res.* **13:** 732–741.

Givens, M.L., Kurotani, R., Rave-Harel, N., Miller, N.L., and Mellow, P.L. 2004. Phylogenetic footprinting reveals evolutionarily conserved regions of the gonadotropin-releasing hormone gene that enhance cell-specific expression. *Mol. Endocrinol.* **18:** 2950–2966.

Gupta, M. and Liu, J.S. 2005. De novo *cis*-regulatory module elicitation for eukaryotic genomes. *Proc. Natl. Acad. Sci.* **102:** 7079–7084.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. 2004. The Gene Ontology GO database and informatics resource. *Nucleic Acids Res.* **32:** D258–D261.

Hartman, S.E., Bertone, P., Nath, A.K., Royce, T.E., Gerstein, M., Weissman, S., and Snyder, M. 2005. Global changes in STAT target selection and transcription regulation upon interferon treatments. *Genes & Dev.* **19:** 2953–2968.

Howard, M.L. and Davidson, E.H. 2004. *cis*-Regulatory control circuits in development. *Dev. Biol.* **271:** 109–118.

Ihmels, J., Bergmann, S., and Barkai, N. 2004. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20:** 1993–2003.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409:** 533–538.

Johansson, O., Alkema, W., Wasserman, W.W., and Lagergren, J. 2003. Identification of functional clusters of transcription factor binding motifs in genome sequences: The MSCAN algorithm. *Bioinformatics* **19:** i169–i176.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P.A., and Gingeras, T.R. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296:** 916–919.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res.* **31:** 51–54.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309:** 1564–1566.

Kerber, B., Fellert, S., and Hoch, M. 1998. Seven-up, the *Drosophila* homolog of the COUP-TF orphan receptors, controls cell proliferation in the insect kidney. *Genes & Dev.* **12:** 1781–1786.

King, D.C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., and Hardison, R.C. 2005. Evaluation of regulatory potential and conservation scores for detecting *cis*-regulatory modules in aligned mammalian genome sequences. *Genome Res.* **15:** 1051–1060.

Kloster, M., Tang, C., and Wingreen, N.S. 2005. Finding regulatory modules through large-scale gene-expression data analysis. *Bioinformatics* **21:** 1172–1179.

Kolbe, D., Taylor, J., Elnitski, L., Eswara, P., Li, J., Miller, W., Hardison, R.C., and Chiaromonte, F. 2004. Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res.* **14:** 700–707.

Kolchanov, N.A., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Stepanenko, I.L., Merkulova, T.I., Pozdnyakov, M.A., Podkolodny, N.L., Naumochkin, A.N., and Romashchenko, A.G. 2002. Transcription Regulatory Regions Database TRRD: Its status in 2002. *Nucleic Acids Res.* **30:** 312–317.

Krivan, W. and Wasserman, W.W. 2001. A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res.* **11:** 1559–1566.

Laganière, J., Deblois, G., Lefebvre, C., Bataille, A.R., Robert, F., and Giguère, V. 2005. From the Cover: Location analysis of estrogen receptor α target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl. Acad. Sci.* **102:** 11651–11656.

Lantinga-van Leeuwen, I.S., Leonhard, W.N., Dauwerse, H., Baelde, H.J., Oost, B.A.V., Breuning, M.H., and Peters, D.J.M. 2005. Common regulatory elements in the polycystic kidney disease 1 and 2 promoter regions. *Eur. J. Hum. Genet.* **13:** 649–659.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799–804.

Levine, M. and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424:** 147–151.

Lexa, M., Horak, J., and Brzobohaty, B. 2001. Virtual PCR. *Bioinformatics* **17:** 192–193.

Mach, B., Steimle, V., Martinez-Soria, E., and Reith, W. 1996. Regulation of MHC class II genes: Lessons from a disease. *Annu. Rev. Immunol.* **14:** 301–331.

Mantovani, R. 1999. The molecular biology of the CCAAT-binding factor NF-Y. *Gene* **239:** 15–27.

Martens, J.A., Laprade, L., and Winston, F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae SER3* gene. *Nature* **429:** 571–574.

Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. 2003. TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31:** 374–378.

Medina, J.M. and Tabernero, A. 2002. Astrocyte-synthesized oleic acid behaves as a neurotrophic factor for neurons. *J. Physiol. (Paris)* **96:** 265–271.

Noble, W.S., Kuehn, S., Thurman, R., Yu, M., and Stamatoyannopoulos, J. 2005. Predicting the in vivo signature of human gene regulatory sequences. *Bioinformatics* **21:** i338–i343.

Petz, L.N., Ziegler, Y.S., Schultz, J.R., and Nardulli, A.M. 2004. Fos and Jun inhibit estrogen-induced transcription of the human progesterone receptor gene through an activator protein-1 site. *Mol. Endocrinol.* **18:** 521–532.

Philippakis, A.A., He, F.S., and Bulyk, M.L. 2005. Modulefinder: A tool for computational discovery of *cis* regulatory modules. In *Proc. Pac. Symp. Biocomput.* 519–530.

Pravettoni, A., Colciago, A., Negri-Cesi, P., Villa, S., and Celotti, F. 2005. Ontogenetic development, sexual differentiation, and effects of Aroclor 1254 exposure on expression of the arylhydrocarbon receptor and of the arylhydrocarbon receptor nuclear translocator in the rat hypothalamus. *Reprod. Toxicol.* **20:** 521–530.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309.

Robertson, G., Bilenky, M., Lin, K., He, A., Yuen, W., Dagpinar, M., Varhol, R., Teague, K., Griffith, O.L., Zhang, X., et al. 2006. cisRED: a database system for genome-scale computational discovery of regulatory elements. *Nucleic Acids Res.* **1:** D68–D73.

Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132:** 365–386.

Sabo, P.J., Humbert, R., Hawrylycz, M., Wallace, J.C., Dorschner, M.O., McArthur, M., and Stamatoyannopoulos, J.A. 2004. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl. Acad. Sci.* **10:** 4537–4542.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. 2005. Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132:** 797–803.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. 2004. JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32:** D91–D94.

Segal, E. and Sharan, R. 2005. A discriminative model for identifying spatial *cis*-regulatory modules. *J. Comput. Biol.* **12:** 822–834.

Segal, E., Yelensky, R., and Koller, D. 2003. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics* **19:** i273–i282.

Sharan, R., Ben-Hur, A., Loots, G.G., and Ovcharenko, I. 2004. CREME: *Cis* regulatory module explorer for the human genome. *Nucleic Acids Res.* **32:** W253–W256.

Sharrocks, A.D. 2001. The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* **2:** 827–837.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richard, S., et al. 2005. Evolutionarily conserved elements in vertebrates, insects, worms, and yeast genomes. *Genome Res.* **15:** 1034–1050.

Sinha, S., Nimwegen, E.V., and Siggia, E.D. 2003. A probabilistic method to detect regulatory modules. *Bioinformatics* **19:** i292–i301.

Sinha, S., Schroeder, M.D., Unnerstall, U., Gaul, U., and Siggia, E.D. 2004. Cross-species comparison significantly improves genome-wide prediction of *cis*-regulatory modules in *Drosophila*. *BMC Bioinformatics* **5:** 129.

Sorek, R. and Ast, G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13:** 1631–1637.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **101:** 6062–6067.

Tapscott, S.J. 2005. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132:** 2685–2695.

Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C., et al. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424:** 788–793.

Thompson, W., Palumbo, M.J., Wasserman, W.W., Liu, J.S., and Lawrence, C.E. 2004. Decoding human regulatory circuits. *Genome Res.* **14:** 1967–1974.

Viatour, P., Merville, M., Bours, V., and Chariot, A. 2005. Phosphorylation of NF-κB and IκB proteins: Implications in cancer and inflammation. *Trends Biochem. Sci.* **30:** 43–52.

Wang, W., Cherry, J.M., Nochomovitz, Y., Jolly, E., Botstein, D., and Li, H. 2005. Inference of combinatorial regulation in yeast transcriptional networks: A case study of sporulation. *Proc. Natl. Acad. Sci.* **102:** 1998–2003.

Wasserman, W.W. and Fickett, J.W. 1998. Identification of regulatory regions which confer muscle-specific gene expression. *J. Mol. Biol.* **278:** 167–181.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3:** e7.

Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R., and Young, R.A. 2003. Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113:** 395–404.

Zhou, Q. and Wong, W.H. 2004. CisModule: De novo discovery of *cis*-regulatory modules by hierarchical mixture modeling. *Proc. Natl. Acad. Sci.* **101:** 12114–12119.

Received October 31, 2005; accepted in revised form March 2, 2006.